

國立交通大學

光電工程學系碩士班

碩士論文

提升三維多點互動系統使用者辨識能力之

多圖騰演算法開發



High-freedom 3D Interactive System

for Multi-user/Multi-touch by Multi-mark Algorithms

研究生：黃書怡

指導教授：戴亞翔 教授

黃乙白 教授

中華民國一百零一年七月

提升三維多點互動系統使用者辨識能力之
多圖騰演算法開發

High-freedom 3D Interactive System
for Multi-user/Multi-touch by Multi-mark Algorithms

研究生：黃書怡

Student : Shu-Yi Huang

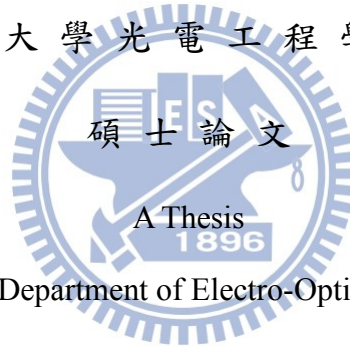
指導教授：戴亞翔

Advisors : Ya-Hsiang Tai

黃乙白

Yi-Pai Huang

國立交通大學光電工程學系碩士班



Submitted to Department of Electro-Optical Engineering

College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

In Electro-Optical Engineering

July 2012

Hsinchu, Taiwan, Republic of China

中華民國一百零一年七月

提升三維多點互動系統使用者辨識 能力之多圖騰演算法開發

碩士研究生：黃書怡

指導教授：戴亞翔教授
黃乙白教授

國立交通大學電機學院

光電工程學系碩士班

摘 要

近年來 3D 顯示科技的蓬勃發展，使得影像跳脫傳統框架，栩栩如生地出現在我們眼前。此外 2D 觸控面板的爆炸性成長，使得多點觸控改變了使用者對於人機介面的使用習慣與期待。鑒於立體影像已蔚為風潮，加上智慧型手機與平板電腦的普及，未來將兩者結合之 3D 智慧型手機與 3D 平板電腦已然成為趨勢。

然而要使 3D 技術普及化，除了 3D 顯示設備外，擁有一個友善的人機互動介面更是推動此項技術的關鍵，因此 3D 互動系統將成為下世代的關鍵技術。為了達成平面與三維空間的連續互動介面，我們選擇了內嵌光感測元件的架構，提出適用於薄型可攜式電子顯示器之多點、以及多人三維立體近距離互動系統。

在以往的 3D 多點/多人互動系統中，使用者間部分重疊的情況始終為一大困擾，因此，經由所提出的演算法，包括時序性互動演算法、多圖騰式演算法等，多位使用者可同時的與系統進行三維互動並被成功的分辨。最後，所提出的演算法被證實於內嵌光感測元件的 4 吋面板上。多環形圖騰演算法提升了使用者的辨識能力，解決了使用者間 40% 之重疊情形，同時獲得個別使用者之二維的位置 (x, y) 以及深度資訊 (z) ；另外，藉由多 T 形圖騰演算法，不僅是使用者的三維資訊 $(x, y, \text{ and } z)$ ，其傾斜與旋轉角度 (θ, ϕ) 等資訊也能夠被準確地截取出來。

High-freedom 3D Interactive System

for Multi-user/Multi-touch by Multi-mark Algorithms

Student: Shu-Yi Huang

Advisors: Ya-Hsiang Tai, Yi-Pei Huang

Institute of Electro-Optical Engineering

National Chiao Tung University

Abstract

Recently, the widespread of 2D multi-touch user interfaces has become a part of our lives. Meanwhile, the increasing popularity in 3D images dramatically changes the user experience. Therefore, by the fact of great mass fervor in smart phone and tablet computer, 3D smart phone and 3D tablet is undoubted the battle field of next generation technology.

However, a portable 3D device cannot be all-pervading if there is no friendly 3D multi-interactive user interfaces. Hence, in order to establish a continuous interaction space from 2D surface to 3D near field distance, the embedded optical sensor based structure is chosen in our approaches. However, for multi-user/multi-touch interactive systems, the users' overlapping has been a serious issue in user recognition. Therefore, by the proposed methods, the sequential-lighting method and the multi-mark method, user overlapping issue might be solved. Three algorithms, the sequential-lighting algorithm, multi-ring mark algorithm, and the multi-T mark algorithm, and the accordant light marks are proposed to provide distinguishable 3D user interfaces. Finally, the feasibility of the proposed algorithms was demonstrated on a 4" panel with in-cell optical sensors. The ability of user identification can be improved to 40% overlapping by the multi-ring mark algorithm. Meanwhile, 5-axis information (x , y , z , θ , and ϕ) including 3D coordinate and orientation can be obtained through the multi-T mark algorithm.

誌 謝

誠摯的感謝指導老師戴亞翔教授、黃乙白教授，給予我進修碩士的機會，讓我能藉此機會提升自我能力，不僅在於研究上、更對於未來的方向提供了寶貴的意見。也感謝謝漢萍教授在碩士生涯的諸多指點。此外，由衷的感謝各位口試委員提供的精闢見解，使本論文更加完備。

在此感謝王國振學長兩年來亦師亦友的教誨，其對於研究的嚴謹態度，實為學弟妹的典範，不僅讓論文更加完善流暢，也讓我在研究方法以及報告技巧上學習成長了許多。尤其要感謝馬明青學姐、董軒宇學長、以及馬志堯學長在我碩一時不厭其煩的教導，帶領我渡過一次又一次的考驗。也感謝宣賀及正昌學弟在實驗及研究上提供了十分寶貴的建議以及幫助，讓論文及實驗能順利完成。

在實驗室的日子裡，特別感謝綺文及博凱對我的照顧與陪伴，也感謝有林芳正、陳致維、蔡韻竹、張育誠、蔡柏全、王奕智、許精益、丁志宏、任台翔、簡又儀、謝博元等博班學長姐，提供專業的意見及教導，也感謝思頤、博詮、昌毅、立偉、濟宇、禹辰、子寬等學長姐在我碩一這年的照顧。而碩二一起認真與玩樂的同學們，白諭、秉彥、柏皓、上翰、哲軒、博鈞，也非常感謝你們在這兩年生活上的幫助與分享。且要對實驗室的雅惠、純敏、蓮方、以及穎佳四位助理小姐和學弟妹表達感謝，讓我們能沒有後顧之憂的做研究且讓實驗室充滿歡樂的氣氛。

最後，我要感謝我們家人，謝謝你們這些年來的支持與鼓勵，並感謝我的朋友們冠緯、祿盛、雅晴等，不論在何時都願意為我加油打氣，讓我能夠再次向前出發，朝目標邁進。感謝我所認識的各位，讓我能順利完成碩士學位，這份喜悅與榮耀我將與我認識的各位分享。

Contents

摘 要.....	i
Abstract.....	ii
誌 謝.....	iii
Figure Captions.....	vi
Chapter 1 <i>Introduction</i>	1
1.1 Preface.....	1
1.2 2D Multi-touch Technologies	2
1.3 Motivation and Objective	10
1.4 Organization of the Thesis	12
Chapter 2 <i>Prior Arts of 3D interactive Systems</i>	13
2.1 2D Touching of 3D Stereoscopic Images	13
2.2 3D Virtual Touch Technologies.....	15
2.3 Machine-based	16
2.3.1 Haptic Workstation™.....	16
2.3.2 Wii.....	17
2.4 Camera-based.....	19
2.4.1 Microsoft 3D Gesture Interface.....	20
2.4.2 WorldViz PPT	21
2.5 In-cell Optical-based.....	23
2.5.1 ThinSight.....	23
2.5.2 Sensible Backlight	24
2.5.3 Directional Image Sensor.....	26
2.5.4 Color-filter-based Sensing	27
2.5.5 Multi-mark Based	28
2.6 Summary of 3D Virtual Touch Systems.....	30
Chapter 3 <i>Structure and Algorithms</i>	32
3.1 Overall Structure	32
3.2 Sequential-lighting Method	34
3.2.1 Synchronization	36
3.2.2 Noise Suppression.....	36
3.2.3 Full Search Method.....	37
3.2.4 Adaptive Window Method.....	37
3.3 Light-mark Method.....	40

3.3.1	Light-mark Design	41
3.3.2	Multi-ring Mark: Overlap Improvement.....	43
3.3.3	Multi-T Mark: 5-axis Information (x, y, z, θ , ϕ).....	49
Chapter 4	<i>Experiments and Results</i>	63
4.1	Experiment Setup.....	63
4.2	Sequential-lighting Method	64
4.3	Multi-ring Mark	67
4.3.1	Resolution in Z-axis.....	68
4.3.2	Overlapping with Normal Incident	70
4.3.3	Overlapping with Tilt Light-pen	73
4.4	Multi-T Mark	77
4.4.1	5-axis Information and User Identification.....	78
4.4.2	Resolution in Depth (z).....	78
4.4.3	Resolution in Orientation (θ , ϕ).....	79
Chapter 5	<i>Conclusion and Future Work</i>	82
5.1	Conclusion	82
5.2	Future Work	84
Reference	87

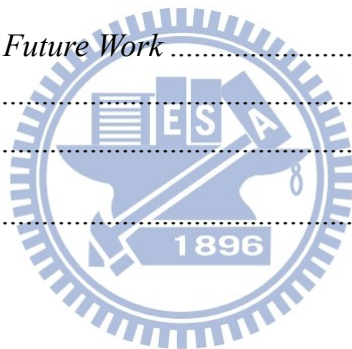


Figure Captions

Fig. 1.	Trend of interfaces and 3D displays technologies.....	2
Fig. 2.	First 2D touch product with multi-touch user interface - JazzMutant’s Lemur music.[4]	2
Fig. 3.	Principle of capacitive sensing: when a voltage apply, (a) there exist a capacitance between X electrode and Y electrode. (b) once a finger touch the surface, there is a distortion in electrostatic field and the mutual capacitance is reduced.	3
Fig. 4.	Working principle of position determination in projective capacitive sensing system. [6].....	4
Fig. 5.	Illustration of a resistive touchscreen. [6]	4
Fig. 6.	A schematic representation of conventional optical touch technology.	5
Fig. 7.	General set-up of a Frustrated Total Internal Reflection (FTIR) system which is based on optical total internal reflection within an interactive surface; as a user touches the screen, the light escapes and is reflected at the finger’s point of contact, which is then detected by an IR camera at the back of the pane. [8]	6
Fig. 8.	General set-up of a Diffuse Illumination (DI) system which is similar to that for FTIR but with IR lighting placed behind the projection surface; moreover, DI is capable of tracking and identification of objects by using their shape or fiducially printed on their bottom surfaces. [8]	7
Fig. 9.	General set-up of a Diffused Surface Illumination (DSI) system which utilizes small number of (two or three) infrared illuminators and evenly distributed the light across the screen surface. [8]	7
Fig. 10.	Schematic of a camera-based optical touch technology. [13]	8
Fig. 11.	General cross-section of LCD in-cell optical touch panel. [14].....	9
Fig. 12.	2D in-cell light-sensing can be achieved by (a) sensing the shadow of touching objects caused by blocking the regional ambient light where (b) the objects’ positions reveal lower gray level [14].....	9
Fig. 13.	(a) The prototype of capturing the reflection of IR backlight is employed in 2D in-cell light-sensing where (b) the objects’ positions exhibit higher gray level. [15] 9	
Fig. 14.	Continuous interaction space from 2D touch surface to 3D virtual touch. [16].....	10
Fig. 15.	A 3D virtual touch prototype demo – 3D Fishing Game. The video link: http://youtu.be/ipJeMAdZVCI	11
Fig. 16.	Depth perceptions are (a) behind and (b) in front of the display plane. [17]	14
Fig. 17.	Main issue in 2D touch interface with stereoscopic data is that a user is either focus on the finger or the 3D virtual objects which both degrades the visual perception. [18].....	14

Fig. 18.	5 axes (x, y, z, θ , and ϕ) information in position and orientation.	15
Fig. 19.	Haptic Workstation and Head-Mounted Display (HMD) for rendering the virtual cockpit. [19]	17
Fig. 20.	Wii utilizes an infrared camera to sense a series of LEDs in the sensor bar, orienting the controller in space.[22].....	18
Fig. 21.	Acceleration Sensor: this chip provides all the motion sensing. A piece of silicon is evenly placed between two capacitors in electric field. Once the user accelerates the controller in one direction, it causes the silicon to bend in the other and causes the electric field to change, which translates into motion.[22]	18
Fig. 22.	Limited field of view in cameras causes the blind range, unworkable region, in near-distance.	19
Fig. 23.	Intrinsic characteristics in camera-based system render a tradeoff between discontinuous working space and large system volume which forbids the integration to a display.	20
Fig. 24.	Schematic of the Microsoft 3D Gesture Interface.	21
Fig. 25.	Microsoft 3D Gesture Interface demonstrates that the “draw” command is activated when a “close loop” of shadow image is detected.	21
Fig. 26.	3D coordinate is acquired by tracking (a) active LED markers by at least two to (b) four tracking cameras.	22
Fig. 27.	(a) Basic construction of ThinSight, and (b) applications in ThinSight display including 2D finger multi-touch, data transformation between a mobile device and the display, and 3D gesture interaction through a device which can cast a beam of infrared light onto the display. [15].....	23
Fig. 28.	(a) Working principle and optical structure of the sensible backlight system, and (b) layout of backlight with RGB LEDs integrating with IR photo transmitters. [27] ..	25
Fig. 29.	(a) Sensor image of touch, (b) sensor image of hover, (c) extracted touch points, and (d) extracted hover points. [27]	25
Fig. 30.	(a) Concept of a 3D input device using a directional image sensor, and (b) structure of thin-film lateral diode and light shielding layers to create a directional field-of-view. [28]	26
Fig. 31.	One dimensional representation of the response of two directional sub-pixels. [28]	27
Fig. 32.	System structure of color filter based sensing.	28
Fig. 33.	A detected image under red color filter. Undesired blue light penetrates a red color filter and overlapping of red and blue spots occurs.	28
Fig. 34.	Designed light marks of solid T shape, X shape, and circle.....	29
Fig. 35.	Complex overlapping conditions need to be processed case by case which results in heavy computation requirement.	29

Fig. 36.	Different categories of 3D interactive systems have different characteristics that match for different applications.	31
Fig. 37.	Hardware structure of the embedded optical sensors display.....	32
Fig. 38.	Concept of the proposed 3D virtual touch system with multi-user collaboration... ..	33
Fig. 39.	A 3D interactive system model with sequential mark (a) light source one was turned on during frame = $n \times k + N = 3k + 1$ where n equals to the number of users and N is the sequence of the user (b) light source two was turned on during frame = $3k + 2$ (c) light source three was turned on during frame = $3k + 3$, $k = 0, 1, 2, \dots$	34
Fig. 40.	Flow chart of Sequential-lighting algorithm.	35
Fig. 41.	Full search method: 2D coordinate (x and y) is defined by the maximum accumulation within the N by N search block.	37
Fig. 42.	Adaptive window method is employed to detect the depth value (z), where a look-up table for window sizes is built according to the experiments.	38
Fig. 43.	Hand gesture models. [33]	40
Fig. 44.	Concept of (a) multi-touch for single user through light gloves [36] with Multi-ring mark and (b) multi-user interaction with 5-axis information by Multi-T mark approach.....	41
Fig. 45.	Principle of mark design including out-mark and in-mark, where out-mark determines the perceivable degrees of freedom, and in-mark is utilized for user identification. (a) Multi-ring mark design, (b) Multi-T mark design.....	43
Fig. 46.	Flow chart of Multi-ring mark algorithm.....	43
Fig. 47.	Structuring elements for morphological thinning. Ones and zeros stand for foreground and background pixels and the blanks can be either one or zero which we don't care about. At each term, the image is first thinned by B^1 , then B^2 , as soon. The process is repeated until none of the thinning produces any further change.	45
Fig. 48.	Light pens are positioned at 50mm and detected by the embedded optical sensors. Output images after image enhancement steps and thinning operation from the 1st term until convergence are shown.....	45
Fig. 49.	(a) Ring filter are constructed according to the size of out-mark at different heights, and (b) shows the structure element of ring filter at $z=0$	46
Fig. 50.	In candidate detection, the accumulation results at different ring filter (z) are divided by the corresponding normalizing factors. The local maximum with its value close to 1 will be defined as the candidates.....	47
Fig. 51.	In-mark window for in-mark extraction.....	48
Fig. 52.	Ranking process is able to define different users.	48
Fig. 53.	Flow chart of Multi-T mark algorithm.	49
Fig. 54.	Illustration of filling holes operating on a multi-T mark image.....	51

