A VISION-BASED HUMAN HAND GESTURE RECOGNITION INTERFACE FOR IMAGE BROWSING APPLICATION

by

CHAN LIH YANG

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

December 2009

Acknowledgement

I would like to acknowledge my supervisor, Dr Khoo Bee Ee for her unfailing guidance and advices throughout my project. Her support and constructive suggestions are vital and key to the success of this project completion.

I would like to extend my appreciation to Lau Jian Yap, Hoo Swee Chuan, Loo Chun Hou, Leong Wei Chian and Norrish for generously lending their “hands” as the training sample and tester in this project.

Finally, last but not least, a big thanks to all others especially my family who have helped, encouraged and supported me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATION

ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
HCI	Human Computer Interaction
PC	Personal Computer
LUT	Look-up Table

PEMBANGUNAN ANTARA MUKA BERASASKAN PENGENALAN ISYARAT TANGAN SECARA PENGLIHATAN UNTUK APLIKASI PERLAYARAN GAMBAR

ABSTRAK

Tetikus komputer amat berguna sebagai perkakas masukan untuk komputer sudah sekian lama ini. Akan tetapi, ia menghadkan kebebasan pengguna dan kebiasaannya menjadi tempat pembiakan bakteria and punca penyebar penyakit. Tesis ini mencadang dan membangunkan satu antara muka berasaskan pengenalan isyarat tangan pengguna untuk aplikasi pelayaran gambar di komputer. Dengan ini, pengguna dapat memberi perintahan kepada komputer tanpa sebarang sentuhan dengan perkakas input. Projek ini dicapai dengan mengimplikasikan cara pengenalan objek Viola-Jones yang menggunakan ciri-ciri Haar and aturcara pembelajaran AdaBoost. Cara ini membolehkan aplikasi ini mencapai tepatan pengenalan isyarat tangan pengguna sebanyak 94% secara masa nyata dan 89% dalam percubaan memberi perintahan kepada komputer untuk melayari gambar-gambar. Untuk menpercepatkan kelajuan pengenalan tangan, sekmentasi dengan warna kulit dicadangkan untuk mengurangkan kawasan pencarian. Cadangan ini berjaya mengurangkan 19% masa pemprosesan dalam pengenalan. Tambahan lagi, ciri-ciri melatih-semula di dalam aplikasi membolehkan pengguna mengemaskini klasifikasi apabila diperlukan.

A VISION-BASED HUMAN HAND GESTURE RECOGNITION INTERFACE FOR IMAGE BROWSING APPLICATION

ABSTRACT

Computer mouse has been an efficient input device. However, the mouse usage limits user's freedom. Besides, the devices are easily contaminated with bacteria and spreading disease among users. The contactless vision-based hand gesture recognition is one of the solutions to the freedom and hygiene problem. But it faces challenges of usability in term of cost and environmental variation like lighting. This thesis proposes and implements hand gesture recognition methods in image browsing application, to allow users views pictures contactless from input device in real time. The lower level of the approach implements the posture recognition with Viola-Jones object detection method that utilizes Haar-like features and the AdaBoost learning algorithm. With this algorithm, real-time performance and high recognition accuracy up to 94% detection rate can be obtained. The application system yield average of 89% successful input command in a series of evaluation. Moreover, the application requires only common PC and webcam to address the concern of deployment cost. To further enhance the speed of hand detection in real-time application, an idea to reduce the area of search window by incorporating skin colour segmentation is proposed in this thesis. A reduction of 19% of processing time is achieved with the proposed method, comparing to the processing time without skin colour segmentation. In addition, the re-training feature in the application enables users to update the classifier easily whenever needed.

CHAPTER 1

INTRODUCTION

1.1 Background

The rapid growth of computerization has made human-computer interaction (HCI) essential part in daily life. Nowadays, it becomes so important that it has been deeply embedded in modern human life, ranging from shopping, banking, to entertainment and medication. According to Jenny (1994), HCI is the study of how people interact with computers and to what extent computers are or are not developed for successful interaction with human beings. The study of HCI considers large number of factors, including the environmental factors, comfort, user's interface and system functionality.

For the case of personal computer (PC), the input method of human-computer interaction has evolved from primitive keyboard, to high precision laser mouse and today's advanced multi touch screen panel. However, there is a drawback as these devices are easily contaminated with bacteria as user's physical contact is required especially in public computer such as hospital (Ciragil et al., 2003). The study of Schultz et al. (2003) reports 95% of keyboards in clinical areas are contaminated with harmful microorganism. As the result, the input devices have became a media in spreading disease from one user to others.

