Ph.D. Thesis

Hand Gesture Recognition Based on Convex Defect Detection and Density Distribution

June, 2017

Graduate School of PaiChai University

Department of Game Multimedia Engineering

Yanan Xu

Hand Gesture Recognition Based on Convex Defect Detection and Density Distribution

Supervisor: Dong-Won Park

June, 2017

Graduate School of PaiChai University

Department of Game Multimedia Engineering

Yanan Xu

The undersigned have examined the dissertation entitle: "Hand Gesture Recognition Based on Convex Defect Detection and Density Distribution" presented by Yanan Xu, a candidate for the degree of Ph.D. in Game Multimedia Engineering and hereby certify that in their judgment it is worthy of acceptance.

Advisor

June, 2017

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dong-Won Park Ph.D.

Professor

Dept. of Game Multimedia Engineering

Graduate School of PaiChai University

Principle Examiner

June, 2017

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Sung-Ohk Ahn Ph.D.

Professor

Dept. of Game Multimedia Engineering

Graduate School of PaiChai University

Examiner

June, 2017

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Gouchol Pok Ph.D.

Professor

Dept. of Liberal Arts

Liberal Arts of PaiChai University

Examiner

June, 2017

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Wook-Ho Son Ph.D.

Scientist

Electronics and Telecommunications Research Institute

Examiner

June, 2017

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Young-Ho Sohn Ph.D.

Professor

Dept. of Computer Engineering

Yeungnam University

Hand Gesture Recognition Based on Convex Defect Detection and Density Distribution

Yanan Xu

*Department of Game Media Engineering*

*Graduate School of PaiChai University, Daejeon Korea*

(Supervised by Professor Dong-Won Park)

Human motion analysis is an important research direction in the field of human-computer interaction and recognition, wherein, hand gesture is a kind of natural and intuitive interaction mode that can conveniently express rich meanings. Vision based hand gesture recognition has become one of the key techniques in the realization of new human-computer interaction currently. However, due to the diversity and complex of hand gesture itself as well as the temporal and spatial differences, coupled with the visual inadaptability, research on hand gesture recognition has become a challenging issue.

On the basis of the existing techniques of hand gesture recognition, a new method based on density convex defect detection and density distribution of hand gesture recognition is proposed in this paper. Firstly, detect skin area and segment hand part with color features. Secondly, use convex defect to detect the fingertips, and represent a gesture through the count and orientation of the fingertip combined with gesture contour length, area and other geometric features to extract the convex defect features. Thirdly, extract hand gesture distribution features that composed of density distribution feature and relative distance between fingers, where density distribution is a region based shape representation. Finally, integrate all the features to calculate similarity distance for recognition. Result shows that proposed approach is invariant to the rotation, scaling and translation of hand gesture images, and offers some robustness to scene illumination changes.

**Keywords**: hand gesture recognition, skin detection, density distribution, convex detection, similarity distance

# Table of Contents

Abstract i

Table of Contents ii

List of Figures iv

List of Table v

I. Introduction 1

1.1 Background and significance 1

1.2 Research status 3

1.2.1 Review of Literature 3

1.2.2 Application and foreground 6

1.2.3 Difficulties in research 9

1.3 Organizational structure of dissertation 11

II. Related Work 12

2.1 Hand gesture recognition overview 12

2.1.1 Definition of hand gesture 12

2.1.2 Classification of hand gesture recognition 13

2.2 Vision based hand gesture recognition 17

2.2.1 General flowchart of vision based hand gesture recognition 17

2.2.2 Sample image capturing 19

2.2.3 Image preprocessing 20

2.2.4 Feature extraction 21

2.2.5 Classification and recognition 22

III. Hand gesture segmentation 24

3.1 Skin color detection 24

3.1.1 Color balance 24

3.1.2 YCgCr color space conversion 26

3.2 Hand gesture region location 29

IV. Hand gesture features extraction 32

4.1 Gesture convex defect detection 32

4.1.1 Convex defect and fingertip detection 32

4.1.2 Gesture convexity and relative position of fingertips 36

4.2 Gesture distribution feature 38

4.2.1 Density distribution feature 38

4.2.2 Improved gesture density distribution feature 40

V. Similarity measurement 44

VI. Result and conclusion 46

References 51

Korean abstract 58

Acknowledgement 59

# List of Figures

[[Figure 1-1] Immersion CyberGrasp data glove 3](#_Toc480558734)

[[Figure 1-2] Canesta virtual keyboard 7](#_Toc480558735)

[[Figure 2-1] The static gesture and dynamic gesture 14](#_Toc480558742)

[[Figure 2-2] Classification of hand gestures based on expressive meaning 15](#_Toc480558743)

[[Figure 2-3] Hardware based data acquisition equipment of hand gesture 16](#_Toc480558744)

[[Figure 2-4] Architecture of hand gesture recognition 17](#_Toc480558745)

[[Figure 2-5] The process of hand gesture image preprocessing 20](#_Toc480558746)

[[Figure 2-6] Gesture recognition process 22](#_Toc480558747)

[[Figure 3-1] Processing effect of color balance 25](#_Toc480558748)

[[Figure 3-2] Skin color clustering region in YCbCr (a) and YCgCr (b) 28](#_Toc480558749)

[[Figure 3-3] Gesture segmentation of image with simple background 30](#_Toc480558750)

[[Figure 3-4] Gesture segmentation of image with complex background 31](#_Toc480558751)

[[Figure 3-5] Gesture segmentation of image with noisy skin area contained 31](#_Toc480558752)

[[Figure 4-1] Detected contour of hand gesture 33](#_Toc480558753)

[[Figure 4-2] Convex hull of hand gesture 33](#_Toc480558754)

[[Figure 4-3] Convex defects of hand gesture 34](#_Toc480558755)

[[Figure 4-4] Definition of length and width of hand gesture contour 35](#_Toc480558756)

[[Figure 4-5] Tightness of different gestures to convex hulls 37](#_Toc480558757)

[[Figure 4-6] Relative position of fingertips of different hand gestures 37](#_Toc480558758)

[[Figure 4-7] An instance of image region division 40](#_Toc480558759)

[[Figure 4-8] DDF of different images with similar gestures 41](#_Toc480558760)

[[Figure 4-9] Sketch map of skin area region 42](#_Toc480558761)

# List of Table

[[Table 6-1] Hand gesture location result 47](#_Toc480558588)

[[Table 6-2] Features used in this paper 48](#_Toc480558589)

[[Table 6-3] Recognition result of 5 number gestures with simple background 48](#_Toc480558590)

# I. Introduction

## 1.1 Background and significance

Human-computer interaction is an essential part of most people's daily life. The traditional human-computer interaction mode from the original keyboard to the current mouse, joystick, and wireless input devices, greatly facilitates the interaction between people and computers and makes it easier for people to operate the computer and improve work efficiency [[[1]](#endnote-1)]. However, these interactions rely on the additional input devices, and do not confrom to the people's interaction habits, therefore, this kind of interaction mode cannot completely meet the demands of human-computer interaction [[[2]](#endnote-2)]. In addition, the communication between people and machines often has no expression, no action and too much mechanical, so, it will bring great significance to improve the level of human-computer interaction and enhance the practicability of human-computer interface if the machine can understand the human languages.

In recent years, with the rapid development of machine vision technology, especially the progress of image processing and recognition technology, people's attention is no longer limited to the improvement of traditional input methods, how to make use of the biological characteristics of human beings to study more natural interaction technologies, so that human and computer can interact directly is becoming the current research focus of human-computer interaction. These researches include gesture recognition, facial expression recognition, face recognition, lip reading recognition, limb movement tracking, eye gaze tracking and pose recognition [[[3]](#endnote-3)], etc. According to the current development trend of human-computer interaction technology, human centered human-computer interaction technology is bound to replace the computer centered ones.

Hand is the most flexible part of human body, it can express rich and various meaning, and it is more natural, direct compared to other human characteristics. Hand gesture is vivid, intuitive, and contains a lot of interactive information, with same expression ability as natural languages such as spoken language and written language, it is an important part of body language [[[4]](#endnote-4)], and therefore, it can act as a means of natural communication between human and machine, and plays an important role in the field of human-computer interaction. However, due to the complexity of hands, the diversity, ambiguity and uncertainty of gesture, hand gesture recognition is becoming a challenging interdisciplinary research topic and a hot and difficult research topic in the field of human-computer interaction, studies on hand gesture recognition are very important for improving human-computer interaction.

In order to allow computers to "understand" the gesture information, the initial gesture recognition was through the use of wearable technology, allowing users to do some hand gestures with special data gloves on [[[5]](#endnote-5)] [[[6]](#endnote-6)], the data gloves would transfer users’ gestures and location information back to the computers and make them comprehend the gestures and behaviors of users. Figure 1-1 shows a multi-function virtual reality device composed of many sensors on the glove called Immersion CyberGrasp. Through the software mapping, the virtual objects can be shifted, clutched and rotated by the glove with the ability of “reach into the computer”. The latest release of this product is able to register bends for each finger. The glove can transmits hand gesture to the computer in real time accurately, and then receives feedback from the virtual environment to the operator. It provides users with a direct and universal human-computer interaction mode [[[7]](#endnote-7)].

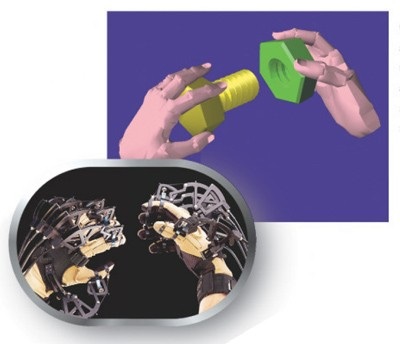


Figure 1-1. Immersion CyberGrasp data glove

The hand gesture recognition based on wearable technology has the advantages of high accuracy, simple data and fast processing speed, etc., but because of the shortcomings of expensive equipment, inconvenient operation, and not suitable for long-distance control, this kind of interaction model is hard to get promotion. Vision based gesture recognition takes use of the camera to capture the image sequence, and through the image processing and analysis to get further gesture recognition. The greatest advantage of vision based hand gesture recognition is that the input is simple with lower dependence on equipment, and it is in line with the people's daily interaction, therefore, vision based hand gesture interaction is bound to be the pursuit of human-computer interaction.

## 1.2 Research status

### 1.2.1 Review of Literature

On the algorithm of hand gesture recognition, researchers have proposed a lot of different solutions. At present, the most commonly used algorithms include HMM (Hidden Markov Model) model based on statistics [[[8]](#endnote-8)] and algorithm based on genetic algorithm [[[9]](#endnote-9)] and artificial neural network [[[10]](#endnote-10)].

Statistics based HMM takes advantage of causal relationship between visual features obtained from prior knowledge to deal with the inherent problem of uncertainty in video processing, not only can build the probabilistic modeling of dependencies among different features corresponding to multiple random variables in every moment, but also consider the transition probability between every moment, which can well reflect the timing relationships between features. However, it needs to maintain a certain size of the sample library and ask for large computational quantity, even though the larger the size of the sample library, the closer to the actual situation and the higher the accuracy of hand gesture recognition will be, moreover, this method also needs to use data smoothing technology to enlarge the value of small probability. The genetic algorithm is used to discretize images, control the discrete points and then convert image recognition problems into combinatorial optimization problems of a series of discrete points. But it is not able to take use of the feedback information from network in time, in addition, this method is troubled with slow search speed, large training sample and long training time. Artificial neural network has a large number of simple processing units (neurons) which are widely connected to form a complex information processing network, it mimics the function of information processing, storage and retrieval of human brain neural system in a certain extent and level, its processing efficiency is high with small samples can be satisfied, but the training process needs the participation of human, and the accuracy of recognition is influenced by the subjective factors.