Fig. 55.	A practical example for Filling holes. (a) is the binary image by thresholding the captured image, and (b) shows the complement of the binary image in (a) for use as a mask image. (c) is the marker image, F , generated by Eq. (6). Finally (d) shows the image with all the holes filled.	52
Fig. 56.	Connected-component labeling operator: (a) Row data of detected image converts into binary image. After the first pass, the following labels are generated as (b). The label equivalence relationships are given in (c) and then (d) merging the equivalent labels with the label value that is the smallest for a given connected region. Final result (e) in color to clearly see two different regions that have been found	53
Fig. 57.	Unique label on each connected regions are utilized to separate different marks to different binary images.	54
Fig. 58.	2D coordinate of a T mark is acquired by the full search method, where the origin of the maximum accumulation in search window (i_{max} and j_{max}) is assigned to the 2D coordinate (x and y).	55
Fig. 59.	(a) Orientation angles about z-axis is called the rotation angle (ϕ), and the tilt angle (θ) is rotated about the axis parallel to the short branch. (b) A normal direction of a multi-T mark with the definition of short and long branches.	55
Fig. 60.	(a) Rotation filters from 0 to 170 degree with 10 degree's step where a normal T with a long branch parallel to x-axis is defined to 0 degree. (b) Matrix of the rotation filter at 0 degree.	56
Fig. 61.	Branch filters from 0 to 170 degree is able to define the orientation of short branch in a multi-T mark.	57
Fig. 62.	Multi-T marks with rotation filter 30 degree indicate the rotation angles of (a) 30 degrees or (b) 210 degrees.	58
Fig. 63.	Image rotation and length of braches acquisition.....	59
Fig. 64.	Under different tilt angles, the branch ratio L/S changed accordingly. Hence $L1/S1 \neq L2/S2 \neq L3/S3$	60
Fig. 65.	A comparison of various edge detectors with marks at different heights, which is a function of z , and orientation, which is a function of ϕ	61
Fig. 66.	User identification is achieved by normalizing (N) each multi-T mark to the same size and processing Prewitt edge filter (E) to extract in-mark feature. Finally the no. of user can be identified by ranking (R) the accumulation.	62
Fig. 67.	Prototype of the 3D multi-interactive system with 4-inch sensible LCD notarizes the proposed algorithms.	64
Fig. 68.	Model of three users cross over: User 1 and 2 meet at position B and User 1 and 3 meet at position C where the overlapping might occur.	64
Fig. 69.	(a) Without applying sequential-lighting method, three light marks are captured simultaneously. At (b) point B and (c) point C, the overlapping occurs. Different	

users cannot be identified.....	65
Fig. 70. Images captured by the optical sensors with 30 Hz sampling rate under sequentially illuminating method. Each light-pen was set to be 10 Hz and synchronized with the optical sensors.....	65
Fig. 71. Multi-ring mark includes two parts: one is out-mark for tracking 3D coordinate (x, y, and z), the other is in-mark for user identification.	67
Fig. 72. Experiment results of 3-axis detection and user recognition in Multi-ring mark system.	68
Fig. 73. Look-up table (LUT) for depth determination in multi-ring mark system.....	69
Fig. 74. 5 positions for height measurement.	69
Fig. 75. Experiment results in depth detection with 5 mm step.	69
Fig. 76. Percentage of overlapping is defined as l / D , where l represents the length of overlap section and <i>Dmin</i> is the minimum diameter of the projected light marks.	70
Fig. 77. In normal incident condition, the percentage of overlapping increase as the height of the light pens increase.....	71
Fig. 78. Overlap of user1 and user2 in vertical movement.....	72
Fig. 79. Overlapping of user1 and user3 in vertical movement.	72
Fig. 80. Overlapping of user2 and user3 in vertical movement.	73
Fig. 81. Serious overlapping conditions can be attained by tilting the light pens.	73
Fig. 82. Result of two marks overlapping experiment from 0 to 100%	75
Fig. 83. Result of three marks' overlapping experiment.	76
Fig. 84. Multi-T Marks are utilized to determine 5-axis information by out-mark T and meanwhile achieve user identification by in-mark, the inner blocking strips.....	77
Fig. 85. Schematic of orientation coordinates where rotation angle (θ) and tilt angle (ϕ) are the angles rotate about the short bar of T and z-axis respectively.....	77
Fig. 86. Experiment results of 5-axis detection and 3-user recognition.	78
Fig. 87. Look-up table for depth value detection by using the length of the short branch. ..	78
Fig. 88. Result of depth measurements.....	79
Fig. 89. Result of tilt angle measurements.	80
Fig. 90. Result of rotation angle measurements.	80
Fig. 91. Summary of the proposed systems.....	82
Fig. 92. Structure for the bare-hand 3D multi-touch system.	85
Fig. 93. Working principle of directional sensor by shielding layer.....	85

Chapter 1 *Introduction*

Introduction

1.1 Preface

Recent years, the battle field of technology competition has moved from the hardware into the software. The software competition, to be more precise, refers to pursuit of a friendly user interface and a better user experience. As shown in Fig. 1, in the early 1870s, C. Sholes invented the QWERTY keyboard [1] as an input device for a computer. Until 1964, the first prototype computer mouse [2] was made to use with a graphical user interface. In the next year, the first two-dimensional (2D) touch screen was developed. However, the touch technology does not dramatically change the user habits until the bump up of the smart phone. In the beginning of 21st century, the wide adoption of smart phones and tablets have accelerated a user interface transformation and paved the way to multi-touch technologies. Multi-touch technologies started to play an important role in our lives. It became so unexpendable in today's world since Apple established multi-touch as a "must-have" technology. The result is that people of all ages expect every display to be touchable with multiple fingers.

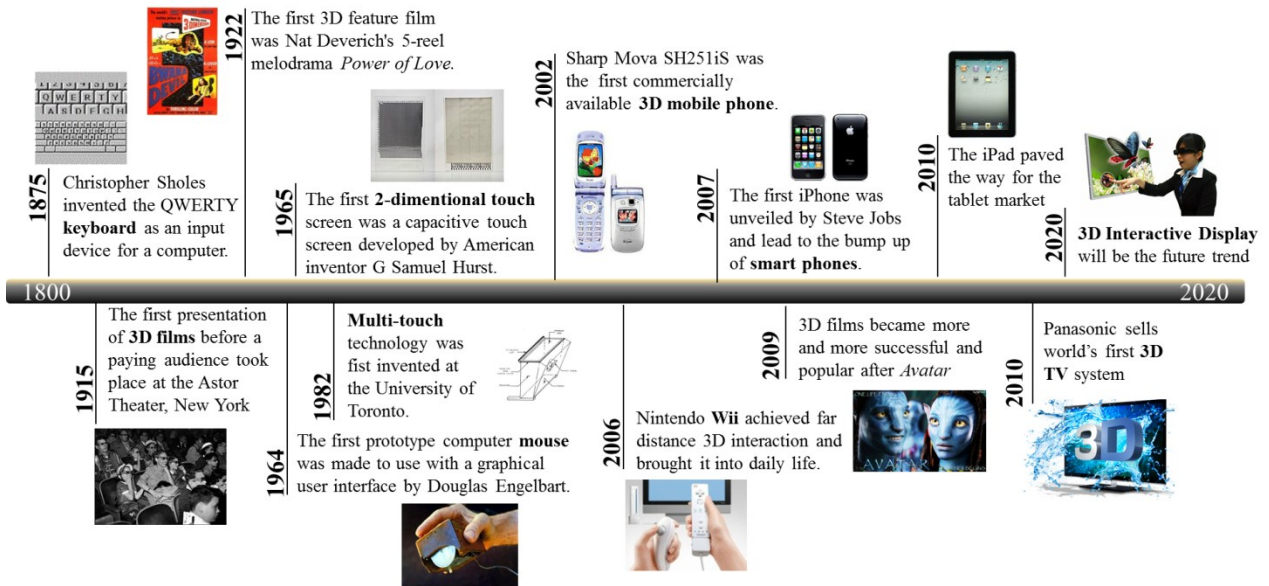


Fig. 1. Trend of interfaces and 3D displays technologies.

1.2 2D Multi-touch Technologies

Multi-touch technology is defined as the ability to recognize two or more contact points at the same time. [3] In 2004, the first commercial product JazzMutant's Lemur music [4], as shown in Fig. 2, brings multi-touch technology into our lives.



Fig. 2. First 2D touch product with multi-touch user interface - JazzMutant's Lemur music.[4]

At present, the mainstream of 2D multi-touch technologies are projected capacitive, analog resistive, and camera-based optical. These technologies increase the flexibility between human and machine by allowing a user to simultaneously control rotation, scaling, and translations (RST) from multi-touch gestures [5]. More

intuitive interactions and user interfaces have been achieved due to multi-touch technologies. In the following paragraphs, we'll briefly describe these technologies.

Projected Capacitive: In the projected capacitive sensing system, there is a capacitor at every intersection of each row and each column. A voltage is applied to an X-Y grid, which is an electrode consisting of a matrix of drive lines and sense lines, to form a uniform electrostatic field. As a finger or conductive stylus touches or close enough to the surface of the screen, it renders a distortion in the local electrostatic field, as shown in Fig. 3. The capacitance change at every individual point on the grid can be measured to accurately determine the touch location, as illustrated in Fig. 4. A single drive line is excited with an ac signal. The capacitance at each intersection between the drive line and each sense line is measured simultaneously. Next, multiplexer outputs the measured values and next being converted into a digital data stream by an A/D converter. Digital signal processing (DSP) is thus employed to interpret the data stream to 2D (x and y) coordinate of the touch location on screen surface.

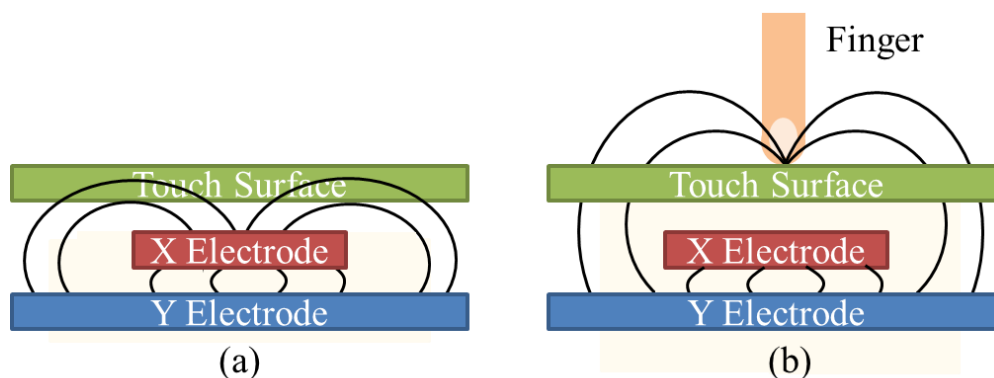


Fig. 3. Principle of capacitive sensing: when a voltage apply, (a) there exist a capacitance between X electrode and Y electrode. (b) once a finger touch the surface, there is a distortion in electrostatic field and the mutual capacitance is reduced.

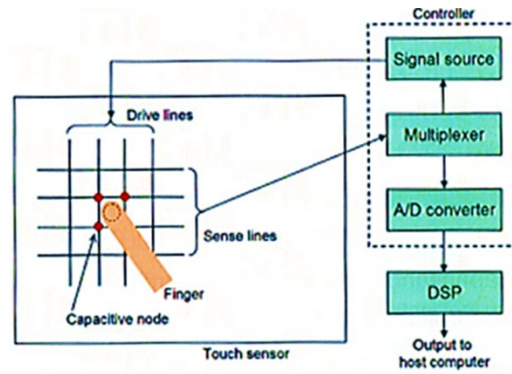


Fig. 4. Working principle of position determination in projective capacitive sensing system. [6]

Analogue Resistive: An analogue resistive touchscreen panel is composed of several layers, as shown in Fig. 5. The sensing layer is constructed of two sheets of material separated slightly by spacers: a sheet of glass providing a stable bottom layer and a sheet of Polyethylene (PET) as a flexible top layer. The two sheets are coated with a resistive material, usually a metal compound Indium Tin Oxide (ITO) and separated by an air gap or microdots. When an object, such as a finger, presses down on a point on the outer surface, the two metallic layers connected at the point. This causes a change in the electrical current, which is registered as a touch event and sent to the controller for processing. Hence the position of the touch on the surface can be measured.

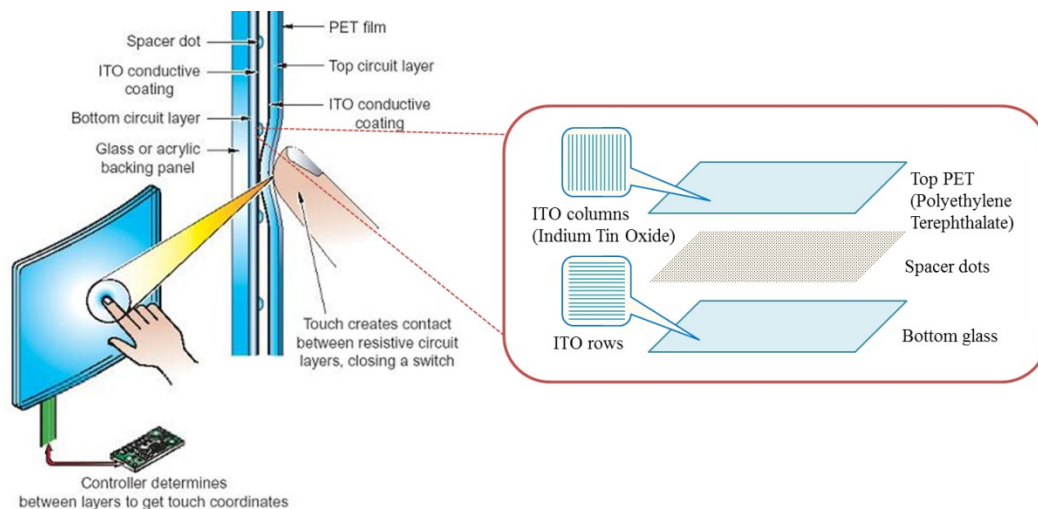


Fig. 5. Illustration of a resistive touchscreen. [6]

Optical: In conventional optical touch technology [7], an array of infrared (IR) LEDs were allocated on two adjacent bezel edges of a display with photo sensors placed on the two opposite bezel edges to determine a touch event. The photo sensor outputs can be used to locate a touch-point coordinate, as shown in Fig. 6.

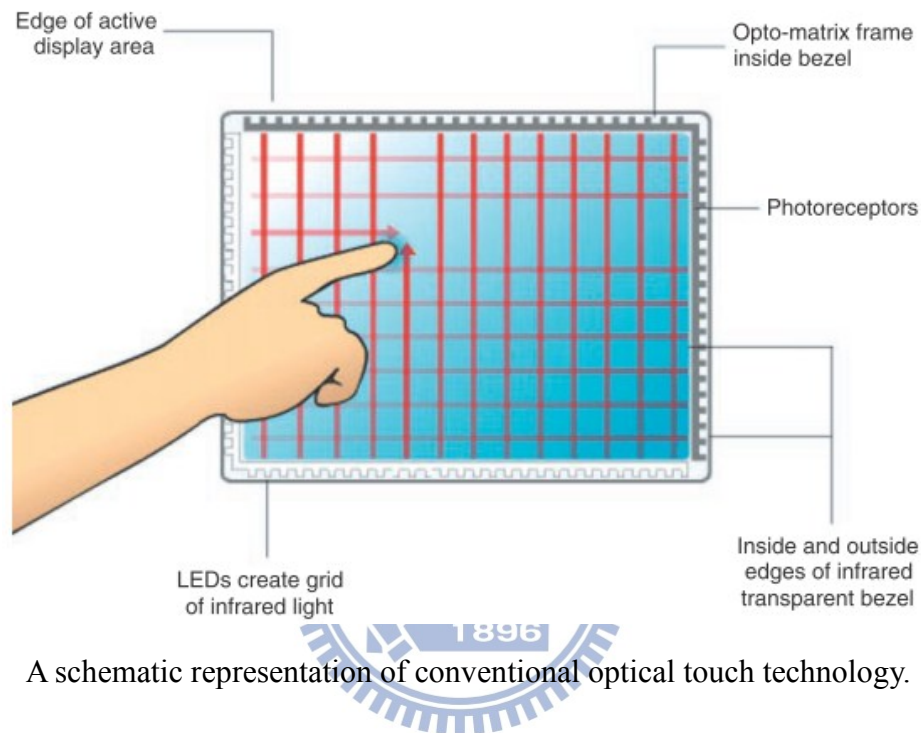


Fig. 6. A schematic representation of conventional optical touch technology.

Although the traditional type of optical touch has been hampered by two factors, the relatively high cost compared to competing touch technologies and the issue of performance in bright ambient light, certain features of optical touch remain desirable and represent attributes of the ideal touch screen. For examples, the option to eliminate the glass or plastic overlay that maintain the display quality, the digital nature of the sensor output, no direct impact of a touch object, and multi-touch implementation are the promising features in optical-touch technology.

There are several new types of optical-touch systems that is able to detect multiple touches, for instance, vision-based, camera-based, and LCD in-cell optical touch.

Vision-based optical touch systems [8] employ one or more IR imaging cameras to capture the image of the entire screen, which usually means that the camera must be located a significant distance away from the screen. Therefore, most vision-based touch systems, like frustrated total internal reflection (FTIR) [9][10], diffuse illumination (DI) [11], and diffused surface illumination (DSI) [12] systems as illustrated in Fig. 7 to Fig. 9, are implemented with the detecting cameras located behind a projection-screen surface and process the captured images to determine the 2D coordinates of touching objects.

