Fingertip Tracking and Hand Gesture Recognition Techniques for Natural User Interface

Unseok Lee

(Master's Program in Computer Science)

Advised by Jiro Tanaka

Submitted to the Graduate School of
Systems and Information Engineering
in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering
at the
University of Tsukuba

March 2013

Abstract

Fingertip Tracking and Hand Gesture Recognition(HGR) is an important research issue in the field of Human-Computer-Interaction, because its extensive application in virtual reality, sign language recognition, and computer games. Existing vision-based approaches are greatly limited by the quality of the input image from optical cameras. Variations in lighting and background clutters would only worsen the problem. Consequently, these systems have not been able to provide satisfactory results for hand gesture recognition. On the other hand, depth-based approaches are able to provide satisfactory results for hand gesture recognition even poor light and cluttered background condition. One of the more popular devices used to do depth-based approach is Microsoft's Kinect, which has sensors that capture both RGB and Depth Data. The advent of relatively cheap image and depth sensors has spurred research in the field of fingertip tracking and gesture recognition. Accordingly, it was required fingertip tracking and hand gesture for natural user interface.

In this thesis, we proposed fingertip tracking and hand gesture recognition method for natural user interface. Implemented interface provides three types of natural user interface by detecting fingertips number both hands. Moreover, performance can be improved by using k-means clustering algorithm and convex hull methods together when tracking more than two hands.

In result, our system provided new natural interface for drawing, image manipulation and video manipulation. It was improved fingertip gesture recognition comparing with vision-based approach as well.

Contents

List of Figures

List of Tables

Chapter 1.

Introduction

1.1 Natural User Interface

The interface using the hand gesture is popular field in Human Computer Interaction(HCI). Recently, a new possibility is provided to HCI field with the development of the sensors and technology [1]. This development has made possible robust recognition which is like to identify fingers and hand gesture recognition in bad condition such as dark light and rough background. At the same time, the interface for natural interaction was required from many users. However, it doesn't reach to this requirement of users.

The term natural interprets often to mean the mimicry of the real world. It is the design philosophy and source for that metric standard and the repetitive process manufactures the product. In this research, we discuss the touch and gestural interaction as one mode making the organization of the natural user interface possible. However, as to we, moreover NUI believes. Actually, we can think new that he/she spread so that he/she could indicate the mouse and keyboard, voice command, and in-air gesture and we are able to make a kind of interface for the cell phone and so forth after the design guide. They don't prescribe that in them oneself or the input and print technique guarantee them to provide the opportunity for making the natural user interface from us. The natural user interface is in UI. It optimizes the route the optimizes in the expert that it knows as the experience and that it is made so that we can use in this technology it has an effect on the potentiality of the new technology to the reflector human ability in which the method we betters than. It is applied to the given context and technical goal of the invention. It satisfies our necessity.

The range of the imagery in which that hears that quickly is awakened. Therefore the term natural is strong. First, and, the important situation for understanding uses that so that we can describe the in fact outside attribute in the product itself. As to the natural ingredient of the natural user interface, there is not about the interface in the least.It is natural and felt that user mentions as the way that we acts and considers while the method they uses that if they make the wound accurately he/she looks at. Following is the design guide lines for natural user interface [2]

- Create an experience that, for expert users, can feel like an extension of their body.
- Create an experience that feels just as natural to a novice as it does to an expert user.
- Create an experience that is authentic to the medium—do not start by trying to mimic the real world or anything else.

- Build a user interface that considers context, including the right metaphors, visual indications, feedback, and input/output methods for the context.
- Avoid falling into the trap of copying existing user interface paradigms.
- Forget past interaction styles. Don't simply transcribe an application rendered in a traditional medium (web or GUI) as a NUI.
- Choose a promising niche for developing a family of NUI apps. Thus far, the NUI has shown the most success in social and entertainment contexts. Its application to other domains requires an analysis of the way in which the interaction would support and teach the rules of the interaction domain.
- When developing a single NUI application, start with the most fundamental interactions. Perfect them through careful design and testing. Then extrapolate those designs to more complex regions in an interaction domain.
- Test the fundamental mechanics of the primary interactions before building out the entire interaction. When these are working well (i.e., users enjoy doing them), build on them.
- In building a family of NUI applications, start with simple ones (i.e., those with few possibilities and a clear and familiar model of interactions) and perfect them. Apply those learnings to more advanced applications.
- Study existing NUI applications. Are they fun to use? Do the interactions seem seamless and intuitive? Did users hate to stop using them? If the answer to all these questions is yes, then apply what you have learned to your application.

A NUI is not a natural *user interface*, but rather an interface that makes your user act and feel like a natural. An easy way of remembering this is to change the way you say "natural user interface"—it's not a *natural* user interface, but rather a *natural user* interface

1.2 Natural User Interface Leverage

While that is made by the Gartner of the natural user interface technology and group, as the part of the human computer interaction technology is treated, the media of the comprised multiple technology is the various channel and interactions. We are provided as the potential solution about the problem of the natural interaction with IT. Many this technologies evolves still. In the reality scenario, that is expected to undergo the complexity of the accuracy and arrangement-capability and important change. Figure lower part news shows the partial outline of the important technology in the applied perspective. That is displayed as shown up. It prescribes that data input in which the base layer is needed is formed so that the little bit of technology can interpret the high level of the natural operation and interaction trigger like the speech recognition, gesture recognition, and touch and so forth. Benefits/knowledge gained from the data collected by these "base" NUI technologies includes [2]

- Behavioral or gestural analytics solutions that can monitor data to identify patterns over a longer time frame

- Device virtual assistants that can emulate a human and provide feedback from the application to the user
- Applications can overlay digital data over video feeds, and use that data to interact with users

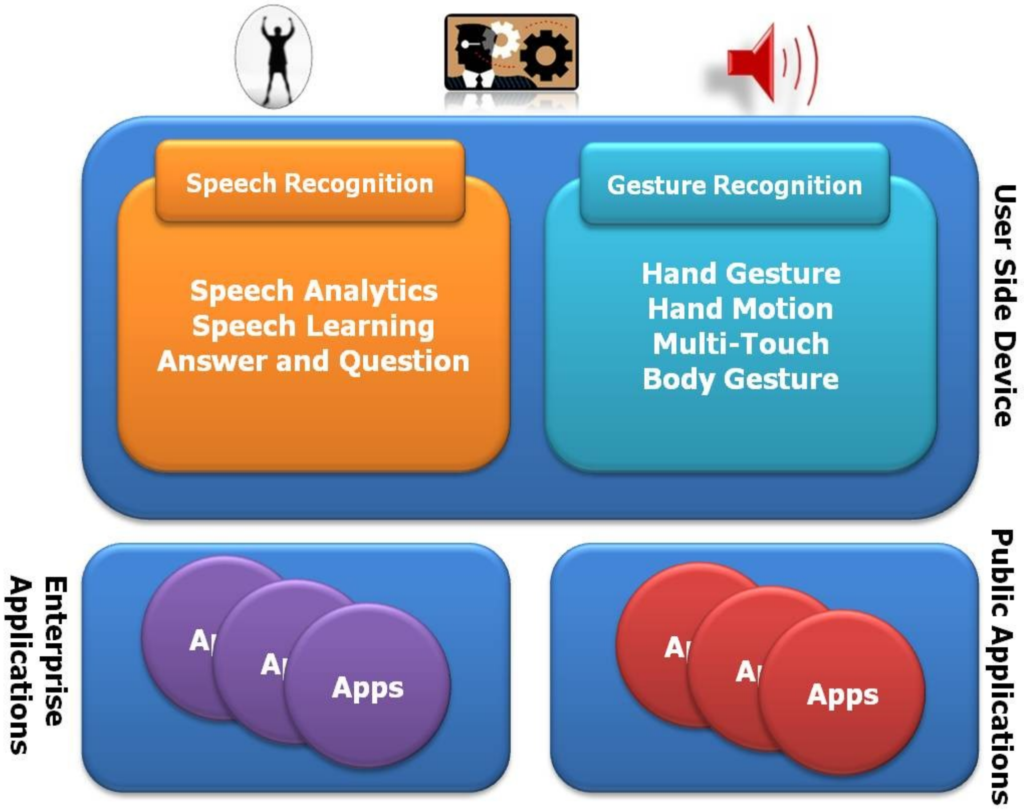


Figure 1.1. NUI Technologies Landscape

When we try to leverage these technologies to provide a sophisticated user-to-application interface, we encounter some difficulties [2]:

- Lack of standardization – Most of the technologies are being developed in isolation and secrecy to protect intellectual property. We find the data coming from different providers varies significantly.
- Differing technologies – Depending on a company's technology preference, the core programming platforms are different. This will require developers to factor in platform compatibility and interoperability as a major application design factor.
- Lack of aggregation – While all the component technologies are critical to solving the problem of a more productive IT interface, it is the combination of these technologies in an appropriate form which provides more value. It is truly a case of the sum being more valuable than the individual components. For example, in the case of patient monitoring, a combination of gesture recognition (frantic hand waving) combined with speech recognition (the words "pain" or "can't breathe") can provide a reasonably accurate alarm for a heart attack victim.

In order to deal with this problem, there is the limited request about the holistic "NUI framework" providing various NUI sources and capability which can gather data from the process and standardize that. That is provided in order to use in the inappropriate application next. The characteristics of such a framework should include, among other things:

The ability to accept and process data from multiple NUI sources, including [2] :

- The same type of device concurrently, e.g. multiple MS Kinect devices
- Different types, e.g. MS Kinect device(s), Sony Playstation Move device(s), etc.
- Different categories, e.g. gesture recognition and speech recognition simultaneously

The source devices should be configurable, and:

- Provide a repository of NUI events – All NUI event data collected from various sources should be stored in a repository and made available for analysis and action.
- Provide capabilities to provide logic to interpret events and derive some sophisticated inferences. It should have built-in rules for common patterns that can be analyzed from the repository data, and allow specification of rules for recognizing new patterns. This will help train the framework to become more relevant and accurate in real-life usage.
- Standardization – Preferably, should provide a standardized nomenclature and data structure for commonly used events, as well as provide the capability to define vertical-specific content (can be XML-based).