Many techniques on HOG (Histogram of Gradient) like [[[11]](#endnote-11)] [[[12]](#endnote-12)] have been proposed in the past which employ edge and gradient based descriptors for hand gesture recognition. But they are only able to detect hand gestures in a simple background and are liable to fail when the background is cluttered. Literature [[[13]](#endnote-13)] proposed the Tortoise model to characterize the basic characteristics of the human hands, trained self-learning gesture pattern database based on the model, combined with the rule of survival of the fittest in the genetic algorithm to achieve the goal of hand shape and pattern matching, hand gesture recognition in the mixed feature space of geometry and texture, it would improve the real-time but only 10 kinds of gestures with large difference from each other were chosen in the experiment carried out. Paper [[[14]](#endnote-14)] extracted the edge pixels of hand gesture, took use of the idea of model-based matching using Hausdorff distance to realize the recognition of Chinese alphabet, the method proposed had the advantages of small computation and strong adaptability and disadvantages of ignoring the situation of rotation, scaling and skin color interfere. [[[15]](#endnote-15)] used the shape feature of hand gesture to do the recognition, due to the limitation of contour feature, it could not able to recognize gestures which are very close to. [[[16]](#endnote-16)] achieved the retrieval and recognition of binary images by using density distribution feature, but it had no idea to distinguish similar gestures. In [[[17]](#endnote-17)] the authors putted up with a method based on hand characteristic curves, the result of combination of color, motion and edge information, this method can reduce the dependence on hand segmentation, but the computation is too complex, and the real-time performance is too poor. Lars and his partners [[[18]](#endnote-18)] selected color as feature, detected the palm and fingers using particle filter, and then used template matching method for hand gesture recognition.

In addition, there are many other research works that successfully harvest some certain achievement of hand gesture recognition. Paper [[[19]](#endnote-19)] used the maximum difference feature to classify the gestures after the segmentation of MDF (The Most Discriminating Feature) space, the algorithm can adapt to the occasion with complex background. In [[[20]](#endnote-20)] an inductive learning system was introduced, this system could extract rules from DNF (Disjunctive Normal Form), and it obtained the recognition rate of 94%. The elastic curve matching method introduced by paper [[[21]](#endnote-21)] was less independent on segmentation, and could reach the recognition rate of 85% in a complex background. Bjorn et al. [[[22]](#endnote-22)] used color and motion features to detect and track the hand, combined the method of template matching and nearest neighbor classifier to recognize hand gesture. In [[[23]](#endnote-23)] the authors divided the hand into 21 different regions and train a SVM classifier to model the joint distribution of these regions for various hand gestures so as to classify the gestures.

### 1.2.2 Application and foreground

In the field of new human-computer interaction, there are many systems that have already been put into use in the world.

In December 2003, the Cybernet company located in Michigan, USA, introduced the world’s first fully functional weather map management system called “GestureStorm” that utilizes both body tracking and gesture recognition technology, with which the meteorologist can control the pace of the visuals and even incorporate spontaneous close-ups with simple hand movements that seem natural to the viewing public. Canesta Company launched a new device in late 2004 as shown in figure 1-2, it can ensure the personal digital assistant (PDA) with the function of gesture recognition. This device can project the image of keyboard on the desktop through the internal lens of PDA, in the meanwhile, it will also shoot a beam of red light over the “keyboard”, by detecting the interval time consumed by the infrared pulse after leaving transmitter, being rebound by user’s finger and returning back to the PDA sensor eventually, this device can accurately sense the position of a user’s finger at any moment. The consumed time of infrared time corresponds to a certain distance, and according to these distances the three-dimensional image of the position of the fingers on the keyboard can be drawn, so that PDA can accurately get the operating information of user’s finger on keyboard. A research team in Georgia Institute of Technology developed a device named as “Gesture Panel” to replace the instrument panel device usually used on automobile, with which, the driver only needs to make a gesture in the designed area, can adjust the temperature or sound volume of the car, and do not need to transfer the attention of the road. This device based on hand gesture recognition can make good contribution to reducing traffic accidents.



Figure 1-2. Canesta virtual keyboard

In the aspect of vision based hand gesture recognition, the representative research results include: the Fujitsu laboratory completed the identification of 46 sign languages symbols in 1991 [[[24]](#endnote-24)]. J. Davis and M. Shah [[[25]](#endnote-25)] used gestures captured from visual gloves with bright marks on the part of fingertips as the input of system, this system can recognize 7 kind of hand gestures. Starner et al. achieved a recognition rate of 99.2% on the recognition of short sentences composed of 40 words with part of speech in American Sign Language. K. Grobel and M. Assam extracted features from video, and used HMM to recognize 262 isolated words, the recognition rate reached to 91.3% [[[26]](#endnote-26)] [[[27]](#endnote-27)]. Also, Vogler [[[28]](#endnote-28)] and Metaxas applied these two methods to the recognition of American Sign Language, by using a position tracker and three mutually perpendicular cameras as the gesture input device they completed the recognition of 53 isolated words with recognition rate of 89.9%.

On the side of method based on sensor dependent devices such as data gloves, Christopher Lee and Xu in CMU developed a gesture control system for manipulator operation [[[29]](#endnote-29)]. M. W. Kadous [[[30]](#endnote-30)] took use of Power Gloves as the sign language input device to identify the vocabulary set composed of 95 isolated words, and the recognition rate was up to 80%.

Gesture is a relatively stable expression system which is composed of hand motion and expressions, it is a special language of communication by action and vision. It is also a kind of body languages with largest amount of information, its expression ability is similar to those natural languages such as spoken language and written language, so hand gestures can be used as a means of human-computer interaction, and it is more vivid, intuitive and has a very strong visual effect. Study on hand gesture recognition is not only useful to improve and enhance the studying and working conditions of deaf peoples, to provide them with a better quality of service, but also can apply to computer aided sign language teaching, TV program bilingual broadcast, virtual human research, special effects production in film, animation production, medical research, games, entertainment, and etc. The main application fields of gesture recognition are as following:

1. Education and life improvement for children, the elderly and deaf-mutes. Through some human-computer interfaces, can complete the natural communication between children, the elderly or deaf-mutes and the computers, and thus improve their ability of education. In addition, by using hand gesture recognition based human-computer interfaces, a communication channel between the normal peoples and deaf-mutes can be established, so that normal peoples can “listen” and understand the “words” of deaf-mutes. Until now, there are a series of hand gesture recognition systems come out in the aspect of deaf-mutes’ education.
2. Environment and virtual objects control in virtual reality environment. Use hands to complete the browsing, selection and manipulation of the virtual environment on the operation interface. Use different definitions of gestures to control the movement and turning of virtual objects, or through the movement of real hands to control the movement of the mirror images of hands in augmented real environment.
3. Application on smart home appliances and control field. Among the computer controlled means, hand is regarded as a flexible and efficient controlling way, it helps advance the research work of robot control and remote robot operation. For the operation in some special occasion in where it is inconvenient for direct manipulation, the requirement for a natural human-computer interface is more urgent. Until now, application on hand gesture in the field of control has also obtained some results, such as the vehicle control system based on hand gesture recognition introduced by Jaguar company [[[31]](#endnote-31)] and video cameras controlled by gesture command such as “zoom”, “panoramic” and “tilt”.
4. Demonstration and study of robot. By studying the mechanism of human visual language from the perspective of cognitive science to improve the human language understanding ability of computer.

### 1.2.3 Difficulties in research

At present, although the study of hand gesture recognition has made great progress and achieved high recognition rate in different areas, it still faces many challenges, such as: the extraction of gesture invariant features, transition model between gestures, minimal recognition units of sign language recognition, automatic segmentation of recognition units, recognition approach with scalability about vocabularies, auxiliary information for hand gesture recognition, signer independent gesture recognition, mixed gestures recognition.

Thereinto, the static gesture recognition based on vision is the current trend of hand gesture recognition and mainly has the following two aspects of technical difficulties:

1. Difficulties in target detection

Target detection is to capture the target from the image stream in a complex background, which is to extract the object of interest. In vision based hand gesture recognition methods, it is always a difficult problem to divide the human hand area and other background areas in the image, this is mainly due to the variety of background and unforeseen environment factors.

2. Difficulties in target recognition

Hand gesture recognition is to explain the high-level implications contained according to the posture and changing process of hand. In view of the following features of hand gesture, the key technology of hand gesture recognition is to extract the geometric invariant features.

1) Hand is an elastic object, there may big differences exist between same gestures and high similarity between different gestures. Human hand has more than 20 degrees of freedom, its movement is very flexible and complex. Therefore, same gestures made by different people may vary, and gestures made by the same person at different time or place may also different.

2) Hand has a lot of redundant information, since the key part of hand gesture recognition is to identify finger features, so palm feature is one of the redundant information.

3) The position of hand refers to the projection of hand from three-dimensional space to two-dimensional, so the projection direction is very important.

4) It is easy to produce shadows due to the non-smooth surface of hand.

Because of these characteristics of hand, the above problems have not been solved well yet, so it is necessary to add some restrictions when doing hand gesture recognition.

## 1.3 Organizational structure of dissertation

On the basis of existing hand gesture recognition technologies, this paper introduced an algorithm for static gesture recognition based on contour and spatial density distribution features of gesture image. In this work, we carried out the analysis and research on vision based gesture algorithm from segmentation and recognition two main aspects. The organizational structure of this paper is arranged as following:

Chapter 1 described the background and significance of hand gesture recognition, reviewed the previous researches and analyzed the current research status and technical barriers.

Chapter 2 introduced the basic concept and classification of hand gesture recognition, and made a brief introduction to the procedure of vision based hand gesture recognition.

Chapter 3 and 4 focused on the keystone of this paper. Chapter 3 through skin color detection and regional segmentation to get hand gesture part, this process mainly refers to color space conversion according to the principle of color balance and threshold segmentation. Chapter 4 calculated the convexity of hand gesture using convex defect detection, and extracted the spatial density distribution features from two aspects of overall and local attitude. Chapter 5 showed the similarity measurement method for hand gesture classification, and chapter 6 presented the study result and conclusion.

# II. Related Work

## 2.1 Hand gesture recognition overview

### 2.1.1 Definition of hand gesture

Because the diversity, polysemy and differences in time and space of hand gesture, and also because the influence of different cultural background, the definition of gesture is also varied. Hand gesture, in a broad sense, is the movement of hand made by people consciously (including the bending and stretching of fingers, rotation of wrist, movement of hands in space), whether it is used to operate a tool or an object, to perform a task or to communicate, hand gesture is always capable of expressing a signer’s intention. From the point of view of hand gesture recognition, hand gesture is defined as: a variety of gestures or movements produced by hands or arms combined.

Hand gesture recognition is to identify the meaning that signer want to express based on gestures made. With human hand as the input device of computer, the communication between human and machine will no longer needs special medium, users can simply define a series of appropriate gestures to control surrounding machines. Compared with other input methods, hand gesture recognition has the following characteristic [[[32]](#endnote-32)]:

Nature: hand gesture is a kind of natural, intuitive and easy-learning human-computer interaction model with human hands as input.

Simplicity and richness: a single gesture can indicate one command or one parameter, and for continuous ones, the gestures and movements of hands and fingers can provide more abundant information to help the understanding of higher level gestures.