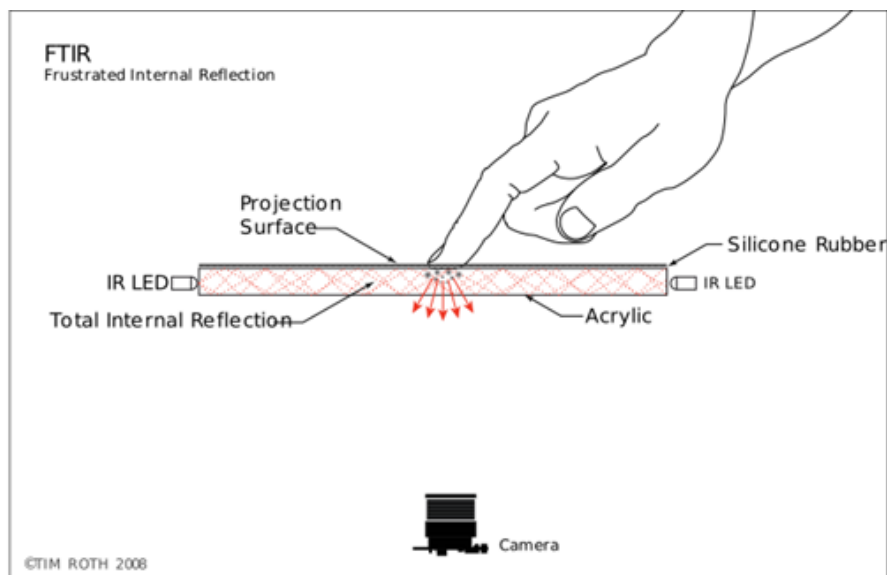


Fig. 7. General set-up of a Frustrated Total Internal Reflection (FTIR) system which is based on optical total internal reflection within an interactive surface; as a user touches the screen, the light escapes and is reflected at the finger's point of contact, which is then detected by an IR camera at the back of the pane. [8]

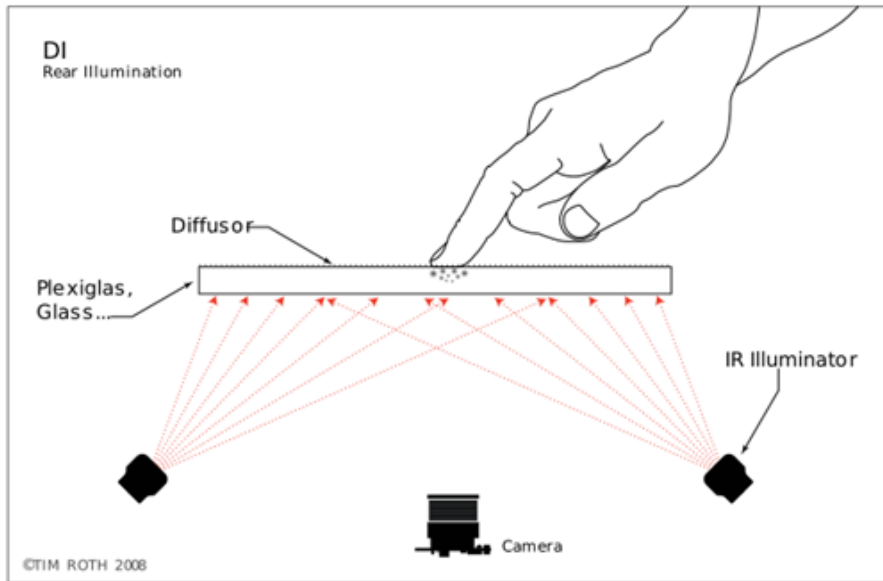


Fig. 8. General set-up of a Diffuse Illumination (DI) system which is similar to that for FTIR but with IR lighting placed behind the projection surface; moreover, DI is capable of tracking and identification of objects by using their shape or fiducially printed on their bottom surfaces. [8]

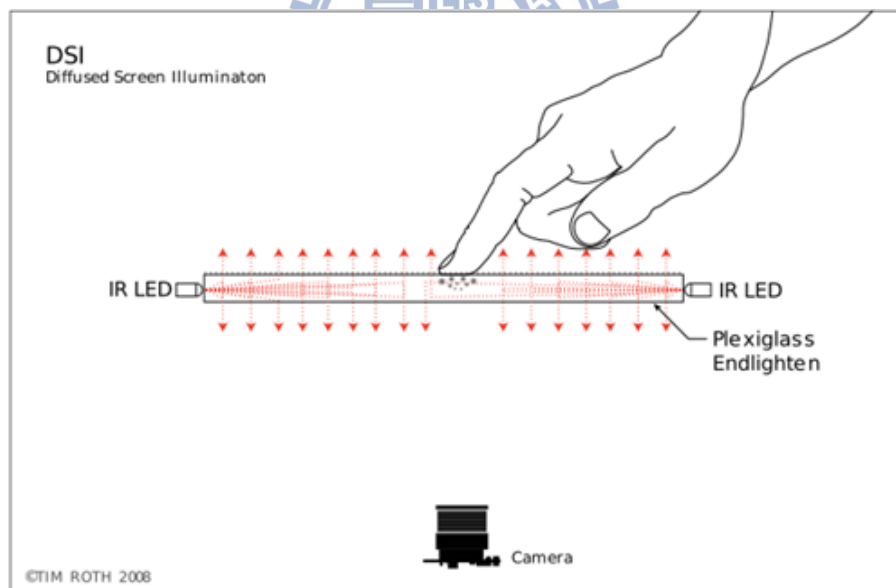


Fig. 9. General set-up of a Diffused Surface Illumination (DSI) system which utilizes small number of (two or three) infrared illuminators and evenly distributed the light across the screen surface. [8]

For a vision-based optical touch system, 2D multi-touch can be successfully achieved. However, the large system volume is essential. Therefore, a camera-based optical touch technology has been developed to reduce the size of the system.

Camera-based touch technology utilizes at least two line-scanning cameras, as illustrated in Fig. 10, located at adjacent corners of a display. The light source is usually one or two IR LEDs that are integrated into each camera assembly. The light is emitted in a plane across the surface of the screen and is reflected back to the cameras by retro-reflecting strips located along three edges of the screen.

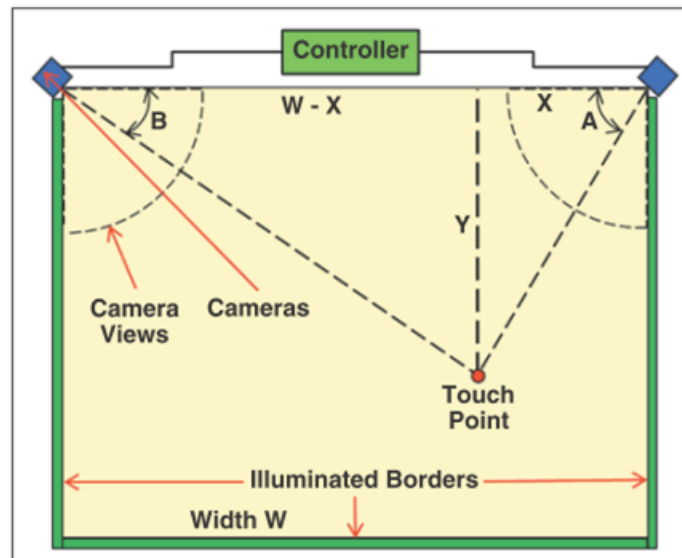


Fig. 10. Schematic of a camera-based optical touch technology. [13]

However, a thick bezel must be fabricated due to the illumination of IR light upon display surface. Hence, LCD in-cell optical touch is proposed to maintain the virtue of thin form factor in a display.

LCD in-cell optical touch, also called “in-cell light-sensing,” establish by integrating light sensing element (photodiode or photo transistor) into some or all of an LCD’s pixels, as illustrated in Fig. 11, which allows the display to act as a large array photosensor. The photo sensors receive light by either sensing the shadow of touching objects causes by blocking the ambient light [14] , as shown in Fig. 12 or the reflected light from infrared emitters embedded in the LCD’s backlight [15], as shown in Fig. 13; a multi-touch controller samples each photosensor and calculates the X-Y coordinates of touching objects.

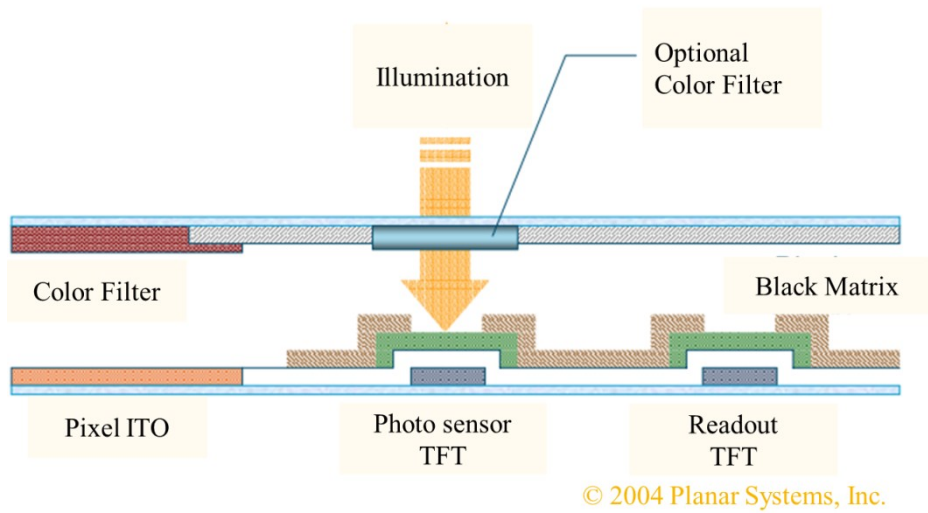


Fig. 11. General cross-section of LCD in-cell optical touch panel. [14]

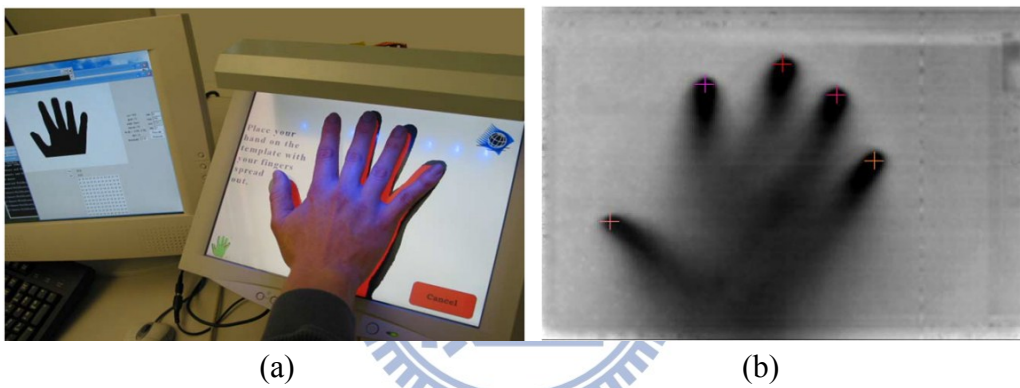


Fig. 12. 2D in-cell light-sensing can be achieved by (a) sensing the shadow of touching objects caused by blocking the regional ambient light where (b) the objects' positions reveal lower gray level [14].

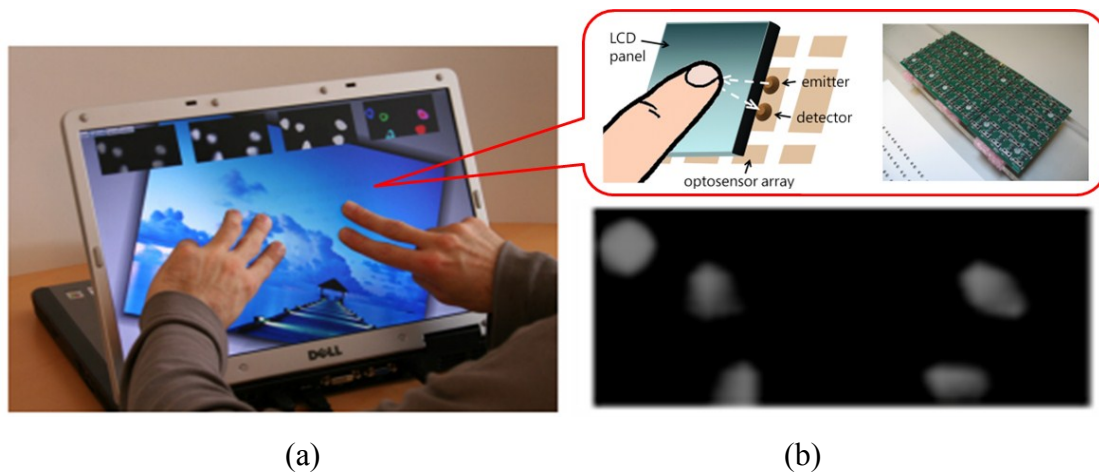


Fig. 13. (a) The prototype of capturing the reflection of IR backlight is employed in 2D in-cell light-sensing where (b) the objects' positions exhibit higher gray level. [15]

1.3 Motivation and Objective

Although multi-touch has not been widely deployed in the past, there has recently been renewed interest in it, driven by new devices such as the iPhone that make multi-touch an integral part of the user interface. Meanwhile, the rapid growth of 3D display technology has a dramatic impact on user experiences. The increasing interest in auto-stereoscopic image has given rise to an attention to 3D near distance interaction. However, the gains in the performance of 3D graphics hardware and rendering system have not been matched by a corresponding user interface to interact with the virtual 3D images we create. Therefore, in order to meet the trend in the future 3D mobile devices such as 3D cell phone and 3D tablet, there is an urgent need of 3D interactive interfaces that can provide multi-user/multi-touch, continuous, high-freedom, friendly interaction and at the same time integrable with portable devices.

As more and more researchers devoted in 3D interaction technology development, different approaches have been proposed which will be examined in the following chapter. In order to establish a continuous working space, as shown in Fig. 14, from a surface to near region, in-cell optical-touch seems to be one of the most promising technologies to extend from 2D to 3D multi-interaction.

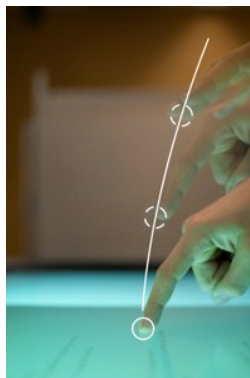


Fig. 14. Continuous interaction space from 2D touch surface to 3D virtual touch. [16]

A 3D virtual touch system for a single user has been successfully demonstrate on a 3D auto-stereoscopic display with in-cell optical-sensor, as shown in Fig. 15, in 2010 FPD show. However for the existing 3D near distance interactive systems, there is still no satisfactory solution for multi-user/multi-touch user interfaces, which experience limitation in number of user, complex signal-processing demands (which adds cost and power consumption), and incomplete user identification for users overlapping. Therefore, our objective is to improve the performance and overcome the present issues in 3D near distance interfaces for multi-user/multi-touch interaction. By increasing the number of users, enhancing the user identification ability, and increasing the acquisition coordinates of a user, a high-freedom and friendly 3D interactive system from multi-user/multi-touch can be achieved.

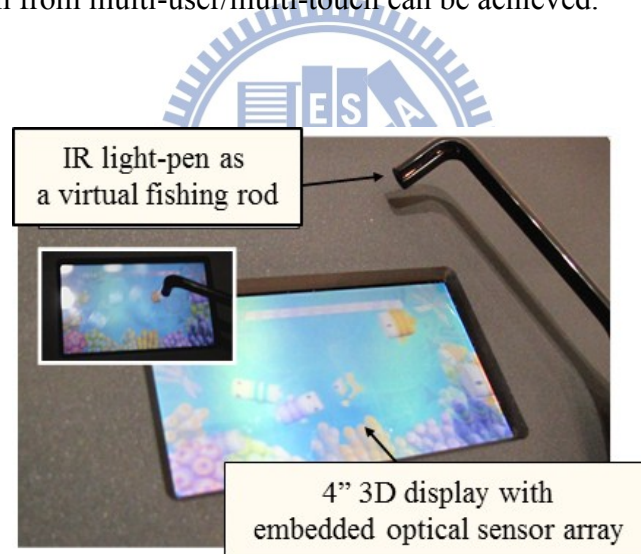
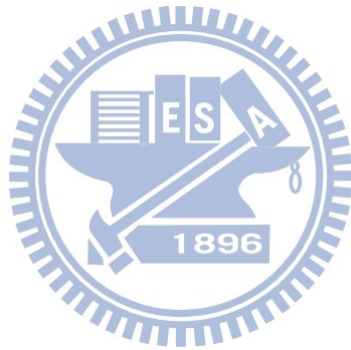


Fig. 15. A 3D virtual touch prototype demo – 3D Fishing Game. The video link: <http://youtu.be/ipJeMAdZVCI>

1.4 Organization of the Thesis

In the following paragraphs, we'll briefly describe different 3D virtual touch technologies in chapter 2 and the reason for choosing in-cell optical-based display as our basic structure. The proposed system and the according algorithms are discussed in chapter 3. Next several experiments had been done to demonstrate the feasibility of the system in chapter 4. Finally the conclusion and future work are given in the last chapter.



Chapter 2 *Prior Arts of 3D interactive Systems*

Prior Arts of 3D interactive Systems

2.1 2D Touching of 3D Stereoscopic Images

In recent years, two technologies have dominated technical exhibitions and the entertainment market: multi-touch interface and 3D stereoscopic displays. A stereoscopic image means we can perceive a 3D virtual object jump out of the display plane. It is achieved by sending each eye a slightly different perspective of the same scene which results in two image projections on the display. By visual processing, our brains can automatically combine the two images and the 3D virtual object with depth perception will appear at the focus point, as illustrated in Fig. 16. Additionally, by considering the comfort of a viewer, up to plus and minus 5 cm of depth value is constructed for an auto-stereoscopic image which means a 3D image can be observe without wearing 3D glasses.

However, an issue occurs when a user wants to interact with a 3D virtual object of stereoscopic image. While in 2D touch interface, the user is either focus on the finger, which makes the selection ambiguous and breaking into two images, or on the virtual object, which disturbs the visual perception of the finger, as shown in Fig. 17.

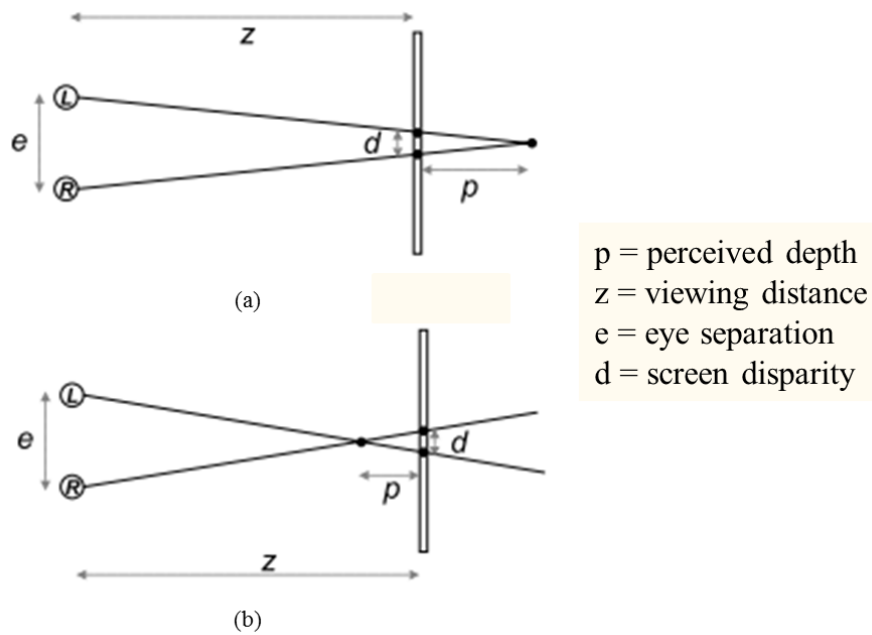


Fig. 16. Depth perceptions are (a) behind and (b) in front of the display plane. [17]



Fig. 17. Main issue in 2D touch interface with stereoscopic data is that a user is either focus on the finger or the 3D virtual objects which both degrades the visual perception. [18]

Therefore, the 3D virtual touch technology is essential to provide a friendly and intuitive user interface to a 3D display or other interactive applications. More details will be discussed as follow.

