UNIVERSITI PUTRA MALAYSIA

GESTURE RECOGNITION USING WEB CAMERA

LEW YUAN POK.

FK 2004 25

GESTURE RECOGNITION USING WEB CAMERA

By

LEW YUAN POK

Thesis Submitted to the School of Graduate Studies,
Universiti Putra Malaysia in Fulfillment of the
Requirements for the Degree of Master of Science

February 2004

DEDICATION

To

My parents :

Lew Kwan Shin & Lee Se Moy

My brothers and sister :

Yuan Huai, Yuan Sing & Yuan Yee

and

Yien Yien

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirements for the degree of Master of Science

GESTURE RECOGNITION USING WEB CAMERA

By

LEW YUAN POK

February 2004

Chairman : Associate Professor Hj. Abdul Rahman Ramli, Ph.D.

Faculty : Engineering

Gesture recognition represents an important ability by which a computer is able to directly accept human gesture as input to trigger different actions just like conventional input devices such as keyboard, mouse, joystick and etc. As the Human Computer Interaction (HCI) progresses over the years, emphasis is placed more on developing input devices which are most convenient and easy to use. Human gesture is not only natural and intuitive to a user, but can also represent motions of high degree of freedom which is of utmost importance in many applications especially in virtual reality.

This thesis presents the design of an offline system which is capable of recognizing hand postures from the visual input of a web camera. The hand segmentation is based on image subtraction technique and a skin color modeling process. Fourier descriptors are used as the features to describe the geometry of different hand postures while the recognition process is based on minimum distance classifier. The results obtained indicate that the system is able to recognize hand postures with reasonable accuracy.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia untuk memenuhi keperluan untuk Ijazah Master Sains

PENGECAMAN PERGERAKAN BERPANDUKAN KAMERA WEB

Oleh

LEW YUAN POK

Februari 2004

Pengerusi : Profesor Madya Hj. Abdul Rahman Ramli, Ph.D.

Fakulti : Kejuruteraan

Pengecaman pergerakan merupakan sesuatu keupayaan yang penting di mana sebuah komputer dapat menerima pergerakan manusia sebagai input untuk melancarkan pelbagai tindakan seperti yang dilakukan oleh peranti input tradisional misalnya papan kekunci, tetikus, kayu bedik dan sebagainya. Seiring dengan perkembangan cara interaksi antara manusia dengan komputer selama ini, lebih banyak penekanan diberikan kepada pembangunan peranti input yang berguna dan senang dipakai. Pergerakan manusia bukan sahaja semulajadi dan intuitif kepada seorang pengguna komputer, tetapi juga dapat mewakili pergerakan yang berdarjah kebebasan tinggi, menjadikannya amat penting dalam banyak aplikasi terutamanya kenyataan maya.

Tesis ini membentangkan proses rekabentuk sebuah sistem luar talian yang mampu mengecam keadaan tangan daripada input visual sebuah kamera web. Segmentasi tangan adalah berdasarkan kaedah penolakan imej dan sebuah proses permodelan

warna kulit. Deskriptor Fourier pula digunakan sebagai penghurai untuk menggambarkan geometri keadaan tangan yang berlainan manakala proses pengecaman adalah berpandukan pengelas jarak minimum. Keputusan yang diperoleh menunjukkan bahawa sistem ini dapat mengecam keadaan tangan dengan ketepatan yang munasabah.

ACKNOWLEDGEMENTS

My deepest thanks to my supervisor, Dr. Hj. Abdul Rahman Ramli, without whom this project would never have materialized in the first place. He has constantly encouraged me and given me clear guidance to make sure that I am able to complete the thesis within a reasonable time frame. Also to both my co-supervisors, Dr. Veeraraghavan Prakash and Mrs Roslizah Ali who have given me valuable advices throughout the whole period of the writing of this thesis.

A very big thanks to Qussay Abbas Salih al-Badri, my senior in the research group, who has been most generous in sharing his ideas and experiences. Also to my colleagues, Koay Su Yeong and Beh Kok Siang who have provided me valuable feedbacks on my thesis.

Not to mention also all the lecturers and staffs of Department of Computer and Communication Systems Engineering as well as members of Multimedia and Intelligent Systems Research Group who have all played their supporting roles.

