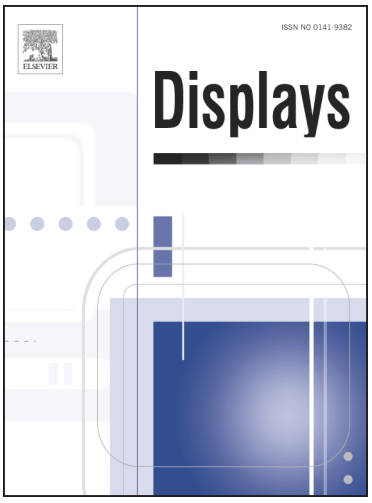


# Fast Hand and Finger Detection Algorithm for Interaction on Smart Display

Donghyeon Lee, Jin-Sung Kim, Hyuk-Jae Lee

PII: S0141-9382(17)30162-2
DOI: https://doi.org/10.1016/j.displa.2018.02.001
Reference: DISPLA 1860

To appear in: *Displays*

Received Date: 2 October 2017
Revised Date: 29 November 2017
Accepted Date: 6 February 2018

Please cite this article as: D. Lee, J-S. Kim, H-J. Lee, Fast Hand and Finger Detection Algorithm for Interaction on Smart Display, *Displays* (2018), doi: https://doi.org/10.1016/j.displa.2018.02.001

# Fast Hand and Finger Detection Algorithm for Interaction on Smart Display

Donghyeon Lee, Jin-Sung Kim and Hyuk-Jae Lee

***Abstract*— This paper presents a fast and accurate hand detection and tracking algorithm for depth images using an object contour analysis for which pairs of horizontal edges of an object are used. The computational complexity for generating the pairs of edges and estimating the object shape is low. A shape analysis is performed in the raster scan order and an object which can be a finger is selected as the finger candidate during the line scan. The initial hand is detected by searching for an object with a shape that corresponds to the predefined shape of the initial hand in the entire depth range. Thus, there is no assumption of the hand location throughout the depth, and the cluttered background has little effect on the detection. This paper also proposes a low-complexity hand tracking method which achieves 97.11% in the detection rate and 10.39 msec in the average processing time for a frame, respectively, when a single hand is detected. For five hands, the average processing time is 21.15 msec. These results show that the proposed method can be applied to the real-time application of vision-based HCI especially for smart display.**

***Keywords*—Object detection, hand detection, finger detection, depth image, human-computer interaction, smart display.**

## I. INTRODUCTION

At present, convenient human-computer interaction (HCI) schemes are required in many applications, such as smart TVs, intelligent vehicles, augmented reality and wearable devices. Especially, applications with various contents in smart TVs are becoming more diverse and important. In order to access various contents on smart TVs, a complicated user control is required. Many methods for user interface of smart TVs have been proposed [1-4], and vision-based HCI is considered to be a promising method for user interfaces. Extensive studies have been made in the detection of a body, face, eye, arm, leg and hand. Many such results have been commercialized successfully in electronic game applications, in which body and hand motion detection is used as a means of control. A hand is very important for human communication, and hand recognition is very important in vision-based HCI as well. In order to achieve precise control, the shape and motion of the hand must be detected precisely. Therefore, hand detection, hand shape recognition and hand tracking have been one of the major research subjects in HCI.

For hand detection, color-based hand segmentation is widely used, but the performance of this method is affected by the illumination and the background. To reduce the effect of the illumination, the HSV color space is used [5 – 8]. However, the effect of the illumination cannot be removed completely, and the performance of the color-based segmentation is limited when the hand color deviates considerably from the skin color model or when there is an object the color of which is similar to the skin color. The foreground-background segmentation method is used for hand segmentation. In this method, the difference between the current frame and the accumulated previous frames is used to detect the foreground [9, 10]. However, the accuracy of background detection decreases when there is motion in the background or when the object in the foreground does not move for a long time. In [11], the color of an area with motion is compared to the skin color. When the color of the area corresponds to the skin color, the area is determined to be a hand. In the learning-based hand detection method, a Haar-like feature and AdaBoost are adopted [10, 12–15]. This method detects a hand and recognizes the hand shape. Unlike face detection for which this method is very accurate, it is not very effective for hand detection due to hand shape changes and hand rotation.

For depth images, foreground-background segmentation is not difficult. In addition, depth images are not affected by variations in illumination and color. As the segmentation in a depth image is straightforward compared to color-based segmentation, depth images are widely used for object detection [16]. Although segmentation is not difficult in a depth image, hand detection is a still challenging operation because a hand is a non-rigid object which has many directions of rotation. In a depth image, hand detection methods which use the foreground-background segmentation have also been proposed. To find the area of the foreground, motion detection is used or the background image is detected by accumulating several previous frames [17, 18]. In many studies, a hand is assumed to be located at the position closest to the camera. Thus, the object with the minimum depth in a depth image is determined to be a hand [7, 19 – 23]. In [24], a depth range of a position of a hand is predefined and the object in this range is determined to be

Donghyeon Lee and Hyuk-Jae Lee is with the Inter-university Semiconductor Research Center (ISRC), Department of Electrical Engineering and Computer Science, Seoul National University, Seoul, Korea (e-mail: dhlee@capp.snu.ac.kr; hyuk_jae_lee@capp.snu.ac.kr)

Jin-Sung Kim is with the Department of Electronic Engineering of Sun Moon University, Asan, 336-708, Korea (e-mail: jinsungk@sunmoon.ac.kr)