Besides of the hygiene concerns, the commonly used human hand gestures are expected to be part of HCI to serve the users better in the sense of higher degree of freedom and natural way compared to device based input method (Mathias et al.,

2004). However, recognizing human hand gesture is a highly complex task which involves many fields of studies including motion analysis, modeling, pattern recognition and gesture interpretation (Ying et al., 1999). As computational power grows exponentially making real time recognition more feasible, the integration of hand gesture recognition into HCI has obtained attention from researchers in recent years. Basically, there are three major categories of hand gesture HCI which are active infrared (IR), glove-based and vision-based gesture interface (Moeslund et al.., 2003). Active IR employs IR camera for detection but sensitive to sunlight, which make it a major drawback. Glove-based interface refers to the HCI where users are required to wear certain type of equipment to track the fingers position and hand motion (LaViola, 1999). Glove-based input interface has been exists since 1980s. The gloves technologies for hand gesture recognition is relatively matured compared to vision based recognition method, where numerous glove-based input devices are available in marketplace, for example: Sayre glove, MIT LED glove and Data glove (Sturman et al., 1994). The gloved-based hand gesture recognition is widely used in virtual reality application and sign language recognition. As an example, local researchers have developed a wireless Bluetooth Data gloves to recognize 25 common words signing in Bahasa Isyarat Malaysia (BIM) successfully (Tan et al., 2007). Besides glove-like equipment, some researchers even develop a sensor array that can be worn at wrist to detect muscle contraction to predict fingers movement and recognize hand gesture (Honda et al., 2007).

Even though glove-based input does allow user to apply hand gestures in HCI, input device attachment at hand or any part of body is required to make it works. Therefore, it still poses certain limit to freedom of usage (Quek, 1994) and can be a

media of disease spreading. On the other hand, the vision based gesture recognition method recognizes hand gesture in real time without any invasive devices attached to the user's hand. The vision based hand tracking is being done using image acquisition and processing with single or multiple cameras. Hence, there is no physical contact needed by users in this HCI method. There are many successful integration of the vision-based gesture recognition HCI into application such as replacing TV remote control with finger tracking, or interpretation of American Sign Language (Pavlovic et al., 1997). Vision-based gesture recognition HCI is one of the HCI methods that offers highest degree of freedom and naturalness (Moeslund et al., 2003), comparing to the commonly used QWERTY keyboard, glove-based gesture and active infrared sensor recognition. However, it also has toughest technical challenges among all.

1.2 Problem statement and motivation

As stated earlier, to tackle the problem of disease spreading through input device, the contactless vision based gesture recognition input is one of the solutions since users do not need to touch or hold any device with this method. Meanwhile, it is a free and natural way of HCI. But, it is challenging to promote the growth of hand gesture-based input application. First of all, the vision-based gesture recognition implementation cost has to be comparably equal or lower than normal input devices cost like keyboard, in order to encourage the gesture-based input application deployment. Besides, the vision-based hand gesture's recognition accuracy is lower and varying under different environment, compared to devices based input which is consistent over different condition (Moeslund et al., 2003).

Noticeably, many works have been done to make vision-based gesture recognition system feasible. But, the better usability of recognition system always comes with higher cost of deployment. For example, Chen et al. (2007) and Hongo, et al. (2000) employs multiple cameras to achieve desirable accuracy but indirectly increases the cost of implementation when more hardware and processing power are required. On the other hand, there are numerous works that successfully implement the gesture recognition HCI at lower cost but usability is compromised where the users are bounded to certain limitation that defeats the purpose of hand gesture recognition HCI. For instance, Gupta and Ma (2001) employs fast feature extraction with single camera but requires users to rigidly aligned hand to camera. On the other hand, Yuanxin Zhu et al. (2000) gesture recognition works is limited for users who wear long sleeve shirts.

Therefore, a HCI system needs to be developed to integrate suitable hand gesture recognition techniques together to achieve better usability and lower cost at the same time. Hence, in order to promote vision-based hand gesture recognition, this thesis's main motivation is to develop a gesture HCI system that balance between usability and cost.

1.3 Objectives of the thesis

The goal of this thesis is to develop a real time vision-based hand gesture recognition user input interface which is low cost but acceptable accuracy. The real time system is expected to run smoothly in PC without noticeable slowing down. The

HCI system comprises many parts, from image processing, segmentation, hand detection, feature extraction to hand gesture classification. Compared to previous similar works, the system should have extra features and improvement to meet the goal. To validate the performance of the HCI, a self-developed image browsing application with vision based gesture recognition input is chosen as test vehicle because the image browsing is one of the most common tasks in PC.

In the nutshell, the thesis objectives are:

- To develop a low cost vision-based hand gesture recognition interface system with comparable accuracy and tolerance under different condition.
- To enhance current existing hand gesture recognition technique to achieve desirable usability.
- To evaluate the performance of vision-based hand gesture recognition interface using a self-developed image browsing application

1.4 Thesis scopes and approach

The scope of this thesis is focusing on development of real time vision based hand gesture recognition HCI, where it requires only modest computing power and webcam. This thesis does not intend to cover the software compatibility issues like variance of camera driver and operating system. In general, this thesis is to study, develop, enhance and evaluate the hand gesture HCI system with the self-developed image browsing application. Since the thesis does not attempt to study human hand gesture behaviour, the hand gesture recognized in the HCI should be a non-intuitive and non-standard hand gesture like American Sign Language. (Wikipedia, 2008)

In order to develop the HCI system and meet the thesis objectives, a step-by-step methodology is outlined. Basically, there are four stages in the development cycles: Planning, Implementation, Optimization and Evaluation.