A logical architecture of the framework we have envisaged is shown below:

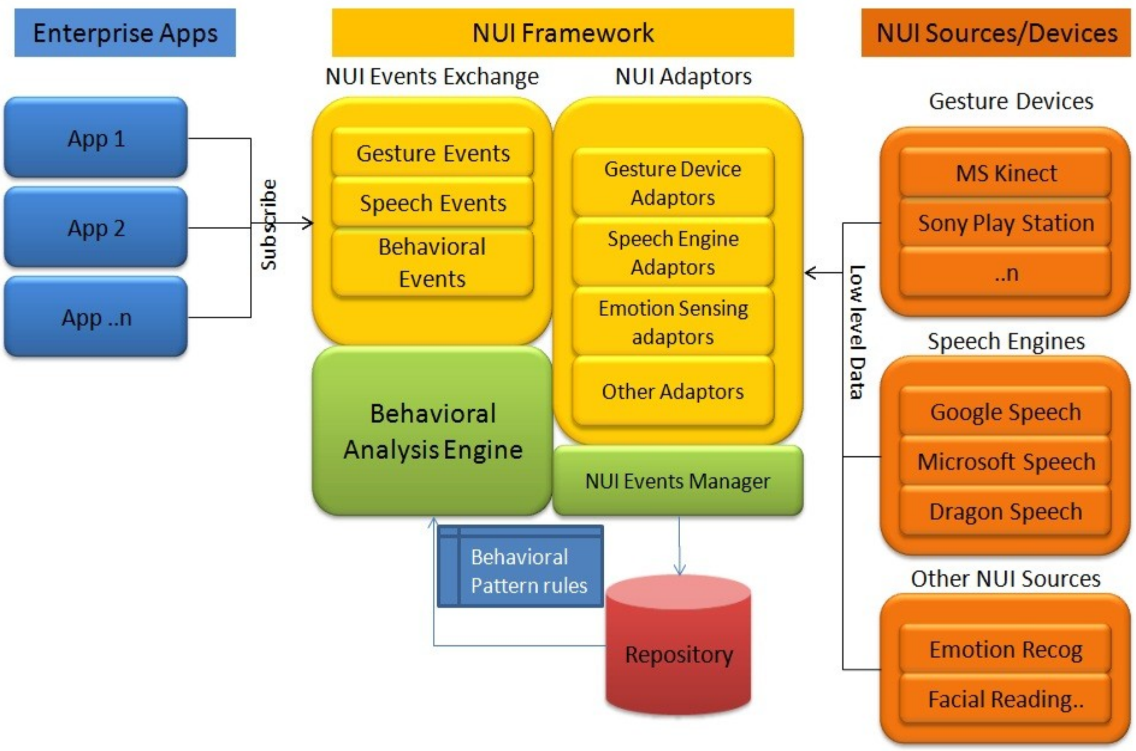


Figure 1.2. Logical Architecture for NUI framework [2]

This is essential to be the development in NUI space, naturally, the natural progression and install the framework in which the service provider is so in us like HCL. As to that, our customer will play the role of the definite enabler and differentiator when adopting NUI technology genuinely.

1.3 Gesture

1.3.1 Gesture Design

Gesture interface design has two types: not pre-defined (perceptual) and pre-defined [3]. Current our research on gestures focuses on pre-defined gesture recognition while perceptual interface design is complicated and difficult.

- Not pre-defined natural gesture interaction

The Perceptive User Interface (PUI) which it can be late is accompanied with the different human expression like the facial expression and body movements that maybe it aims to recognize the natural human gesture. Moreover the interface of kind of this provides to fulfill the human dialogue human computer interaction.

The interface which this necessity is able picks up the gesture command intended from the thing, that is the series of the sports including the gesture which is not many intentions.

Even, furthermore it is controlled including the music raising the telephone about the gesture in which the user is not the little bit of consideration like the automatic volume so that the small bit fish gathering time of the useful information for the small bit of the intelligent situation can select.

- Pre-defined gesture interaction

In the Pre-defined user interface, the hand strikes a pose or generally the specific gesture plans the order. What is natural was not needed. However, the gesture was developed for the situation or could be based on the standard sign language. In this case, as shown in only there is and there is not to catch the mouse for the expression, the gesture command changes to the next slide into the remote operation, the little bit of the special situation regards the place as the several other HCI devices. Even, it can be eaier and to be very implemented it believe much more and PUI which is the new which the passive input mode sound stops much more and, is splendid and has the pre-defined user interface doesn't have the reclamation mistake.

Pavlovic [4] noted that, ideally, naturalness of the interface requires that any and every gesture performed by the user should be interpretable, but that the state of the art in vision-based gesture recognition is far from providing a satisfactory solution to this problem. A major reason obviously is the complexity associated with the analysis and recognition of gestures.

1.3.2 Gesture Classification

The gesture is classified as the various methods. The gesture can be classified from the anatomy view: hand and dug operation and head, facial gesture and body gesture [5]. That is most, the expression. It is natural. Because of being instinctive and being often the usedest and being suitable for the large-scale display interaction, we concentrates on the hand and dug operation in our research. According to their functionalities, the hand movement can be classified as the following [6]: semiotic, ergotic, and epistemic. It is the symbol means for delivering the important information and result from the shared cultural experience. As to ergotic, the human which relates to the things concept and manipulates the virtual reality capacity cooks the artefact. Epistemic learns through the experience or the sense of touch inquiry which tactually the human can know from the environment.

Figure 1.3. Kendon`s gesture continuum

While going to the formality linguistic ingredient of the expression in the body language to the signal language, it is replaced by the denotation which presently it goes from the voice in the body language to the signal language: the body language is the spontaneous movement of the hand and arm for the voice. The body language is not generated in case there is ever nearly no voice. For example, we cannot say that "let's go up" it divides with the finger into the gesture. The gesture which is the same as the language is the special spoken language or gesture replacing the commission.

In this case, the audio stops in order to perform the gesture. The pantomime is a series of gesture having the narrative function of being used in the theater which takes care to talk. The symbol is the hand specific attitude having the fixed format of the meaning.

However, the meaning can change over the various cultivations. The reference numeral-languages is the well defined linguistic communication system. Therefore, based on (hand movement) technical view it is dynamic and as the stillness we can classify the gesture (the hand attitude) the decoding information from the gesture is not important in the pre-defined user interface The hand attitude and hand movement are both prescribed as the finger and sports of the hand and arm. The hand attitude is prescribed in one moment in the appropriate time as the location of the hand and finger. However, generally one means uses the situation where the hand is used as the technology very so that the hand attitude and gesture can show the use of hand for the communication which there is no physical manipulation of any kind of object, is intended.

1.3.3 Gesture Recognition

The gesture is the expression. It is the meaningful body sports, having the intention which conveys the information or which it acts with the environment physical movement of the finger, hand, arm, head, and face or body. In HCI literature, the word gesture was used in confirming many hand movements for the control of the computation.
However that doesn't contribute the necessity to the final product in this way to move your hand. Therefore it doesn't think as the gesture that one hand to the place starts the production or manipulation [7].

According to McNeill [8] the dynamic gesture movement consists of three parts

- Approach: body begins to move
- Stroke: the gesture itself
- Return: return to balanced posture

The gesture is divided to be dynamic with the static thing. However, the little bit of gesture has the dynamic static element. And the pause is there important in one kind or greater to the step gesture. The vision basis hand gesture recognition was determined and to be the useful technology the number of systems were proposed. There is the following type of the technology of the vision basis: The based marker and skin color detection and motion detection technology. However the static gesture is the hand structure and special pause indicated by one image. The dynamic gesture is the mobile gesture indicated by a series of image. It is the core technology for the body language recognition system to obtain the feature having the dominant information from the variant and composite the hand sports. The static hand attitude can play the role of the special turning state of the dynamic gesture and moreover, the dynamic gesture can be captured and the dynamic can be analyzed as a series of static gesture. The temporary which admits the recognizes and which the temporary extracts gesture from the sports handles by the finite state machine technology having the series sports having the start signal as the input.
The surrounding noise was the obstacle for the gesture recognition based on the computer vision. Which model the task by the adequacy normal and it was performed in the optimal environment. The noise background having the color which is similar to the Work Environment having many moving objects the human cadaver skin (nearly, even, that is the eye of the person for finding out the anole and sameness, is directly, your front) or the observational region obstructs considerably the performance of system admitted the recognizes. The error rate will be improved. Human body language is abundant and able to convey lots kind of expressive information. However, in the vision based human computer gesture interaction scenarios, not all of gestures intend to express controlling intentions, even gestures which have already been defined as gesture control commands can also convey some non-control expressions in some kind of situations. Ignoring un-intended gestures for gesture commands controlling systems is important for error control, also can prevent the system from dealing with useless

information, and reduce the workload of the system.
In the number of used 1) camera and speed of 2) its and latency,3, the view foundation technique changes by us, the structure of the speed of the environment like the limit and sports, any kind of user needs like any kind of limit for 4) clothes, and the low pressure area-level is different from the edge, area, silhouette, moment, and histogram used 5) and 2D or 3D is used. Therefore, in the smart environment, this limit limits the application of the vision-based. More, more concretely, let's assume what you are decreased and it enjoys the audition movie in your right-foot shoot with all lights. If it determines to change the volume of the television in which you have the gesture, that is proven that it is difficult to recognize your gesture under the tribal lighting condition rather using the vision-based. Moreover, if that is directly confronted by the camera so that you can complete the gesture, it will be very uncomfortable and be unnatural [5].
Very, the promising alternative plan appeals to the other sensing technique like the technology of the acceleration base or the electromyogram base (EMG-based) technology. The acceleration-based gesture control distinguishes the scale remarkable large gesture having the various passive orbits. However, if it talks about detecting the delicate finger, it is effective very, it is not. The gesture recognition which it is based on data from the accelerometer is the newborn technology for the body language-based interaction after the rapidity progress of the MEMS technique. In the application range wide, the accelerometer works for the works. It is mostly built in of the production personal electronic device for being new like the Apple iPhone, that is the Nintendo Wii remote which the home appliances, and, the video game provides the new possibility for the interaction, [5].
The natural and awareness of the successive gesture require the segments gesture temporally. It is difficult. By demanding the start range, and, (in other words) space from the appropriate time often the gesture which is adroitly processed in the present system or is disregarded is automatically divided. It is similar to this: The problem of distinguishing the different "arbitrary" sports and intentional gesture. The gesture recognition and various expression and for doing the classification scheme any kind of standard method was not used. However, the most of gesture recognition systems shares the little bit of generic structure.

1.3.4 Gesture Recognition using Depth Data

Using 2D video-based algorithms for tracking hands leads to confusion when there is bad occlusion. If color-based segmentation is used, the hands are not easy to be separated from the face due to their similar skin colors. Moreover, its performance suffered when the lighting conditions changing or the background is rough. Data glove devices that provide hand positions may be a good way to augment color-based segmentation, but they cause uncomfortable User Experience (UE), because of the physical problem. Providentially, a new contact-less approach came out in the past ten years using depth discontinuities to separate hands from the natural background. In this way, depth-based systems avoid all these problems mentioned above.
Nanda et al. [9] implemented a method to track hands in highly cluttered environments

using 3D depth data provided by a sensor that was based on the time-of-flight (ToF) principle to capture depth and color information simultaneously using the same optical axis. The potential fields of possible hands or face contour were calculated by three algorithms: getting potential fields by using distance transform, k-components based potential fields with weights and basin of attraction.

The system was tested in head tracking and hand tracking on ten people with good results.

However, because the contour in which there is lots of the hand shape is changed in order to be needed for the awareness, that is not suitable for the real time system.