2.2 3D Virtual Touch Technologies

To fulfill the demand in 3D (x , y and z) interaction applications such as in 3D stereoscopic display, much research has been devoted to 3D virtual touch technologies. Beside 2D coordinates (x and y) on the surface plane, extra depth information (z) is compulsory to specify a point in a 3D environment. However, rather than using points, many tasks in 3D systems require a user to manipulate entire objects which also have an orientation specified by orientation (θ and ϕ) information. Thus there are 5 axes, as illustrate in Fig. 18, need to be defined in a 3D interactive system.

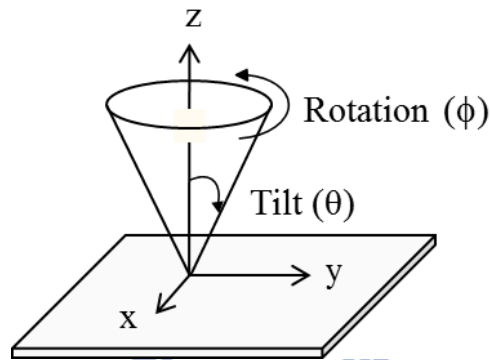


Fig. 18. 5 axes (x , y , z , θ , and ϕ) information in position and orientation.

3D virtual touch technologies, which are most widely used today, can be classified into three major groups in accordance with the employed techniques. Following paragraphs give further details of 3D virtual touch technologies in machine-based, camera-based, and in-cell optical-based.

2.3 Machine-based

The machine-based system is a robust 3D information detection approach where a user has to wear additional equipment to track his or her movement. Potentiometers or optical encoders are usually used to measure the rotation of a joint between rigid linkages of the equipment. Given the angle of each joint and length of the rods, the position of the tracked object can be calculated geometrically. Therefore, the position and orientation information can be obtained.

In machine-based 3D virtual touch systems, higher resolution information can be achieved by a more delicate and complex mechanical design. By wearing the solid equipment onto the body, it is possible to render force-feedback to the user; however, it is on the contrary uncomfortable for long time wearing and impractical to carry around with portable devices. Therefore, the machine-based system is often indoor fixed and collaborating with applications such as desktop set-up or virtual reality chambers.

2.3.1 Haptic WorkstationTM

The Immersion Haptic WorkstationTM [19] is an example of implementing machine-based virtual touch system to interact with virtual environments, as shown in Fig. 19. It is composed of data gloves [20], 5-axis hand tracking and force-feedback on the wrists. A pilot can see virtual images by the Head-Mounted Display (HMD) to simulate the flying conditions inside a blimp capsule through data gloves, and can also feel how fast he is going by evaluating the force applied on his hands. In [21], Ott et al have shown that the teleoperation of a vehicle using haptic devices is more efficient when having a gesture interface. However, one disadvantage of the Haptic

Workstation™ is that it is uncomfortable during long sessions, because the user arms must be kept outstretched while supporting the weight of the exoskeleton.

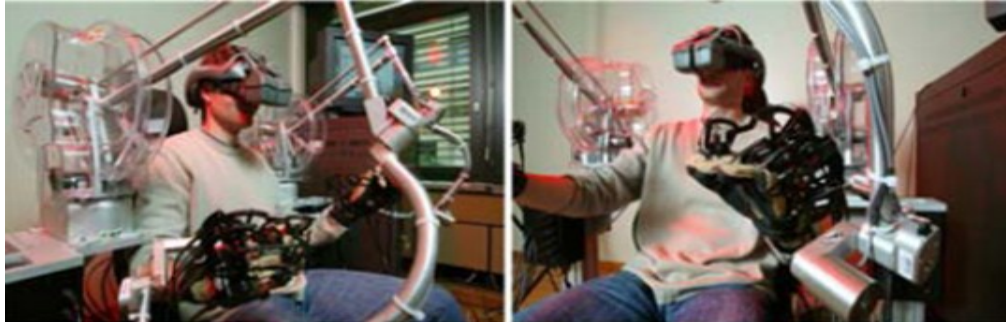


Fig. 19. Haptic Workstation and Head-Mounted Display (HMD) for rendering the virtual cockpit. [19]

2.3.2 Wii

The Nintendo Wii video game console redefines the way how video games are played. The controller, Wii-mote [22], is able to translate a player's motion, such as swinging and punching, into game playing on the screen. The working principle of Wii is hybrid both the camera-based and machine-based techniques to obtain the translation and rotation information. As shown in Fig. 20, there is an infrared camera in a Wii-mote. Thus, only the relative position can be calculated by sensing a series of corresponding LEDs in the sensor bar but without obtaining the absolute depth value. To increase the freedom in gestures, a machine-based technique, a three coordinates accelerometer [23], as shown in Fig. 21, is employed to detect the acceleration along three axes (x , y , and z). Hence different gestures, like wave, twist, push and pull can be achieved. To take the field of view in IR camera into consideration, a user must stand at 1 to 2 meter form the sensor bar in order to capture the full motion of the user. Hence, it is neither applicable for on screen surface nor in near distance region interaction.

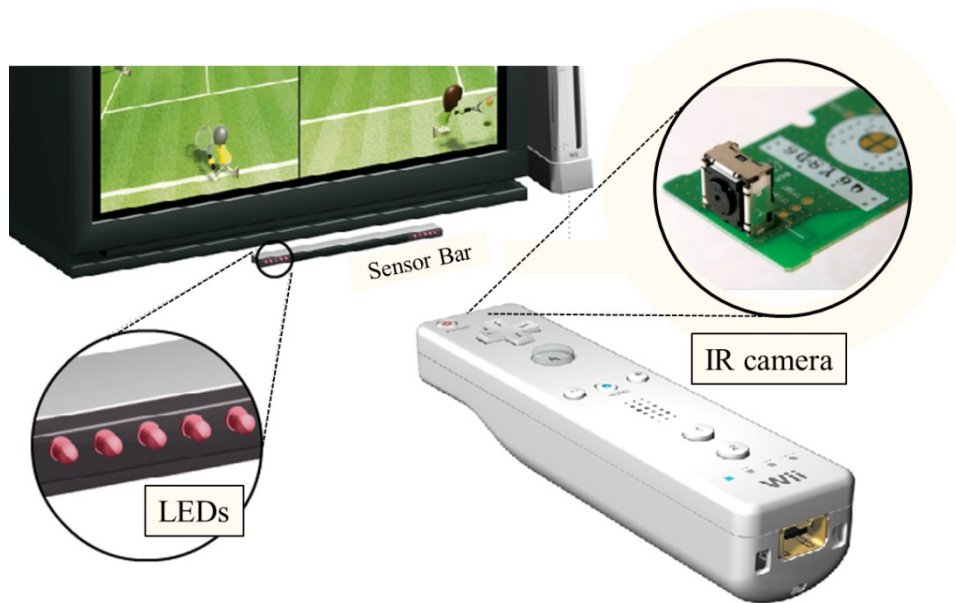


Fig. 20. Wii utilizes an infrared camera to sense a series of LEDs in the sensor bar, orienting the controller in space.[22]

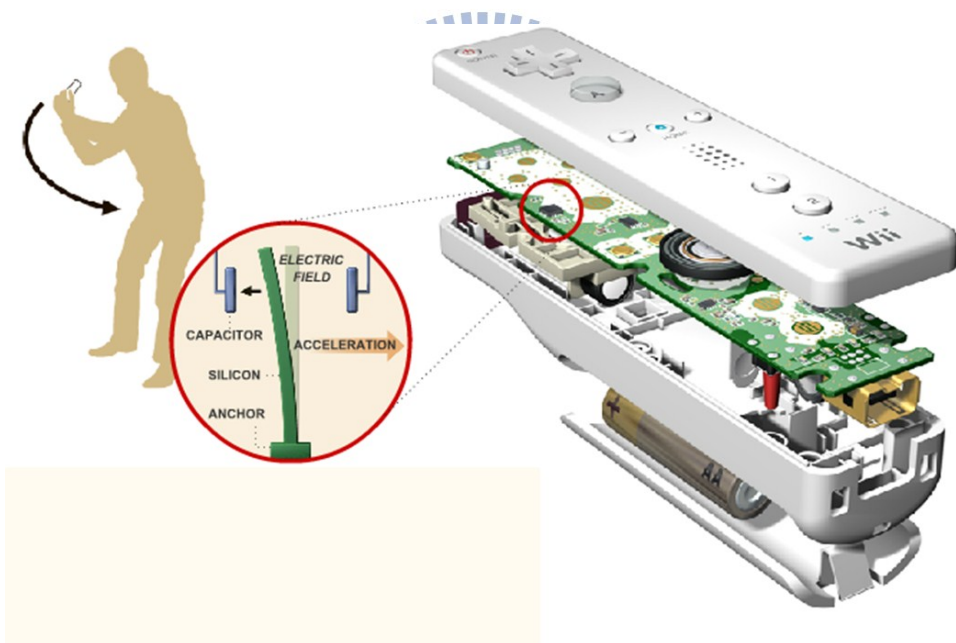


Fig. 21. Acceleration Sensor: this chip provides all the motion sensing. A piece of silicon is evenly placed between two capacitors in electric field. Once the user accelerates the controller in one direction, it causes the silicon to bend in the other and causes the electric field to change, which translates into motion.[22]

2.4 Camera-based

Commercial products collaborating with camera-based 3D virtual touch technology have dominated recent entertainment market. In many applications, the camera-based 3D interactive system is becoming pervasive and much more physical and tangible. Various designs of camera set-up and different sensing methods have been proposed, such as shadow sensing H.J. Luinge, P.H. Veltink, "Inclination measurement of human movement using a 3-D accelerometer with autocalibration," *Neural Systems and Rehabilitation Engineering, IEEE Transactions*, vol.12, no.1, pp.112-121, 2004.

[24] and active light sensing [25] techniques. Complex image processing is applied on the high resolution images captured by the cameras to calculate the transition and/or rotation information.

However, limited field of view of cameras, as shown in Fig. 22, prevents the camera-based technology from integrating into portable devices. Objects proximity to the display cannot be detected which makes it difficult to perform a continuous interaction space from 2D to 3D. Hence, the camera-based 3D interaction system is more suitable for far-distance applications, such as to TVs or tabletop displays. Otherwise, the system volume must be increased for capturing the full vision of the interaction region, as shown in Fig. 23. Meanwhile, since the camera-based system requires high resolution data to calculate 3D information, a larger size of CCD is needed which impedes the system being integrated into portable devices.

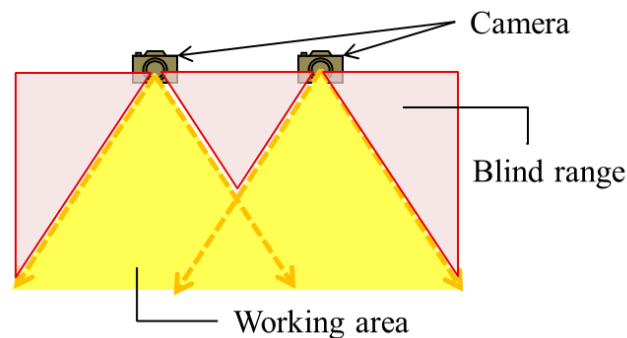


Fig. 22. Limited field of view in cameras causes the blind range, unworkable region,

in near-distance.

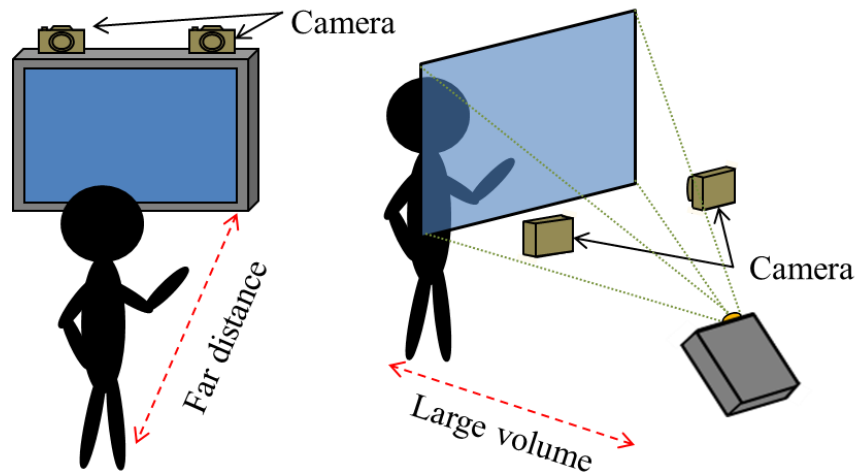


Fig. 23. Intrinsic characteristics in camera-based system render a tradeoff between discontinuous working space and large system volume which forbids the integration to a display.

2.4.1 Microsoft 3D Gesture Interface

In 2011, Microsoft set out a gesture control interface that a user can interact beyond the touchscreen H.J. Luinge, P.H. Veltink, "Inclination measurement of human movement using a 3-D accelerometer with autocalibration," *Neural Systems and Rehabilitation Engineering, IEEE Transactions*, vol.12, no.1, pp.112-121, 2004.

[24]. The system composes of a retro-reflective screen, an image projector, and an infrared camera, as depicted in Fig. 24. The incident light from the projector is reflected back by the retro-reflective screen, which consists of many retro-reflectors, along a vector parallel to but opposite in direction from the source. Therefore the reflected light is captured by the camera with minimum scattering. Once a user put hands above the display, the shadow of hands can be detected. Without touching on the display surface, the 3D virtual touch can be achieved by gesture recognition, as shown in Fig. 25.

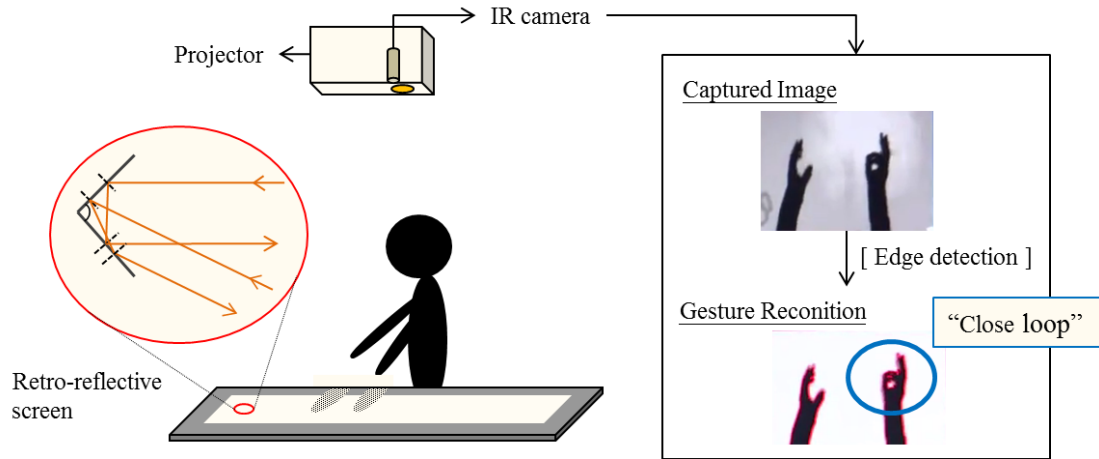


Fig. 24. Schematic of the Microsoft 3D Gesture Interface.

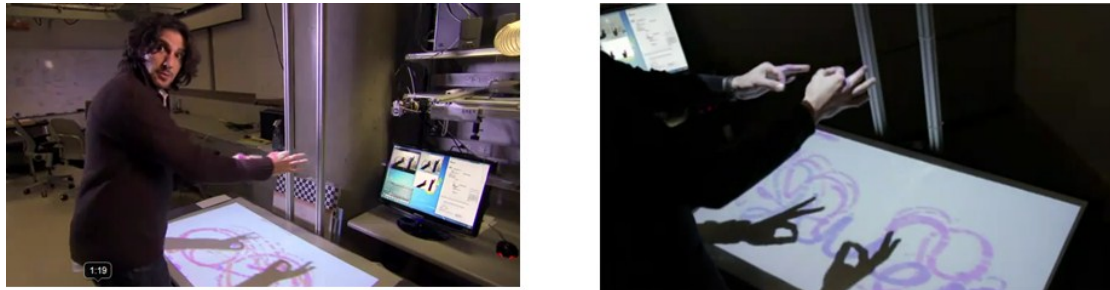


Fig. 25. Microsoft 3D Gesture Interface demonstrates that the “draw” command is activated when a “close loop” of shadow image is detected.

However, the absolute depth information cannot be obtained by the shadow method where only a camera is employed. Moreover, the projector and the IR camera must be located a significant distance away from the screen in order to display and capture the entire working space. The cost of retro-reflective screen is relatively high. The 3D gesture interface proposed by Microsoft is not suitable for thin-form display integration.

2.4.2 WorldViz PPT

WorldViz PPT (Precision Position Tracking) system [25] is available to obtain 5 axes data from the active LEDs which are put on the object, as shown in Fig. 26. Two or more cameras which are deployed in the workspace are able to track up to 32

visible infrared LED markers simultaneously. As the object with markers makes its way through the workspace, the cameras acquire data. Through image processing, the captured data are converted into accurate 3D coordinates, and the rotation information can be obtained through the relative movements between the markers.

