Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 86 (2016) 257 – 260

Procedia Computer Science

2016 International Electrical Engineering Congress, iEECON2016, 2-4 March 2016, Chiang Mai, Thailand

Hand Gesture Recognition for Thai Sign Language in Complex Background Using Fusion of Depth and Color Video

Chana Chansri[a], Jakkree Srinonchat[b]*

[a,b]Signal Processing Research Laboratory, Department of Electronics and Telecommunication Engineering, Faculty of Engineering Rajamangala University of Technology Thanyaburi, 39 Moo1, Klong6, Thanyaburi, Patumthani, 12110 Thailand

Abstract

Hand detection and gesture recognition are the active research area in the computer vision. The main purpose to develop the sign language recognition and Human Computer Interaction (HCI). This article investigates and develops the technique to recognize hand posture of Thai sign language in a complex background using fusion of depth and color video. The new technology of sensors, such as the Microsoft Kinect, recently provides the depth video which helps researchers to find the hand position in the scene. This advantage is used to segment the hand sign in the color video without the environment interference such as skin color background. The histograms of oriented gradients are used to extract the image features of hand sign. These features are then pass to the artificial neural network for training and recognition. The result showed that the proposed method is robust to detect the hand gestures in the complex background. It provides the accuracy recognition for the Thai fingerspelling of 84.05%

Peer-review under responsibility of the Organizing Committee of iEECON2016

*Keywords:*Thai Sign Language; Hand Detection; Depth Camera; Artificial Neural Network.

1. Introduction

The sensors technology are recently developed to operate with the real time system. This is very effective for the researchers and machine design. The hand detection and gesture recognition are also exploited the new technology of sensors to operate which based on the machine vision and human-computer interaction. The human body is non rigid object, it make hand shape varied with different time. This makes the challenges to a lot of research works for

* Corresponding author. Tel.: +66 2549 3588.
E-mail address: jakkree.s@en.rmutt.ac.th.

Peer-review under responsibility of the Organizing Committee of iEECON2016
doi:10.1016/j.procs.2016.05.113

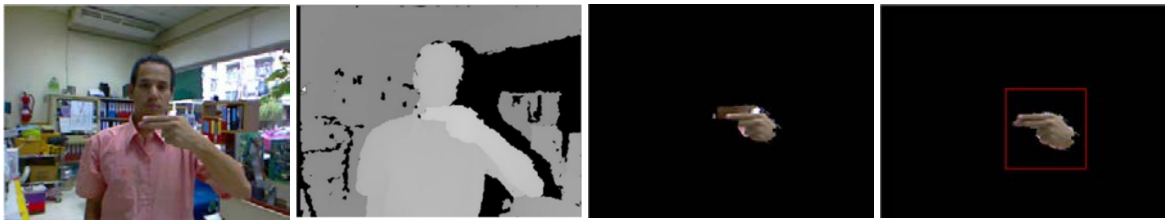
achievement of hand detection [1]. Also the complex background and instability of lighting conditions are directly effect to the hand detection in various working environment. This article proposes the technique of hand detection which applies to recognize the Thai sign language. It is designed to operate in front of the Kinect camera in this experiment, it is used to detect the hand posture in the scene. There is total absence of a learning phase in this section to reduce the computational time and without the training data which compare to other techniques[2-4] such as Viola-Jones Detector. The hand detection is also based on color and depth information combination. It is used to obtain a specific threshold that defines the regions of interest on the scene. The signer is always in foreground of the image so when the depth information was processed with color information, it will eliminate the background of color image and limbs such as face, arm, elbow etc. and this system is robust against complex environment. After the signer's hand is detected, the Histograms of Oriented Gradients (HOG) are used to extract the hand images features that algorithm provides good results with human and hand detection [5]. These features will be used as input for neural network to perform the database training and gesture recognition.

2. Hand Detection and Segmentation

The Microsoft Kinect using a diffractive optical device and infrared laser diode for generating an irregular dot pattern. The color and two megapixel grayscale chip was incorporated with an infrared (IR) filter used for determining the disparities between the emitted light dots and their observed position. After that, the depth of an object in the scene are triangulated which the identity of an observed dot on the object must be known. The irregular pattern obtained better performance than uniform pattern. This light is positioned with 640 dots horizontally and 480 dots vertically and its displacement is 3 mm, at the two meters distance from the source. The reflected light will used to obtain the depth of image in z-axis then the 3 dimensional environments can be simulated. The depth value d is mapped to a depth value in meters with a formula [6] by

$$\delta\tan(d) = 0.1236\tan(d/2842.5 + 1.186) \quad (1)$$

The Kinect sensor has two cameras, these cameras are installed in different locations. Thus the color and depth image will have a certain deviation in the range of visual angle. The color and depth image are captured with the Kinect sensor as shown in the Fig. 1(a) and Fig. 1(b). The color image and the depth image cannot be completely overlapped according to the pixels as shown in the Fig. 1(c). The image is captured the corresponding coordinates in the depth image (0,0) and color image (640,480) from Kinect sensor with image size 1280x480 pixels. The object position is referred to find the difference between the column and row in depth and color image. It bases on the same reference in depth and color image showing the value position of depth image is few less than color image. Those different values are the compensation index to use the location correction in the color image.



(a) (b) (c) (d)

Fig. 1. (a) RGB Image, (b) Depth image, (c) Overlay image between color and depth, (d) Hand segmentation correction.

The hand segmentation based on depth image relate to the distance from the signer's hand to Kinect sensor, the object closer a camera will get more brightness and if it is less brightness representing the object is far. This features make it possible to locate the hand in color image due to the hand is always in front of the scene generating the maximum brightness in the depth image. The threshold is used to define the regions of interest using the blob detection technique to determine the hand location. With blob detection technique, the center of signer's hand will