In planning stage, the methodology, techniques selection and system architecture is drafted. The application mainly requires a gesture recognition methodology that works fast enough for real time implementation. So, to meet this requirement, literature review in the field of hand gesture recognition, tracking and segmentation are carried out to identify any suitable methodology for this thesis; including papers on appearance and model based recognition, eg: Kalman filter, HMM, Boosting method as well as active contour and colour segmentation. The architecture of the system and functionality of each component are drafted.

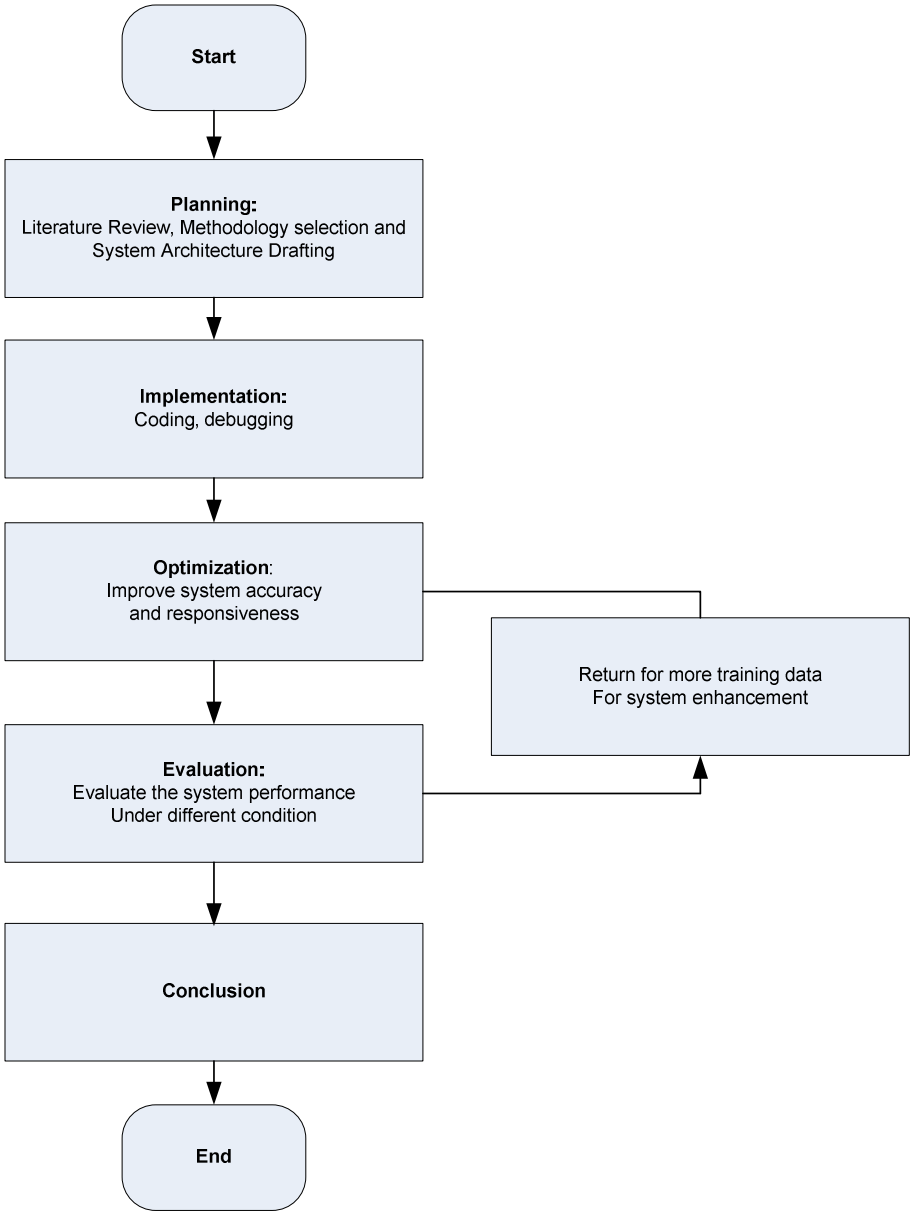

Figure 1.1: Flow chart of the overall work stages

When the overall architecture is well defined and methodology is selected, each component or module in the system is built independently and combined them together later, based on system architecture planning. The core of the system is the gesture recognition based input which is also the major focus of this thesis. So, to build the recognition system, first of all, numerous hand gesture images from one

individual are collected under an optimal lighting condition. Then, features extraction coding and classifier trainer are developed with this set of data. In this project, four types of Haar-like features are implemented with over hundreds of instances. The resulted classifier from training is verified for syntax error and bugs free in real time under similar lighting condition by the same individual whose hand gestures are training sample set.

In Optimization stage, the skeleton of the system is basically ready but performance enhancement is required. The methodology and coding are then further optimized to reduce the processing time as much as possible, including the skin colour segmentation which accelerate the hand localization process.