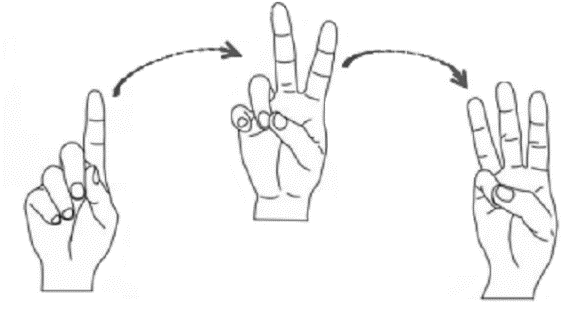
Directness: with human hand as the direct input of computer, it will free the intermediate medium between human and computers, users can just simply define one or some appropriate gestures to control the machine.

### 2.1.2 Classification of hand gesture recognition

There are many classification methods for hand gesture recognition, different literatures have different classification methods, and they are mainly include:

1) Classification based on real variable characteristic

Gesture implementation is a dynamic process, its characteristics are manifested in the changing of hand shape caused by fingers bending, and position and orientation’s changing of hand in space, so the description of hand gesture needs to be considered from two aspects of time and space. According to the real variable characteristic, hand gestures can be divided into static and dynamic ones. As shown in figure 2-1, a static gesture (a) refers to the space attitude of hand at a certain point of time, includes shape, orientation and relative position to the body, a dynamic gesture (b) is a sequence of attitudes in a period of time. Static gesture corresponds to a point in space, and dynamic gesture corresponds to a trajectory in the space of model parameters which needs time-varying spatial characteristics as descriptors. The recognition methods of these two kinds of gestures are also different [[[33]](#endnote-33)] [[[34]](#endnote-34)].

(a) Static gesture (b) Dynamic gesture

Figure 2-1. The static gesture and dynamic gesture

2) Classification based on expressive meaning

Figure 2-2 shows the classification of hand gestures based on expressive meaning proposed by psychologist Quek and Pavlovic in last 1940s. Quek Pavlovic defined the movement of hand to two cases, meaningless action and meaningful action with the ability of user's intention conveyance. The second case refers to the gesture, and gestures can also be divided into operation and communication ones. Operation gesture dose not express any information but just performs tasks such as grasping, gripping and playing pianos, communication gesture is also called interactive gesture and composed with action gesture and sign gesture. Action gesture refers to hand action such as imitation gesture (e.g. action imitation) and pointing gesture, sign gesture such as language used by deaf-mutes has the function of language description and contains instruction gesture (i.e., direct instructions, such as "you, me, him, head, eye, nose") and modal gesture. In the process of human-computer interaction, it is required to agree on the definition of gestures, that is, to define the interactive gesture set and to separate the meaningless actions and meaningful actions from.

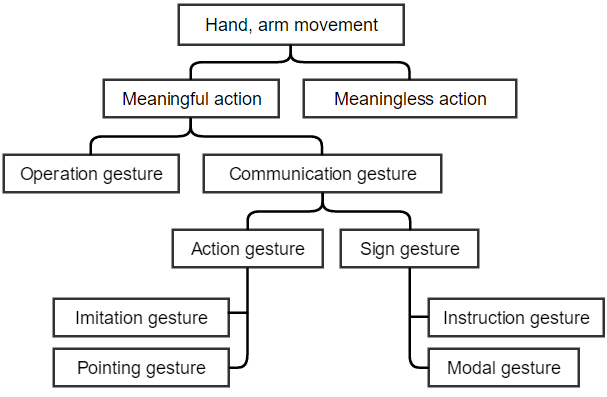


Figure 2-2. Classification of hand gestures based on expressive meaning

3) Classification based on input equipment

Depending on the different ways of gesture image acquisition, hand gesture recognition can be divided into sensor included acceleration hardware based and computer vision based ones.

Figure 2-3 shows some sensor included acceleration hardware assisted in hand gesture recognition, which include mouse, pen (a) and data glove (b), etc. By taking advantage of sole dependence on software algorithm, the recognition system based on mouse or pen is applicable to general desktop systems, but its shortcoming is that only the overall movement of hand can be recognized but not the movement of fingers, only when some movement or orientation change of mouse cursor or nib occur can regard them as gesture expression tools, this kind of hand gesture recognition system is more used in the application of text-proofing. The gesture recognition system based on data gloves takes use of data gloves and position tracker to detect the motion trajectory and timing information of gesture in space. The advantage of this method is less data quantity requirement, high processing speed and recognition rate, it can get the 3D information of hand and finger gestures directly, and can do real-time identification, however, due to the limitation of complex data gloves and position tracker, the gesture recognition system based on data gloves is expensive and difficult to be popularized.



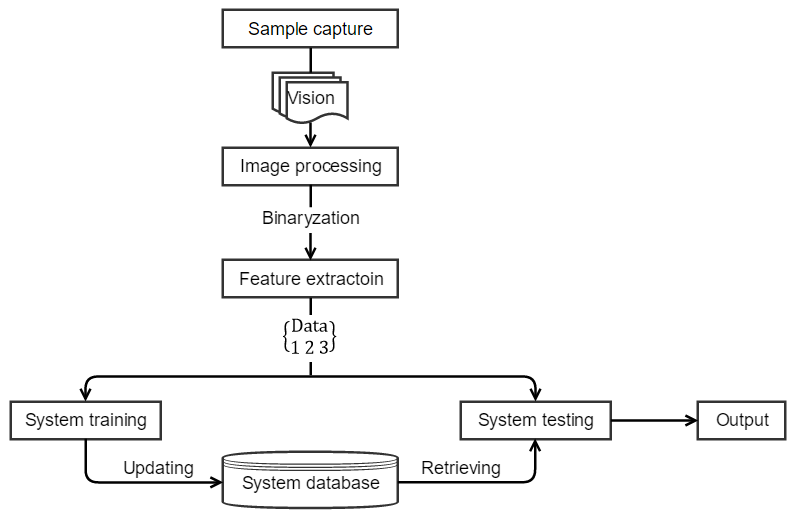
(a) Mouse and pen (b) Data glove

Figure 2-3. Hardware based data acquisition equipment of hand gesture

The gesture recognition system based on visual information uses the camera to collect continuous 2D image sequence, through the extraction and analysis of the image to perform gesture recognition, the biggest advantage of this method is that the input is simple and easy to operate, does not rely on device and interfere with uses. Computer vision based hand recognition systems include tag-based ones and unlabeled ones [[[35]](#endnote-35)]. By making marks in the hand, such as stick or draw dots with special color to achieve gesture recognition by tag-based hand gesture recognition system, although this approach brings more convenience than hardware based systems, but also brings trouble to experimenters. Unlabeled hand gesture recognition system uses human hand as natural input directly, by extracting skin color, shape features to identify gestures, its greatest strength is that the input device is relatively cheap and little user restrictions, human hands are in natural state so that people can interact with machines in a natural way, and because of this natural and intuitive interactive way, it becomes the trend and goal of the development of gesture recognition technology. However, it is difficult to extract the original gesture parts from visual information completely, therefore, its identifiable gesture set is relatively small, has low recognition rate and poor real time performance as well, but these shortcoming are also the reasons why vision based hand gesture recognition system has become a hot research topic in recent years.

## 2.2 Vision based hand gesture recognition

### 2.2.1 General flowchart of vision based hand gesture recognition



Classification

Figure 2-4. Architecture of hand gesture recognition

The general flowchart of vision based hand gesture recognition as shown in figure 2-4, mainly includes: sample images capturing, image preprocessing, feature analysis and identification parameters extraction, classification and recognition four parts [[[36]](#endnote-36)].

1) Sample image capturing: the inputting hand gesture image can be acquired from camera(s), videos or even data glove instrumented devices.

2) Image preprocessing: mainly includes image space conversion, image binarization, filtering denoising and threshold segmentation.

3) Image feature extraction: features are the most useful information that leads to classification, some features have properties of invariance with no leading to misclassification by rotation, scale, translation or illumination.

4) Classification and recognition: in view of the extracted parameters, composites a set of feature vectors to do training and classification jobs using some special recognition algorithm. In academic researches, the recognition results are mostly stored in the form of data in text file, and then calculate the recognition rate of each kind of gestures accordingly.

By inputting hand gestures to system using camera equipment, and after detection, tracking, recognition of extracted image sequence to understand and describe its behavior. When one or more cameras obtain the video stream of user gestures, the system will monitor whether there are hand gestures contained in the data stream according to the interactive mode of gesture, if there are, separates them. Then choose appropriate gesture mode to detect and extract features, and choose appropriate classifier to recognize the gesture in current image. A static gesture only needs to recognize one image and extract features without motion trajectory information, while in the recognition system of dynamic gestures, we need to identify the sequence of gestures and track and split the next frame of data after processing the current ones until the complete of whole series of gestures.

### 2.2.2 Sample image capturing

Gesture recognition is generally divided into static gesture recognition and dynamic gesture recognition, static gesture recognition is the recognition of hand shape, read out the meaning of hand expression, and dynamic hand gesture recognition is the recognition of hand motion trajectory in space, and then perform the corresponding operation based on obtained trajectory parameters, such as playing courseware on the projection, hand gestured can be used to flip up and down, pause, start, etc.

Image capturing is to acquire an image (static gesture) or a sequence of images (dynamic gesture), which is then processed next. This process is mostly done using a single camera with a frontal view of human hand performing some gestures. There are also systems use multiple cameras in order to get more detailed information of hand gesture [[[37]](#endnote-37)] [[[38]](#endnote-38)], this kind of system has the advantage of allowing the recognition of occluded gestures benefited from the scene captured by the other camera from another perspective. There is another system with the camera mounted on a hat presented in [[[39]](#endnote-39)], it is only set to capture the area in front of the hat wearer, the main advantage of this system is that the position of camera is always adapted however the people moves or turns around.

Sample image capturing and dataset establishment are very important for the study of image processing and recognition, if the captured images can well adapt to background, illumination, distance, angle or other objective factors, it will be much easier and more accurate for the extraction of features describing the hand gesture. Although there are several systems that seem to reliable on complex background [[[40]](#endnote-40)], however, the image capturing is often performed in a cleaned up environment with a uniform background [[[41]](#endnote-41)].

### 2.2.3 Image preprocessing

The basic aim of image preprocessing is to optimally process the sample image, extract the binary image and prepare for the feature extraction next. As shown in figure 2-5, the hand gesture image preprocessing is mainly divided into color space conversion, noise filtering and threshold segmentation.

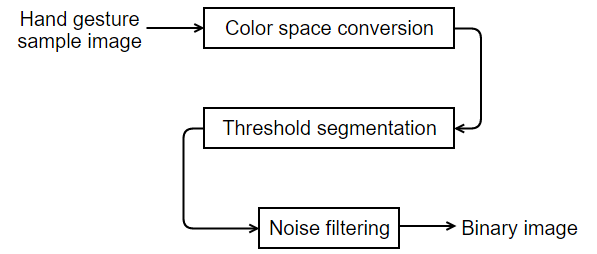


Figure 2-5. The process of hand gesture image preprocessing

Segmentation is the most important step of image preprocessing, it divides the input image into regions separated by boundaries. Dynamic gestures need to be located and tracked and static gestures only have to be segmented [[[42]](#endnote-42)].

Since human skin is easy and invariant to the changing of scale, translation and rotation, it is common used for segmentation [[[43]](#endnote-43)]. The process of hand gesture image segmentation generally includes: creating a probabilistic model of skin color to calculate the probability of each pixel, finding the interested coarse regions with comparing with the threshold value, more analysis and filtering being carried out for example involve the size or perimeter of the located regions in order to exclude noisy regions such as faces.

### 2.2.4 Feature extraction

Good image preprocessing leads to perfect features extraction result and play a significance role in a successful classification and recognition process next. The aim of feature extraction is to find and extract features that can be used to determine the meaning of a given gesture, a feature or a set of features should have the abilities of describing the gesture uniquely and be robust to the shift and rotation of hand gesture in order to achieve a reliable recognition.