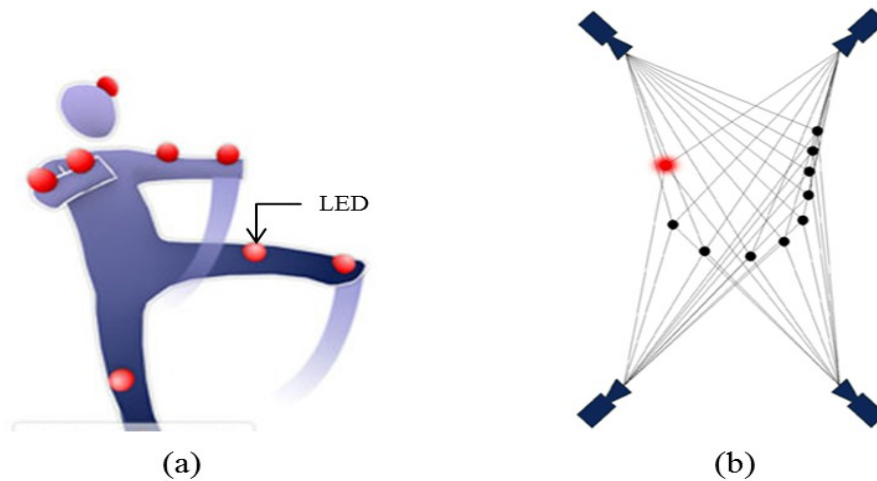


Fig. 26. 3D coordinate is acquired by tracking (a) active LED markers by at least two to (b) four tracking cameras.

WorldViz PPT is a reliable solution for wide area 3D coordinate tracking which can be further applied on 3D far-distance interaction applications. However, it is not an ideal sensing system for portable devices because of the blind range limitation which enlarges the system volume, and the acquisition of high resolution cameras which distinctly increases the cost.

2.5 In-cell Optical-based

The in-cell optical-based 3D virtual touch system is the extension of the 2D in-cell multi-touch system. The continuous space from 2D surface to 3D virtual touch can be achieved without hybrid other mechanisms since they share the similar structure which was firstly proposed by W.D. Boer et al at 2003 [26]. Meanwhile, objects can be sensed under similar working principles; a hovering object can be detected either by the reflected light which is emitted by infrared LEDs in the LCD's backlight, or by the object direct illumination on the display. Several techniques were proposed to determine the 3D coordinate (x , y , and z) and/or orientation (θ , and ϕ) information.

2.5.1 ThinSight

The idea of extending 2D multi-touch to 3D interaction in an in-cell optical-based system was first proposed by S. Hodges et al in 2007 [15]. A hardware structure with infrared sensors and emitters integrated into a thin form-factor display was suggested to detect multiple fingers placed on or infrared-emissive objects near the display surface, as shown in Fig. 27 (a).

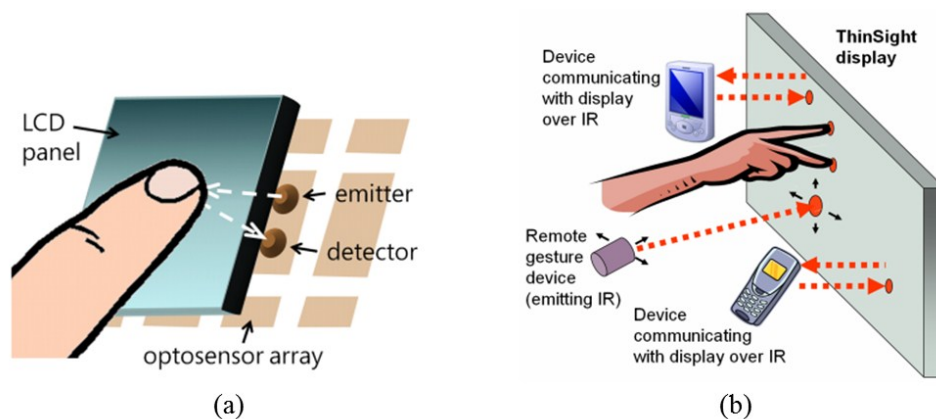


Fig. 27. (a) Basic construction of ThinSight, and (b) applications in ThinSight

display including 2D finger multi-touch, data transformation between a mobile device and the display, and 3D gesture interaction through a device which can cast a beam of infrared light onto the display. [15]

The layer of diffusing film between the LCD itself and the brightness enhancing film is removed because it cause too much attenuation of the infrared signal, especially for passive objects that reflect the light emitted from the display. However, the removal of diffuser has a detrimental effect on brightness and viewing angle of the LCD panel. Besides, it is not able to provide depth information. Still, based on the structure, a number of promising applications, such as data transformation, and 3D gesture interaction through an IR emitting device are proposed, as shown in Fig. 27 (b).

2.5.2 Sensible Backlight

A multi-touch LCD display architecture with hover sensing capability was proposed by K. Yi et al in 2010 [27]. Instead of changing an LCD manufacturing process to insert photo sensors in cell, the system proposes a sensible backlight where a backlight unit is integrated with an IR sensor array, as shown in Fig. 28. IR light sources for touch and hover detection are positioned on the bezels of the display. The 2D multi-touch is achieved based on frustrated total internal reflection (FTIR). For recognizing simple hovering objects, side IR illuminators emit light with tilt angles. Hence the approximate hovering positions can be obtained by sensing the reflected light by the IR sensors sensible backlight.

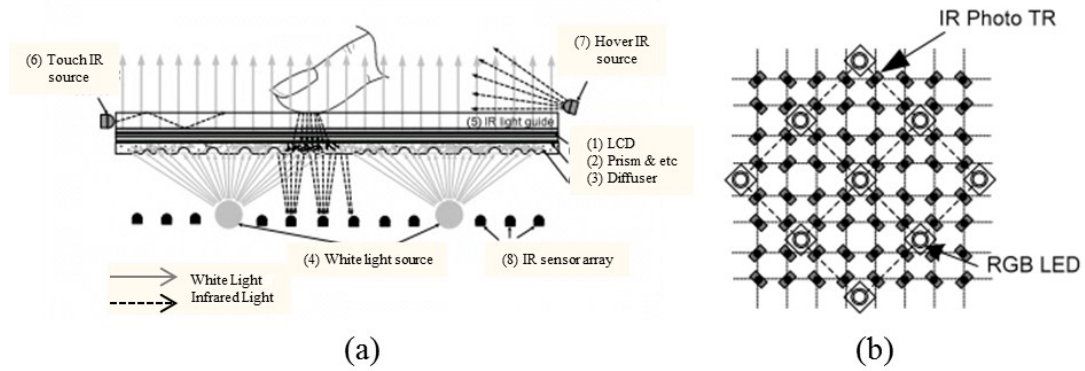


Fig. 28. (a) Working principle and optical structure of the sensible backlight system, and (b) layout of backlight with RGB LEDs integrating with IR photo transmitters. [27]

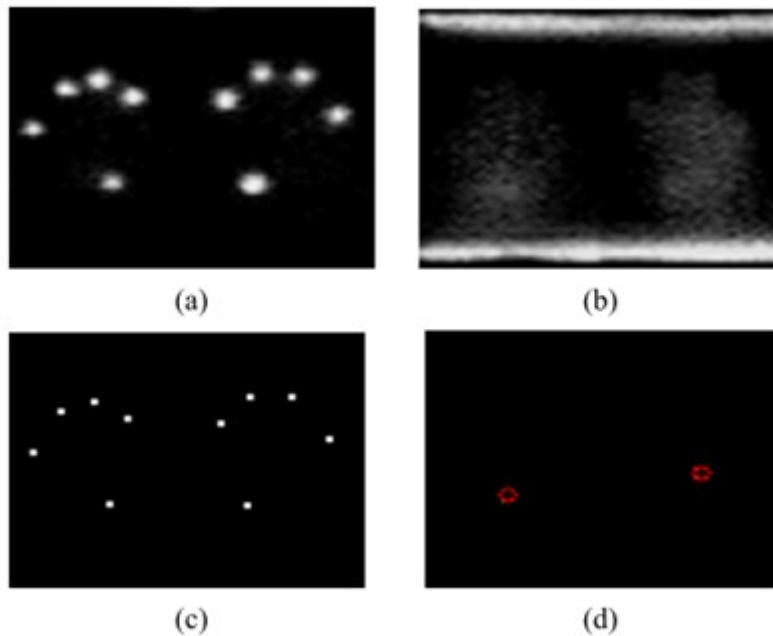


Fig. 29. (a) Sensor image of touch, (b) sensor image of hover, (c) extracted touch points, and (d) extracted hover points. [27]

The 2D multi-touch and hover positions can be obtained by utilizing the sensible backlight system. However, the depth information cannot be found, as shown in Fig. 29 (d). The approximate hover position, which is determined by the center of mass of each captured image, outputs a 2D coordinate on the surface. Moreover, a thicker border is needed for illuminated infrared light on the hover objects, and the panel size is limited by the intensity of side illuminated light. Therefore, the system is more

applicable on middle size displays with near-distance interaction.

2.5.3 Directional Image Sensor

A LCD system with integrated 3D input device was proposed by C. Brown et al in 2010, [28] and the concept is illustrated in Fig. 30 (a). The 3D input function is successfully achieved by employing directional image sensors integrated onto the TFT substrate. The directionality of the image sensor is created by an upper light shield formed in a second additional metallization layer, as depicted in Fig. 30 (b), to allow only incident light within a specific angle being detected. A set of four orthogonal direction sensors is able to generate four unique directional sub-images from the detected light. Sub-images are then processed to extract planar coordinates (x,y) corresponding to the location of the peak output signal, as shown in Fig. 31. Finally, by examining the relative displacement (d) of the object's 2D coordinate in each directional sub-image, the depth value (z) can be calculated by Eq. (1).

$$d = \frac{1}{2} z \cdot \cos \theta$$

Eq. (1)

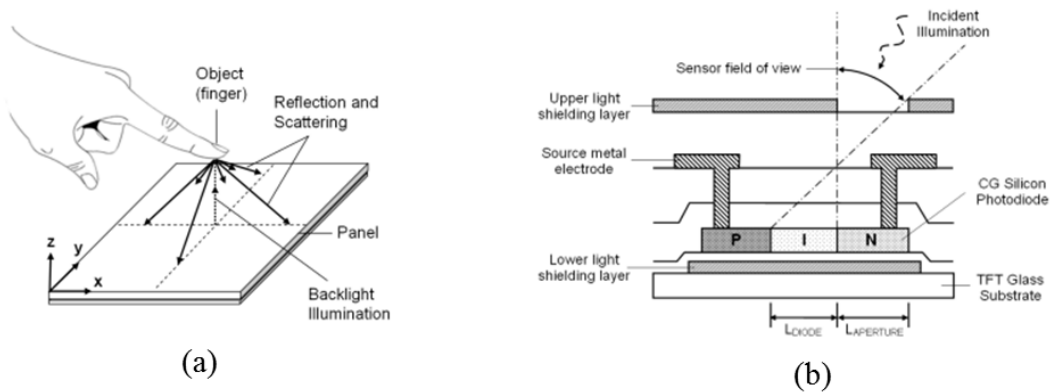


Fig. 30. (a) Concept of a 3D input device using a directional image sensor, and (b) structure of thin-film lateral diode and light shielding layers to create a directional field-of-view. [28]

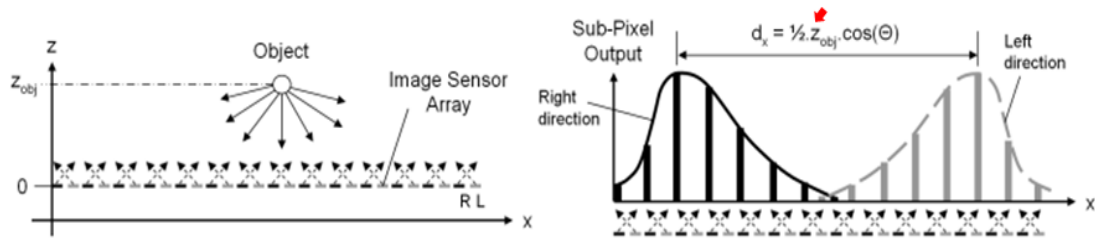
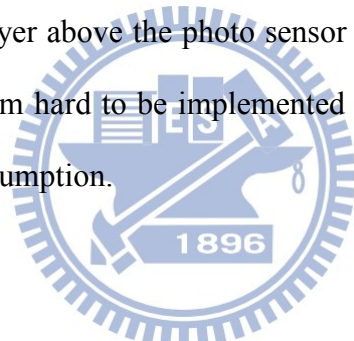


Fig. 31. One dimensional representation of the response of two directional sub-pixels. [28]

However, due to the construction of the panel, limited field-of-view of the directional sensor, and the insufficient sensitivity of the image sensor, detected z-axis exhibits a limited linear response from 0 to 20 mm. It is not enough for near field 3D interaction where ± 50 mm depth can be perceived in an auto-stereoscopic display. Meanwhile, the shielding layer above the photo sensor significantly reduced aperture ratio which makes the system hard to be implemented on mobile devices because of the considerable power consumption.



2.5.4 Color-filter-based Sensing

A 3D multi-interactive system achieved by color filter based sensing was proposed by H.Y. Tung, et al [29] under in-cell optical based structure where photo sensors are embedded on a TFT substrate in a LCD. As illustrated in Fig. 32, multi-wavelength of light sources, red, green, and blue are utilized as interaction illuminators. By using a color filter as a band-pass filter, different users can be identified by extracting the sensor outputs under accordant color filters. After separating light sources, center and radius of each circle are calculated and used to determine the 2D coordinate (x and y) and height (z) respectively. Therefore, 3D multi-touch interaction with user recognition can be achieved.

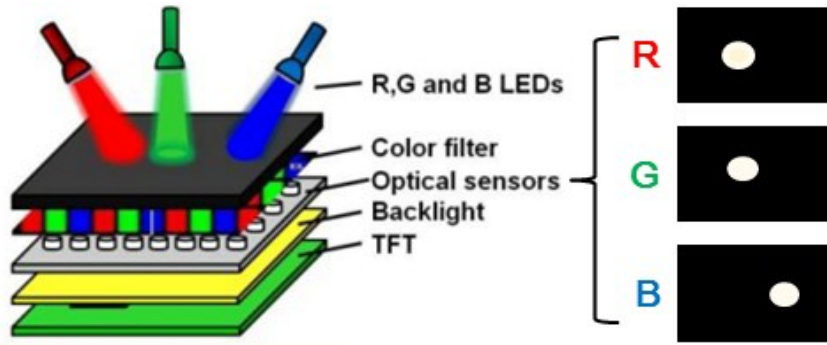


Fig. 32. System structure of color filter based sensing.

However, due to the usage of visible light sources, the display quality is degraded. The limited number of three users can be achieved due to the limited kinds of color filter in a display. Moreover, the imperfection of color filter would result in light leakage where light spots are too close and overlapping happens, as shown in Fig. 33. Nevertheless, a complex image processing, Hough transform, needs to be applied for circle detection which results in heavy computation.

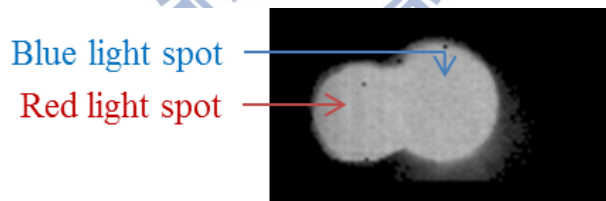


Fig. 33. A detected image under red color filter. Undesired blue light penetrates a red color filter and overlapping of red and blue spots occurs.

2.5.5 Multi-mark Based

A 3D multi-mark interactive system was proposed by C.C. Chao, et al [30] to establish an interface between multiple users and 3D images. Based on the in-cell optical-sensor structure, illuminators such as IR light pens covering with designed marks, as shown in Fig. 34, is utilized to provide significant features. The display quality can be maintained by utilizing IR LEDs. By the proposed algorithm, different

users and their 3D coordinate (x, y, and z) can be identified. Meanwhile, the overlay of a circle with one of other marks, T or X, can be distinguished for up to 40% overlapping. Finally, the proposed algorithm was successfully demonstrated on a 4” panel.

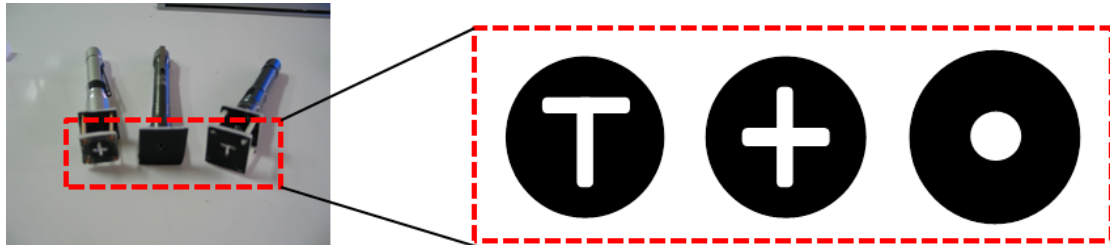


Fig. 34. Designed light marks of solid T shape, X shape, and circle.

However, the limited of 3 users can be achieved due to the specific designed mark. Meanwhile, the complex overlapping cases, as illustrated in Fig. 35, dramatically increase the loading on a circuit, which also increases the cost and power consumption. Moreover, due to the similar characteristic of T shape and X shape, the overlapping of T and X cannot be identified in the system where only two users overlapping with one of them must be a circle can be deal with.

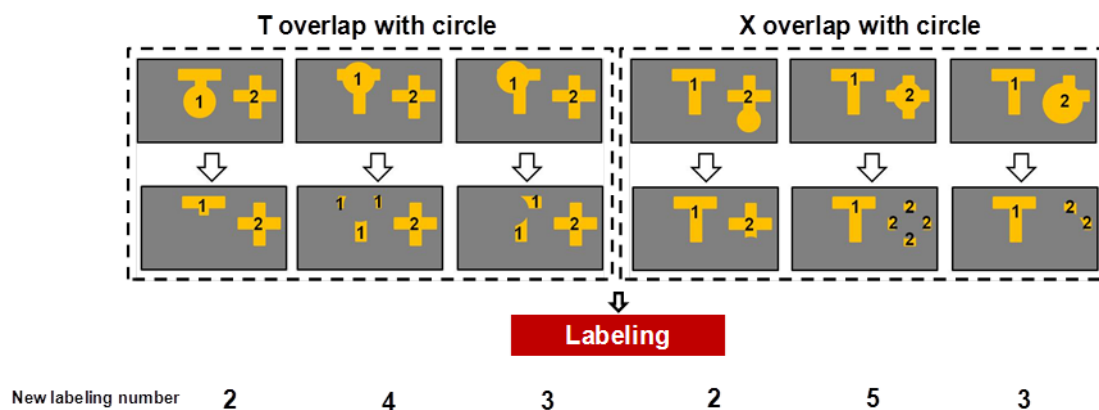


Fig. 35. Complex overlapping conditions need to be processed case by case which results in heavy computation requirement.

