CHAPTER 1

INTRODUCTION

1.1 Rationale

The interaction between human and machine can be achieved by implementing human gesture applications. One of the most popular methods in human gesture application is hand gesture recognition. Gesture recognition offers a natural way of communication between humans and machines. It was developed several years ago, and now it can be easily found in many applications in real world e.g. in virtual reality, games, robotics, sign language recognition, etc. In the field of computer vision, human gestures are the natural way of communication between humans and machines or human-computer interaction (HCI). Human gestures can be expressed by the whole body or a part of the body such as face or hand as shown in Figure 1.1. Among these human gestures, the hand is the most frequently used. Hand gesture recognition (HGR) has been interesting topic research in past decades. Many important methods have been developed to improve the interaction and to make more interactive between human and computer in [4], [5], [6], [7], [8], [9] and [10]. The improvement of those methods in human gestures is still interesting and challenging.

(a) (b) (c)

FIGURE 1.1: The human gestures (a) Hand gestures by [1] (b) Facial gestures by [2] (c) Whole body gestures by [3]

Hand gesture recognition is challenging because there are problems on some uncontrolled environments such as the complex background, non-gesture moving object, and insufficient lighting as shown in Figure 1.2. The question is how to solve these problems? To overcome these uncontrolled environments, many methods have been developed for hand gesture recognition. As explained in [4], [5], [6], [7], [8], [9] and [10], researchers have developed methods to recognize hand gestures under uncontrolled environments, but some other problems caused by the uncontrolled environment are still exist. In fact, it is almost impossible to eliminate all these uncontrolled environment problems. But at least, the trade-off between the accuracy and the computation complexity among those methods to solve the uncontrolled environment problems can be evaluated. In [4], [5], [6], [7] and [10] the authors explain that the complexity of background problem in hand gesture can be overcome by using skin color & RGB Depth information. Their methods achieve the accuracy ranging from 85% to 95%. Under other uncontrolled environment like uncontrolled lighting and cluttered background, She et.al [8] and Chen et.al [9], proposed the evaluation of features points in hands such as palm center, fingertips, and joints as well as using depth sensor like Kinect sensor. Their experiments resulted the accuracy of 85-93.2%. So far, some of them have been implemented in human life but there are still opportunities to improve those methods.

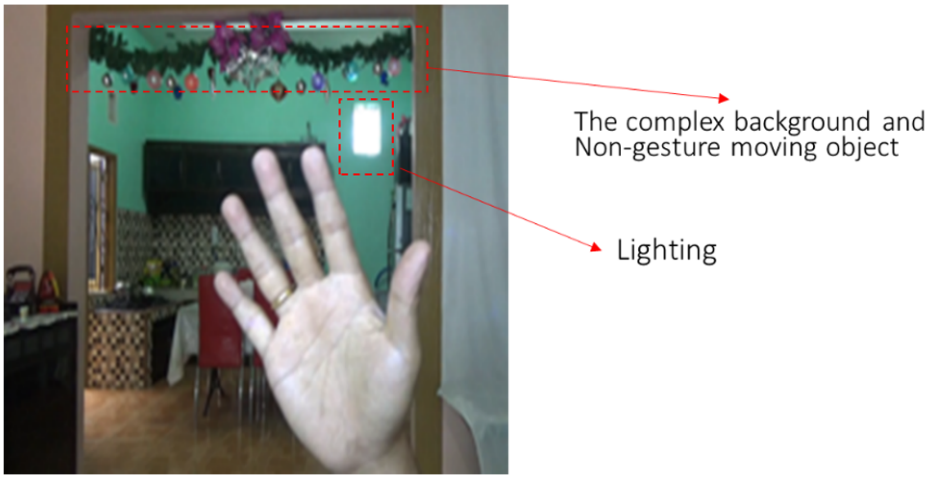

FIGURE 1.2: The uncontrolled environments

1.2 Theoretical Framework

According to [11], gestures are defined as the expression of body movement involving a whole human body or partially (head, face, arm, hand, and finger) for bringing meaningful information and interacting with the environment including machines. Among those human gestures, the hand gesture is the most frequently used in many research. Hand gesture recognition is a process to recognize the meaningful expression of hand movement. It was developed several years ago and it has been interesting topic research in the past decades.