What features should be extracted will vary from different gesture models, and has a direct relationship with the purpose and method of recognition. According to the requirement of hand gesture recognition, the features extracted from the hand image should not only be able to maintain a good non deformation in the same type of gesture, but also can distinguish different types. Various methods can be applied for representing the features that can be extracted, the commonly used features of static hand gestures include: gradient histogram, image subspace projection, shape features, etc. [[[44]](#endnote-44)]. The traditional gradient histogram is easy to calculate and implement, it has the invariance of translation but not rotation, with same gesture image, its histogram will be different after rotation. In addition, the gradient histogram does not have the property of uniqueness, that is, different gesture images may have similar histogram. Image subspace projection (e.g. PCA, ICA) is a kind of statistical signal processing technique, this method is able to removing the correlation of higher-order statistics and making relatively comprehensive representation of the local features of training sample images. However, the feature invariance of this method is acquired in the training of a large number of samples, that is to say, once the training samples are not able to cover all the position, scale and rotation angle, such method cannot achieve the extraction of invariant features. Moreover, most of this kind of algorithms belong to unsupervised algorithms, and pay more attention to obtain the features with less error, so that cannot guarantee the best classification performance of features.

Some methods used hand shape such as hand contour and silhouette while other utilized fingertips position, palm center, etc. Shape based method has fully considered the invariance of translation, size and rotation in the process of features extraction, so it is the most commonly used feature extraction algorithm for static gestures currently. Thereinto, method relies on the outline of a given hand region is a simpler one, by using some edge tracking algorithm to extract the outline of hand, and then there will be two different kind of extreme of the outline: the peaks and valleys. The peaks are usually the cue to find out finger tips and valleys to regions where two fingers join the palm of the hand.

### 2.2.5 Classification and recognition

The classification represents the task of assigning the extracted features to some predefined classes in order to recognize the hand gesture. The classification mainly aims to find the best matching reference features of the features extracted, the recognition process is affected by the proper selection of features and suitable classification algorithm [[[45]](#endnote-45)]. Figure 2-6 shows the basic recognition process, before recognition, the system should be trained with enough data so that a new feature vector can be classified with good accuracy [[[46]](#endnote-46)].

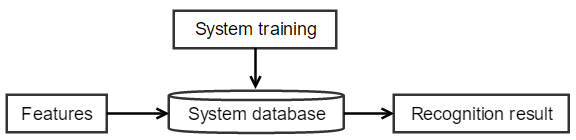


Figure 2-6. Gesture recognition process

The most commonly used hand gesture recognition techniques are shown as following:

1) Template matching: template matching is one of the simplest recognition methods, it matches the inputting gesture with each standard one, by measuring the similarity between to accomplish recognition.

2) Artificial neural network (ANN): through widely connecting a large number of small processing units (neurons) together to constitute a complex information processing network, ANN imitates the functions of information processing, storage and retrieval of human brain in varying degrees, it has the abilities of learning, memorizing and calculation.

3) Hidden Markov model (HMM): HMM has been successfully applied to some continuous recognition systems such as speech and handwriting, and has been the mainstream approach in the field of dynamic identification. Using HMM to model forms of sign languages makes it possible to deal with the high randomness of hand gestures.

4) Geometric features based recognition: recognition techniques based on geometric features take advantage of edge and regional features as recognition factors, the specific implementation depends on the specific significance of hand gesture actually.

# III. Hand gesture segmentation

## 3.1 Skin color detection

### 3.1.1 Color balance

In the skin color based hand gesture detection methods, skin color model is usually established under standard light source, however, this kind of skin color model cannot correctly detect the skin area when the light source of image deviates from the standard light source, which then will affect the follow-up works of hand gesture recognition. With the method of color balance the problem of color deviation can be better solved.

The basic aim of color balance is mainly to eliminate the color deviation problem when doing skin region detection under colored light source. The process of color balance is to determine the illumination intensity of scene, adjust the value of R, G, B, and restore the original color property of the image scene. Since it is difficult to determine the illumination condition for most images, so here we use “Gray World” assumption method [[[47]](#endnote-47)] to correct color.

“Gray World” is an assumption based on the composition of image natural color, it is assumed that the average value of the three components R, G and B of an image with rich color tends to be the same gray value. In the physical sense, “Gray World” assumes that the mean value of the reflection of light in nature is a fixed value in general, and this value is approximately as “gray”. By enforcing this assumption to the implementation of image processing, color balance can eliminate the influence of ambient light from the image, and restore the original scene of image.

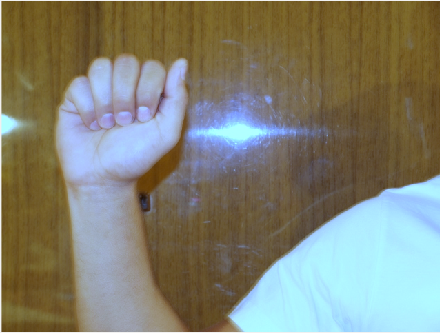
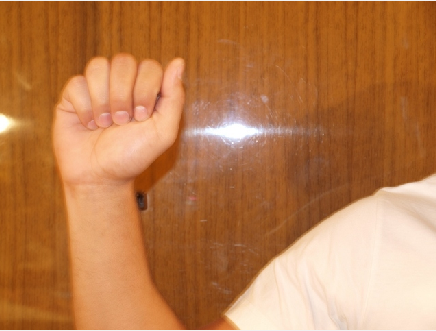
The calculation steps of color balance are as follows:

1) Calculate the respective mean and total mean of R, G, B of hand gesture image, where

2) Calculate the adjustment coefficient of each component, let

3) Adjust the values of R, G, and B with adjustment coefficients:

4) Change values that greater than 255 to 255.



(a) Original image (b) Image after color balance

Figure 3-1. Processing effect of color balance

Figure 3-1 (a) shows an original hand gesture image with color deviation, the color of the whole image seems reddish, and (b) is the image after color balance, we can see that after color balance the hand skin color has almost been recovered and can be used for skin detection.

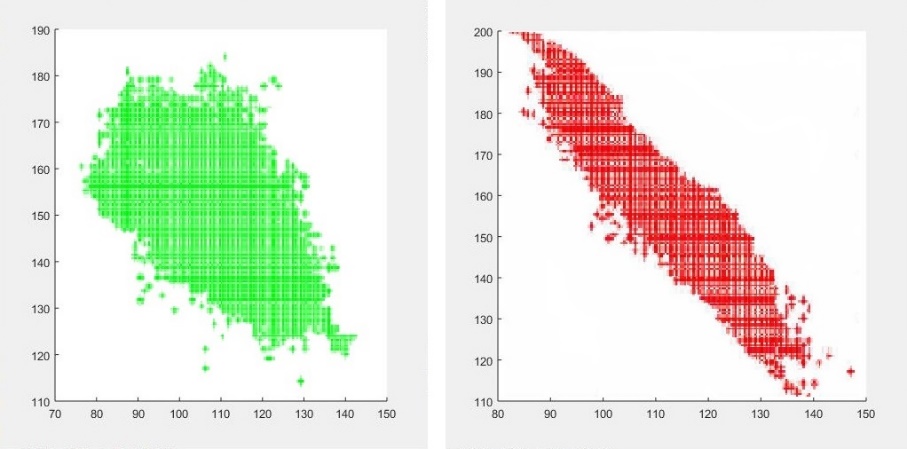
### 3.1.2 YCgCr color space conversion

Color contains more information than gray, and it will not change with the changing of direction, size of hand gestures, so using skin color information to detection skin area in color images is a fast and effective method. At present, there are a lot of color spaces which can be applied to skin detection and segmentation, such as normalized RGB [[[48]](#endnote-48)], YCbCr [[[49]](#endnote-49)], YUV [[[50]](#endnote-50)], HSV [[[51]](#endnote-51)], HIS/GIHS [[[52]](#endnote-52)] color space, etc.

The most commonly used color space is RGB color space, in RGB color space, the color is composed of three components which are R, G and B, these three components all contain the luminance information and have great correlation between each other, and this correlation is not conductive to the detection and segmentation of skin color. Some hand gesture detection algorithms based on skin color adopt normalized RGB color space, but the normalized RGB color space only can remove the relative part of brightness in R, G, and B but still has brightness information left, so the brightness adaptability of this kind of color space is not ideal for skin color detection. Therefore, it has important significance to choose a suitable color space for skin color detection, and one way is to convert the RGB color space represented image to color spaces that have brightness and chrominance value independent of each other.

YCbCr color space has the similar advantage of separating the luminance component from image color with HIS and some other color spaces, but better than HIS and other color formats, the calculation process and space coordinate form of YCbCr color space is relatively simple. YCbCr color space is consistent with human visual perception, and has the characteristic of good color clustering and color mutual independence. Considering the YCbCr color space, a human skin color model can be considered practically independent on the luminance and concentrate in a small region of the Cb-Cr plane. In the YCbCr color space, Y is the luminance component and Cb and Cr are the blue-difference and red-difference chroma components [[[53]](#endnote-53)]. Because of the blue-difference (B-Y) in skin color is relatively small, so the YCbCr color space is often used for the building of skin color model.

YCgCr color space was derived from YCbCr color space that used the color difference G-Y instead of B-Y. Since the skin color clustering effect of YCgCr space is better than YCbCr space, this paper will use the new color space YCgCr, which is based on and similar to YCbCr color space. Firstly, YCgCr color space is not sensitive to the changing of illumination. Secondly, the Y channel on behalf of the brightness information, if there is somewhere needs to extract the features of gray image, it can be directly carried out on the Y channel, which can reduce the amount of computation. Finally, [[[54]](#endnote-54)] pointed out that the YCgCr color space has stronger skin color clustering effect, shown as figure 3-2, it found that a 2D projection of YCgCr skin color into Cg-Cr subspace, in which the skin color clustering effect of Cg-Cr color space (a) is better than Cb-Cr color space (b), and the Cg and Cr components can effectively distinguish skin and non-skin color.



(a) (b)

Figure 3-2. Skin color clustering region in YCbCr (a) and YCgCr (b)

YCgCr color space and RGB color space conversion formula can be derived from the YCbCr color space and RGB color space conversion formula [[[55]](#endnote-55)], the conversion formula is as follows:

This paper uses fixed threshold method to determine skin areas, this method is simple, mature and easy to use. Zhang [[[56]](#endnote-56)] presents experiments of 1010 skin color pixels which are in different ages and body areas. Finally, he builds a parallelogram model in Cg-Cr color space for skin color detection, the model for the skin color in the transformed Cg-Cr is described as following:

When the image pixel in YCgCr color space satisfies the above model, it can be regarded as color pixel and reset to 255, otherwise, as non-skin pixel and reset to 0, thus, we can get the corresponding binary image.

## 3.2 Hand gesture region location

For images with complex background, it is inevitable to produce noisy and skin-like regions in the process of skin color segmentation. For the obtained binary image contained noise, we can remove the non-gesture regions and thus only keep the gesture region according to the following conditions.

1) Remove regions whose skin area is less than 400 pixels. Because the gesture features will be lose and difficult to ensure the detection of fingertips if the area is too small, so these parts should be removed as noised. The area is the number of target pixels, its calculation formula is:

Where,

represents the area, *w* is the width of region, *h* is the height of region, and *R* is the target binary region.

2) The width and height of skin region should be bigger than 20 pixels in order to avoid the mistaken extraction of some small regions that meet condition 1) but not human hands.

3) The palm and fingers should appear in the visual window.