Argyros et al. [10] proposed a 3D hand tracking method using a stereoscopic system of two video streams. That could bind the color-based tracking and 3D hand tracking, it operates that on a real time. The image was captured by the camera. In the both video streams, the hand blob was indicated by 2D color-based hand tracker and was together in accordance by the correction annual production. The hand contour was arranged and the hand was re-organized after this in 3D in space. This method was applied to many applications. One experiment operated the CD player by the hand gesture. Here the contour of two hands and forefinger was detected. As capability is the advantage on a real time basis. Rather the predicted depth data using the stereoscope system is noisy. The necessity for the correction increases the system complexity. The Van den bergh and so on, The ToF camera was used instead of the stereo camera in order to improve the awareness. The ToF camera having the low resolution (176x144 pixels) was used and in obtaining the segmentation, deeply, the image that was mated with RGB camera for the hand detection with the high resolution (640x480 pixels). As to the ToF camera and RGB camera, the scale was adjusted in the first stage and the critical value distance was prescribed in order to dispose of the background image according to data deeply. The residue pixel was passed through the skin color detection in order to receive the hand data. The complexion used in the detection was determined by the skin color model which the pre-trained updated by the color information brought from the face is suitable. Three situations were evaluated in the hand detection: the hand was at the face and side of the hand overlapped with the face and there was the second opinion person after the testing machine. Deeply, in all 3 situations, 98% accuracy achieved over. The base detection reduced the accuracy of the color-based detection from 92% of the first situation to 19.8% of the third situation considerably on the other hand. The average adjaceny margin optimization transformation was used in building the classification machine for the gesture recognition. And it was suitable to coincide with the hand gesture in which the Haarlet coefficient is there stored to the database. The RGB-based awareness showed the accuracy of 99.54% and the base awareness showed the accuracy of 99.07% deeply. The combination of 2 methods showed 99.54% on the other hand. This proposes because the base awareness can form the base of the recognition system in kindly, particularly, the hand segmentation deeply.

1.4 Purpose and Approach

Fingertip tracking and hand gesture recognition is an important research issue in the field of Human-Computer-Interaction, because its extensive application in virtual reality, sign language recognition, and computer games [11].

Existing vision-based approaches [12], [13] are greatly limited by the quality of the input image from optical cameras. Variations in lighting and background clutters would only worsen the problem. Consequently, these systems have not been able to provide satisfactory results for hand gesture recognition. On the other hand, depth-based approaches are able to provide satisfactory results for hand gesture recognition even poor light and cluttered background condition.(e.g. [12]) One of the more popular devices used to do depth-based approach is Microsoft's Kinect, which has sensors that capture both RGB and Depth Data. The advent of relatively cheap image and depth sensors has spurred research in the field of fingertip tracking and gesture recognition. Accordingly, it was required various natural user interface for fingertip gesture interactions.

In this paper, we proposed natural user interface by using fingertip gesture tracking. This interface provides three types of natural user interface by detecting fingertips number both hands. Moreover, performance can be improved by using k-means clustering algorithm and convex hull methods together when tracking more than two hands.

In result, our system provided new natural interface for drawing, image manipulation and video manipulation. It was improved fingertip gesture recognition comparing with vision-based approach as well.

1.5 Organization

This thesis is organized as follows. In Chapter 2 we present the related works. In Chapter 3 we outline the details about our fingertip tracking and hand gesture recognition techniques. In Chapter 4 we introduce three types of natural user interface and interactions using proposed techniques. In Chapter 5 we discussed evaluation and result. Finally, Chapter 6 presents the conclusions and future works.

Chapter 2.

Related Works

In this chapter, we present some related works to bare hand tracking and Kinect related works.

2.1 Bare Hand Tracking

Barehanded means that no device and no wires are attached to the user, who controls the computer directly with the movements of his/her hand.
In the last ten years, there has been a lot of research on vision based hand gesture recognition and finger tracking. Interestingly there are many different approaches to this problem with no single dominating method. The basic techniques include color segmentation [14], infrared segmentation [15], blob-models [15], and contours [16]. Typical sample applications are bare-hand game control, and bare-hand television control [17]. Most authors use some kind of restriction, to simplify the computer vision process:

- Non real-time calculations
- Colored gloves [18]
- Expensive hardware requirements (e.g. 3D-camera or infrared camera) [17]
- Restrictive background conditions
- Explicit setup stage before starting the tracking
- Restrictions on the maximum speed of hand movements

Changing the bright condition and background clutter additionally, the most of systems has the problem. Any of presented work doesn't provide the strong tracer technique for the fast hand movement. Moreover, the most of systems demands any kind of the device-step before the interaction can leave.

2.2 Kinoogle: a Kinect interface for natural interaction with Google Earth

In this paper, it introduced Kinoogle, a natural interaction interface for Google Earth using Microsoft Kinect. Kinoogle allows the user to control Google Earth through a series of hand and full-body gestures. It began by describing the software design and modules underlying Kinoogle, then offer detailed user instructions for readers interested in trying it out.

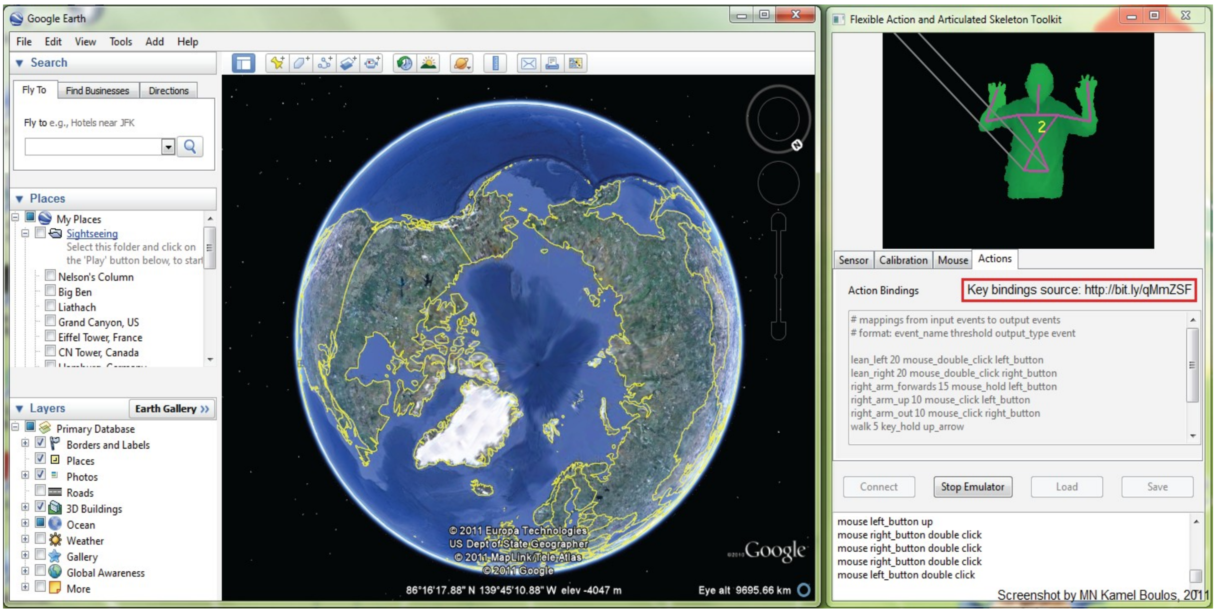


Figure 2.1. Natural Interface for Navigate Google Earth and Street View

The user produced the specific gesture. The main function of the Kinoogle interprets point data from the Kinect control in order to determine to act with the earth control and Kinoogle GUI if it strikes a pose. As to Kinoogle, there is the responsibility of the communication and this is achieved among the other 3 objects controlling the third party software through the thing using the Kinoogle message included in message object.

Moreover Kinoogle handles the detection of the fixed pause. And the user switches the mode in the interval also whether Kinoogle is various. Or not (for example, while it cleans with the fan, it zooms).If only the user activates the hand of his / she, the mean position of the hand is calculated over the last 100 Frame.

And the virtual box (300*200 pixels) is placed in this mean position surrounding. And, there is 3 quadrants which enter upon and in which the user can locate the hand of the / she of it for the pause detection: In the left on the box, in other words, the right of the box.When being positioned in the quadrant which the both hand corresponds to the mean position surrounding, the various pause is activated.

For example, it will be detected as the pause for the tilt to locate the other hand to the right of one hand and mean position on the mean position.

2.3 Related Work using Kinect

After the Microsoft Kinect of on November, 2010 was developed, many they which excited recognition system are based on this device were developed within 18 months. The solution of the camera is deeply very enough 640x480 pixels with its HGB camera under the both many situations.

Yang et al, [19] suggested the body language recognition system which is provided as Kinect and uses the depth information. That could recognize eight gestures in order to control the media player with the maximum confusion rate of 8.0%. The algorithm for the hand tracking finds the hand waving sports based on the assumption that there is the tendency to start the interaction session that the user has such campaign as the first. The civilians shift algorism which is continually befit was applied so that it could trace the hand by using the probability deeply. It revises the histogram until Frame deeply. The passive orbit was tested by 3D feature vector and the gesture was recognized by HMM. This system proves the applicability using the Kinect for the gesture recognition of the contact-less UI.

The previous system could not find the fingertip. Thus, it was limited to recognize only motion gestures like wave, and move up/down, left/right, and forward/backward. The method for tracing the center of the fingertip and palm by using Kinect was described by the Raheja et al. [20]. The definite value was applied to the depth of the hand area for that segmentation. And it entered upon and the palm filtered from the hand and was taken out. The finger was remained in the image.

When the hand was at the front of user under most of situations, the finger is come close to the Kinect having the shallowest depth. Therefore, since determining the minimum depth, the fingertip was discovered. The center of the palm is determined to find the greatest of the distance in image of the hand. When the finger was expanded, the accuracy of the sensing fingertip was nearly, 100% accuracy and that of palm center was 90%. However, this method did not attempt the gesture recognition.

He et al. [21] suggested the other access using the depth data provided as Kinect in order to detect the fingertip. First, that found the threshold hand point in data deeply. And it produced the convex hull including the hand by Graham scan [22].

The fingertip was detected by calculating the angle between the candidate point. After the fingertip was discovered, the movement was recognized and to mouse-click it was tested in the popular game angry birds. However, only, one gesture was recognized.

Chapter 3.

Methods

In this chapter, we present method for tracking fingertips and recognizing hand gestures. The system of the present work consists of three main components: Hand Detection, Finger Identification, and Gesture Recognition.