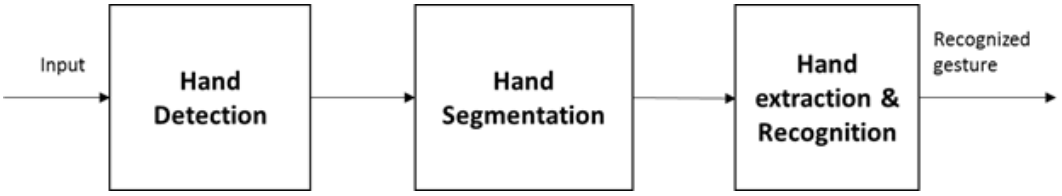

FIGURE 1.3: Diagram Block Hand Gesture Recognition System

Figure 1.3 displays the hand gesture recognition system (HGR). Initially, a hand gesture can be captured by using device sensors like a glove, Kinect or camera. The first step of HGR system is hand detection. This step captures data from an input device and the data can be in the forms of videos or images. The input data will be analyzed to get the interest image or frame video. This process commonly uses color space to separate the hand area from the background. As a part of hand gesture recognition system, hand detection, some literature [4], [5], [12], and [13] are combined with hand segmentation process. In this part, initially, the hand region can be detected through a color filtering because of its good features [14]. Another way for detection hand region is using a gray scale filtering. However, color filtering is more usable than a gray scale filtering because it has a capability to enhance the detection process thereby improving the detection result [15]. On the other hand, the perceived of color filtering varies in luminance and chrominance components depending on the darkness of the human skin as well as among different of human races. Moreover, a human being perceives characteristics of a color component such as brightness, saturation and hue component than the primary color red, green, and blue. Therefore, color filtering can be used to separate luminance from chrominance components.

The next step is hand segmentation process. This is an important and challenging step in hand gesture recognition system. In this step, the hand segmentation process separates the relevant data from the complex background or noise before processing to the next process. In the other hand, an uncontrolled environment like background complex, cluttered environment, and uncontrolled lighting, influence this segmentation process. It is important to remove noise from the image or frame video. Once it gets the clear image or frame video from noise, it is easy to get hand features in hand extraction & recognition process. Several methods have been developed for hand segmentation methods. One of the most popular is color space segmentation e.g. RGB, HSV, CIE LAB / L*A*B, YCrCb, etc. [4], [5], [8], and [16].

Finally, hand extraction step is intended to recognize the hand gesture. Hand extraction is intended to extract some shape features of hand such as hand contour, hand direction, finger position, fingertips, etc. Feature extraction is a critical part of hand extraction which determines exactly the most discriminating information in shape variation, motion, and textures of pattern of interest frame. Although hand shape can be recognized by extracting some shape features such as hand contour, hand direction, finger position, fingertips, these shape features are not always got easily because of uncontrolled environments such as the complex background, non-gesture moving object, and uncontrolled lighting. Moreover, although some other non-shape features are available, such as color, motion, silhouette, and textures, these features are inadequate for recognition. It is clear that the task of hand extraction is not easy. Therefore, this step must be done carefully and the shape features should be as good as possible. The hand extraction step is usually followed by the recognition process using classifiers, which uses the extracted shape feature vector to classify the gesture into its respective class. In [11], there are some classification methods that was employed in hand gesture recognition such as distance similarity, nearest neighbor, hidden Markov models (HMMs), principal component analysis (PCA), artificial neural networks, support vector machines, contour models, feature extraction, Gabor filtering, optical flow, etc.

1.3 Statement of the Problem

The hand gesture recognition methods still have the same problem when they attempt to recognize hand gesture accurately because of the existence of uncontrolled environment problem especially complex background. The complex background problem can be represented by cluttered background including lighting condition that occurs during the hand motion pattern in video duration. The previous studies in [5] and [6] indicate that some methods have been applied to remove complex background. The hand gesture recognition method proposed in [5] used skin color and motion detection. However, using skin color RGB based detection is highly affected by the color of people skin. Moreover, they could only recognize fewer hand gestures. Meanwhile, another hand gesture recognition method proposed in [6] used color information for extraction. However, using color information for extraction is highly affected by the change of lighting. Moreover, they could only recognize 6 hand gestures.

1.4 Objective

The objective of this study is to develop individual hand gesture recognition system by using spatio-temporal analysis under complex background. The expected result of this study is to increase the number of hand gesture recognition including its accuracy compared with references [5] and [6].