4) It is known from 3) that the center of the gesture area should be inner skin region, that is, the center of gravity should locate inside of the palm. The calculation formula of center of gravity) is:

Where,

5) Calculate the aspect ratio of skin color region, because the aspect ratio of a normal person’s gesture is in a certain proportion range. Here we define the aspect ratio of skin region as σ:

Figure 3-3, 3-4, 3-5 are the segmentation results of hand gesture of images captured under different backgrounds, (a) are original images, (b) are results after converting the image from RGB to YCgCr color space, and the conversation will separate the illumination info of images to the component of Y and choma info to Cg and Cr, (c) are binary images with white areas detected as skin and black areas non-skin after binarization by thresholding, and (d) are final segmentation results with only hand gesture regions contained after denoising.



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

Figure 3-3. Gesture segmentation result under simple background

In figure 3-3, we can clearly get from (c) that this image (a) still contains two pieces of skin-like regions in its background even it seems to be captured under a pure black one.



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

Figure 3-4. Gesture segmentation result under complex background

Figure 3-4 is the segmentation result of an image with a more complex background, and it shows that even under a complex background, the hand gesture region can still be well located.



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

Figure 3-5. Gesture segmentation of image with noisy skin area contained

Figure 3-5 can prove that by adopting fixed thresholding method, the algorithm proposed by this paper can successfully remove those skin-like areas or noisy skin areas.

# IV. Hand gesture features extraction

## 4.1 Gesture convex defect detection

### 4.1.1 Convex defect and fingertip detection

In issues involving human hands such as sign language recognition and gesture recognition, fingertip is one of the most popular characteristics, because the number of fingertips can be considered to be the number of fingers and the direction of fingertips can effectively express the stretch information of fingers. Contour analysis is a commonly used method for fingertip detection, which achieves the location of fingertip based on geometric features of contour, such as the edge curvature method used in literature [[[57]](#endnote-57)] for contour detection, and the least square ellipse fitting method used in [[[58]](#endnote-58)] for fingertip detection, this kind of algorithms require high accuracy of contour and a large amount of computation, and are very dependent on the quality of gesture segmentation. In this paper, a method of fingertip detection based on convex defect is adopted.

We first explain some concepts of convex defect.

1) Contour

The contour of hand is a series of points which are the boundary pixels of the hand area. After obtaining the contour, the gesture and its shape then can be detected and recognized by using contours analysis. Contour is constituted by the connection of edges, common edge recognition algorithms contain Sobel, Canny, Prewitt, Roberts, and Fuzzy logic methods [[[59]](#endnote-59)]. Here we use Canny edge detection method to extract the contour of hand gesture, in figure 4-1, the red curve is the contour of hand gesture in the figure.

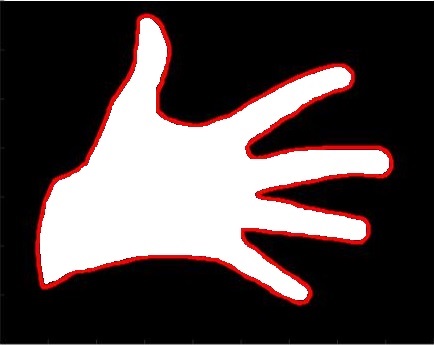
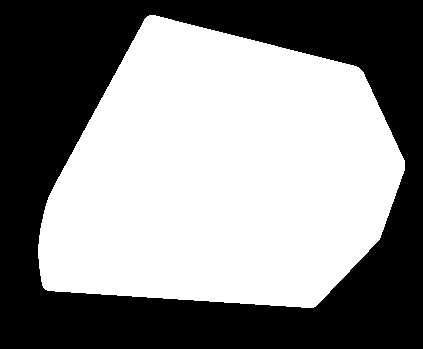
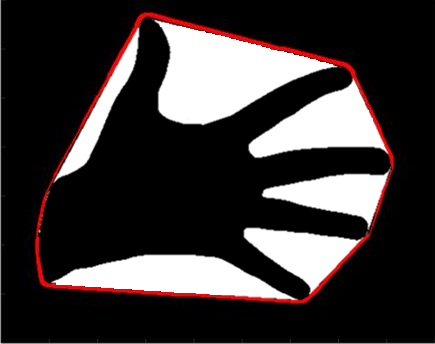


Figure 4-1. Detected contour of hand gesture

2) Convex hull

The convex hull of hand gesture contour is the convex polygon surrounded by all the convex vertices in gesture contour, as shown in figure 4-2 (a), the polygon composed by red curve is the convex hull of hand gesture in the figure, and (b) is the separated convex hull extracted from (a).



(a) (b)

Figure 4-2. Convex hull of hand gesture

3) Convex defect

The convex defect is defined as the difference between gesture convex hull and contour, they are contained in the convex hull but not hand area. As shown in figure 4-3, the white areas ①-⑥ are all the convex defects. The data structure of each of the convex defects contains three components: start contour point, end contour point and concave contour point. For example, for convex defect ②, P1 is its start contour point which is the starting point of the defect，P2 is its end contour point which is the termination point of the defect, and P3 is the concave point which is the furthest point away from the convex hull, and the furthest distance is the depth of convex defect.

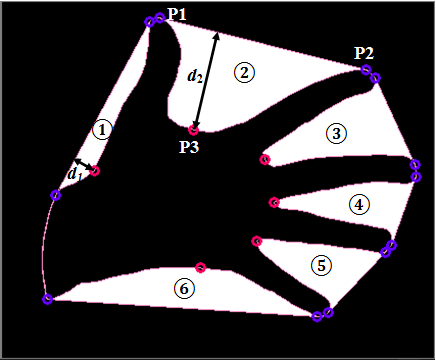


Figure 4-3. Convex defects of hand gesture

It shows from figure 4-3 that the fingertip is closely related to the convex defect, which is close to the start and end contour points of convex defect. Therefore, it is possible to detect the fingertips by using hand gesture contour and convex defects. The count and position of fingertips can be determined as following:

1) Conduct noise elimination on the obtained convex defects. The normalized depth of convex defect cannot be too small or too large, figure 4-3 shows that the depth of convex defects between fingers is more obvious than other convex defects, so the depth of convex defect can be used to distinguish whether a convex defect is a convex defect between fingers or not. Based on the physiological structure of human hand and large numbers of experiments, it is appropriate to define the depth minimum and maximum threshold as big as the 1/5 and 1/2 of the height of hand gesture contour respectively, and the height of hand gesture contour is defined as the length of the quadrilateral formed by points composed of the minimum -coordinate value, maximum -coordinate value, minimum -coordinate value and maximum -coordinate value, namely, points of , , , as shown in figure 4-4. In figure 4-3, is the depth meeting the condition of gesture convex defects while is not.

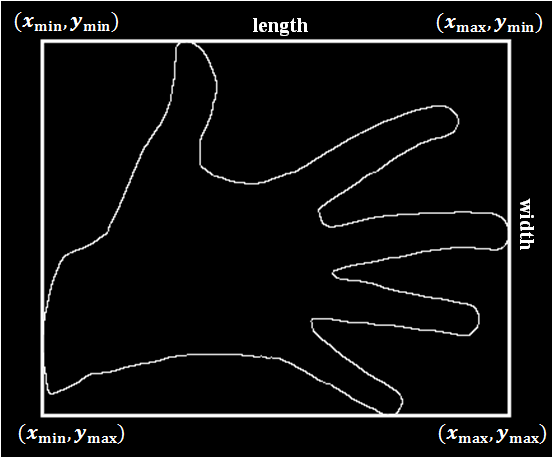


Figure 4-4. Definition of length and width of hand gesture contour

2) Scan filtered convex defects clockwise, take the start contour point of the first convex defect and the end contour point of the last convex defect as the first and last fingertip respectively.

3) Take use of the average position of the end contour point of current convex defect and the start contour point of next convex defect as the position of current fingertip.

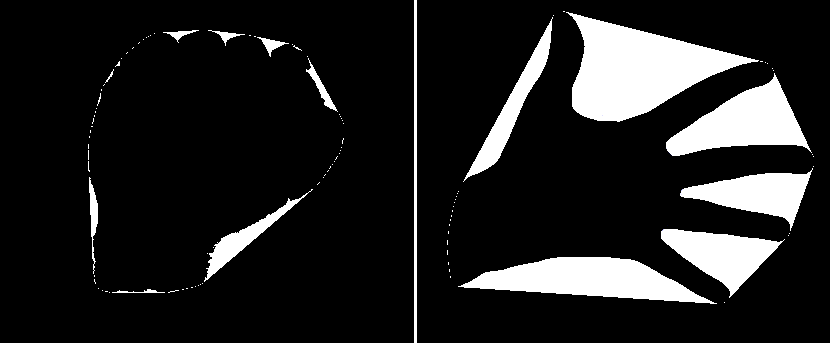
### 4.1.2 Gesture convexity and relative position of fingertips

For the binary image with convex defects, we can extract the convexity of gesture and relative position of fingertips as one of the features used for hand gesture recognition.

Through the observation and analysis of gesture contour and convex hull, we can get that with different gesture the tightness to its convex hull is also different, as shown in figure 4-5, the convex hull of the fist gesture in (a) almost contains the whole gesture contour, but the gesture contour in (b) has big difference with its convex hull, with several depression existing between.

The tightness of hand gesture contour to its convex hull is defined as the gesture convexity, which is denoted by δ, its value is the area ratio of gesture contour and convex hull.

Where, is the area of convex hull, is the area of gesture contour, we can get according to the image that , so , and the bigger the value, the tighter the gesture contour to convex hull.



(a) (b)

Figure 4-5. Tightness of different gestures to convex hulls

In addition, different gestures can be distinguished by the relative position of fingertips, which is composed of two values of *α* and *β*, *α* is the summation of angles with centroid as vertex, lines from centroid to the first fingertip and other fingertips as edges, and *β* is the value of the angle with centroid as vertex, lines from centroid to the first fingertip and last fingertip as edges, that is, *,* , where *N* is the number of fingertips, *θ* is the angle between the first fingertip and the other fingertip to the gesture centroid been considered as vertex. As shown in figure 4-6, different gestures have significantly different relative position of fingertips, so *α*, *β* can be used to the further recognition of different hand gestures.

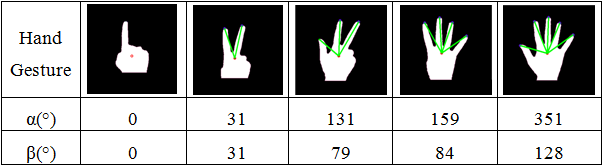


Figure 4-6. Relative position of fingertips of different hand gestures

The hand gesture recognition method based on the convex defect can greatly reduce the count of contour scanning and the amount of computation, but only the use of depth to determine the convex defect may cause the missed or mistaken detection of convex defect, for instance, when the hand gesture is not completely open, the convex defect depth between the adjacent fingers maybe too much small and be filtered out as noise, which could lead to large errors in the gesture recognition, so that the number of the fingertips cannot be correctly detected. So it needs the assistance of other features to more effectively describe the gesture.

## 4.2 Gesture distribution feature

### 4.2.1 Density distribution feature

The shape of an image is represented by the pixel dots distributed in different space regions. The alternative distribution of white pixels (target points) and black pixels (background points) constitutes a variety of target shapes of a binary image, so when describing the shape of a binary image, the regional distribution information of target pixels is very helpful since for two similar binary images their regional distribution of target pixels are also similar. The basic idea of density distribution feature (DDF) is to get the density distribution features through the statistic of distribution situation of target pixels in different regions, so it can be used in the recognition and classification of binary images.