Last but not least, I would like to thank my parents and family members who have been most supportive in each and every way I could think of. Also, I owe many thanks to Yien Yien, who has stood behind me no matter how harsh the situation is.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2 D	:	Two Dimension
3 D	:	Three Dimension
A/D	:	Analogue to Digital
AI	:	Artificial Intelligence
CCD	:	Charge Couple Device
CMOS	:	Complex Metal Oxide Semiconductor
DFT	:	Discrete Fourier Transform
DTW	:	Dynamic Time Warping
FSM	:	Finite State Machine
FD	:	Fourier Descriptors
GUI	:	Graphical User Interface
HMM	:	Hidden Markov Model
HCI	:	Human Computer Interaction
IR	:	Infrared
IC	:	Integrated Circuit
MIPS	:	Millions of Instructions per Second
NIR	:	Near Infrared
PC	:	Personal Computer
QE	:	Quantum Efficiency
RBFNN	:	Radial Basis Function Neural Network
Rx's	:	Receivers
RNN	:	Recurrent Neural Network
RGB	:	Red, Green, Blue
ROI	:	Region of Interest
USB	:	Universal Serial Bus
VLSI	:	Very Large Scale Integration
VE	:	Virtual Environment

CHAPTER 1

INTRODUCTION

1.1 Background

The evolution of the interaction between computer and human or better known as Human Computer Interaction (HCI) has always been interesting. The first generation computer user interface is a text based interface where user interact by typing relevant commands. By typing the commands through keyboard, the computer will perform relevant actions and inform the user of the outcome.

Then came the Graphical User Interface (GUI) where the main medium of interaction are symbolic graphics called icons. Each icon is assigned a specific meaning intended for specific application or certain action (Webopedia, 2003). In order to carry out a specific action or to run a specific application, the user can just click on the relevant icon by using a mouse. Text is still dominant in this second generation computer user interface as it is impossible to convert every line of command into icon, however, the use of text is minimized or avoided when necessary.

Nevertheless, it was soon realized that no matter how powerful the two kinds of user interface are, they will never be able to replace the way human communicate most naturally, *i.e.* through the use of gesture. As applications shift from 2D to 3D and eventually Virtual Environment (VE), the use of conventional input devices such as keyboard, mouse, trackball or joystick becomes more awkward and inconvenient than ever. It was then understood that nothing is able to navigate better in such applications than the human gesture itself. The human gesture is not only natural and intuitive to a user, it can also represent motions of high degree of freedom impossible to achieve using other input devices.

Owing to the many advantages displayed by the use of human gesture as a powerful input device and the potential applications, gesture recognition has become an important research field in recent years. Covering the diverse fields of motion modeling, motion analysis, pattern recognition, machine learning and even psycholinguistic studies, gesture recognition focuses on the study of human posture and movement of different parts of body such as hand and head (Wu and Huang, 1999). These also become the major motivation of this thesis to study various aspects of the vision based static hand gesture recognition system and to implement a system which can eventually be used as input in VE.

1.1.1 Gesture Recognition

Gestures are loosely defined by different researchers in different disciplines. From the point of view of biologists, the notion of gesture is to embrace all kinds of instances where an individual engages in movements whose communicative intent is paramount, manifest, and openly acknowledged (Nespoulous, 1986). This includes pointing at an object to indicate a reference, putting a thumb up to indicate consent or approval or conveying abstract meaning to one another using sign language. Gesture recognition on the other hand can be defined as the ability to identify specific human gestures by a machine or computer and using them to convey information or for controlling a device.

Gestures are important part of human communication where large amount of information is encoded in very compact form such as specific hand shape or orientation accompanied by certain movement. The ability to recognize these gestures represents an opportunity to decode the information conveyed most naturally by human. It is an alternative way of interaction between human and computer where little learning is required on the user as gesture is part of human's daily life communication.

Teaching a computer to recognize human gestures has been a goal of pursuit by various researchers since the 1980's. Earlier efforts in this research are focused on using a mechanical glove to capture the human gestures. On the mechanical glove, sensors are used to measure the bending angle of various joints of the hand and the pressure exerted by the palm. The different bending angle measurement will correspond to different geometric representation of the hand. In order to determine the position and orientation of the forearm in space, tracking sensors are used. However, the use of the mechanical glove is considered intrusive and unnatural to a user.

The focus in gesture recognition research later shifts to vision based gesture in conjunction with the advances made in the field of machine vision and computer vision. Vision based gesture recognition is a gesture recognition approach based on visual input. Imaging device such as camera is used to capture image sequences of human motion. Geometric properties of human hand or head together with their trajectories are then extracted from the image sequences to reconstruct human gesture. The vision based approach proves to be an attractive alternative since it is natural and not intrusive.

Earlier vision based gesture recognition system uses gray level image sequences as the input. In 1992, first attempt was made to recognize hand gesture from color video signals in real time. This was no coincidence as it is also the year where the first frame grabber for color video signal is available (Kohler, 2002). This marks the gradual shift of gesture recognition research from gray level video input to the color video input. The presence of color provides an additional and powerful cue for researchers especially in the area of segmentation.

