Journal of Industrial Information Technology and Application Vol.1 No.2 (2017)
ISSN(Online) : 2586-0852

Real-Time Hand Detection From a Single Depth Image by Per-Pixel Classification

Myoung-Kyu Sohn[1)], Sang-Heon Lee[1)], Byunghun Hwang[1)], Hyunduk Kim[1)], Hyunsoek Choi[1)*]

[1)] Department of IoT and Robotics Convergence Research, DGIST, Daegu, Korea

Abstract. Due to their convenience and naturalness, hand pose recognition or gesture recognition methods are gaining attention as an upcoming complement of traditional input devices such as keyboards, mice, joysticks, etc. Robust hand detection from an image is the first stage to solve the hand gesture recognition. Due to the release of the commercial depth camera, elimination of the cluttered background from a depth image is much easier than from a RGB image. However, accurate hand segmentation from a human body still remains in challenging task. Here, we propose robust real-time hand detection algorithm from a depth image. The algorithm is designed to detect hands with various hand poses in various positions in 3D space. We train Radom Decision Forests to every pixel in the image to detect hand. The pixel in the image has one of the two label, hand or non-hand. We optimize the random decision forests parameters by various experimental conditions. The result shows that the per-pixel classification accuracy is 94% and the RDF with 5 trees requires only 12ms with no help of parallel programming.

Keywords; Hand detection, Decision forests, Hand pose, Gesture recognition

1. Introduction

Hand gesture recognition recently has attracted much interest as a means for interaction between human and computer in a variety of fields such as games, browser, media control, etc. [1, 2, 3]. Implementation of hand gesture recognition system represents a challenging task and relates to resolution of specific problems involved in detection of hand, recognition of hand shapes[4, 5, 6] and hand trajectory classifica-

* Corresponding author: {smk, pobbylee, bhhwang, hyunduk00, choihs}@dgist.ac.kr
Received: 2017.3.15; Accepted: 2017.7.10; Published: 2017.9.1

tion for recognizing movements of hand, etc. Detection of the hand in image has become easier as the background and foreground separation can be obtained simply from depth information due to spread of depth camera that provides depth information. However, the detection is still difficult when a hand is on the body or the distance of the hand from the camera is further than body position[7]. In this paper, we present a robust real-time hand detection method from a single depth image.

2. Propose Method

To detect the of a hand, Random Decision Forests(RDF) classifier[8, 9] was used. The detection problem can be solved as binary classification task. Decision forests was intended to classify a hand from a body. RDF can significantly improve individual stability and accuracy RDT as ensemble of Random Decision Tree(RDT). In addition, RDT itself also has the properties robust to overfitting as it uses the characteristic of data selected randomly. In training phase of the RDF, pixel based feature value is used. The extracted feature value by pixel has simple depth comparison characteristics as shown in Eq. 1. When target pixel is given, feature value of concerned pixel is as shown in Eq. 1:

$$f_\theta(I,p) = \mathrm{d}_I\left(p + \frac{u}{d_I(p)}\right) - \left(p + \frac{v}{d_I(p)}\right) \tag{1}$$

where, $d_I(p)$ is depth value of a pixel, p, in an image, I. Parameter $\theta=(u,v)$ indicates two offset values from p. Offset is normalized by depth, $d_I(p)$, which makes feature value depth invariant.

Training is to set splitting function of the node in such a way that nodes with similar class are gathered together when pixel data values that have entered parent nodes are distributed to children nodes. Each node split the date to left child node and right child node in accordance with the rule specified in Eq. 2 and 3. Before splitting the data, candidates of feature value and threshold value are extracted randomly. Then, information gain is calculated on every candidates of feature and threshold value (Eq. 4). The value of the candidate parameters is stored at each node when information gain is maximized. Each node uses that as split function in case of testing phase.

$$Q_l(\phi) = \{(I,p) \mid f_\theta(I,p) < \tau\} \tag{2}$$

$$Q_r(\phi) = Q \setminus Q_l(\phi) \tag{3}$$

$$G(\phi) = H(Q) - \sum_{s\in\{l,r\}} \frac{|Q_s(\phi)|}{|Q|} H(Q_s(\phi)) \tag{4}$$

$$\phi^* = argmax G(\phi) \tag{5}$$

where, H is Shannon entropy for entire input data of concerned node, and $\theta=(u,v)$ is the set of feature value and threshold value which were extracted randomly on each node. $\theta=(u,v)$ is the offset value as shown in Eq. 1. In training, feature value and

threshold value resulting in the highest information gain are stored in each node, which applies equally to all nodes. In other words, the objective of training is to store both feature value and threshold value maximizing the information gain in respective nodes.

In testing phase, the pixel value of test image is given as input value of trained RDF. When this value arrives at final leaf node based on split function of each node, the probability of the leaf node can be taken for final classification. By performing this for all trees and obtaining the mean value, final classification probability value can be calculated asthis for all trees and obtaining the mean value, final classification probability value can be calculated as

$$P(c \mid I, p) = \frac{1}{T}\sum_{t=1}^{T} P_t(c \mid I, p) \tag{6}$$

while P_t is the probability value in each tree, and T is the number of trees used in RDF. P is final value of probability that estimated class of target pixel is class, c.