DDF is a kind of feature that can reflect the spatial distribution of target pixels, and can be defined as

It contains two *M*–dimension vectors, the first vector represents the relative density of target pixels within each sub image region, and is the first order difference of in the direction of radial coordinates [[[60]](#endnote-60)].

The division process of the M sub image regions includes:

1) Calculate the maximum distance from target pixels to the centroid of the binary image, namely center of gravity, which can refer to Chapter 3.2.

2) With as center, as radius to draw the circumscribed circle of target region, and then divide the image to *M* sub image regions using the region division method of equidistant or equal areas [[[61]](#endnote-61)].

(a) Equidistance division method

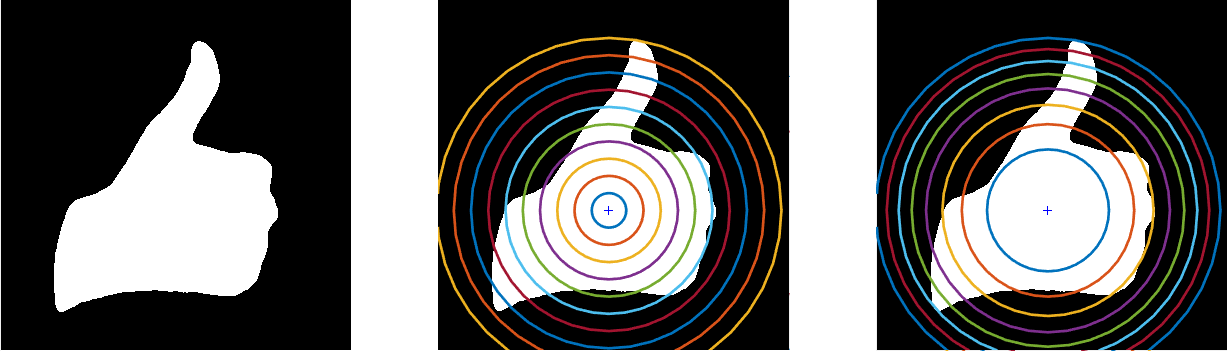
That is, the interval length covered by each sub image region is equal, which means:

(b) Equal area division method

That is, the area of each sub image region is equal, which means:

Where,.

Figure 4-7 is an example of region division of binary image, (a) is the original image, (b) is the division result using method of equidistant, the internal length (the width of annulus or the radius of circle) of each sub image region (annulus or center circle) is equal, (c) is the division result using method of equal area, and the area of each sub image area is equal.



(a) Original image (b) Equidistant region division (c) Equal area division

Figure 4-7. An instance of image region division

DDF can be regarded as a kind of effective shape feature that can well reflect the general shape information of a binary image, it has following virtues:

(1) DDF can grasp the overall shape info of image.

(2) DDF has the property of shift invariant. It benefits from the choice of gravity as centroid and taking care of target area only for the partition of sub image regions, the position of target in the image does not affect the value of DDF.

(3) DDF is not sensitive to scale deformation. Since the target area of all images are divided into same number of sub image regions, so the DDF of the new image obtained after the enlargement or reduction is basically same as the original one.

(4) DDF is invariant to rotation. Because of the adopting of circular division method, the change of rotation has little effect on DDF.

### 4.2.2 Improved gesture density distribution feature

Human hand is articulated complex deformable object consisting of one palm and 5 adjacent fingers, each finger is composed with several segments and joints. From a whole point of view, the gesture is a joint structure, with the movement of joints, the shape of the hand is constantly changing, and the different attitude of gesture can be described by means of the changing of spatial state of segments and joints.

In this paper, we will extract the spatial distribution feature from two aspects of the overall and local attitudes of hand gesture, and we name it as gesture distribution feature (GDF).

On the one hand, the different shape of gestures can be described by the distribution of pixels in the space, different images with similar gestures usually have similar distribution information as shown in figure 4-8. So we can extract DDF of hand gesture, that is, the distribution of skin pixels in different spatial regions as one of the basis of hand gesture recognition.

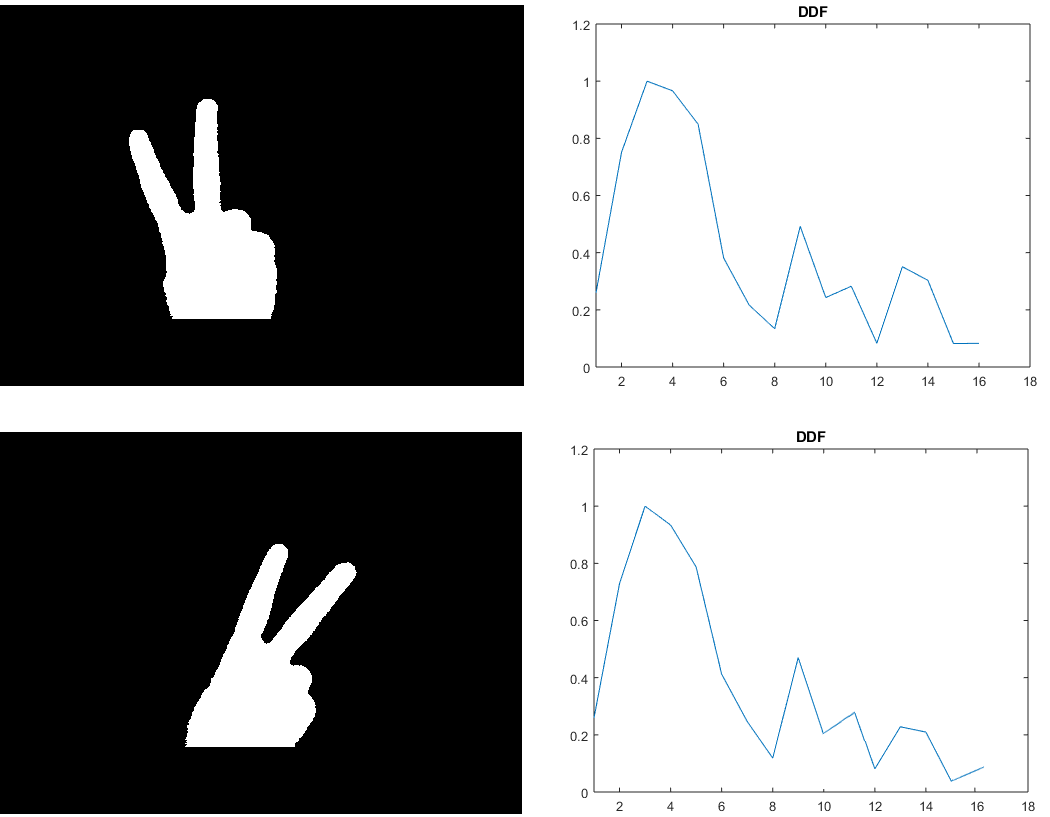


Figure 4-8. DDF of different images with similar gestures

On the other hand, the different shape of hand gesture is derived from local attitude of gesture, that is, connecting relationship of fingers such as joint angles, the selection of joint angles in different regions can well describe the changes in the joints of human hand. In this paper, we use normalized arc length (e.g. red arc shown in figure 4-9), as can be seen from figure 4-9, different have different arc length and arc segments.

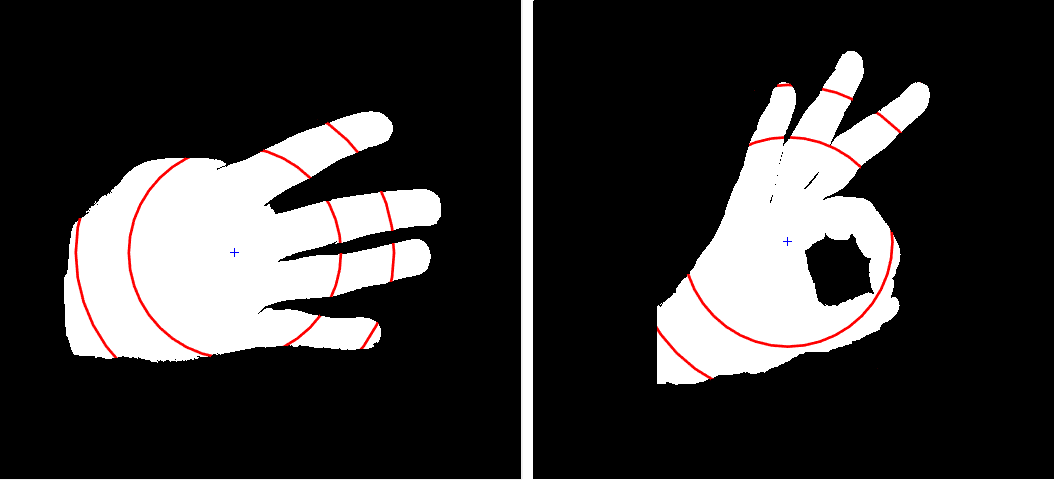


Figure 4-9. Sketch map of skin area region

The above features can be described using one feature vector:

.

Where, the first vector refers to the relative density of target pixels in each region after diving the image into *M* sub image regions, the second vector is the first order difference of in the direction of radial coordinates, and the third vector represents the arc length of skin region in each sub region outer edge.

As the third component of GDF, is mainly relies on the distribution feature of fingers, so when certain constraints are satisfied, only sub image regions that contain fingers can be sampled to reduce the computational complexity and improve the gesture recognition speed. According to the physiological structure of human hand, the sub image regions in which the distance from target pixels to the centroid of the binary image is bigger than but not equal to the half of maximum distance are sub regions that meet requirement. So, the improved GDF is expressed as:

.

The calculation steps of GDF are as follows:

1) Segment the binary image into sub image regions using the region division method of equidistant, which can refer to Chapter 4.1.1.

2) Conduct the statistics for each sub image region, that is, count the amount of target pixels in each sub region and find out the max value .

3) Calculate each component of the first feature vector of GDF:

4) Calculate each component of the second feature vector :

5) Conduct the statistics on the length of skin outer edge of each sub image region that contains fingers , and find out the max value .

6) Calculate each component of the third vector of GDF:

# V. Similarity measurement

By chapter 4, we obtained the feature vector used for the recognition of hand gesture in this paper, that includes the convexity of hand gesture, relative position of fingertips and gesture distribution feature, that is , *M* is the number of sub images of GDF, in order to classify different hand gestures to be recognized, we use Euclidean distance metric to measure the similarity distance between gestures.

First calculate the Euclidean distance of each sub feature, that is separately, and then compute the integrated distance of them, since each sub feature has its different feature space, it is necessary to normalize the corresponding distance before merging. Different from the normalization of each internal components of the feature, it is the normalization of the similarity distance of each sub feature, the Gaussian model [[[62]](#endnote-62)] is adopted to the normalization of similarity distance:

1) Calculate the corresponding Euclidean distance of each sub feature between the query image and each standard image in the image database.

Where, is the corresponding sub feature of the query image and standard image, and *n* is the dimension of the sub feature, such as for *δ*, *n*=1, and for *r*, *n=M*.

2) Calculate the mean value *μ*, standard deviation *σ*, and normalized distance.

The normalized distance of each sub feature is  respectively with range belongs to [0, 1].

3) Calculate the integrated distance of the similarity measurement.

Where, are weights of each sub feature, and, is the similarity distance between features of two hand gesture images, the smaller the value, the higher the similarity between gestures.

# VI. Result and conclusion

This paper focuses on the study of static hand gesture recognition based on monocular vision. Firstly, detect skin color areas in YCgCr color space and exclude non-human hand regions using prior knowledge; Secondly, take use of gesture contour and convex defect detection to detect fingertips, extract the first part of the features used for hand gesture recognition and classification, which includes the tightness of gesture contour to its convex hull, that is convexity, and the relative position of fingertips; Thirdly, in order to make up the limitation of gestures that not completely open of convex defect detection, based on the non-rigid characteristic of human hand, we extract the hand gesture distribution feature that includes spatial density distribution feature and arc length of skin area as another part of features used for hand gesture recognition and classification; Finally, identify the test hand gesture by measuring the similarity distance between the test one and standard ones in database.