3.1 System Introduction

Particularly, there is enough open source framework which is developed for HCI like the hearing sense for the communication between physical device or motion sensor and UIs.
OpenNI provides the API to the cross-platform due to NI application of the thing which is one and them many languages. Presently, that supports 3 kinds of major platform: Windows 7, MAC OSX and Linux.
The application comprised at OpenNI is usually the portable and to be used in the other middleware it is easy. Figure 3.1 describes 3 abstract layers of the OpenNI concept [23].
Each, the layer shows the integrated type element：

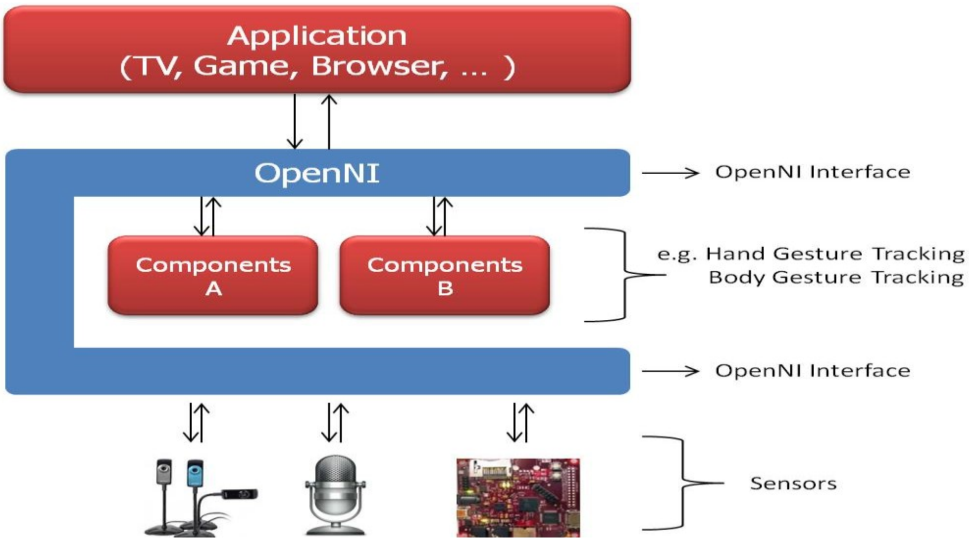


Figure 3.1. OpenNI Abstract Layers

- The top layer represents the applications implemented for natural interactions.
- The middle layer represents the OpenNI framework, in which it not only interacts with physical sensing devices as well as software, but also communicates with middleware components.
- The bottom layer represents all kinds of sensing devices, including visual and audio sensors.

OpenNI provides two kinds of production nodes. That is the production node and the sensor relating production node the middleware relating. Each, the node is a series of ingredient producing the fixed type of data about the NI-based application. Presently, there are five sensor-related production nodes supported with OpenNI. There are deeply the generator, image generator, and IR generator and audio generator with the device. As to the current system, the device node uses their two firsts in the place making the device configuration possible and the generator produces the depth-guide deeply.

3.2 Fingertip Tracking Techniques

3.2.1 Getting the Depth Image from KINECT

The internal architecture of Kinect is shown in Figure 3.1. It has the infrared camera and PrimeSense sensor while it is used when RGB camera charms the image in order to calculate the depth of the object.

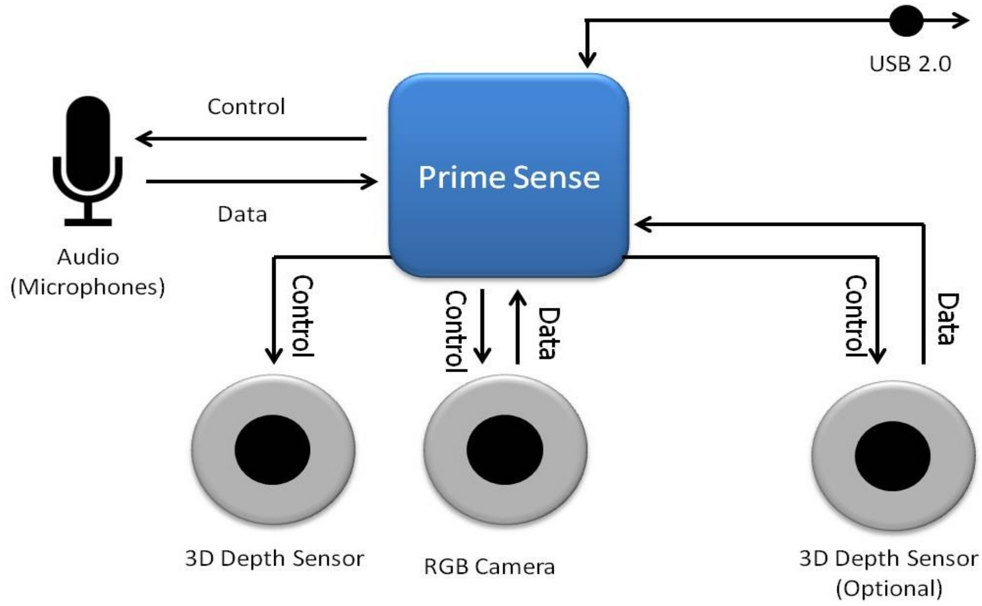


Figure 3.2. MS Kinect Architecture

Frati [24] stated "it has the web cam-exterior structure. While the user gestures and it uses the voice command or seen object and image, it controls through the natural user interface with the virtual reality and acts". Kinect is the steel device. It is clear to could be used in the various and complicated applications. At the same time, the object, deeply, the image and RGB image can arrive. It is called as the optical coding in which this 3D scanner system uses the variant of the image base 3D reconfiguration. The depth output of Kinect is of 11 bit with 2048 levels of sensitivity. The depth value d_{raw} of a point in 3D can be defined as calibration procedure

$$d = K \operatorname{tang}(Hd_{raw} + L) - O$$

where d is the depth of that point in cm, H is 3.5x10-4 rad, K=12.36 cm , L = 1.18 rad and O=3.7 cm.

Figure 3.3. Depth image acquired using Kinect

3.2.2 Hand Detection

The first stage separates the hand from the background. Deeply, as the gesture is passively recognized in advance, the critical value specifies the depth range that it has to display. As to the detection range and zRange, there is between 0.5m and 0.8m and the hand can be detected till this range, most, 2. The pixel having the external zRange deeply is disregarded in the rest of the gesture detection process. Henceforth, moreover the hand pixel can be mentioned as the point. This hand pixel is projected with 2D space for the sequential analysis.. Distance between two pixels p_1 (x_1, y_1) and $p_2(x_2, y_2)$ is defined as:

$$D(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

The k-method is applied in order to divide all pixels in two groups. K-means clustering is a method to partition *n* observations into *k* clusters, $C_1, C_2, \ldots, C_k$. Each observation is identified with the nearest mean, μ_i (x,y), which is calculated as the mean of points in C_i. K-means clustering minimizes the within-cluster sum of squares:

$$\frac{\arg min}{C} \sum_{i=1}^{k} \sum_{p_j(x,y) \in C_i} \left| |p_j(x,y) - \mu_i(x,y)| \right|^2$$

In this system, the value of the k is 2. The system operates on a real time basis and whenever there is the change of the input data source, the K- method is continually applied. In the early in all repetitions, the empty cluster is initialized to the arbitrary point positioned as the average in the zRange. After K-means collects, it is bundled to belong to the hand. If the distance is the pre-defined or less between two centers of the hand, 2 clusters become one.

After the point of the hand is bundled, we detect the convex hull of the hand like the contour of the hand. If the set is bulging, it is included in the line segment done with the point set of all pairs and pair in moreover, the set. As to the Let S, in a series of point or {P0, P1, P2,…, Pn} and 2D airplane, the convex hull of S is the convex polygon which the head of a family including all points in S is small. Graham scan algorithm [22] is used in calculating the convex hull of the searched hand cluster. The method is explained on figure 3.4 and as follows, works for that.

Graham Scan Algorithm

1) P_0 (x, y) will be found to have the large-scale y-axis. If there is one point or more with the y-axis which the same head of a family is big, this we having the smallest X-axis will be selected.
2) An imaginary line is comprised in $P_0(x,y)$ and each different between point. The angle is measured between this line and X-axis. It goes up, it is classified in the order. In the implementation, the cosine as to classified angle, is replaced by the corresponding to cosine of the angle in 0 with the reason why the function reduces uniformly in 180 degree domain.
3) The first spot of the sequence aligned with the line segment between $P_0(x, y)$ is formed. And $P_1(x,y)$ is in the convex hull.
4) Let $P_0, P_1, \ldots, P_s$ be the partial convex hull constructed. For each of the remaining points in the sorted sequence, P_t, if the line segment $\overline{P_s P_t}$ makes a counterclockwise turn from the line segment $\overline{P_{s-1} P_s}$,P_t is added to the partial hull, and this step is repeated. If $\overline{P_s P_t}$ makes a clockwise turn from the line segment $\overline{P_{s-1} P_s}$,Ps is removed from the partial convex hull. The process backtracks to the partial hull $P_0, P_1, \ldots, P_{s-1}$, and this step is repeated.

To implement Step 4, the direction from Ps-1 and Ps to Pt is determined by the cross product of the two vectors defined by the three points: $\overrightarrow{V_1} = (P_{s-1}, P_s)$ and $\overrightarrow{V_2} = (P_s, P_t)$.

$$CroVec = \vec{V_1} \times \vec{V_2} = (x_s - x_{s-1})(y_t - y_{s-1}) - (y_s - y_{s-1})(x_t - x_{s-1})$$

CroVec of in counter-colckwise direction or that yes and direction the negative inside and direction is the co-linear. The hand contour is detected by the amended Douglas Moore-neighbor tracing algorithm. The point of one group it clusters is discovered and is stored. Figure 3.5 shows the example which substantiates the algorithm in order to detect the contour. The DefineN for being the adjaceny of 8 of the pixel a (a). The Let P indicates the appearance pixel. The q will indicate the beginning pixel of the neighboring checking. The Let C indicates the set searched of the appearance point which is initialized in order to be the empty set.

