Chapter 2

Literature Review

Due to the limitations in computing technology and lack of powerful computer vision techniques in early days, tracking of of single hand in less cluttered or non-cluttered scene with high resolution imagery had been relatively difficult problem during that time. However, a lot of robust, real-time and useful hand trackers [4, 5, 6] had been introduced. Visual hand tracker [4], called DigitEyes can track a single hand with 27 degree of freedom in real-time from gray scale images at the speed of 10 frames per second. Early hand trackers are not limited only to 2D and some are capable of tracking hand in 3D space. One of the good example of 3D hand tracker is the real-time 3D hand tracker of Ahmad [5], which can operate at very fast speed of 30 frames per second. However those systems are not suitable for detecting and tracking hand in security application since they need high resolution imagery and detail of hand must be visible in the image.

Some early hand tracking systems like Pfinder [7] would be applicable to the hand detection problem for security. Pfinder is quite distinct in the fact that it attempts to follow the way humans look for the hand in images. Instead of directly detecting hands in an image, Pfinder looks for human bodies first and then easily segments out hands from the rest of the body by using skin color. However, since detecting humans in a cluttered video sequence is itself a very difficult problem, and the human body could easily be partially occluded in the scene, I try to bypass the human detection problem in our work by finding hands directly, without any attempt to find the entire human body first.

Over past 10 years, computer hardware and software technology is becoming more advanced, high computational power is available for researchers to implement powerful, sophisticated and computational costly algorithms in machine learning and image processing. Those powerful algorithms enable many robust and practically usable hand detectors and trackers. Most of the modern hand detectors and trackers are intended for human-computer action applications and very few researches had been done for security applications.

There are several approaches to hand detection and tracking. The first approach uses skin color information to segment hands from the background and then tracks segmented hands between frames using a tracking algorithm. The face and hand tracking system for sign language recognition [8] first segments the image into skin and non-skin regions using an elliptical model for skin pixels in C_bC_r space. Then face detection is used to locate the face skin blob ideally leaving only the skin blobs of hands. The system constructs a template for each hand then in subsequent frames, finds the region best matching that template using a minimum mean-squared error cost function. A similar approach is used by a real-time hand gesture system based on evolutionary search [9] to detect and track hands for human-robot interaction. The hands and face tracking system for VR application [10] also follows this

approach but this system is extended to track hands and faces in 3D for a virtual reality application.

Some hand trackers used slightly different approach. The hand tracker of Shamaie and Sutherland [11] does not use skin color information, so it works on monochrome video sequences but slightly high resolution imagery and less clutter background seems to be required. Hands are extracted from the background using a blob analysis algorithm then tracked using a dynamic model from control theory. Unfortunately, the techniques used in these systems to locate hand is just a image segmentation based on regional properties and relying more on tracking. Since tracking algorithms normally require more than two consecutive frames, these approaches are not suitable when the goal is to extract hands from single images.

The approach used in open hand detection in a cluttered single image using finger primitives [12] is quite distinct from previously mentioned approaches. This system makes use of the geometric properties of the hand, such as parallel edges of fingers, without the use of skin color or motion information. Their proposed system is robust to the size and the orientation of hands with the limitation that one or more fingers must be visible. So, it is not applicable to my case due to its needs for high resolution imagery, although this system can detect hand in cluttered single image robustly.

However, there is another approach, where a detection window is scanned over the image and each of the scanned image patches are classified as hand or non-hand. In this approach, various object detection techniques are used for classification of scanned image patches and detector is able to locate hands in static images. This approach is used by robust hand detector [13], which is able to detect upright hand in pretty high resolution image. Their system utilized boosted classifier cascade object detector [14, 3] and poses of detectable hands are limited to six fixed postures. Because of limitation on the postures of hand, their system is not directly usable for my desired system where hand is assumed to be in arbitrary posture. A boosted classifier tree for hand shape detection [15] uses similar approach, but it constructs classifier tree instead of normal classifier cascade. Their system is able to not only detect hand but also classify the shape and posture of hand. But their system experienced the limitation of the constraints on the shape and orientation of hand making less applicable for my problem.

Very robust object detector, boosted classifier cascade [14, 3] is made even more powerful by introducing new set of Haar-like filters [16] and several variations of AdaBoost learning algorithm [17]. This improved system was demonstrated as a hand detector, which is able to detect a hand in infinite number deformations and poses, in research work for human-robot interaction based on Haar-like features and eigenfaces [18]. According to the illustration of hand detector output sequences, it seems that the detector was tested on the image sequence, in which only a single high resolution hand is present, in contrast to our case, in which multiples hands and entire humans' body may present in the scene. This makes system less suitable for applying to my problem without any modification and improvement.

The system intended for real-time hand tracking in crowded scenes [19] tried to utilized the motion information between two adjacent frames addition to the appearance information. The main idea behind this system is from pedestrian detector [20], in which AdaBoost learning algorithm learn to recognize not only in appearance pattern in a single image but also the pattern of motion between two consecutive frames. Unfortunately, this system does

not work as good as in the case of detecting pedestrian and is not practically usable due to its high false positive rate. The main reasons for having high false positive rate is that the speed of hand motion is varied widely and the learning process is not able to generalize the motion pattern of hand.

From above literature review, some approaches can not be used for hand detection problem in security applications, although they are very efficient for their intended applications. However, approach based on general object detector of Viola and Jones [14, 3] and improved version of it [16, 17] are suitable starting point for my aimed system. Hand detector, mentioned in this thesis, is based mainly on this system with additional modules to improve the performance.