In order to verify the accuracy of gesture location in YCgCr space, we carried out the color space conversion and hand gesture region location experiments under simple background and complex background respectively.

Table 6-1. Hand gesture location result

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

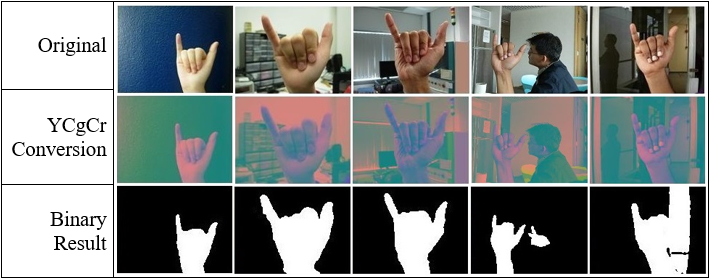
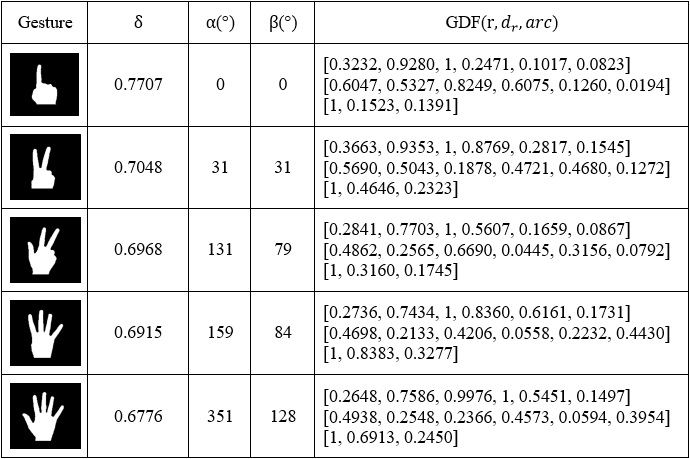


Table 6-1 shows the result that the YCgCr color space is more ideal for the detection of hand gestures under simple background (a) and complex background that does not contain skin-like regions (b) or contains skin-like but not hand-like regions (c), however, the binarization of hand gesture images under background that contains hand-like regions (d) or continuous skin-like regions (e) is less desirable, which is one of the main problems should be solved in the future.

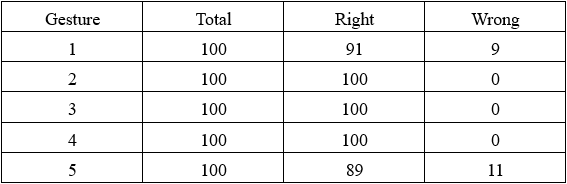
We also implemented experiments of 5 different kinds of number gestures (one to five) under simple background to show the recognition result of the hand gesture recognition algorithm proposed in this paper. Table 6-2 lists features adopted of some standard number gestures, it shows that these features have a certain ability to distinguish the number gesture templates used in this paper.

Table 6-2. Features used in this paper



In order to verify the desirability of recognition features used in this paper, we selected 20 representative images for each kind of gesture as query images for hand gesture recognition experiment, for each image, first remake it by translating, rotating and scaling in the experiment, thus, there would be total 100 images for each gesture in the query image database. Table 6-3 shows the recognition result.

Table 6-3. Recognition result of 5 number gestures with simple background



We can get from Table 6-3 that in the case of the ideal segmentation of the hand gesture region, the recognition accuracy of the static number gesture based on simple background can basically reach the expected standard. By analyzing the misidentified gestures in the experiment, we found that the cause of the error of gesture recognition is the difference in the length of the wrist compared with the corresponding standard gesture, for example, the nine misidentified hand gestures in series of gestures of number 1, the whist length contained in these gestures is longer than their responding standard gesture, and the wrist length contained in the 11 misidentified gestures in series of gestures of number 5 is shorter than their corresponding standard one. And this is one of the limitations of the hand gesture recognition algorithm used in this paper, which is another problem needs to be resolved in the future work.

In vision based hand gesture recognition approach, gesture recognition system composed of four stages, gesture capture, and isolation of region of gesture from the image, feature extraction and then gesture classification. Methods which are chosen in each stage highly affect the recognition performance.

Take advantage of the feature of not sensitive to light of YCgCr color space to convert the color space of hand gesture image, remove non-skin and non-hand regions and obtain binary image. Combined with the characteristics of multi-joint and non-rigid of human hands to extract the features based on convex defects and hand gesture distribution, the extract method is simple, easy to operate, and has strong practicability.

The convex defect detection aims to extract the hand gesture area size and relative position of fingertips, it is not influenced by the direction and position of the gestures, the extraction of hand distribution feature uses the center of gravity as centroid and performs the circular division based on the axial symmetry and central symmetry, so that the hand gesture distribution feature does not change with the change of the rotation, translation and scaling of the image. Therefore, the hand gesture recognition algorithm proposed in this paper has good robustness to image rotation, translation, scaling and illumination.

The method used in this paper needs to be improved as follows: 1) limited to deal with the hand-like noise contained in the hand gesture image; 2) limited to recognize the similar gestures with different wrist lengths; 3) the recognition rate will be reduced for gestures with small discrepancies or large areas of pixel loss while hand gesture segmentation. These should be focused on in our next work.

# References

# 볼록한 결함 검정과 밀도 분포의 손짓 식별에 관한 연구

서아남

지도교수 박동원

배재대학교 대학원 게임공학과

사람 신체의 운동 분석은 인간-컴퓨터 상호 작용 분야의 중요한 연구 방향 중의 하나이다. 그 중에서 손짓은 직관적이고 자연스러운 상호 작용을 반영할 수 있는 양식이며 풍부한 의미를 편하게 표현할 수가 있다. 그러나 손짓 자체가 시간과 공간의 다양성 및 복합성, 그리고 시각의 불적응성의 성질을 가지기 때문에 이에 관한 연구가 도전성이 있는 과제가 된다.

본 연구에서는 전에 손짓 식별 기술의 기초에서 볼록한 결함 검정과 밀도 분포의 손짓 식별 방법을 제시한다. 먼저, 색채 특징을 이용해서 피부 영역을 구분하고 손짓 부분을 분할한다. 둘째, 볼록한 결함을 이용해서 손가락 끝에 관련 부분을 검정한다. 손가락 끝의 개수, 방향 및 손짓 윤곽의 길, 면적과 기타 기하학적 특징을 통해서 볼록한 결함의 특성을 추출하고자 한다. 셋째, 밀도 분포 특징과 손가락 사이에 상대적 거리에 인하여 구성된 손짓 분포 특징을 추측하고자 한다. 마지막으로 이상의 모든 특징을 결합해서 유사성 거리를 계산해서 손짓을 식별한다. 결과에 의하여 본 연구에서 논의된 방법을 이용해서 식별된 결과가 손짓 그림을 회전하거나, 이동과 줄임에 따라 변하지 않은 것이 나타낸다. 그리고 장면 조명의 변화에 대해 견고성을 미치는 것이 보인다.

**키워드**: 손짓 식별, 피부 검정, 밀도 분포, 검정, 유사성 거리

# Acknowledgement

Time flies, for the learning period of master and doctoral program in the department of Game & Multimedia Engineering, I studied game development and computer science systematically, and key focused on the research of object tracking and image processing, while the learning and research also increased the experience of my life.

First of all, I have to say thank you to my advisor Prof. Dong-Won, Park, his knowledgeable and rigorous scholarship has motivated and spurred me deeply. In the learning period of master and doctoral program, not only for the patient instruction but also for the providing me with the opportunity to take part in several projects, which created a strong academic and practical platform for me. Here, I would like to express my heartfelt gratitude to Prof. Park, hoping for your greater success in future studies and cultivating more outstanding disciples.

Meanwhile, thanks to Prof. Gouchol Pok for his professional advice, it was his professional knowledge helped me find out the right direction of this paper. And I also have to thanks to all the professors for sharing their knowledge that led me to walk into the hall of knowledge and enriched my vision.

Finally, thanks to my family for their persistent expectations and inclusiveness, I would like to say to my family that I will always strive to live and will never disappoint you.