When the training and classifier coding are confirmed working, more training samples are collected from 5 individuals in an uncontrolled lighting condition to create variation and simulate the data randomness. The classifier is then re-trained with more samples. The accuracy of each instance of the feature is evaluated and weight is assigned based on AdaBoost methodology. As the result of including more training samples, the classifier is more robust and accurate, compared to the initial classifier which is meant for coding debug.

The completed application is tested under different condition based on experiment requirement. The result is reviewed and compared to similar projects. The reasons on failure of recognition under certain condition are also studied, understood and fixed if possible, for instance, going back to Optimization stage for more comprehensive training sample.

Finally, conclusion of the project is summarized and documented in this thesis. The overall project work flow is illustrated in Figure 1.1.

1.5 Thesis Outline

This thesis is organized as follow. This Chapter 1 introduces the background of HCI and the motivation to apply visual based hand gesture recognition for HCI. Besides, the problem statement, application overview and scope of this thesis are described as well in this chapter.

The Chapter 2 discusses the two hand gestures type: static and temporal. Review of current available approaches to recognize the hand gesture is revealed. Besides, several hand detection techniques with segmentation method like colour, contour and differencing segmentation are presented. Lastly, the recognition algorithm – Viola-Jones and AdaBoosting that is being chosen to be implemented in this thesis is explained.

The Chapter 3 describes how each the classifier in the system is being built, through the sample collection, classifier training and evaluation process. It starts with description of Haar-like feature, integral image construction and feature extraction components. Then the way of implementing AdaBoost training is explained and being evaluated with control data. Receiver Operating Characteristic (ROC) curve is plotted to show performance of each type of classifier after training. In addition, the skin colour segmentation technique that being used to localize hand is explained as well at the end of chapter.

The Chapter 4 introduces the image viewing application in this thesis and explains its system architecture behind the application. First, the system requirements and setup is defined. Then, the application layout, features and hand gesture commands are presented. As the system architecture is based on states machines, the flow and connection of each state are described in detail as well.

In Chapter 5, the completed application is tested for usability with different users, and environmental background. The hand orientation and lighting are also artificially changed to simulate different environment. Then, the successful rate in each experiment is tabulated and discussed. At the end, the responsiveness of the system is evaluated in term of number of classification executed per seconds.

Finally, Chapter 6 draws a conclusion and major contributions of this project. Then, based on the findings throughout the project, future improvements are suggested and discussed.

1.6 Summary

In Chapter 1, the problem of HCI and possible solutions are being reviewed. Then the objectives, approach and scope of this project is discussed. The goal of this thesis is to develop a real time vision-based hand gesture recognition user input interface which is low cost but acceptable accuracy. At the end of the chapter, the outline of this thesis is described.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Implementation of vision based hand gesture recognition as HCI is a very wide scope of field for study. The main focus in this chapter is to review on the hand gestures, hand detection and recognition methodology, and recent related works. Through the study, we will be able to understand and hence identify the suitable techniques for the implementation in the thesis. This chapter is organized as followed. Firstly, the definition and category of hand gesture are explained. Then, techniques of recognition for the hand gesture are reviewed and suitable methods are chosen for implementation. Finally, the recent works on visual based hand gesture recognition system are reviewed before the chapter ends.

2.2 Hand Gestures

A hand gesture is a form of non-verbal communication made using hand. According to Ying and Thomas (2001), the hand gestures can be classified into several categories: controlling gestures, conversational gestures, manipulative gestures and communicative gestures. Controlling gestures is the navigating gesture which uses hand orientation and movement direction to navigate and pointing in virtual environment or some display control applications. The example of controlling gesture is the virtual mouse interface, which enable users to use hand gesture to navigate the mouse cursors instead of using a physical mouse on the desk (Tsang et al., 2005). Conversational gestures are part of human interaction, for example emphasizing certain part of conversation with hand gesture. Manipulative gesture is a

way to interact with virtual objects such as tele-operation (Hasegawa et al., 1995) and virtual assembly.

Basically, a meaningful hand gesture can be represented by both temporal hand movements and static hand postures (Ying et al., 2001). Further explanation of temporal and static hand gesture as followed.

2.2.1 Temporal Hand gesture

The temporal hand movements or dynamic hand gestures represents certain actions by hand movements. For example, the conductor in orchestra is using temporal hand movement gesture to communicate music tempo to the team.