2.6 Summary of 3D Virtual Touch Systems

There is no perfect touch technique since the diversified products demands, as shown in Fig. 36. Machine-based systems are the most precise and reliable 3D interactive systems that can further render the force-feedback. However, the high cost due to the delicate apparatus and the inconvenience of device-wearing make the machine-based system mostly apply to military or medical training system or other elaborate applications. On the other hand, camera-based systems are more cost effective and produce rich data which provide enough reliability. Most of them by nature rely on projection and well-distance cameras capturing. Therefore, long-distanced and wide-area position tracking applications, such as tabletop display, TV, and interactive billboard, are more collaborated with camera-based systems.

The bump up in smart phone and tablet computer bring the mobile devices into the major consumer electronic products. Meanwhile, since the increasing interests in 3D applications and the gains in the auto-stereoscopic image performance, portable devices with 3D technology might soon be ubiquitous. Nonetheless, the technologies on the market neither match the requirements in mobile devices nor establish a robust continuous interaction space from 2D to 3D. Therefore, a potential solution refers to the in-cell optical-based technology which is able to maintain the thin form factor, to detect multiple touches, to preserve optical performance, and easy to carry with.

However, there are several issues in present multi-user/multi-touch user interfaces, such as limitation in number of users, complex signal-processing demands, and insufficient user identification ability. Therefore, in the following chapters, 3D virtual touch systems with multi-touch and user calibration (which means identifiable users) will be presented. Meanwhile, the proposed algorithm shall further increase the number of users, reinforce the user identification ability, and/or increase the detected

axes of users. Hence, a high-freedom 3D interface for multi-user/multi-touch could be achieved.



Fig. 36. Different categories of 3D interactive systems have different characteristics that match for different applications.

Chapter 3 *Structure and Algorithms*

Structure and Algorithms

3.1 Overall Structure

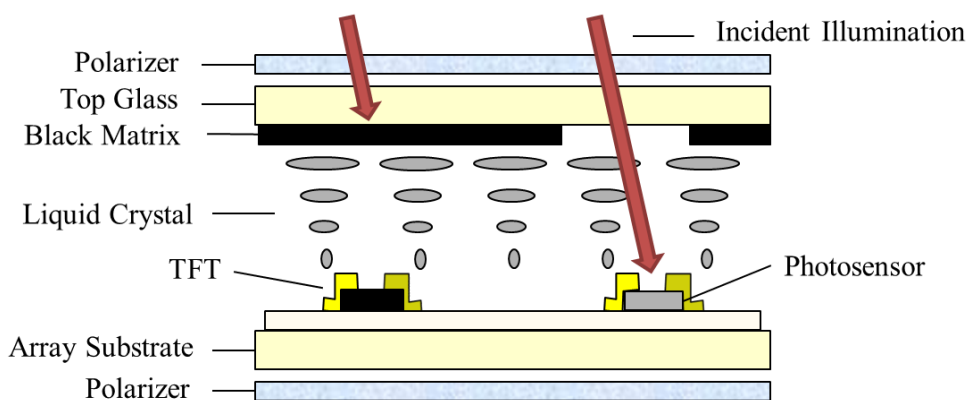


Fig. 37. Hardware structure of the embedded optical sensors display.

In order to maintain thin form factor as well to interact in near region to a display, the in-cell optical-based structure is chosen. The optical sensors are integrated onto the same layer with TFT substrate to form a sensor array all over the display, as shown in Fig. 37. For an in-cell optical-based display, the bare-finger interaction currently has some limitations, such as insufficient sensor resolution, serious ambient light effect, and inadequate depth sensing range. Therefore, an infrared LED light pen is utilized as the input device. The concept of the proposed 3D virtual touch system is illustrated in Fig. 38.

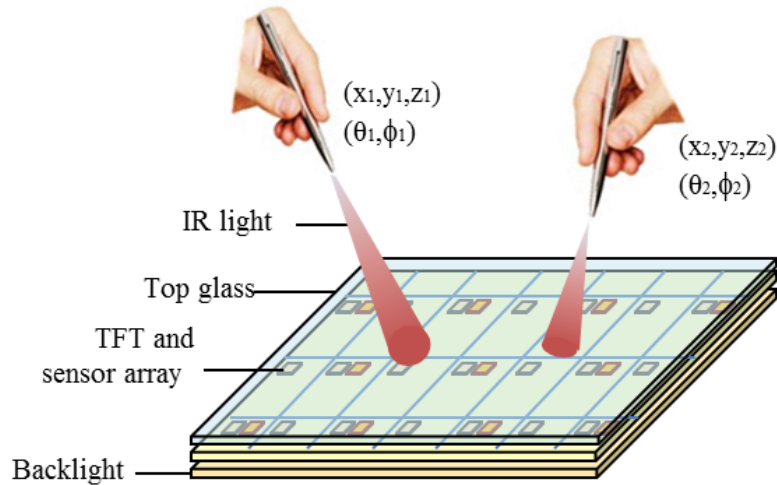


Fig. 38. Concept of the proposed 3D virtual touch system with multi-user collaboration.

In 3D virtual multi-touch technologies, three major issues need to be conquered. The first one is the limitation in number of users; there should be a more plastic method for the extension amount of users. The second issue is user recognition; instead of sacrificing the display quality or increasing the complexity of signal-processing to identify different users, a more simple but solid solution should be proposed. The last but not least one is the unsatisfactory user identification ability especially during overlapping cases. Therefore in the following paragraphs, two methods, Sequential-lighting method and Light-mark method which includes Multi-ring mark and Multi-T mark algorithms, will be presented to achieve a high-freedom 3D interaction system with 3D coordinate $(x, y, \text{ and } z)$ and/or orientation $(\theta \text{ and } \phi)$ information detection and multi-user calibration. Meanwhile, the performance of the system shall be superior to the previous methods by solving the issues as mention.

3.2 Sequential-lighting Method

In Sequential-lighting method, the light pens are synchronized to the panel with embedded optical sensors. Each light pen emits light which is match to a frame of the sensor respectively. By sequentially turning on and off the light sources at $1/n$ frame rate of the optical sensor, where n equals to the number of the users, there will be only one light spot captured at a sensor frame. For instance, we match three light pens, which meant three users at a time, to the proposed panel, as illustrated in Fig. 39. For the first sensor frame, only the light pen one is turned on, while two others are turned off. For the second sensor frame, the light pen two is turned on while others are off, and so forth. Under sequentially turned the light pens on and off, there is always only one light spot received by optical sensors, which therefore overcome the overlapping issue.

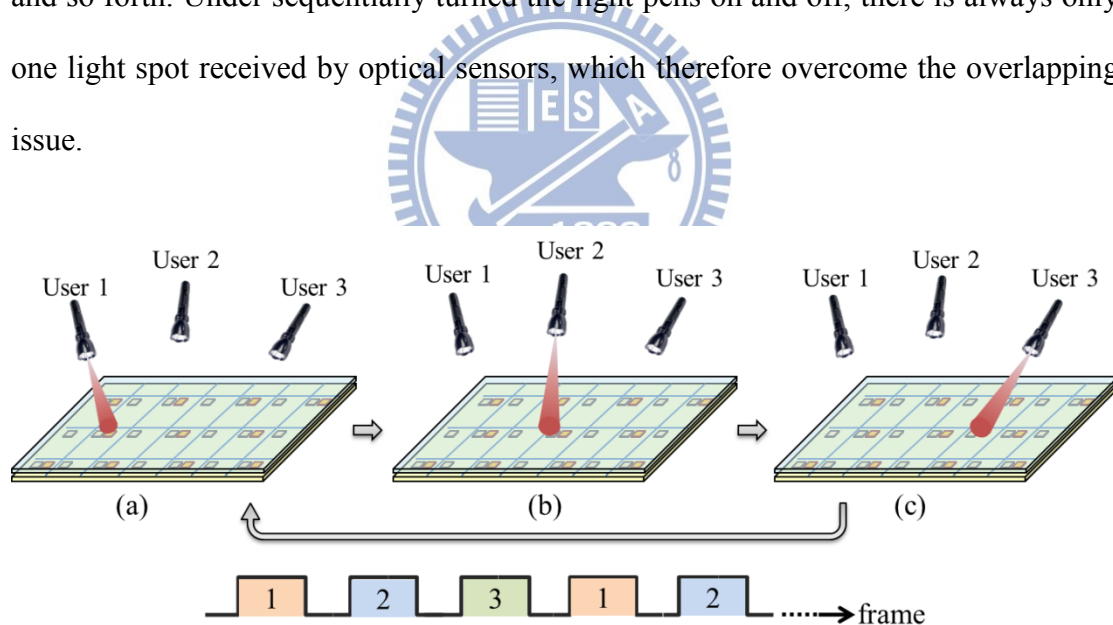


Fig. 39. A 3D interactive system model with sequential mark (a) light source one was turned on during frame $= n \times k + N = 3k + 1$ where n equals to the number of users and N is the sequence of the user (b) light source two was turned on during frame $= 3k + 2$ (c) light source three was turned on during frame $= 3k + 3$, $k = 0, 1, 2, \dots$

Furthermore, the user identification can also be achieved by employing sequential method. As soon as we synchronized the light pens with the panel, the time coordinates will operate in unison. Hence, each raw data can be allocated to the

corresponding user. In the example above, raw data in $3k+1$ frames will be regarded as user 1, raw data in $3k+2$ frames will be regarded as user 2, and so on. Next, Sequential-lighting algorithm was proposed to acquire the 3D information (x , y , and z) of each user. The flow chart of the algorithm is shown in Fig. 40.

After the incident illumination is detected by the optical sensors, the row data in frame $n \times k + N$, where n is the number of users, N is the sequence of the user, and k equals to integers started form zero, are allocated to user N respectively. Hence the user identification is achieved. The complex overlapping issue in multi-user interaction is simplified to single virtual touch conditions. Our next step is to acquire 3D coordinate (x , y , and z) form the row data.

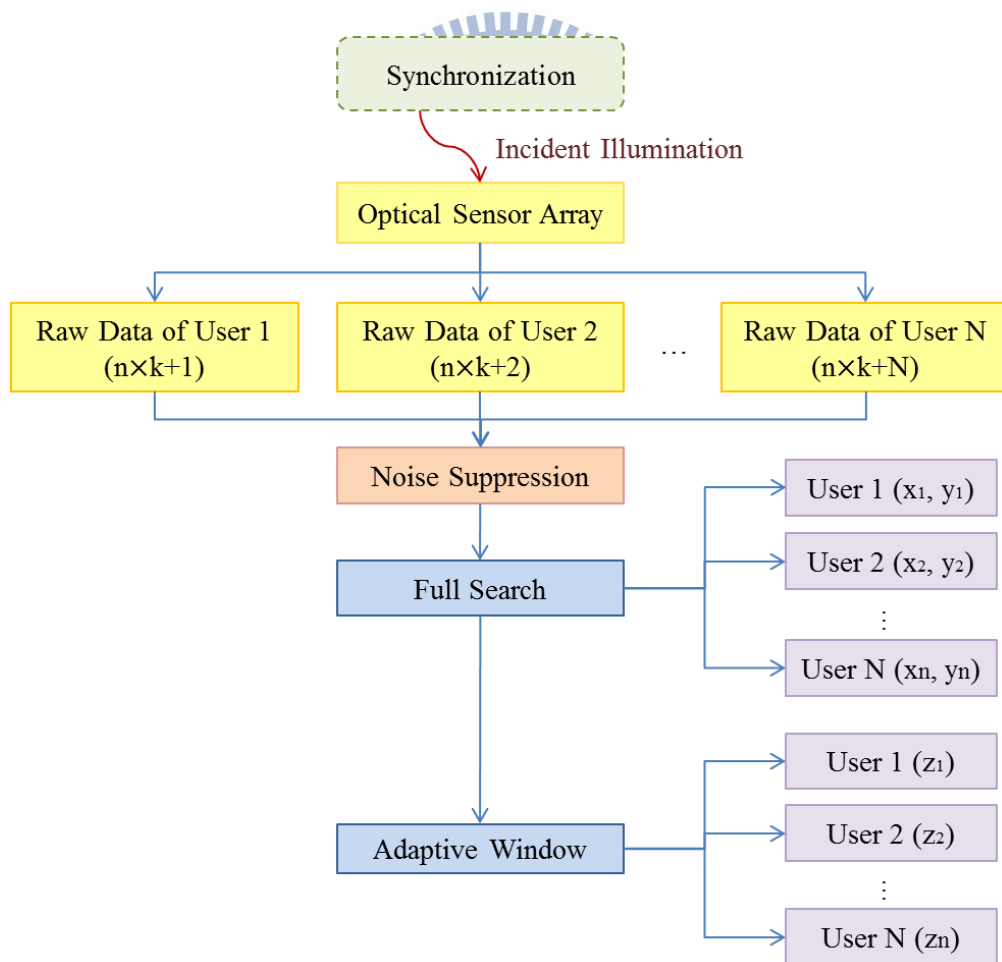


Fig. 40. Flow chart of Sequential-lighting algorithm.

There are three main steps in 3D coordinate acquisition. First, the noise suppression process can reduce the system noise and ambient light effect. Second, the 2D coordinate (x and y) can be calculated by Full-search method. [31] Afterwards, Adaptive window method is utilized to determine the depth information (z). Following, we'll discuss these methods in detail.

3.2.1 Synchronization

The principle of synchronization is similar to that in 3D shutter glasses. [32] The infrared light pen are controlled by an infrared, radio frequency, or Bluetooth transmitter that sends a timing signal which allows the light pen to alternately turn on and off in synchronization with the sensor frame of the panel. Hence the illumination and the sensor frame will operate in unison. Not only the overlapping issue can be eliminated but also different users can be identified according to the frame sequence.

3.2.2 Noise Suppression

Once incident light is detected, the optical sensors will generate photo current which is then converted into a grayscale image. Before calculating the 3D coordinate, the noise suppression process is essential for decreasing ambient light effect and the impact on the panel defects, such as the system noise and bright spots. Accordingly, if a pixel's gray level is lower than the threshold, it will be regarded as a noise and set to be zero. However, a bright spot cannot be eliminated by the previous method. Hence, if a pixel's 4-connectivity is all zero, it will be regarded as a bright spot and set to be zero as well. Therefore, background and system noises have been effectively depressed.

3.2.3 Full Search Method

Full search method is developed for 2D coordinate (x and y) detection. By utilizing a light pen with infrared LED as an input device, the projection of light would indicate the object's position. In ideal cases, the pixel with maximum gray level of the capture image would be the center of the projection and thus be regarded as the 2D coordinate (x and y) on the display surface. However, noise which cannot be fully eliminated in real cases might affect the result seriously. Therefore, Full search method is proposed to acquire 2D coordinate and meanwhile suppress the noise effect. An N by N search block is employed to accumulate the total intensity within the matrix, as illustrated in Fig. 41. Instead of finding a maximum point, a center of an area with maximum accumulation (i_{max} and j_{max}) will be regarded as the 2D coordinate of the object.

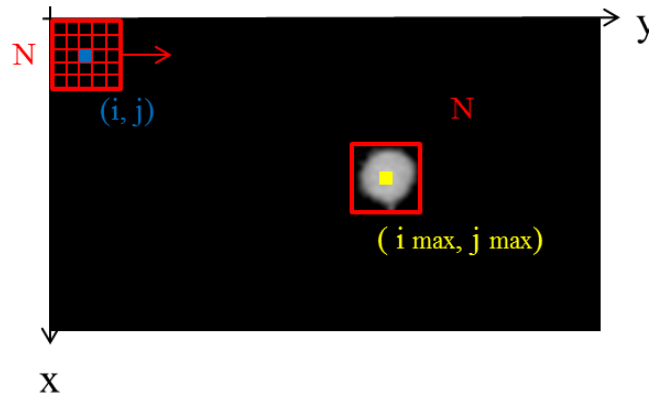


Fig. 41. Full search method: 2D coordinate (x and y) is defined by the maximum accumulation within the N by N search block.

3.2.4 Adaptive Window Method

As soon as the 2D coordinate (i_{max} and j_{max}) is obtained, the adaptive window method is employed for depth information (z) acquisition. The dispersion character in

LED results in size variation in the light projection, which increases while the light pen moving away from the panel. In other words, the depth information (z) can be determined by constructing a look-up table according to the projection sizes in according heights.

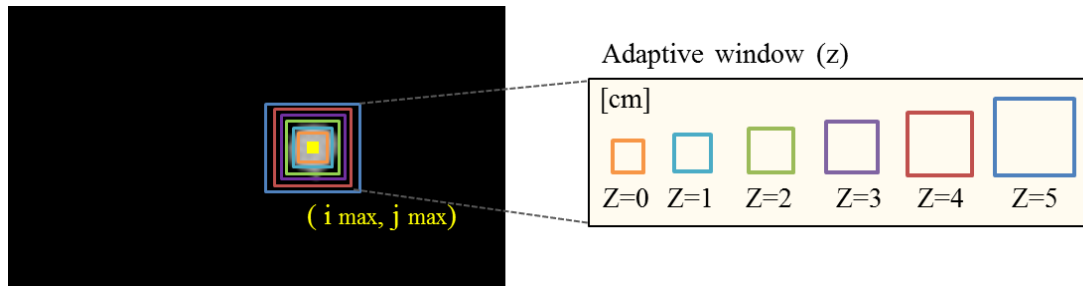
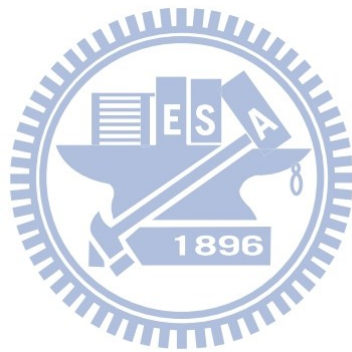


Fig. 42. Adaptive window method is employed to detect the depth value (z), where a look-up table for window sizes is built according to the experiments.