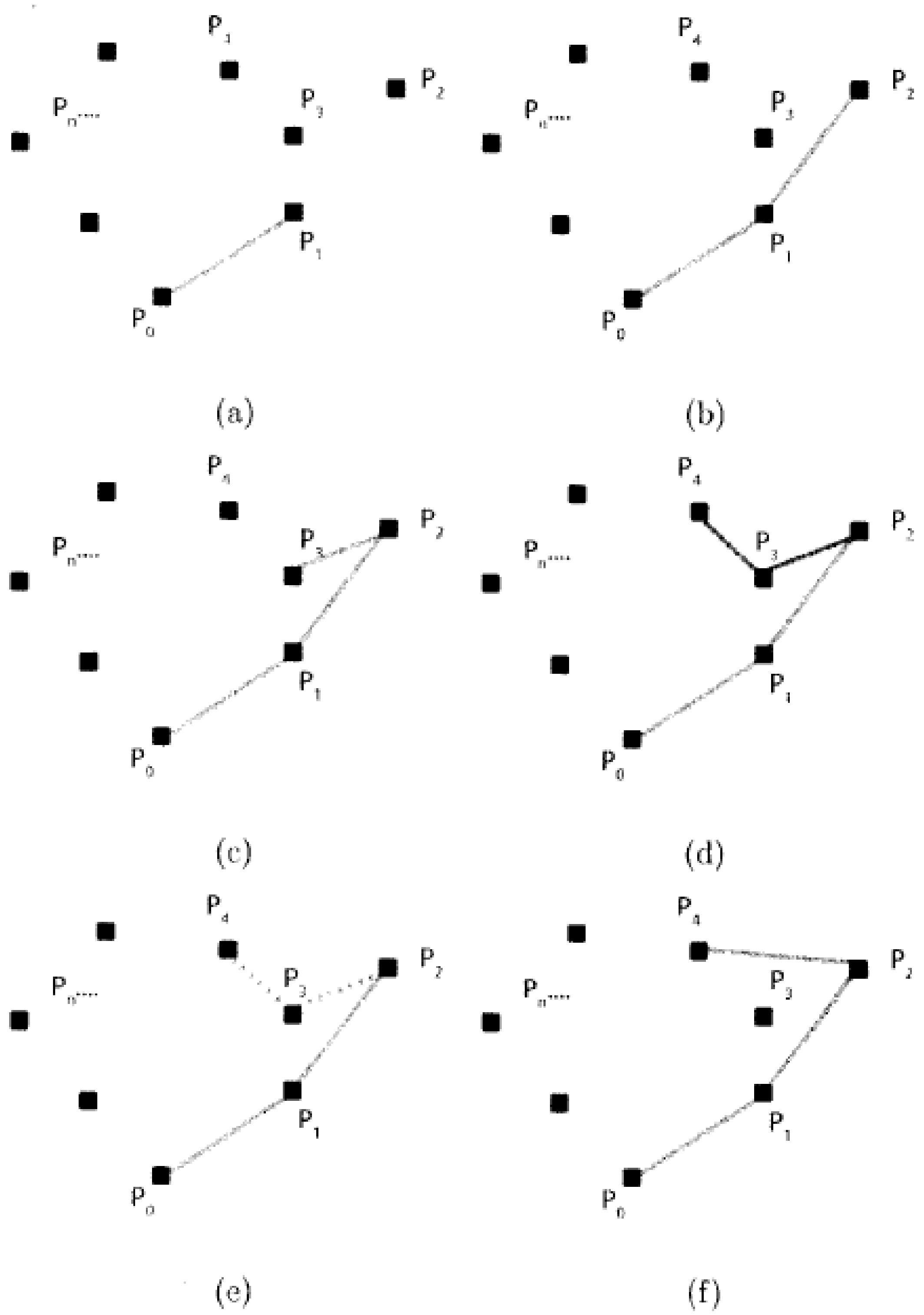


Figure 3.4. Graham Scan [22]

Next, the hand contour is detected by the amended Douglas Moore-neighbor tracing algorithm. The point of one group after the K- method is discovered and is stored.
Figure 3.5 shows the example which substantiates the algorithm in order to detect the contour.

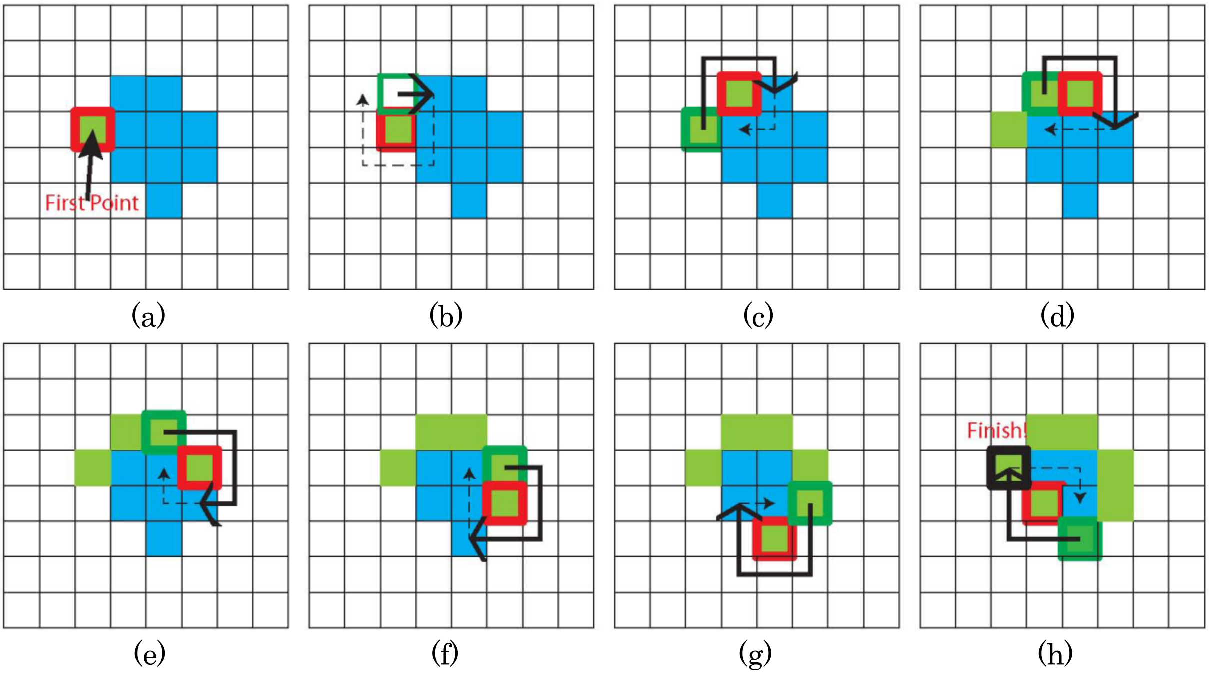


Figure 3.5. Contour Tracing [25]

The contour of the blue pixel is traced. The appearance pixel in which the green pixel is searched will be shown. The pixel which has the blue boundary that pixel having the red boundary indicates the appearance pixel indicates the beginning pixel of the neighboring checking and excluding to start last appearance pixel.
As to the black arrow, the next appearance pixel was already detected to show the clockwise path of the neighboring checking. Therefore arrow shows the route which the algorithm not needs to finish. The blue dot is the center of this cluster (See, figure 3.6).
The performance benefit, only, all 25 numbers points are measured.

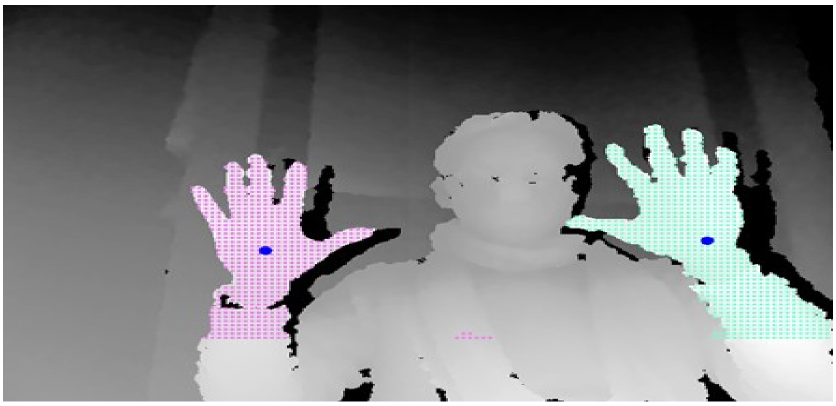


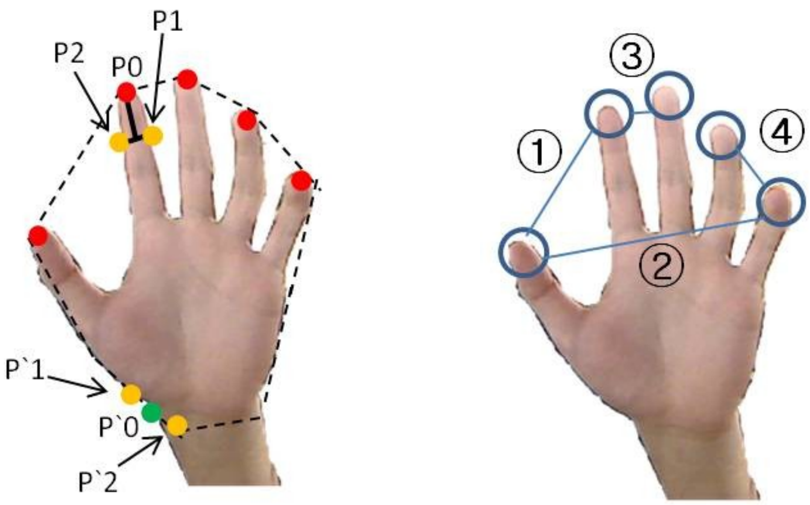
Figure 3.6. Hand tracking with k-means clustering

3.2.3 Finger Identification

N (a) will be the adjaceny eight points of the pixel a. The p will indicate the appearance pixel. The q will indicate the beginning pixel of the neighboring checking. The Let C indicates the set searched of the appearance point which is initialized in order to be the empty set.

Contour Tracing Algorithm

1) From top to bottom, and left to right, It is set up as it finds all pixels from the screen to the pixel s and that starts if the hand point is discovered.
2) Set the contour pixel *p* to be *s*. The beginning pixel of the adjaceny checking q is the point and that is directly, the north of the s.
3) The p is inserted into C. The adjaceny N (p) will be calculated.
4) Let's start in the q and until the hand pixel r is discovered, the adjaceny N (p) will be rotated in the clockwise direction.
5) P for being the q and for being P new appearance pixel r will be placed. When the number of the s reaches again or to start it is searched exceeds the maximum value, it installs. And 3 steps will be repeated.



(a) (b)

Figure 3.7. (a) Fingertip Detection. The red dots are real fingertips. and the green dot is not a fingertip. The yellow dots are used to check for three point alignment. The distance between P_0 and the line made by P_1 and P_2 is apparently larger than that of P_0' and the line made by P_1' and P_2' Thus a threshold distance could be used to differentiate fingertip points and non fingertip points.(b) Finger Identification.

After the hand contour is detected, the center of the palm carves of the hand contour, it is calculated as the center in the circle. The location of the palm center is used in calculating the direction of the finger described in the next section. The center in the circle it carves is much more secure than the center of the hand cluster. Even when opening and spreading the finger, the latter is due to recognize without any problem. The fingertip is detected by checking three point alignment relationship. The candidate point is in the both convex hulls and hand contour. The method shows up in figure 3.7 (a).

Three-point Alignment Algorithm

1) C will be baggily made C of all candidate sets in the hull and hand contour into.
2) For each point P_0 in C, take the two points P_1 and P_2 in the two opposite directions along the contour from P_0.
3) The center point of the P_1 and P_2 let's be found and the distance be calculated between this center point and P_0. If the distance is bigger than the specific value, three point knows to arrange three point, that is, the collinear is not ever different. And it enters upon and this P_0 is distinguished the fingertip. On the other hand, let's return to two steps and the next candidate material point will be checked.

After fingertips are detected, their direction vectors are easy to calculate by subtracting the position of the palm center

$$P_c(x_c, y_c) : \vec{V} = (x - x_c, y - y_c)$$

The vector indicates in the palm center till the finger.

The finger name is determined according to their relative distances which open the palm to the user and which it requires in order to expand the finger. Four procedures for confirming the name of all fingers is shown in figure 3.5 (b). The confirmation of all individual case fingers did not perform ever in the literature having the Kinect platform which I know.