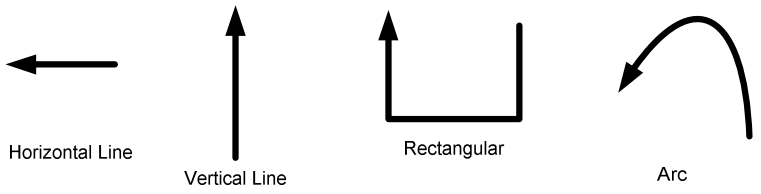

Figure 2.1: Some examples of temporal hand gesture movement (Ying et al., 2001)

To recognize the temporal hand gesture, HCI researchers need to track the hand movement in video sequences with sets of parameters like coordinates and direction. For simple hand gesture, Kalman filter is often employed to estimate, interpolate and predict the hand motion parameters for modeling and recognition (Aditya et al., 2002). Kalman Filter is a feedback control system that consists of time update equations and measurement update equations (Greg et al., 2004). Both set of equation forms an ongoing cycle where time update projects states ahead of time and measurement update adjusts the projection by actual current measurement (Ying et al.,

1999). Besides, Quek (1994) implements vector flow field method to find the velocity of the moving edges of hand. The vector field computation result correlates to the hand movement but the method only able to detect the direction and velocity only.

Hence, these methods are insufficient when dynamic hand gesture grows complex. So, the Hidden Markov Model (HMM) technique is utilized to model and recognize large variation of temporal hand movement gesture (Stoll et al., 1995). Basically, HMM is a statistical model which is being modeled and assumed to be a Markov process with unknown parameters. The challenge of HMM is to determine the hidden parameters from the observable parameters. Different from typical Markov Chain which is shown in Figure 2.2(a), the hidden state in the hidden Markov model is not directly visible by observer. However, the variables which are influenced by the state are visible as shown in Figure 2.2(b). HMM has been widely applied in speech recognition for years (Lawrence, 1989). Due to similarity between speech recognition and temporal gesture recognition, therefore HMM is also employed to recognize human motion in recent years. For instance the HMM technique is implemented in a real time gesture recognition system by Ozer et al. (2005) and Byung et al. (1997) where the body parts are tracked and analyzed by HMM with over 90% of the activities are correctly classified in their system.

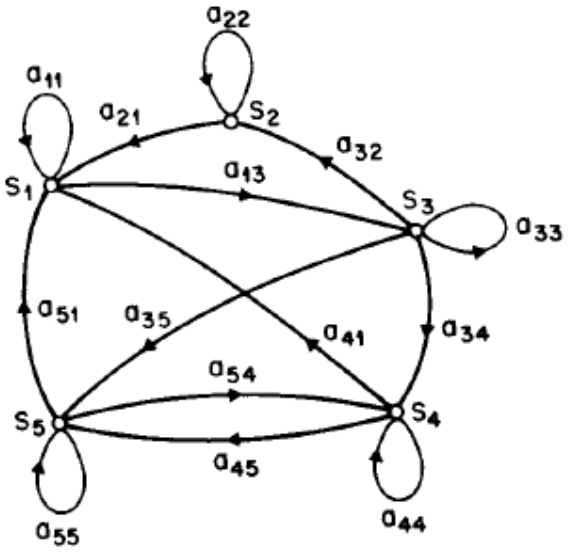

Figure 2.2(a): A typical Markov Chain with 5 states (Labeled from S1 to S5) and a_{ij} represents the state transition probability. (Lawrence, 1989)

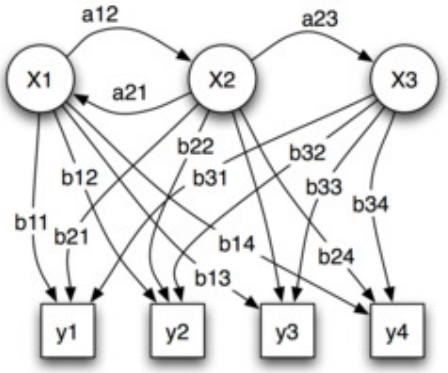

Figure 2.2(b): A Hidden Markov Model where Xn are hidden states and Yn are observable states. (http://en.wikipedia.org/wiki/Hidden_Markov_model)

For higher complexity temporal hand gestures, Wang et al. (2008) proposed implementation of hierarchical dynamic Bayesian networks through low-level image processing instead of HMM. Their experiment shows a slight accuracy improvement and dynamic Bayesian network is recommended for complex hand gesture over HMM. However, the experiment is done offline and further improvement on accuracy is suggested to enhance the method.

2.2.2 Static Hand Postures

Different from temporal hand gesture, the static hand postures express certain thought through hand configuration, instead of movement. Some common examples of static hand postures are “thumb up” to express good feeling or “pointing extended finger” to show direction. In general, the static hand posture detection methods are categorized into two categories of approach, which are appearance based and model based. (Lee J. et al., 1995)

2.2.2.1 Appearance based approach

Appearance-based approaches use image features to model the visual appearance of the hand and compare these parameters with the extracted image features from the input video. The features can be a wavelet, intensity gradient or a brightness difference between two areas like Haar-like feature (Viola et al., 2001). Kolsch M. and Turk M. (2004a) proposes frequency analysis method which instantaneously estimate the posture appearance suitability for classification, enable researchers to predict the classification rate of the hand posture upfront.