3. Simulation and Resutl

The algorithm is verified on the public dataset [10]. The accuracy is a value of the number of correct estimated pixel divided by the number of pixel of hand. Fig. 1 and Fig. 2 shows the accuracy on the effect of decision forests parameters. The accuracy increases in accordance with tree height. Increasing the tree number has a positive effect on accuracy as expected. The resolution of the image is 320. Computation time is 12ms without any parallel programming in 2.4GHz CPU. The final result image is shown in Fig. 3.

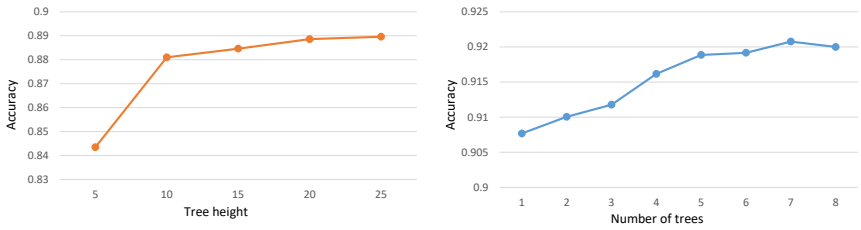

Figure 1. Per pixel classification accuracy. (a) effect of tree height, (b) effect of tree number

4. Future Work

In this paper, we propose a hand detection method for a single depth image. Per-pixel classification by random forests gives the label estimation for each pixel in the

depth image. For better performance in accuracy, various pose of the hand and various location of the hand are required. This detection algorithm can be used the initial value of the hand tracking. The detection performance would be better with tracking algorithm.

Acknowledgement

This work was supported by the DGIST R&D Program of the Ministry of Science, ICT and Future Planning. And, this research project was supported by the Sports Promotion Fund of Seoul Olympic Sports Promotion Foundation from Ministry of Culture, Sports and Tourism (s072015r2112015A0).

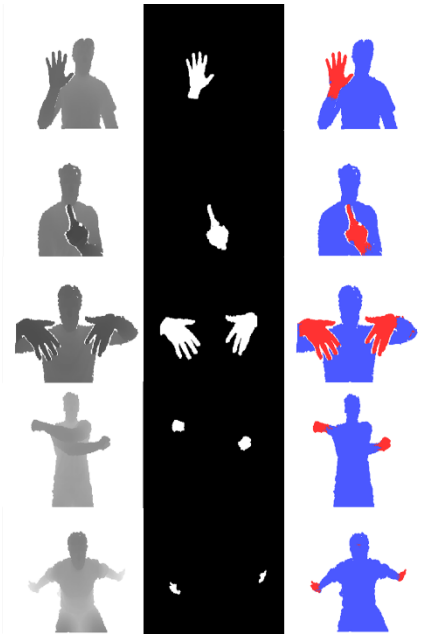

Figure 2. Samples of result images. First column is input depth image and second column shows the corresponding ground truth label image. The third column shows the estimated pixel of a hand (red pixel : hand, blud : not hand).

References

[1] Rautaray, Siddharth S., and Anupam Agrawal. "Vision based hand gesture recognition for human computer interaction: a survey." Artificial Intelligence Review 43.1 (2015): 1-54.

[2] Pavlovic, Vladimir, Rajeev Sharma, and Thomas S. Huang. "Visual interpretation of hand gestures for human-computer interaction: A review." Pattern Analysis and Machine Intelligence, IEEE Transactions on 19.7 (1997): 677-695.

[3] Prisacariu, Victor Adrian, and Ian Reid. "3D hand tracking for human computer interaction." Image and Vision Computing 30.3 (2012): 236-250.

[4] Erol, Ali, et al. "Vision-based hand pose estimation: A review." Computer Vision and Image Understanding 108.1 (2007): 52-73.

[5] I. Oikonomidis, N. Kyriazis, and A.A. Argyros, \Efficient model-based 3d tracking of hand articulations using kinect.," BMVC, p.3, 2011.

[6] Paul Doliotis et al, “Comparing gesture recognition accuracy using color and depth information”, Proc. Pervasive Technologies Related to Assistive Environments(PETRA), Crete, Greece, May. 2011.

[7] Keskin, Cem, et al. "Hand pose estimation and hand shape classification using multi-layered randomized decision forests." Computer Vision–ECCV 2012. Springer Berlin Heidelberg, 2012. 852-863.

[8] L. Breiman, Random Forests, Machine Learning, vol. 45, pp. 5-32, 2001.

[9] V. Lepetit, P. Lagger, and P. Fua. Randomized trees for real-time keypoint recognition. In Proc. CVPR, pages 2:775–781, 2005.

[10] Tompson, Jonathan, et al. "Real-time continuous pose recovery of human hands using convolutional networks." ACM Transactions on Graphics (TOG) 33.5 (2014): 169.