1.5 Hypotheses

In hand segmentation process, the frames of video have been transformed into slices of hand motion patterns in spatio-temporal domain. The hand motion patterns look different from the noisy background in spatio-temporal domain. The complex background appears mostly in the horizontal direction, meanwhile, the hand motion pattern appears in the vertical direction. By applying filtering in this process, it can eliminate the complex background in the horizontal direction from the hand motion pattern. After the complex background is clear, it is easier to obtain hand gesture in hand extraction process. Furthermore, it can recognize number of hand gestures.

1.6 Conceptual Framework

The proposed main concept in this study can be viewed in Figure 1.4. This proposed main concept contains three steps:

1. Hand detection, in color filtering process which has input video data that are analyzed using enhanced frames sequence in V color channel of HSV color space;

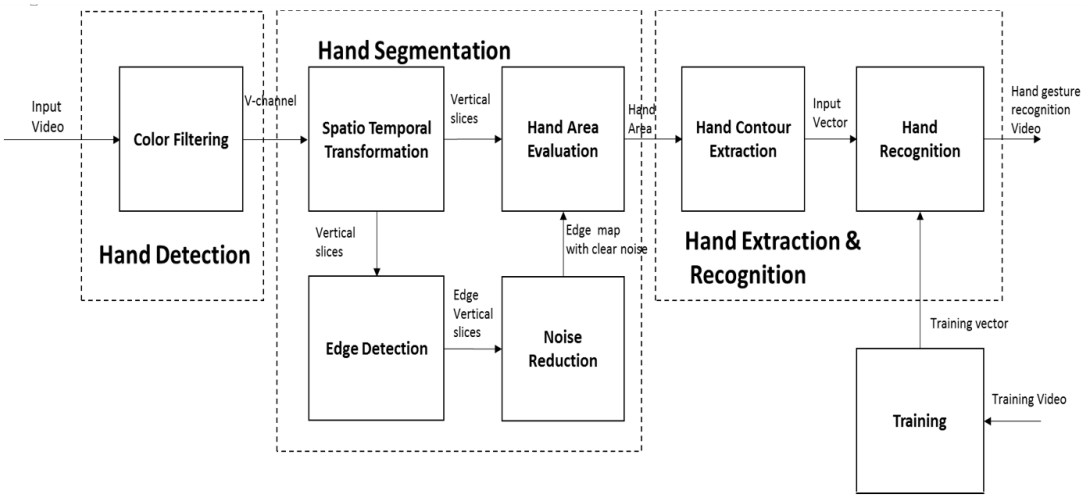

FIGURE 1.4: The Proposed of Main Concept of Hand Gesture Recognition System

2. Hand segmentation, this is the important part to remove complex background. This study introduces a scheme for removing the complex background in hand gesture recognition system using the spatio-temporal domain. As described in [17], spatio-temporal domain consists of two representations, horizontal and vertical domain. Information in vertical domain can be used to get the pattern of interest frame in hand detection process. The spatio-temporal domain process is used to construct some the vertical spatio slices in vertical direction from the interest frame. By using edge detection method, the image spatio slices will be processed to get some edge slices. Thus, in clear noise process, the edge slices will be processed to remove the complex background. In the hand area process, the edge clear noise will be processed to get hand area;

3. Hand contour extraction and hand recognition processes, are the final step to recognize the gesture. All these process will be discussed further in chapter 3.

1.7 Scope and Delimitation

In this research, the scope and delimitation of this study are:

1. The complex background is not a moving situation or static,
2. The complex background is represented by cluttered background and has sufficient lighting condition (lighting condition in the room where the camera can capture images clearly by using additional lighting behind the camera),
3. The environment is in indoor area,
4. Hand gesture moves from left to right side of camera observation and vice versa.

1.8 Significance of the Study

The hand gesture recognition system plays important role in computer vision science especially in an interaction between human and machine known as human computer interaction (HCI). Moreover, human gestures are the natural way of communication between humans and machines. Recently, many applications of hand gesture recognition are found in real world e.g. in virtual reality, games, robotics, sign language recognition, and etc. In the real world, human gesture recognition system is affected by the uncontrolled environments such as complex background and insufficient lighting. By using vertical representation in spatio-temporal domain, the complex background which is used for the system to increase the number of recognized hand gesture can be removed.