An appearance based detection, so called Viola-Jones detection method which is extremely fast and almost arbitrarily accurate approach (Viola et al., 2001) has been popular in faces and hand detection field especially in the real time application implementation. The method requires less computing power (Viola et al., 2001) and even feasible to be implemented in mobile platform like camera and handphone, which lack of processing speed (Jianfeng et al., 2008). Proposed by Viola and Jones, this method uses Haar-like feature extracted from image as the input to the classifier. Haar-like feature is a very simple feature based on intensity comparisons between

rectangular image areas. The method proposed a new image representation called Intergal Image that allows very fast feature extraction. The integral image can be constructed from an image using a few operations per pixel. Once the integral image is computed, these Haar-like features can be computed at any scale or location in constant time. The Haar-like feature instances with various sizes at different location are used as weak classifiers to separate the two classes. A weak classifier or weak learners means the feature with an unclear parameter boundary for two classes. Overlapping between two classes of parameters makes the feature weak in distinguish one to other. However, under certain special condition, there are some of the instances within the pool have better ability to separate two classes. Hence, AdaBoost is suggested by Viola and Jones to be implemented as part of Viola-Jones method to identify the combine many weak classifiers into a strong classifier.

There is one type of feature called Eigenpicture proposed by Kirby and Sirovich (1987, 1990), which is able to represent images at smaller dimension of feature for classification. Turk, M. and A. Pentland (1991) implement the idea into automatic face recognition system, which is well-known as Eigenfaces method. Eigenface approach is derived by applying Principle Component Analysis (PCA) on covariance matrix of an image dataset to find vectors that is best account for the representation of images. These vectors are called eigenvectors of covariance matrix that is corresponding to original face images.

Besides, there are some other appearance based features are implemented successfully in hand gesture recognition. For instance, Elena et al. (2003) experimental study on the template based detection system using Hausdorff distance

shows recognition rate up to 90% and fast enough for real time implementation. The system captures image of user's hand posture with webcam and segmented the hand blob using colour segmentation technique. Then the segmented image is compared to pre-processed template by calculating the bidirectional partial Hausdorff distance. However, the real-time implementation is limited to only four reference template per posture, which means robustness to variation like hand rotation angle is limited as well.

2.2.2.2 Model-based approach

The hand model-based approach is depending on a 3D hand model to estimate the hand parameters by comparing the input images to the possible 2D appearance projected in 3D hand model. Some researchers attempt to create a highly detailed 3D computerized hand model that simulates the articulation of the hand (Huan Du et al., 2007; Pavlovic, et al., 1997). Figure 2.3 shows some examples of hand model commonly used. The hand pose is represented in by a set of parameters which usually acquired by recovering user's palm, fingers, joints and fingertips from input images. The parameters can be an angle between joints, orientation of fingers and etc. Neural networks are often implemented to recognize hand gesture with the set of parameters that representing hand. For example, Berci et al., (2007) implements skeleton hand model to recognize hand postures. Their algorithm performs skeletonization on a hand silhouette to obtain the model of the hand posture.

Although 3D methods provide a more accurate modeling of a human hand, their deployment in augmented environment is challenging as the method is highly sensitive to image noise and hand segmentation errors. Also, this type of method

usually consumes higher computational power and therefore limits it implementation in real-time running application (Siu, 2005).

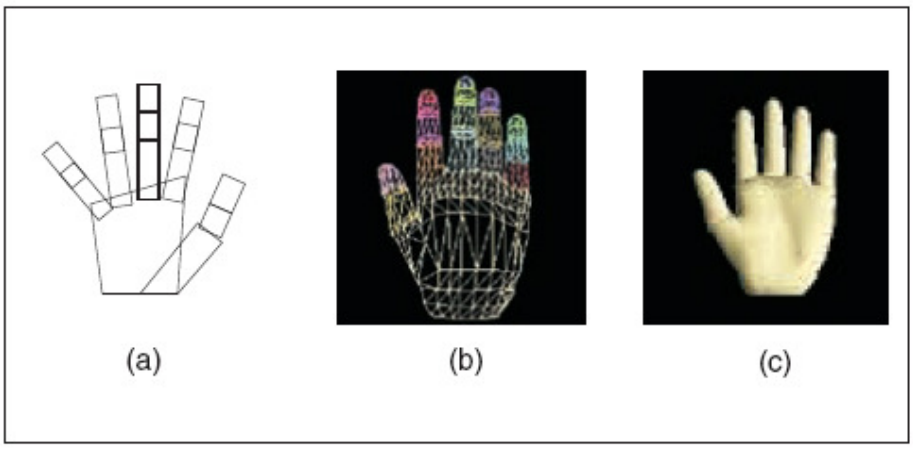

Figure 2.3: Three common types of hand model. (a) Cardboard ,(b) Wire frame, and (c) Contour . (Ying et al., 2001)

However, leveraging on computational power growth in mainstream PC, there are some successful model-based approach hand gesture recognition which is running real time. For example, a project by Vámossy et al. (2007) implements neural network classification on the skeleton hand model input parameter. Their experimental project achieves 81% recognition rate up to 22 frames rate under 320x240 pixels camera with simple background.

2.3 Classification

With the features extracted, classification is needed to differentiate or recognize the class of the input image; either it is model based or appearance based. Here, neural network and AdaBoost classification are being discussed because these methods are extensively implemented in gesture classification.