1. [] L. Gupta, S. M. Gesture-based Interaction and Communication: Automated Classification of Hand Gesture Contours [C]. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2009, Vol. 31(1): 114-120. [↑](#endnote-ref-1)
2. [] M. R. Malgireddy, J. J. Corso. A Framework for Hand Gesture Recognition and Spotting Using Sub-gesture Modeling [C]. 20th International Conference on Pattern Recognition (ICPR), 2010, 3780-3782. [↑](#endnote-ref-2)
3. [] N. D. Binh, T. Ejima. Real-Time Hand Gesture Recognition Using Pseudo 3-D Hidden Markov Model [C]. 5th IEEE International Conference on Cognitive Informatics, 2006, Vol. 2: 820-824. [↑](#endnote-ref-3)
4. [] M. A. Moni, A.B.M. Shawkat Ali. HMM Based Hand Gesture Recognition: A Review on Techniques and Approaches [C]. 2th IEEE International Conference on Computer Science and Information Technology, 2009, 433-437. [↑](#endnote-ref-4)
5. [] Y. Han. A Low-Cost Visual Motion Data Glove as an Input Device to Interpret Human Hand Gestures [J]. IEEE Transactions on Consumer Electronics, 2010, Vol. 56(2): 501-509. [↑](#endnote-ref-5)
6. [] R. Y. Wang, J. Popović. Real-Time Hand-Tracking with a Color Glove [J]. ACM Transaction of Graphics, 2009, Vol. 28(3): 1-8. [↑](#endnote-ref-6)
7. [] L. Dipietro, A. M. Sabatini, P. Dario. Survey of Glove Based Systems and Their Applications [J]. IEEE Transaction on Systems, Man and Cybernetics, Part C: Applications and Reviews, 2008, Vol. 38(4): 461-482. [↑](#endnote-ref-7)
8. [] X. Y. Wang, G. Z. Dai, X. W. Zhang. Recognition of Complex Dynamic Gesture Based on HMM-FNN Model [J]. Journal of Software, 2008, Vol. 19(9): 2302-2312. [↑](#endnote-ref-8)
9. [] X. H. Wang, H. J. Bao. Gesture Recognition Based on Adaptive Genetic Algorithm [J]. Journal of Computer-aided Design & Computer Graphics [J], 2007, Vol. 19(8): 1056-1062. [↑](#endnote-ref-9)
10. [] F. F. Zeng, Y. Wang, G. J. Huang. Fuzzy Neural Network Based Hand Gesture Recognition [J]. Mini-Micro Computer Systems, 2000, Vol. 21(7): 706-709. [↑](#endnote-ref-10)
11. [] W. T. Freeman, M. Roth. Orientation Histograms for Hand Gesture Recognition [C]. IEEE International Workshop on Automatic Face and Gesture Recognition, 1995. [↑](#endnote-ref-11)
12. [] H. J. Lee, J. H. Chung. Hand Gesture Recognition Using Orientation Histogram [C], TENCON 99. Proceedings of the IEEE Region 10 Conference, 1999, Vol. 2: 1355-1358. [↑](#endnote-ref-12)
13. [] X. H. Wang, H. J. Bao. Gesture Recognition Based on Adaptive Genetic Algorithm [J]. Journal of Computer-aided Design & Computer Graphics, 2007, Vol. 19(8): 1056-1062. [↑](#endnote-ref-13)
14. [] L. G. Zhang, J. Q. Wu. Hand Gesture Recognition Based on Hausdorff Distance [J]. Journal of Image and Graphics: A, 2002, Vol. 7(11): 1056-1062. [↑](#endnote-ref-14)
15. [] X. W. Guo, Y. Ge. Classification and Recognition Algorithm of Alphabet Gesture Based on Shape Feature [J]. Computer Engineering, 2004, Vol. 30(18): 160-132. [↑](#endnote-ref-15)
16. [] C. M. Huang, L. L. Zhou. Density Distribution Feature and Application in binary image retrieval [J]. Journal of Image and Graphics, 2008, Vol. 13(2): 307-311. [↑](#endnote-ref-16)
17. [] H. B. Ren, G. Y. Xu. Hand Gesture Recognition Based on Characteristic Curves [J]. Journal of Software, 2002, Vol. 13(5): 987-993. [↑](#endnote-ref-17)
18. [] L. Bretzner, I. Laptev, T. Lindeberg. Hand Gesture Recognition Using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering [C]. Proceeding of the 5th IEEE International Conference of Automatic Face and Gesture Recognition, 2002, 423-428. [↑](#endnote-ref-18)
19. [] Y. Cui, D. Swet and J. Weng. Learning-based Hand Sign Recognition Using SHOSLIF-M [C]. Proceeding of 5th International Conference on Computer Vision, 1995, 631-636. [↑](#endnote-ref-19)
20. [] F. Quek, T. Mysliwiec, M. Zhao. Finger Mouse: a Free-hand Pointing Interface [C]. IEEE International Conference on Automatic Face and Gesture Recognition-FGR, 1995, 372-377. [↑](#endnote-ref-20)
21. [] J. Triesch, C. V. D. Malsburg. A System for Person-independent Hand Posture Recognition against Complex Backgrounds [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2001, Vol. 23(12): 1449-1453. [↑](#endnote-ref-21)
22. [] B. Stenger. Template Based Hand Pose Recognition Using Multiple Cues [C]. Lecture Notes in Computer Science, 2006, Vol. 3852: 551-560. [↑](#endnote-ref-22)
23. [] C. Keskin, F. Kirac¸, Y. E. Kara. Real Time Hand Pose Estimation Using Depth Sensors [C]. In Proceedings of the IEEE International Conference on Computer Vision Workshops, 2011, 1228–1234. [↑](#endnote-ref-23)
24. [] T. Takahashi, F. K. Shino. Hand Gesture Coding Based on Experiments Using a Hand Gesture Interface Device. ACM SIGCHI Bulletin, 1991, Vol. 23(2): 67-73. [↑](#endnote-ref-24)
25. [] J. Davis, M. Shah. Visual Gesture Recognition [J]. IEE Proceedings on Vision, Image and Signal Processing, 1994, Vol. 141(2): 101-106. [↑](#endnote-ref-25)
26. [] T. Starner, A. Pentland. Real-time American Sign Language Recognition from Video Using Hidden Markov Models [C]. International Symposium on Computer Vision Proceedings, 1995: 265-270. [↑](#endnote-ref-26)
27. [] K. Grobel, M. Assam. Isolated Sign Language Recognition Using Hidden Markov Models [J]. In Proceeding of the IEEE International Conference on System, Man and Cybernetics, 1997, Vol. 1: 162-167. [↑](#endnote-ref-27)
28. [] C. Vogler, D. Meraxas. Adapting Hidden Markov Models for ASL Recognition by Using Three-dimensional Computer Vision Methods [J]. Computational Cybernetics and Simulation, 1997, Vol. 1: 156-161. [↑](#endnote-ref-28)
29. [] C. Lee, Y. Xu. Online Interactive Learning of Gesture for Human/Robot Interfaces [C]. In Proceeding of IEEE International Conference on Robotics and Automation, 1996, Vol. 4(1): 2981-2987. [↑](#endnote-ref-29)
30. [] M. W. Kadous. Machine Recognition of Auslan Signs Using PowerGloves: Towards Large-lexicon Recognition of Sign Language [C]. Proceeding of the Workshop on the Integration of Gesture in Language and Speech, 1996: 165-174. [↑](#endnote-ref-30)
31. [] C. A. Pickering, K. J. Burnham. A Research Study of Hand Gesture Recognition Technologies and Applications for Human Vehicle Interaction [C]. 3th Institution of Engineering and Technology Conference on Automotive Electronics, 2007: 1-15. [↑](#endnote-ref-31)
32. [] S. S. Ge, Y. Yang, T. H. Lee. Hand Gesture Recognition and Tracking Based on Distributed Locally Linear Embedding [J]. Image and Vision Computing, 2008, Vol. 26(12): 1607-1620. [↑](#endnote-ref-32)
33. [] W. N. Chan, R. Surendra. Real-time Gesture Recognition System and Application [J]. Image and Vision Computing, 2002, Vol. 36(9): 993-1007. [↑](#endnote-ref-33)
34. [] A. Ramamoorthy, N. Vaswani. Recognition of Dynamic Hand Gestures [J]. Pattern Recognition, 2003, Vol. 36(9): 2069-2081. [↑](#endnote-ref-34)
35. [] Q. Chen, N. D. Georganas. Hand Gesture Recognition Using Haar-like Features and a Stochastic Context-free Grammar [J]. IEEE Transactions on Instrumentation and Measurement, 2008, Vol. 57(8): 1563-1571. [↑](#endnote-ref-35)
36. [] B. Tusor, A. R. Varkonyi-Koczy. Circular Fuzzy Neural Network Based Hand Gesture and Posture Modeling [C]. Instrumentation and Measurement Technology Conference (I2MTC), 2010: 815-820. [↑](#endnote-ref-36)
37. [] J. Segen, S. Kumar. Fast and Accurate 2D Gesture Recognition Interface [C]. Proceeding of the 14th International Conference on Pattern Recognition, 1998, Vol. 1: 86-91. [↑](#endnote-ref-37)
38. [] H. Zhou, D. Lin. Static Hand Gesture Recognition Based on Local Orientation Histogram Feature Distribution Model [C]. Computer Vision and Pattern Recognition Workshop, 2004, Vol. 10: 161-161. [↑](#endnote-ref-38)
39. [] T. Starnet, A. Pentland. Real-time American Sign Language Recognition Using Desk and Wearable Computer Based Video [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1998, Vol. 20(12): 1371-1375. [↑](#endnote-ref-39)
40. [] V. Athitsos, S. Sclaroff. Estimating 3D Hand Pose From a cluttered image [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2003， Vol. 2: 432-439. [↑](#endnote-ref-40)
41. [] C. C. Chang, J. J. Chen. New Approach for Static Gesture Recognition [J]. Journal of Information Science and Engineering, 2006, Vol. 22(5): 1047-1057. [↑](#endnote-ref-41)
42. [] N. A. Ibraheem, R. Z. Khan. Comparative Study of Skin Color Based Segmentation Techniques. International Journal of Applied Information Systems, 2013, Vol. 5(10): 24-38. [↑](#endnote-ref-42)
43. [] S. G. Wysoski, M. V. Lamar. A Rotation Invariant Approach on Static-gesture Recognition Using Boundary Histograms and Neural Networks [C]. Proceedings of the 9th International Conference on Neural Information Processing, 2002: 2137-2141. [↑](#endnote-ref-43)
44. [] A. Utsumi, J. Ohya. Image Segmentation for Human Tracking Using Sequential-image Based Hierarchical Adaptation [C]. Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1998: 911-916. [↑](#endnote-ref-44)
45. [] S. Mitra, T. Acharya. Gesture Recognition: a Survey [J]. IEEE Transaction on Systems, Man and Cybernetics, Part C: Applications and Reviews, 2007, Vol. 37(3): 311-324. [↑](#endnote-ref-45)
46. [] Y. Kuno, M. Sakamoto. Vision-based Human Computer Interface with User-centered Frame [C]. Proceeding of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems, 1994, Vol. 3: 2023-2029. [↑](#endnote-ref-46)
47. [] G. Buchsbaum. A Spatial Processor Model for Object Color Perception [J]. Journal of Franklin Institute, 1980, Vol. 310(1): 1-26. [↑](#endnote-ref-47)
48. [] T. Jebara, K. Russell, Mixtures of Eigen features for Real-time Structure from Texture [C]. Proceeding of the 6th International Conference on Computer Vision, 1998: 128-135. [↑](#endnote-ref-48)
49. [] N. Tsapatsoulis, Y. Avrithis, Efficient Face Detection for Multimedia Applications [C]. IEEE International Conference on Image Processing, 2000, Vol. 2: 247-250. [↑](#endnote-ref-49)
50. [] F. Marques, V. Vilaplana. A Morphological Approach for Segmentation and Tracking of Human Faces [C]. Proceeding in 15th International Conference on Pattern Recognition, 2000, Vol. 1: 1051-4651. [↑](#endnote-ref-50)
51. [] C. Garcia, C. G. Tziritas. Face Detection Using Quantized Skin Color Regions, Merging and Wavelet Packet Analysis [J]. IEEE Transaction on Multimedia, 1999, Vol. 1(3): 264-277. [↑](#endnote-ref-51)
52. [] J. C. Terrillon, M. David. Automatic Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and of Invariant Moments [C]. Proceeding of the 3th International Conference on Face & Gesture Recognition, 1998: 112-117. [↑](#endnote-ref-52)
53. [] YCbCr. http://en.wikipedia.org/wiki/YCbCr. [↑](#endnote-ref-53)
54. [] K. H. B. Chazali, J. Ma. An Innovative Face Detection Based on Skin Color Segmentation [J]. International Journal of Computer Applications, 2010, Vol. 34(2): 6-10. [↑](#endnote-ref-54)
55. [] J. J. de Dios, N. Garcia. Face Detection Based on a New Color Space YCgCr [C]. International Conference on Image Processing, 2003, Vol. 3: 909-912. [↑](#endnote-ref-55)
56. [] Z. Z. Zhang, Y. X. Shi. Skin Color Detecting Unite YCgCb Color Space with YCgCr Color Space [J]. Computer Engineering and Applications, 2010, Vol. 46 (34): 167-170. [↑](#endnote-ref-56)
57. [] A. A. Argyros, M. I. A. Lourakis. Vision-based Interpretation of Hand Gestures for Remote Control of a Computer Mouse [C]. Proceeding of the International Conference on Computer Vision in Human-computer Interaction, 2006: 40-51. [↑](#endnote-ref-57)
58. [] T. Lee, T. Hollerer. Handy AR: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking [C]. 11th IEEE International Symposium on Wearable Computers, 2007: 83-90. [↑](#endnote-ref-58)
59. [] A. Akoum, N. A. Mawla. Hand Gesture Recognition Approach for ASL Language Using Hand Extraction Algorithm. Journal of Software Engineering and Applications, 2015, Vol. 8: 419-430. [↑](#endnote-ref-59)
60. [] H. Liu, S. Q. Feng. Document Image Retrieval Based on Density Distribution Feature and Key Block Feature [C]. Proceeding of the 2005 Eight International Conference on Document Analysis and Recognition, 2005: 1040-1044. [↑](#endnote-ref-60)
61. [] L. Guo, X. H. Sun. Distance Distribution Histogram and Its Application in Trademark Image Retrieval [J]. Journal of Image and Graphic, 2007, Vol. 7(10): 1027-1031. [↑](#endnote-ref-61)
62. [] N. Z. Liu, J. Y. Yang. Color Image Retrieval using Both Color and Texture Features in Object Region [J]. Computer Engineering and Applications, 2003, Vol. 39(32): 187-203. [↑](#endnote-ref-62)