The size of the adaptive window is a function of depth value (z), as shown in Fig. 42. The convolution of an adaptive window with a captured image at the obtained 2D coordinate (i_{max} , j_{max}) is operated to define the depth value. For the adaptive windows which are larger than the size of projection, the convolution results are equal to the total intensity accumulation of the whole projection. Hence the differences between them are close to 0. Once the adaptive window is smaller than the projection, a smaller accumulation would be obtained. By detecting the significant difference increment, the depth value is defined by the previous adaptive window and its corresponding depth value. For example, for an object at 3 cm above the display, similar accumulations will be obtained for adaptive windows from window 5, which symbols the adaptive window at 5 cm, to window 3. Due to a smaller size of window 2, a lower accumulation is found. If the difference in accumulation between window 2 and 3 is greater than the threshold, the depth value is defined as 3 cm.

By utilized the proposed structure and Sequential-lighting algorithm, the 3D virtual multi-touch with user collaboration can be achieved. However, the hardware

limitation in insufficient sensing rate is one of the main issues. As we divide the lighting rate to $1/n$ of the sensor frame rate, the sensing rate for each user is divided by the number of users as well. For instance, for a display with sensing frame rate 60 frame/s, it is better to have users less than or equal to two to prevent discontinuity or line segmentation. The maximum tolerance in sensing rage per user is about 30 frames per second. If the sensing rate can be further enhanced, more users can experience virtual touch concurrently with smooth and continuous output responses.



3.3 Light-mark Method

Considering the feasibility in hardware implementation, the light-mark method is proposed to achieve multi-touch/ multi-user interaction in 3D virtual touch systems. Instead of separating the incident light in time domain, the light-mark method is able to identify different users within a sensor frame. The infrared light pens which are covered with different masks in front project coded light-marks on the panel. Through properly design the light-marks; the user identification can be achieved. Additionally, different masks and algorithms are proposed for different user demands. Generously, for a single user, there is a greater chance of marks overlapping when “gestures” [33], as shown in Fig. 43, are operated with multiple fingers. Hence, Multi-ring mark is proposed to enhance the recognition ability during overlapping conditions with identifiable touches. On the other hand, for multiple users each with a light pen, the user experience and freedom can be enhanced by perceiving not only 3D coordinate but also orientation information. Therefore, Multi-T mark is presented to track 5-axis (x , y , z , θ , and ϕ) information of different users. The concept of Multi-ring mark and Multi-T mark is illustrated in Fig. 44.



Fig. 43. Hand gesture models. [33]

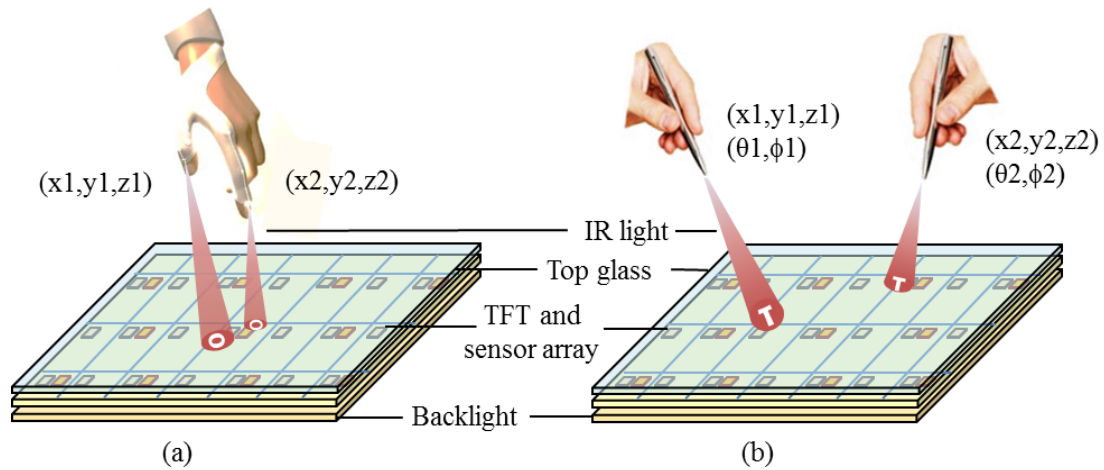


Fig. 44. Concept of (a) multi-touch for single user through light gloves [36] with Multi-ring mark and (b) multi-user interaction with 5-axis information by Multi-T mark approach.

3.3.1 Light-mark Design

To recognize or identify the objects, many kinds of Augmented Reality Tag (ARTag) [34] or markers have been developed. However, our target is not only to identify the objects, but also to deal with the overlapping conditions and to find the 5-axis (x , y , z , θ , and ϕ) information. Additionally, the resolution of embedded optical sensors is relatively lower compared to that of cameras, thus the mark must keep simple. Consequently, the light-mark comprises two parts, called out-mark and in-mark. Out-mark determines the perceivable degrees of freedom and is chose according to the request functions: either to enhance the ability in user identification or to further obtain the orientation information of users. Thus, out-mark with a symmetric pattern is able to provide the 2D coordinate (x , y) and the depth value (z) of users. Hence, a ring mark is designed for the following merits: first, instead of a solid circle mark, a ring mark is designed with intent to reinforce the identification ability because more details can be maintained by utilizing hollow and thinner patterns. Second, without acquiring orientation information, the shape of a ring mark

barely changes while experiencing a planner rotation. Hence a more solid and less memory-cost algorithm can be produced. Otherwise, out-marks with an asymmetric pattern, where a T-shape is chosen, is able to provide not only 3D coordinate (x, y, and z) but also orientation information including tilt angle (θ) and rotation angle (ϕ), which are derived from pattern's directional features.

On the other hand, in-mark, which is also called the characteristic mark, is designed for user identification. Different circular patterns are tendered by considering overlapping conditions of the ring-marks. Different numbers of blocking strips are presented corresponding to the T-marks. Simply by change the pattern or increasing the number of blocking strips, the number of recognizable users can be extended. The overall light-marks are illustrated in Fig. 45. Following, Multi-ring-mark algorithm and Multi-T-mark algorithm are proposed in compliance with the designed marks.

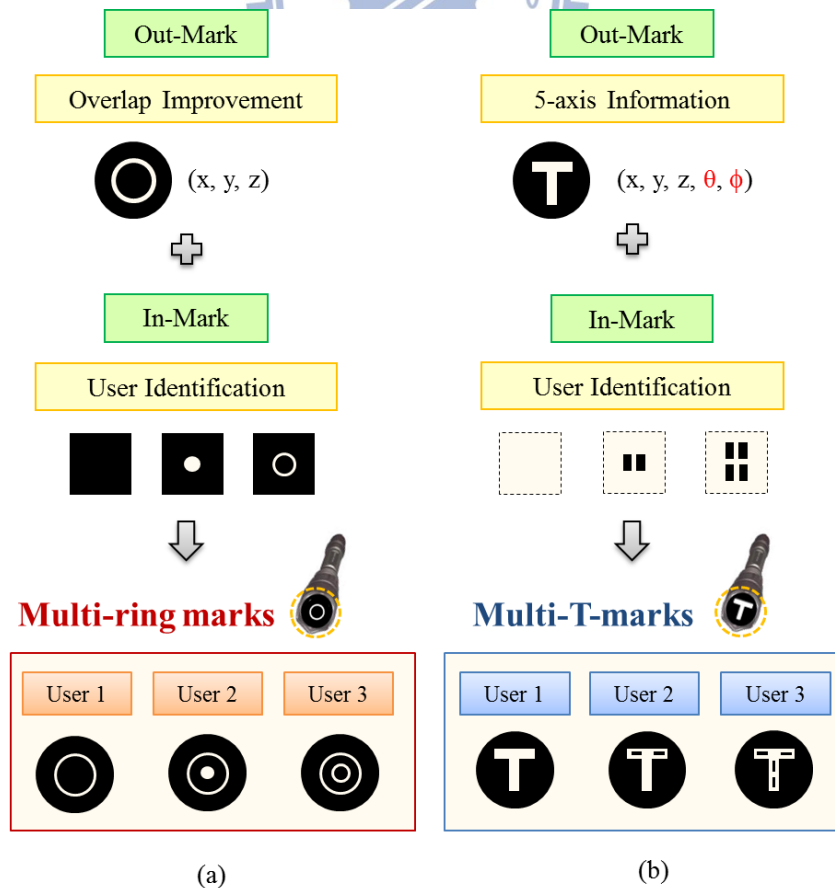


Fig. 45. Principle of mark design including out-mark and in-mark, where out-mark determines the perceivable degrees of freedom, and in-mark is utilized for user identification. (a) Multi-ring mark design, (b) Multi-T mark design.

3.3.2 Multi-ring Mark: Overlap Improvement

The intention for multi-ring-mark algorithm is to enhance user recognition during marks overlapping. The flow chart of the algorithm is shown in Fig. 46. The projected light marks are detected by the embedded optical sensors and converted into a gray level image. After noise suppression, the system noise and ambient light effect can be reduced. There are two main stages in the algorithm: first is to locate the 3D coordinate (x , y , and z) by extracting the out-mark feature. Second is to differentiate users through the in-mark character.

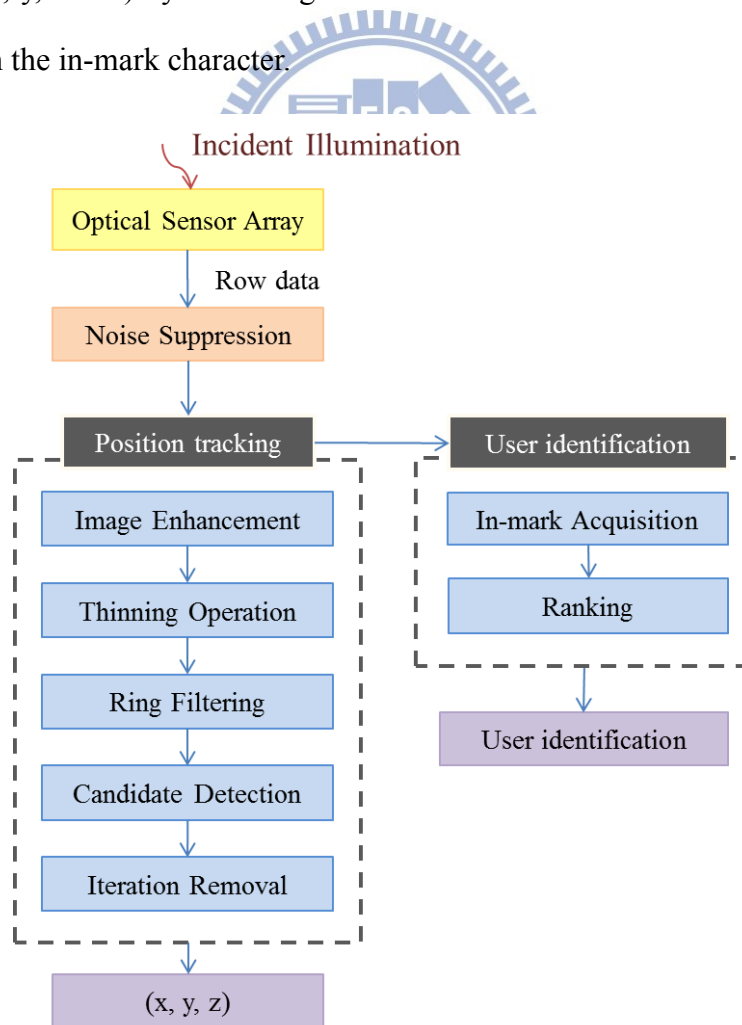


Fig. 46. Flow chart of Multi-ring mark algorithm.

Position tracking: In the first stage, our goal is to define 3D coordinate (x, y, and z) of the users. To achieve the target, five image processing steps, including image enhancement process, image thinning operation, Ring filtering, peak detection, and iteration peak elimination, are employed in turn. In the following, we will discuss each process in detail.

Image enhancement process is executed after noise suppression, hence only the main information which means the projected light marks would be reinforced. Due to the decay of light, a vague shape will be captured for a higher object. Hence in the image enhancement process, we convert the captured image to a binary image. For pixels with gray level larger than the threshold, they are reassigned to ones. And others are reassigned to zeros. By the image enhancement process, the image becomes a binary image. Besides, the intensity difference which results from the decay of light from unlike heights and non-uniform sensitivity of the display is eliminated. Additionally, the image enhancement process happens to fit the requirement in our next step thinning operation where the input image should be a binary image.

Thinning operation [35] is used to reduce the line width of ring mark to single pixel thickness. The thinning of an image A by a sequence of structuring element {B}, denoted $A \otimes B$ [36], can be defined as

$$A \otimes \{B\} = ((\dots((A \otimes B^1) \otimes B^2) \dots) \otimes B^n)$$

Eq. (2)

In every term, the thinning operation is calculated by translating the origin of the structuring element, as shown in Fig. 47, to each point in the image, and comparing it with the underlying image pixels. If the foreground (one) and background (zero) pixel, in the structuring element exactly match to the image, then the image pixel underneath the origin of the structuring element is set to background (zero). Otherwise it is left

unchanged. After a single term of a thinning operation over the image, the operator is applied repeatedly until convergence, which means no further changes to the image. By thinning operation, down to one pixel's line width can be obtained, as shown in Fig. 48.

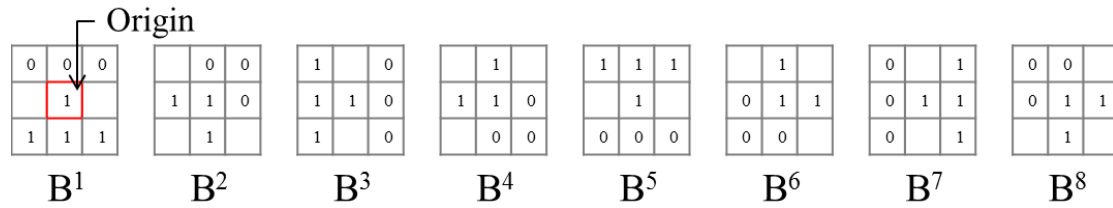


Fig. 47. Structuring elements for morphological thinning. Ones and zeros stand for foreground and background pixels and the blanks can be either one or zero which we don't care about. At each term, the image is first thinned by B^1 , then B^2 , as so on. The process is repeated until none of the thinning produces any further change.

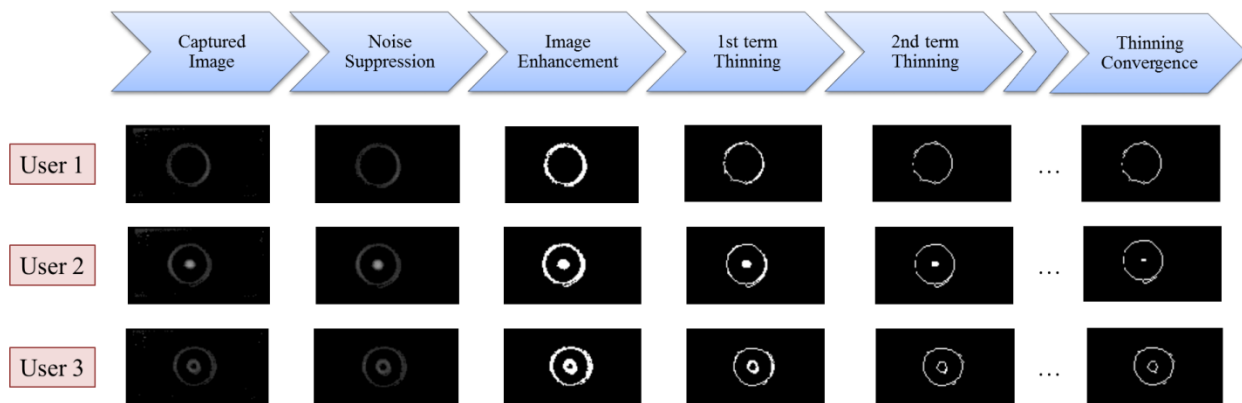


Fig. 48. Light pens are positioned at 50mm and detected by the embedded optical sensors. Output images after image enhancement steps and thinning operation from the 1st term until convergence are shown.

Ring Filtering is able to determine the possible 3D coordinates (x, y, and z) of users, called candidates. Ring filters are constructed according to the size of out-mark at different heights, as demonstrated in Fig. 49. By translating the origin of the ring filter to all points in the image, and processing convolution the ring filter with the underlying image pixels. The numbers of the foreground pixels in the ring filter that

match to the ones in the image are added up to the pixel underneath the origin of the ring filter.

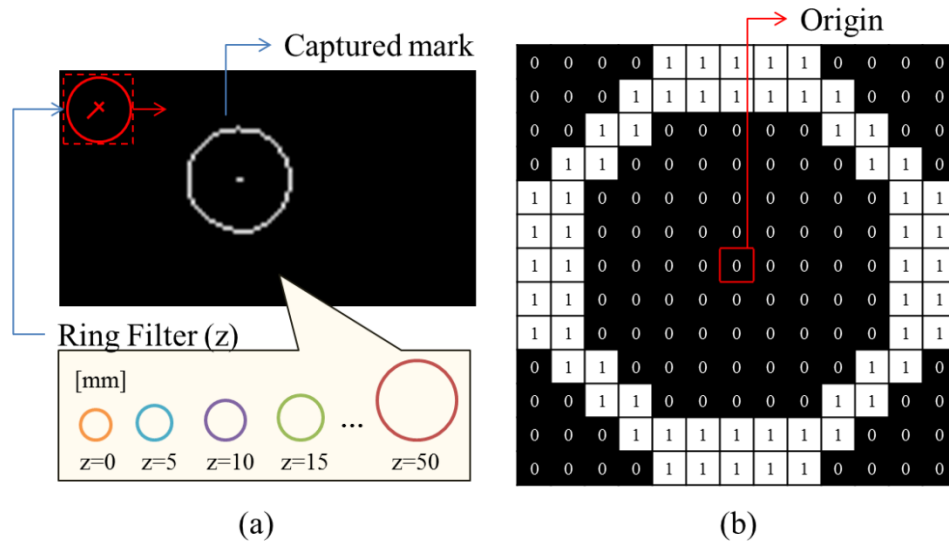


Fig. 49. (a) Ring filter are constructed according to the size of out-mark at different heights, and (b) shows the structure element of ring filter at $z=0$.