2.3.1 Neural Networks

Neural network is a non-linear statistical data modeling tool for complex relationship between input and output of a model. Basically, the architecture of neural network can be classified into two main categories, feed-forward and recurrent neural network as shown in Figure 2.4. Different from feed-forward neural network, the recurrent neural network propagate data from later processing to earlier stage, making it able to recognize time-varying pattern. (Samir, 2000) The neural network often implemented in pattern recognition for classification due to its advantages in noise immunity. To train a feed forward neural network, back propagation learning algorithm is one of the commonly known method. There are other learning algorithms like Delta Rule and Perceptron available for training neural network. (Alsmadi et al. 2009)

Neural network is widely implemented in temporal hand gesture recognition. For example, Vafadar et al., (2008) examine neural network classification with back propagation training on temporal hand gesture under simple background. Colour segmentation in HSV colour space and image morphology operators is implemented to extract hand contour. The classification of the test data yield 99.98% detection rate in noiseless data set, 92.08% in the noisy data. The experiment is done offline where data is captured upfront and processed.

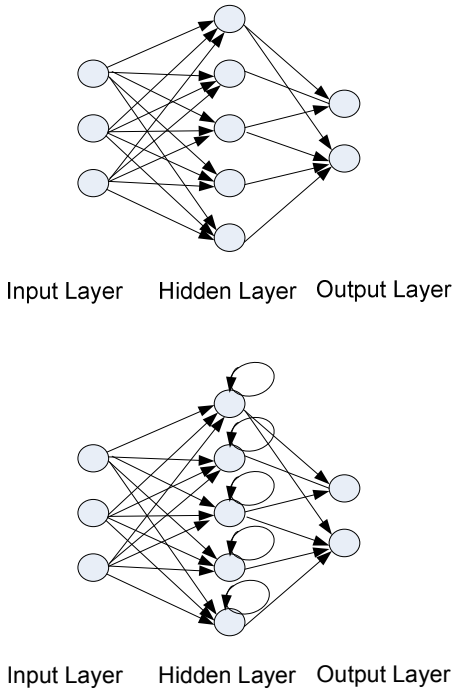

Figure 2.4: Feed forward and recurrent neural network

Deyou (2006) developed a DataGlove hand gesture recognition system that enables users to perform driving tasks through virtual reality concept. The system recognition core is based on a single hidden layer neural network which is trained with supervised learning algorithm. A test set of 100 hand gestures from trained users is tested on the trained neural networks, yields 98% of recognition rate. The recognition rate drops to 92% when the system is tested with alien users whose data is not included during training.

Mu-Chun et al. (1998) utilizes neural network to classify spatial-temporal signal extracted from hand gesture. In the experiment, 51 hand gestures are collected from 4 persons. The experimental result shows correct recognition rate is up to 92.9%. Ho-Joon Kim et al., (2008) combines a convolutional neural network with a weighted fuzzy min-max neural network to perform feature analysis. Then the feature data is

process with a modified convolutional neural network. Six different temporal hand gestures are tested. The experimental result shows lowest recognition rate at 80% for "Thumb up" gesture and highest at 97.5% for a 'wave up" gesture. Then, the weighted fuzzy min-max neural network is applied to select significant feature to reduce the number of feature. After reducing the number of feature 50% less, the recognition rate is still comparable to the initial condition.

Paulraj et al., (2009) presents a method based on neural network to translate "Kod Tangan Bahasa Melayu" into voice. The hand gesture is recorded using webcam under simple background. Then segmentation is done to extract the hand movement. Then discrete cosine transform is applied for feature extraction from the video sequence. A double hidden layer neural network is employed to classify the gesture. Their experimental results show 81% of recognition rate, out of 140 samples.

2.3.2 AdaBoost

Boosting is a method to improve the accuracy of learning algorithm. The AdaBoost algorithm is introduced in 1995 (Yoav et al., 1999) to solve the problems of the boosting algorithms. The AdaBoost algorithm takes a training set (X1; Y1) to (Xm; Ym) as input, where X is the data set and Y is the label of class. The m is the sample size of training set. In this thesis, we assume the label, Y to be 1 or -1, which represents two classes of data. Pseudo code for AdaBoost is shown in Fig 2.5.

Given: $(x_1, y_1), \ldots, (x_m, y_m)$ where $x_i \in X$, $y_i \in Y = \{-1, +1\}$
Initialize $D_1(i) = 1/m$.
For $t = 1, \ldots, T$:

- Train weak learner using distribution D_t.
- Get weak hypothesis $h_t : X \to \{-1, +1\}$ with error

$$\epsilon_t = \Pr_{i \sim D_t} \left[h_t(x_i) \neq y_i \right].$$

- Choose $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)$.
- Update:

$$\begin{aligned} D_{t+1}(i) &= \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{if } h_t(x_i) \neq y_i \end{cases} \\ &= \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t} \end{aligned}$$

where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

Output the final hypothesis:

$$H(x) = \text{sign} \left(\sum_{t=1}^{T} \alpha_t h_t(x) \right).$$

Figure 2.5: AdaBoost Pseudo code (Yoav and Robert E, 1999)