Finger Identification Algorithm

1) The easiest method checks the thumb and index finger. Because, the distance of the thumb and forefinger is due to be the biggest among all adjacent fingers.
2) The baby finger is distinguished the finger which the head of a family which is apart from the thumb is far. The central finger is distinguished one which is the closest to forefinger in the meantime.
3) The remaining one is the ring finger.

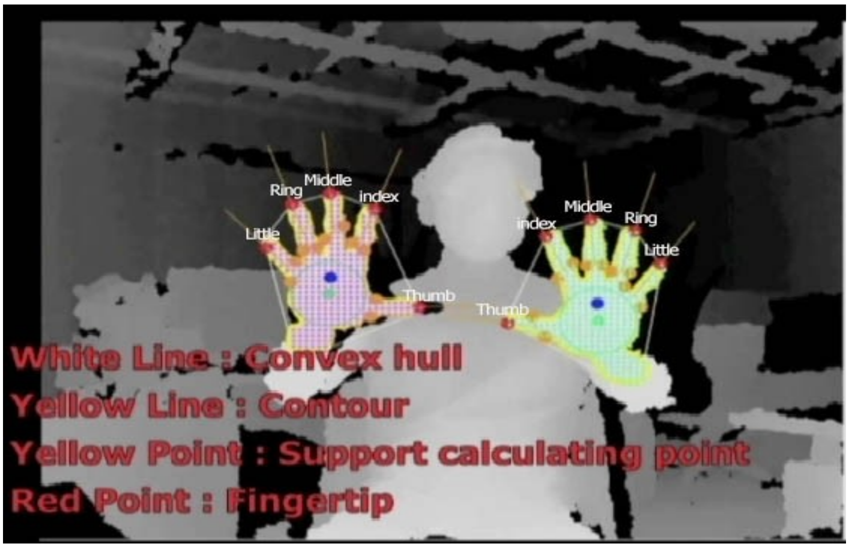


Figure 3.8. Finger Identification Results

When the input data source changes, it is always performed to check the process of the sensing hand and finger. If the next frame in which the same object has the little bit of transformation compared with the previous frame exists still, all attributes of this object are mapped in the old Frame. And the results is shown in figure 3.8.

3.3 Hand Gesture Recognition Techniques

3.3.1 Hand Gesture Recognition

Generally, the signal language is comprised of 3 parts. It is its finger spelling and level word code vocabulary and non-passive feature.

Therefore, the hand gesture awareness (HGR) forms the base in translating the signal language. Recently, the gesture recognition technology binds the depth of 3D image which produces the object distance information normally and is captured by 3D camera or a series of camera.

In [10], Van den Bergh et al. used a Time-of-Flight camera to enhance recognition. Their system could recognize six gestures: open hand, fist, pointing up, L-shape, pointing at the camera, and thumb-up. The RGB-based recognition showed an accuracy of 99.54%, and the depth-based recognition showed an accuracy of 99.07%, while the combination of the two methods showed 99.54%. As the combination of two methods did not provide the improvement which is important in the accuracy, it proposed because the base awareness can form the base of HGR system in kindly, particularly, the hand segmentation deeply.

Numerous researches have been done with Kinect since its launch in 2010. However, only a few systems were developed for HGR, and only a few gestures were recognized. Yang et at. [19] proposed a gesture recognition system using depth information provided by Kinect, and implemented in a media player application. The passive orbit was tested by 3D feature vector and the gesture was recognized by the hidden Markov model (HMM). This system proved the applicability using the Kinect for the gesture recognition of the contact-less user interface (UI). The system could not find the fingertip. Therefore, that was limited in order to recognize the sports gesture after the wave and up/down, left / right, and the forward / backward only.

The method for tracing the center of the fingertip and palm by using Kinect was described by the Raheja and so on. When the finger was expanded, the accuracy of the sensing fingertip was nearly, 100% and that of palm center was 90%. However, this method did not attempt the gesture recognition.

He et al. [21] suggested the access using the depth data provided as Kinect in order to detect the fingertip. After the fingertip was discovered, the sports was recognized and to mouse-click it was tested by the popular game Angry bird and only that recognized one gesture.

3.3.2 Recognition Methods

First of all, if the finger is confirmed as their names, we become the readiness for the gesture recognition. The gesture is passed through the finger intensity, that is three unit layers of the classification machine, and finger name collecting and vector Y-matching. The angle A between two vectors V_1 and V_2 is calculated as

$$A = \arccos \frac{\vec{V_1} \cdot \vec{V_2}}{\left\|\vec{V_1}\right\| \left\|\vec{V_2}\right\|}.$$

Three Layers of Classifiers

1) Finger counting classifier: the gesture was classified by the number of expansion fingers as the first. And it sent to the second opinion corresponding to floor of the classification machine.
2) Finger name collecting: The gesture more and more furthermore is classified as the name called the expansion finger. If the combination of the expansion finger of one more gesture is unique among all gestures, the know process is concluded and the meaning of the gesture is displayed. On the other hand, the gesture is sent to the corresponding to third layer of the classification machine.
3) Vector matching: The direction vector of all expansion fingers is measured and all pair-wise angles are calculated between the expansion finger. The meaning of the gesture is classified according to this angle now. And it enters upon and the meaning of the gesture is allocated and is displayed in the screen.

An example of recognizing "I Love U" is illustrated in Figure 3.9.

Figure 3.9. Recognition Example

First the system determines that the number of fingers in this gesture is three. Then the finger names of "thumb", "index finger" and "little finger " are found. The classifier in the third layer calculates the angle of each two finger vectors are within a range of 30 degree to 60 degree.(see A and B of figure 3.9) Thus the gesture is recognized as "I Love U" in the Popular Gesture scenario, while in the Numbers scenario it is recognized as the number "Three" through a similar process.

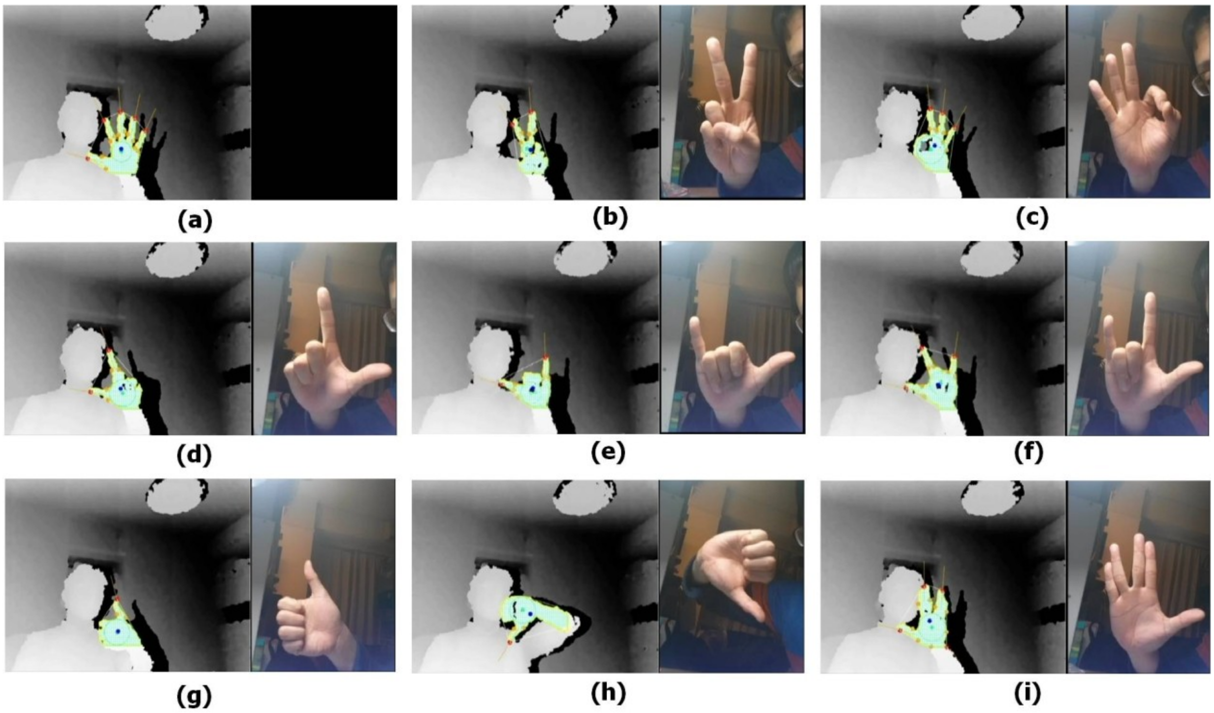


Figure 3.10. Recognition Result. (a) is the "Start Gesture" for the purpose of calibration; (b) is "Victory"; (c) is "Okay"; 5(d) is the letter "L"; 5(e) is the letter "Y"; (f) is "I love you"; (g) is "Thumb-up"; (h) is "Thumb-down"; 5(i) is the Star Trek gesture, which means "Live long and prosper".

The Popular Gesture scenario consists of nine gestures: " Start Gesture" (open hand), "Thumb-up", "Thumb-down","Victory", " Okay", "I love you", the letter "L", the letter "Y", and "Live long and prosper", The HGR results of Popular Gesture are shown in Figure 3.10.

Chapter 4.

Natural User Applications

4.1 System Overview

The KINECT is comprised of IR light, Depth Image CMOS for depth and color Image CMOS for RGB data (see Figure 4.1). Sensors compute the depth of the hand while the RGB camera is used to capture the images. The system is comprised of display and Kinect sensor. Because of Kinect's resolution limitation(640x480), the distance from hands to sensor is limited(see Figure 4.1).

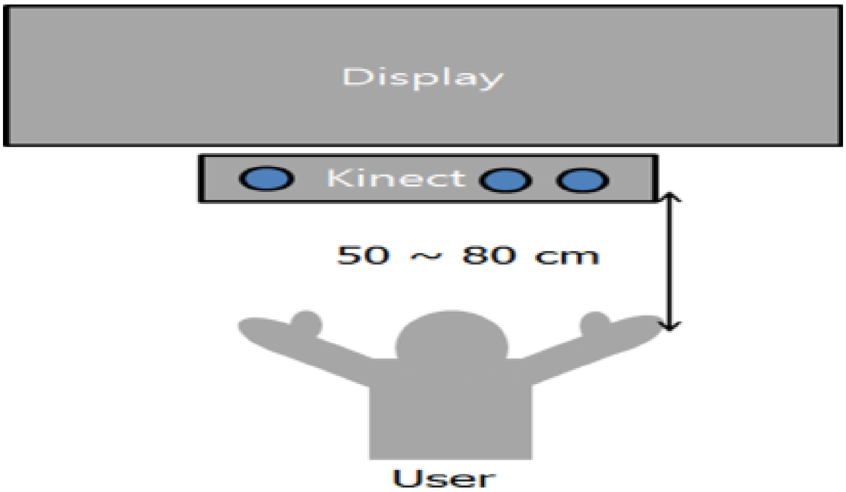


Figure 4.1. System Configuration

In our interface, there are three type applications (see figure 4.2). First is drawing interface using fingertip. Drawing interaction is made when two fingertips are detected. Second is image manipulation interface. It is well mapped multi-touch interface using fingertip.