Presently, gesture recognition still remains an active area of research which is still in infancy. As different applications are put to the drawing board, more gesture recognition techniques are still being improved, developed and created. This includes areas in segmentation, tracking, motion modeling, feature extraction and recognition techniques. Since there is no single known technique which can cater for all the situations to be encountered, gesture recognition is highly problem orientated.

1.1.2 Web Camera

As explained in the Merriam-Webster Online dictionary (2003), a web camera or webcam is a camera used in transmitting live images over the World Wide Web. Generally speaking, any video camera or digital camera that can perform this function qualifies as a web camera. The term refers more specifically to the function perform rather than the hardware itself as suggested by the definition.

Over the years, as hosting live images in the World Wide Web becomes more and more popular, a class of camera which is dedicated to this function mainly plus other additional features begins to emerge in the market and redefine the term, web camera. It is a class of low cost camera usually in the range from above RM100 to RM 400 which is connected to a PC through parallel port or USB port also known as pc camera.

Web camera is a class of camera which either use Charge Couple Device (CCD) or Complex Metal Oxide Semiconductor (CMOS) as the photo sensor. It usually has optical resolution of not more than 640×480 pixels and can produce digital still image or video which can be directly stored inside a pc. Compared to the conventional CCD camera which still use analogue signal, web camera proves to be more convenient as no extra image capture card is needed to do the Analogue to Digital (A/D) conversion. However, lower sensitivity and higher noise level are the two major drawbacks of web camera compared to higher quality CCD camera as justified by its lower cost.

Despite the shortcomings in the ability of web camera, it is still an attractive tool for performing gesture recognition especially when it is readily available in the market at a substantially lower price compared to other CCD camera. Web camera is still adequate to perform real-time gesture recognition as the key to its successful implementation lies not in the camera but in the processing speed of the computer carrying out the necessary manipulation of the image and the algorithm which determines how to manipulate the image effectively and efficiently.

1.2 Objectives

Vision based static hand gesture recognition is a complex task which is consisted of a series of structured steps. The aim of this thesis is to study the necessary steps of vision based static hand gesture recognition which will eventually lead to successful recognition of the human hand gestures off line using web camera.

The main objective of this thesis is to design a vision based static hand gesture recognition system using MATLAB which is able to perform recognition of six different sets of predefined hand postures.

The secondary objective of this thesis is to study the techniques involved in various stages of hand gesture recognition task and to propose modification whenever necessary.

1.3 Organization

The main issue to be addressed in this thesis is vision based static hand gesture recognition. Offline recognition of six different sets of hand posture is implemented using MATLAB 6.1 on a Pentium III 667 with 256MB of memory. A web camera is employed as the imaging device for capturing different hand postures. The thesis is organized as follows:

- Chapter 2 lays a background on the whole subject of computer vision and machine vision. Detailed discussions are then given one gesture recognition namely on vision based gesture recognition including facial gestures and hand gestures. Hand gesture recognition is then divided into two major areas which are static hand gesture recognition and dynamic hand gesture recognition respectively. Major differences between the two areas are highlighted especially in terms of the different approaches adopted. Different techniques used for the various stages of the hand gesture recognition ranging from segmentation, features extraction and gesture recognition technique are then treated in details.

- Chapter 3 presents the methodology involved in constructing the vision based static hand gesture recognition system. Discussions however focus on the three most important major steps of the system namely: (1)hand segmentation; (2)representation and description and (3)recognition. Hand segmentation technique is adopted to extract hand region from the visual input. Segmentation is done in gray level image, Red, Green, Blue (RGB) color space and normalized RGB color space. A skin color modeling techniques using the property of human skin color in color space is then proposed. Representation and description on the other hand explains the methods used starting from preprocessing of the extracted hand region until a set of useful features are extracted to represent the different hand postures. Recognition 5 presents the techniques used in classifying the feature vectors for the six different hand postures. Details on the steps used for implementing K-means clustering and the recognition method used are presented as well.

- Chapter 4 shows the results obtained from segmentation in different color space. The results for segmentation using the proposed skin color model are illustrated followed by the preprocessing, representation and eventually recognition based on the set of features proposed.

- Chapter 5 summarizes the thesis by restating the common problems arise in different stages of the gesture recognition task, thus laying down the requirements for a successful hand gesture recognition task as a whole and to check how much this thesis has helped to answer these requirements. For the questions and problems not addressed in other part of the thesis, they will be addressed in the future research directions.