In AdaBoost learning algorithm, the training data is applied repeatedly to a base learning algorithm in a given number of iteration. Initially, all weights are set equally and get updated in each time of training iteration. On each round of the training, the weights of incorrectly classified examples are increased so that it will be focused on next round of training. A weak hypothesis or learner, h_t is selected in each training iteration, where the goodness of a weak learner is measured by its error, e_t. The error is measured with respect to the distribution, D_t on which the weak learner was trained. Alpha, α_t measures the importance of h_t where α_t gets larger when e_t is smaller. The distribution D_t is next updated using the rule shown in the figure 2.5. The effect of this rule is to increase the weight of examples misclassified by weak learner, h_t and to decrease the weight of correctly classified examples. The final outcome of

the training is the classifier with a combination of weak learners from each training iteration.

The AdaBoost algorithm has been implemented widely in numerous pattern recognition projects (Mathias et al., 2004), (Juan et al., 2005), (Qing et al., 2007), (Qing et al., 2008) and it is tested empirically successful by many researchers (Harris et al., 1996), (Jeffrey et al., 1996), (Richard et al.. 1997). Besides, it is a key ingredient of Viola-Jones detection method as this boosting method helps to determine strongest feature within a very large pool of data.

There is another boosting learning algorithm that similar to AdaBoost, which is called Support Vector Machine (SVM). SVM is proposed by Vapnik, (1998) to solve general pattern recognition problem. When given a set of points belonging to two different classes, SVM finds a hyperplane that separates the largest possible fraction of points of the same class. Yen-Ting et al., (2007) implemented SVM in their Multiple-angle Hand gesture recognition system which achieves over 95% of detection rate. However, there's difference in the computation requirement between SVM and AdaBoost. SVM corresponds to quadratic programming, while AdaBoost corresponds only to linear programming. As quadratic programming is more computational demanding, it makes SVM less feasible in real time application compared to AdaBoost (Yoav et al., 1999). A facial expression recognition experiment is carried out (Yubo et al., 2004) to compare AdaBoost and SVM processing time. Testing on a face sample database with Pentium IV 2.53GHz processor, the AdaBoost method is 300 times faster than SVM method.

2.4 Hand segmentation methodology

The aim of hand detection is to detect and localize the hand regions in image sequences. Artificial object detection such as specifically coloured object as described in Wilson et al. (2003), can achieve very high detection rates despite low false positive rates. Yet, the same is not true for faces and even less for hands because users are naturally reluctant to colour their hands. So, segmentation has becoming one of a crucial part to ensure the success of hand detection in vision-based recognition. Hand detection has attracted a great amount of interest and many methods relying on shape, texture, or temporal information have been thoroughly investigated over the years. Besides the traditional edge-based segmentation, the segmentation techniques like active contour, colour segmentation and differencing are being discussed here.

2.4.1 Active Contour

Active contours or so-called "Snakes" is commonly used in segmenting objects and deformable contour tracking in an image. The segmentation with active contours is done with minimization of the three energies in the active contour equation which are internal energy, image energy and external energy. Usually, the active contour is initialized near the object of interest and attracted toward the contour of the object by the intensity gradient in each iteration (Kass et al., 1987). However, the classic active contour algorithm will not operate well if there are large differences in the position or form of the object between successive images (Yuliang, 2003).

In Kim et al., (2001) work, the tracker utilizes the image flow, which gives rough information on the direction and magnitude of the moving objects. The correlation process between two images makes the snake tracks the object of interest.

The success of tracking is largely based on the calculation of the image flow. Unfortunately, it could become complicated in active vision, for example, the situation with moving cameras. The whole image is moving including both the foreground and background. It is hard to distinguish the motion of the object of interest when it moves in a similar speed as the camera (Yuliang, 2003).

2.4.2 Colour Segmentation

Some researchers have used human skin colour information to extract face and hand regions. That compelling results can be achieved merely by skin colour properties. For example, Schiele B. and Waibel A. (2005) who used it in combination with a neural network to estimate gaze direction. Kjeldsen R. and Kender J. (1996) demonstrates interface-quality hand gesture recognition solely with colour segmentation. Colour space can be mathematically represented by three dimensional coordinate systems. The colour space that is used for segmentation is RGB, HSV, CIELAB and YIQ as shown in Figure 2.6.

In RGB colour space, the three axes perpendicular to each other represents red, green and blue. HSV stands for Hue, Saturation and Value. YIQ is based on luminance and chrominance where Y is the luminance or brightness component, I and Q are the decoupled component of chrominance. The CIELAB colour space has three components as well, which are lightness (L*) and two colour components that position between green/red (a*) and yellow/blue (b*). The colour segmentation method that uses an HSV colour space is debatably beneficial to skin colour identification. The appearance of skin colour varies mostly in intensity while the chrominance remains fairly consistent according to Saxe D. and Foulds R. (1996).