It provides selecting, resizing and rotating natural interaction with images. Third is video manipulation interface. It provides Mapping, Selecting, Time shift and removing surface interaction with movie media objects.

All of them implemented natural interact with object and well mapped interface.

Figure 4.2. Applications

4.2 Drawing Interface

This interaction is mapped of finger paint in smart phone. It can make finger paint in air. This interactions are made when user's fingertip enter detection area(see Figure 1), the system tracks the fingertip. It can make drawing interaction in air when one fingertip is detected(see top image of figure 5). Thereafter, system starts to capture user's fingertip gestures.

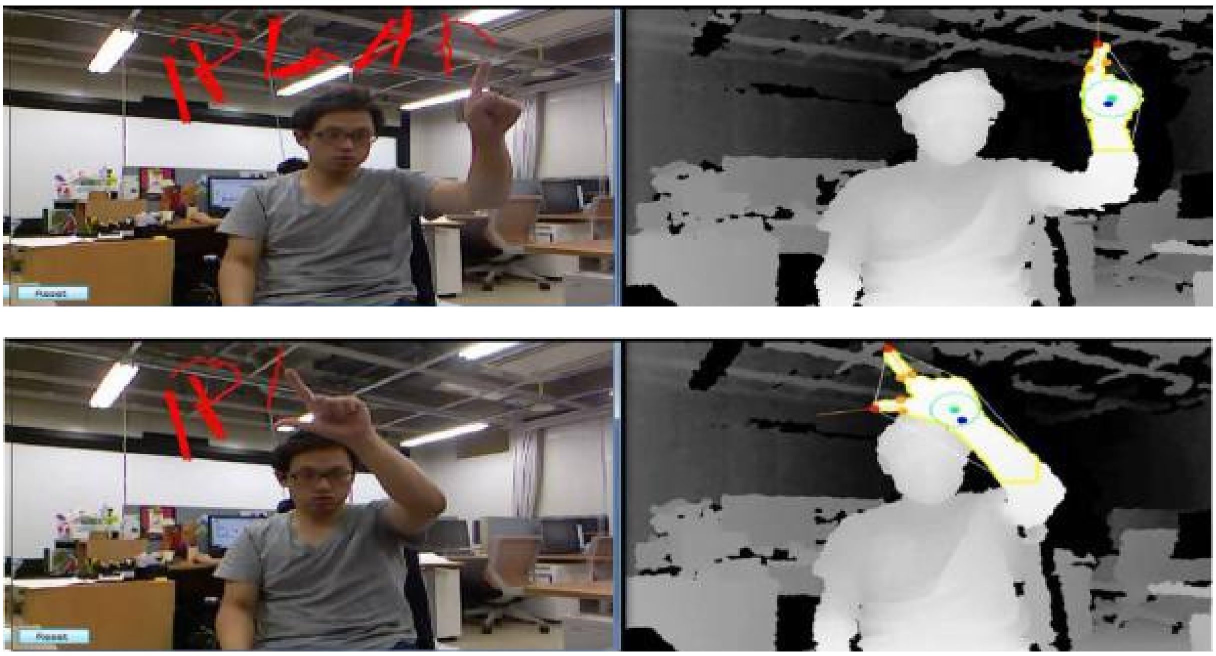


Figure 4.3. Drawing Interaction with fingertip tracking

User can make drawing line interaction in air when more two fingertips are detected(see bottom image of figure 4.3). It is possible even in case of being both hands.

4.3 Image Manipulation Interface

In order to resize the image, system tracks fingertips number. It changes to resizing interaction mode when the system detects two fingers. In resizing interaction mode, system computes distance of each fingertip. It implements multi-touch function of the recent smart device with the hand gesture base. As shown in Figure 4.4, if the distance of each fingertip becomes far to the certain value,

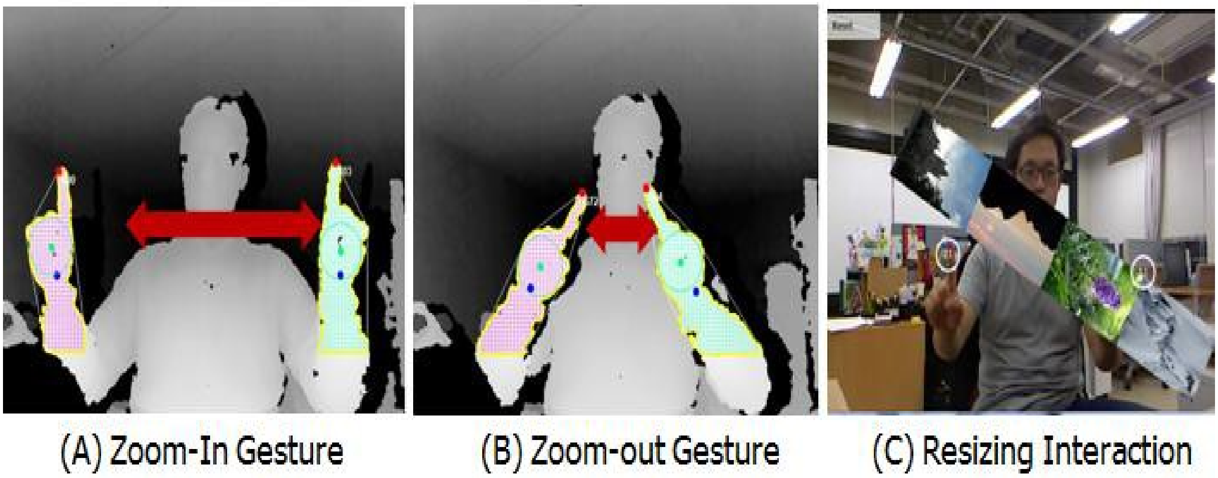


Figure 4.4. Resizing Gesture and Interaction

(to be found experimentally), zoom-in interaction will be performed. Consequently, size of the image is expanded. On the contrary, as shown in Figure 4.4, if the distance of each finger becomes close at the certain value, zoom-out interaction will be performed. Size of the image is reduced.

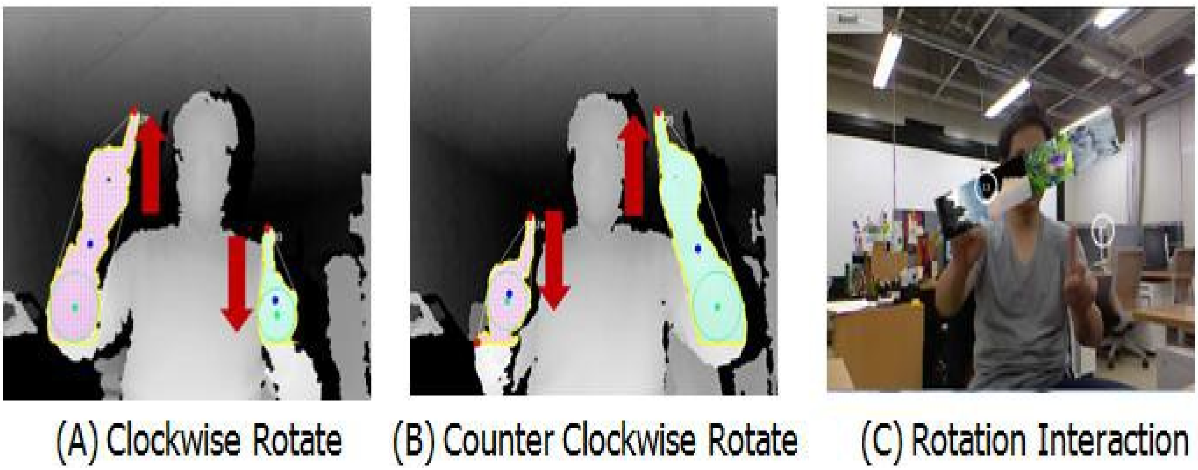


Figure 4.5. Rotate Gesture and Interaction

In order to rotate image, same processes with resizing mode are required. For rotate interaction, detection of two fingers are required. System computes each fingertip's position. If the left hand is raised to the upside and the right hand is downwards let down, then the image is rotated by the clockwise. Even in case of two fingertips in one hand, It is able to make rotate interaction as well. On the contrary, as shown in Figure 4.5, If the right hand is raised to the upside and the left hand is downwards let down, then the image is rotated by the counter clockwise.

4.4 Video Manipulation Interface

This is an interface that focuses on videos. It lets you place that predefined videos in 3D space on top of the RGB image captured by the Kinect. A simple gesture lets you time-shift the video and another closes the video. These are the input gestures that are currently implemented.

First is defining a video space with 4 fingers that 2 on each hand starting a new video inside that space. It is mapping interaction (see figure 8). Second is Opening the right hand while pointing at a video with one finger of the left hand closes the video. It is removing surface interaction. Third is Opening the left hand puts the video into time-shift mode. Moving the right finger on the x-axis forwards or rewinds the video. It is time shift interaction. For making interaction with each surface, it has to select specific surface. Selecting interaction is made by one fingertip detection. If the fingertip is on the specific surface, the surface is selected. Then the surface border color changed to white. System can make interaction with the surface.

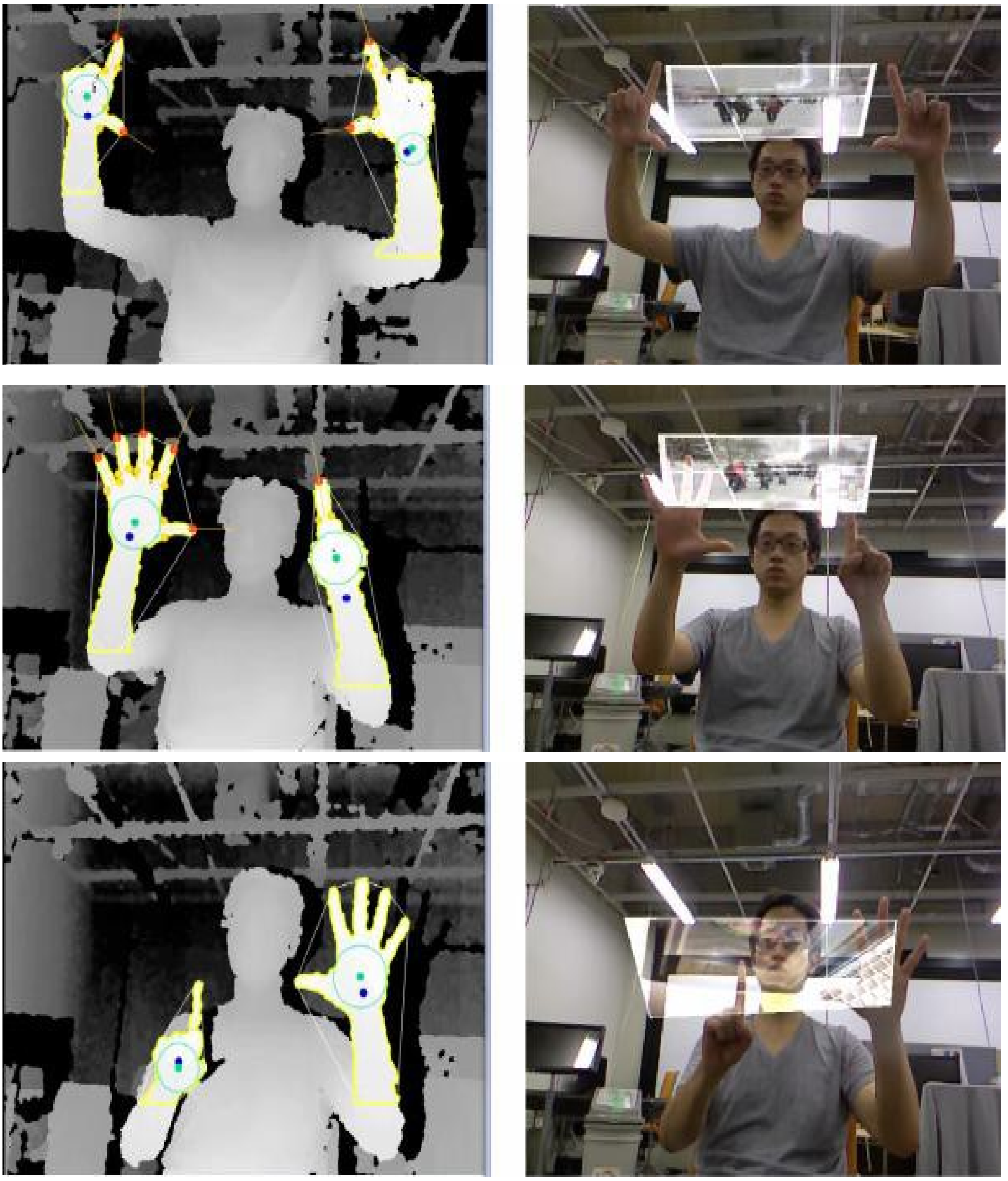


Figure 4.6. Mapping, Time shift and Removing Interaction

Chapter 5.

Evaluation

5.1 User Questionnaire

We did questionnaire for evaluating each interface. The evaluation was progressed with 15 men and 5 women.
We illustrated simply about the function of each interface. 20 users evaluated each interface on three items such as Intuitive use, Ease to learning and Natural interaction. Three kinds of choices were made for each item(Good, Average, Bad).

Table 1. Percentage of ‘Good’

Item \ Interface	Air-Drawing	Image Manipulation	Video Manipulation
Intuitive usage	80%	90%	85%
Ease to learning Gesture	85%	95%	80%
Natural interaction	90%	95%	90%

Table1 shows that each interface is equipped with the element for the natural interaction. Particularly, ‘Image Manipulation’ was high percentage than the other interface. Because of the multi-touch interaction is widely used and familiar to users.
Result shows three interfaces were intuitive and easy to use on average. Therefore, we conclude that the proposed interfaces use natural interaction with finger gesture.

5.2 Gesture Experiment

To test this system, four people are asked to perform each of the nine gestures for 100 times: 50 times with the left hand and 50 times with the right hand.

Table 2. HGR Accuracy

Gestures	Start	Thumb-up	Thumb-down	L	Y
Accuracy(%)	99	89	88	94	90
Gestures	Okay	Victory	Star Trek	I Love U	
Accuracy(%)	84	89	80	85	

Each, the mean accuracy of the gesture is calculated and shows up in the table 2.
The accuracy is nearly 100% and the name of the finger is completely confirmed for "Start Gesture".
The best case is the character "L" having the accuracy of 94 %. The worst case is the Star Trek gesture having the accuracy of 80%. The Star Trek gesture the number of fingers changes looked at in Figure 3.10 according to the user and the angle calculation is needed. Therefore, the performing speed is slower than the other gesture. The accuracy is low. However, in the arbitrary environment, all gestures had more than 80 percents result. Furthermore, when making the same gesture with the both hands, the accuracy is more increased.

Chapter 6

Conclusion and Future work

In this thesis, we presented fingertip tracking and hand recognition techniques. It presented natural user interfaces that allows user to drawing, image manipulation and video manipulation through fingertip gestures as well.
In the developed system, there is particularly, the detection range in 0.5m between 0.8m that hand is distinguished by the depth information with the background. If the distance is included between 2 duster means within the critical value to be used in obtaining two clusters of the hand pixel, the K- method unites two clusters and only one is detected. And the contour tracing algorithm is used that it enters upon and Graham scan algorithm is used in determining the convex hull of the hand in detecting the hand contour. Then, the candidate fingertip is collected by calculating the common pixel in the both convex hulls and hand contour. And it is detected since the fingertip applies the three-point alignment algorithm. The finger name is determined according to their relative distances and the direction vector is allocated to the finger. By passing them through a series of classification machine including three unit layers the gesture is recognized: the finger intensity classification machine and finger name collecting and vector Y-matching Two scenarioes are comprised in order to be perceived: as to we, the gestures or consisting of the nine popular gestures, other thing is the numbers consisting of '1' to '9'. The awareness has the accuracy of the smallest 80% for one hand.
Our work contributes a novel natural user interface for controlling three type of interface as well. The merits of our system are as follows. First, we implemented multi-touch image manipulation using fingertip gestures. Because of using the fingertip gestures, our system provided more intuitive and natural way to control images. Second, we implemented well mapped video manipulation and drawing interaction using fingertip gestures. It is intuitive and natural way as well. Third, our system overcame limitation of Vision-based tracking problem under some natural conditions like fast fingertip motion, cluttered background, poor light condition. Finally, performance was improved by using k-means clustering algorithm and convex hull methods together when tracking more than two hands.In future work, we plan to make marker-less augmented reality 3d interface using depth data. It will need body tracking such as head, arm, eye and PTAM tracking for augmenting in marker-less. Finally, we plan to make 3d augmented interface and interaction for more natural way to interact with computer.

Acknowledgements

I am heartily thankful to Professor Jiro Tanaka, my thesis supervisor, for his many valuable suggestions, precise directions, and kind encouragement.

I am also grateful to all the members of the IPLAB, Interactive Programming Laboratory, University of Tsukuba, for giving me many opportunities to discuss my research with them and helping me in my experiments.

I would like to acknowledge the financial support provided by Docomo Scholarship Foundation.

Reference

[1]Z. Ren, J. Meng and Z. Zhang. “Robust Hand Gesture Recognition with Kinect Sensor”, MM’11, Proceedings of the 19th ACM international conference on Multimedia, 2011.

[2]D. Wigdor and D. Wixon. “BRAVE NUI WORLD”, Published by Morgan Kaufmann,pp.9-15, 2011.

[3]S.Lenman, L.Bretzner and B.Thuresson. “Computer Vision Based Hand Gesture Interfaces for Human-Computer Interaction”, Technical Report CID-172,Center for User Oriented IT Design, pp.3-4, 2002.

[4]V. Pavlovic, R. Sharma and T. Huang. “Visual Interpretation of Hand Gestures for Human-Computer Interaction”, A Review. In IEEE Transactions on Pattern Analysis and Machine Intelligence. pp.677-695, 1997.

[5]A. Ahmad. “A Novel Accelerometer-based Gest’ure Recognition System”, University of Toronto, pp. 2-3, 2010.

[6]M. Axel. “Hand Gesture for HCI”, University of Simon Fraser University, 1996. http://xspasm.com/x/sfu/vmi/HCI-gestures.htm

[7]H. Caroline, S. Gerda and O. Kees. “An Intuitive Two-handed Gestural Interface for Computer Supported Product Design”, pp.2.

[8]D. McNeill. “So you think gestures are nonverbal?”, Psychological Review, vol92 (3), pp. 350-373, 1985.

[9]H. Nanda and K. Fujimura. “Visual tracking using depth data”, Conference on Computer Vision and Pattern Recognition Workshop, pp.37, 2004.

[10]A. Argyros and A. Lourakis. “Binocular hand tracking and reconstruction based on 2d shape matching”, ICPR, pp.207-210, 2006.

[11]J. Wachs, M. Kölsch, H. Stern, and Y. Edan. “Vision-based hand-gesture applications”, Communications of the ACM, Vol. 54 Issue 2, pp.60-70., 2011.

[12]C. Chua, H. Guan, and Y. Ho. “Model-based 3d hand posture estimation from a single 2d image”, Image and Vision Computing , pp.191-202., 2002.

[13]B. Stenger, A. Thayananthan, P. Torr and R. Clipolla. “Filtering using a tree-based estimator”,In Proc. of IEEE ICCV, 2003.

[14]B. Mohamed and B. Thomas. “A Low-Cost Natural User Interaction Based on a Camera Hand-Gestures Recognizer”,HCI2011.

[15]Y. Sato, Y. Kobayashi and H. Koike. “Fast Tracking of Hands and Fingertips in Infrared Images for Augmented Desk Interface”, International Conference on Automatic Face and Gesture Recognition, Grenoble, 2000.

[16]J. Rehg and T. Kanade. “Digiteyes: Vision-based human hand tracking”, Technical Report CMU-CS-93-220, School of Computer Science, Carnegie Mellon University, 1993.

[17]H. Christian and B. Francois. “Bare-Hand Human-Computer Interaction”, Technische Universitat Berlin, pp. 2-4.
[18]Y. Iwai, K. Watanabe, Y. Yagi and M. Yachida. “Gesture revcognition using colored gloves”, IEEE Int. Conf. Pattern Recognition, vol.A, pp.662-666, Viena, 1996.
[19]Y. Cheoljong, J. Yujeong, B. Jounghoon, H. David and K. Hanseok. “Gesture recognition using depth-based hand tracking for contactless controller application”, In Consumer Electronics (ICCE), 2012 IEEE International Conference on, pp. 297 -298, 2012.
[20]J. Raheja, C. Ankit and S. Singal. “Tracking of fingertips and centers of palm using KINECT”, 2011 Third International Conference on Computational Intelligence Modelling Simulation, pp. 248-252, 2011.
[21]G. He, S. Kang, W. Song and S. Jung. “Real time gesture recognition using 3D depth camera”, In Software Engineering and Service Science (ICSESS), 2011 IEEE 2nd International Conference on, pp.187-190. 2011.
[22]R. Graham. “An efficient algorithm for determining the convex hull of a finite planar set”, Information Processing Letter, pp. 132-133, 1972.
[23]OpenNI organization. OpenNI User- Guide, 2011.
[24]V. Frati and D. Prattichizzo. “Using Kinect for hand tracking and rendering in wearable haptics”, IEEE World Haptics Conference, pp. 317-321, 2011.
[25]R. Pradehan and S. Kumar. “Contour Line Tracing Algorithm for Digital Topographic Maps”, International Journal of Image Processing (IJIP), Volume 4, Issue 2, 2010.