Candidate Detection can locate the possible 2D coordinate (x and y) and meanwhile determine the depth value (z) according to the results in ring filtering. A series of normalizing factors is constructed by accumulating thinning ring marks at different heights. Each ring-filtered image is divided by accordant normalizing factor. The center of ring mark can therefore be defined with value close to 1 due to the similar accumulation of the captured ring mark to the normalizing factor. An example of the normalized accumulation within each ring filter is shown in Fig. 50. A local maximum which is found by a pixel greater than its 8-connective with its value close to 1 will be regard as a candidate. However, when the number of candidates that fit the requirements is larger than the number of the users, it indicates that some error or iterated candidates have been detected. As shown in Fig. 50, 4 candidates have been found. Hence in the next step, we'll try to remove the redundant candidates.

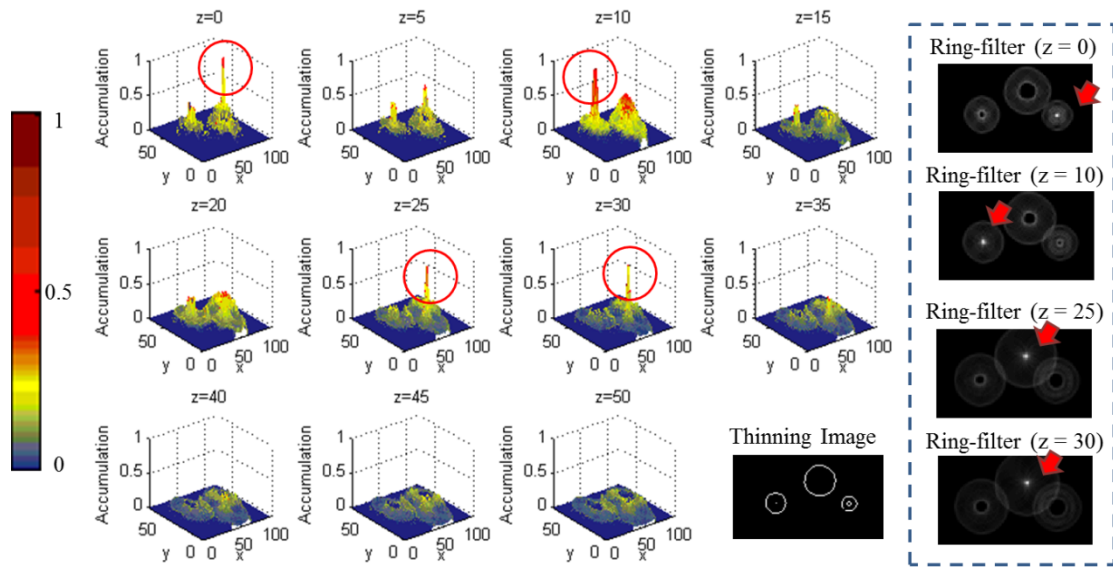


Fig. 50. In candidate detection, the accumulation results at different ring filter (z) are divided by the corresponding normalizing factors. The local maximum with its value close to 1 will be defined as the candidates.

Iteration Removal process is presented to effectively eliminate the redundant candidates. If a distance between two candidates is smaller than the according in-mark window, which will be discussed in the next paragraph, the candidates might actually represent the same ring where only a user exists. Hence, the candidate with value closest to 1 will be reserved while the other will be removed. Such as the example in Fig. 50, 2D coordinates (x and y) of the candidates in ring-filtered image $z=25$ and $z=30$ are almost the same. Therefore, by iteration removal process, only candidates at $z=0, 10$, and 30 are remained. The 3D coordinates (x , y , and z) of users can be obtained.

User identification: Once the 3D coordinates of the users are found, the next stage is to assign each coordinate to the corresponding user.

In-mark acquisition is a step to extract characteristic mark in order to achieve user identification. As shown in Fig. 51, the size of in-mark window is as well a

function of z . Thus, by knowing 3D coordinate (x , y , and z) of the user, the origin of in-mark window (z) is positioned on the image at accordant 2D coordinate (x , and y). Therefore, the user character can be obtained without affect by outer ring mark.

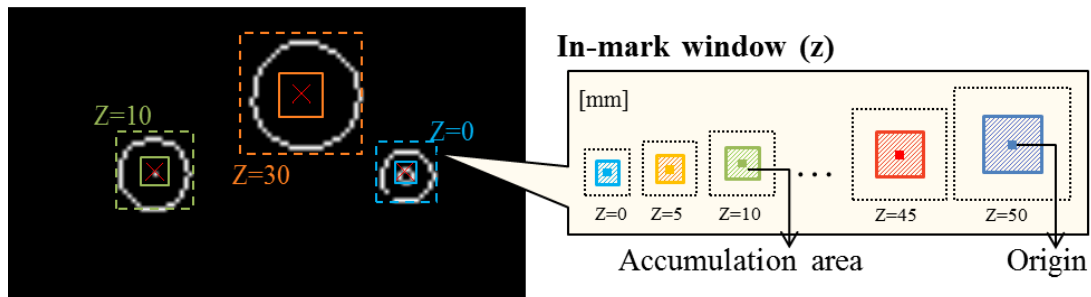


Fig. 51. In-mark window for in-mark extraction.

Ranking process is the last step to match the 3D coordinates to the users. After in-mark acquisition, users can be identified easily by ranking the accumulation of in-marks without normalizing windows to the same size. Due to the previous thinning operation, a solid circle converges to a single point while a ring stays a ring shape with radius equals to the average of the original image. As demonstrated in Fig. 52, user 1 with a blank in-mark can be recognized by minimum in-mark ranking. User 2 with a solid circle as in-mark has middle ranking due to the thinning operation. And User 3 with a ring is defined by maximum in-mark ranking. Therefore, 3D coordinates are assigned to accordant users successfully.

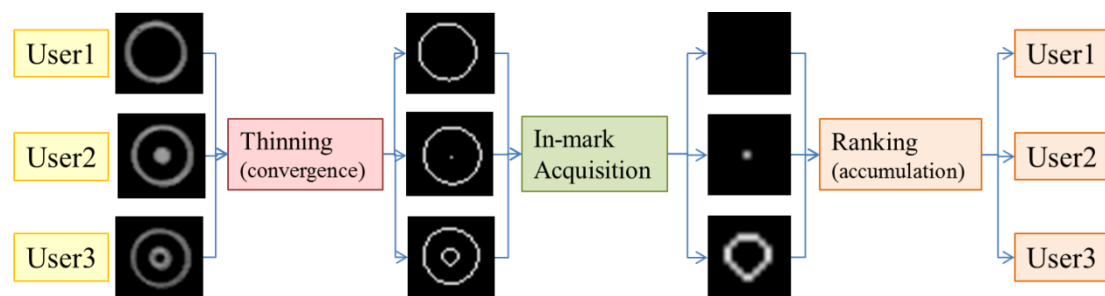


Fig. 52. Ranking process is able to define different users.

3.3.3 Multi-T Mark: 5-axis Information (x, y, z, θ, ϕ)

Multi-T mark algorithm is proposed in compliance with the Multi-T mark design. The main algorithm can be divided into three parts, 2D position tracking, depth and orientation tracking, and the user identification. The flow chart of Multi-T mark algorithm is demonstrated in Fig. 53 where the number of users and their 5-axis (x, y, z, θ, ϕ) information are acquired in compliance with the user.

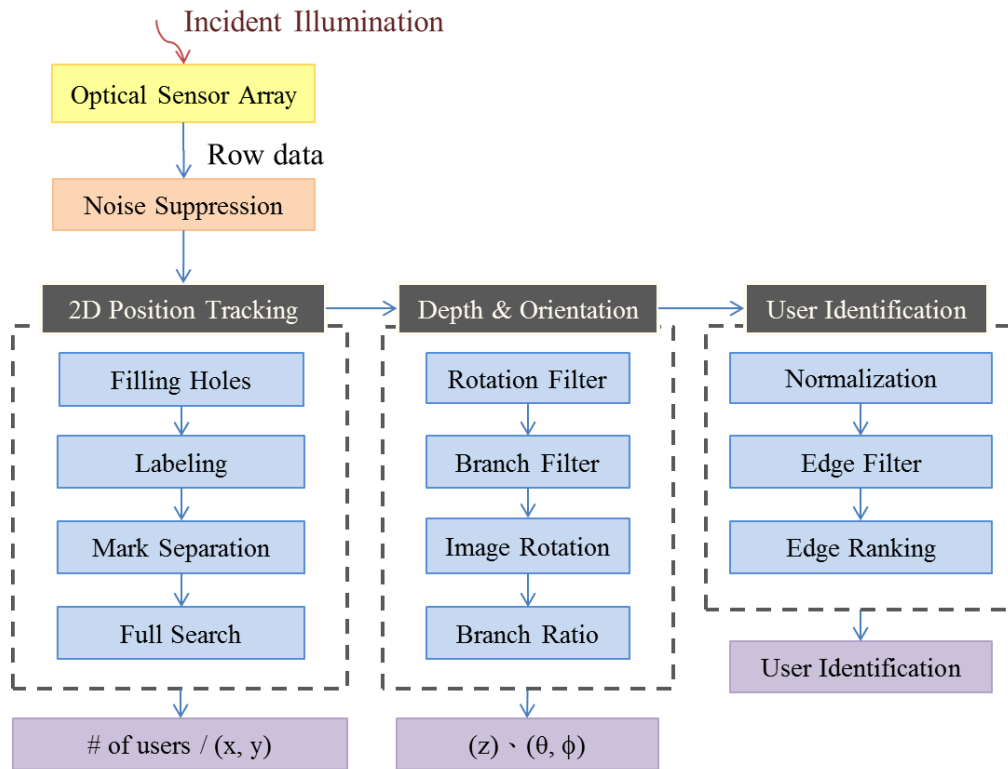


Fig. 53. Flow chart of Multi-T mark algorithm.

2D (x and y) Position Tracking: For each detected row data, the noise suppression process which shares the same mechanism to that in the Sequential-lighting method is conducted to minimize the noise effect. Our first target is to define the 2D coordinate (x and y) of the objects by out-mark T. In order to erase the in-marks effect, the blocking strips are temperately filled up by filling holes operator.

Filling Holes Operation [39] is a procedure based on morphological reconstruction where geodesic dilation and morphological reconstruction by dilation are utilized. The geodesic dilation of size 1 of the marker image (F) w.r.t. the mask (G), denoted by $D_G^{(1)}(F)$, is defined as

$$D_G^{(1)}(F) = (F \oplus B) \cap G \quad \text{Eq. (3)}$$

Where B is a structure element of 3 by 3 ones matrix, and \cap denotes the set intersection which may be interpreted as a logical AND. The geodesic dilation of size n of F w.r.t. $D_G^{(n)}(F)$ is defined as

$$D_G^{(n)}(F) = D_G^{(1)}[D_G^{(n-1)}(F)] \quad \text{Eq. (4)}$$

with $D_G^{(0)}(F) = F$. In this recursive expression, the set intersection in Eq. (3) is performed at each step. The intersection operator guarantees that mask G will limit the dilation of marker F .

Based on the preceding concepts, morphological reconstruction by dilation of a mask image G from a marker image F , denoted $R_G^D(F)$, is defined as the geodesic dilation of F w.r.t. G , iterated until stability is achieved; that is,

$$R_G^D(F) = D_G^{(k)}(F) \quad \text{Eq. (5)}$$

with k such that $D_G^{(k)}(F) = D_G^{(k+1)}(F)$.

In filling holes, we suppose that $I(x, y)$ denote a binary image and F is a marker image that is 0 everywhere, except at the image border, where it is set to $1-I$; that is,

$$F(x, y) = \begin{cases} 1 - I(x, y) & \text{if } (x, y) \text{ is on the border of } I \\ 0 & \text{otherwise} \end{cases} \quad \text{Eq. (6)}$$

$$H = [R_G^D(F)]^c$$

is a binary image equal to I with all holes filled.

Let us take one of the Multi-T marks to illustrate the working principle of filling holes operator. A binary Multi-T-mark with four blocking strips, which means four holes, is shown in Fig. 54 (a). Its complement I^c , as shown in Fig. 54 (b), is used as an AND mask to protect all foreground (1-valued) pixels from changing during iteration of the procedure. A marker image F is an array formed according to Eq. (6). Next, the marker F is dilated with a 3×3 structure element B whose elements are all 1s, as shown in Fig. 54 (d), so the dilation of F of the marker points starts at the border and proceeds inward. Fig. 54 (e)-(i) illustrate the iterative geodesic dilation of F using I^c as the mask until another iteration will yield that same result which, when complemented as required by Eq. (7), gives the result in Fig. 54 (j). Therefore, by utilizing the filling holes operator, the out-mark features can be extracted successfully, as shown in Fig. 55.

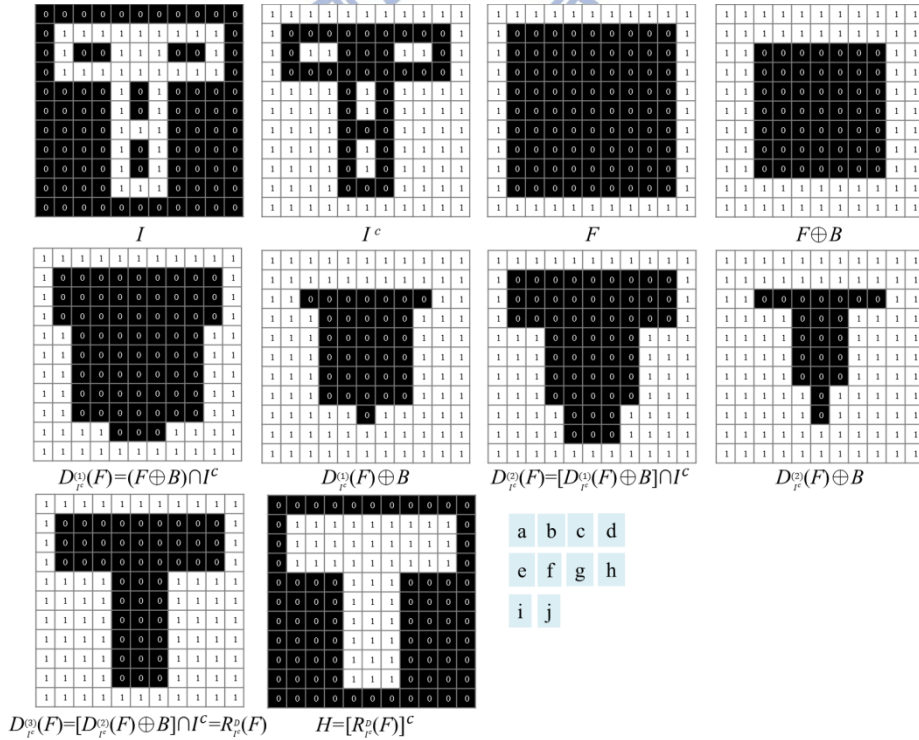


Fig. 54. Illustration of filling holes operating on a multi-T mark image.

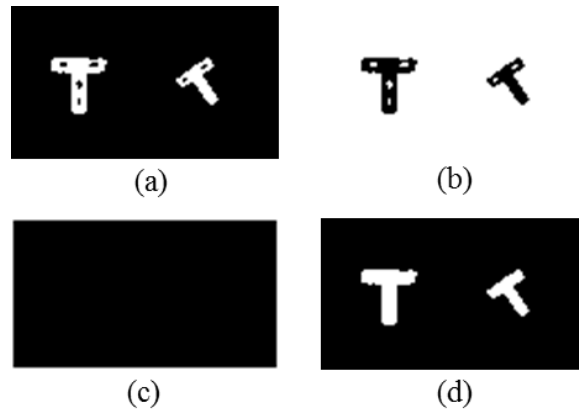


Fig. 55. A practical example for Filling holes. (a) is the binary image by thresholding the captured image, and (b) shows the complement of the binary image in (a) for use as a mask image. (c) is the marker image, F , generated by Eq. (6). Finally (d) shows the image with all the holes filled.

Without overlapping conditions, the number of objects can be determined by the connected-component labeling operator.

Connected-component Labeling [37] operator scans an image pixel-by-pixel in order to identify connected pixel regions. Considering the rotation condition of a projected T mark, 8-connectivity is chosen. i.e. regions of adjacent pixels share the same set of intensity values V , for which $V=\{1\}$ in a binary image. The algorithm makes two passes [38] over the image: One pass to record equivalences and assign temporary labels, and the second to replace each temporary label by the label of its equivalence class. Connectivity checks are carried out by scans the image by moving along a row until it comes to a point p , where p denotes the pixel to be labeled at any stage in the scanning process. When this is true, it examines the four neighbors of p which have already been encountered in the scan. Based on this principle, the labeling of p occurs as follows:

- If all four neighbors are 0, assign a new label to p , else
- If only one neighbor has $V=\{1\}$, assign its label to p , else
- If more than one of the neighbors have $V=\{1\}$, assign one of the labels to p

and make a note of the equivalences.

After completing the scan, the equivalent label pairs are sorted into equivalence classes and a unique label is assigned to each class. As a final step, a second pass is made through the image, during which each label is replaced by the label assigned to its equivalence classes. For display, the labels can be different colors as shown in Fig. 56.

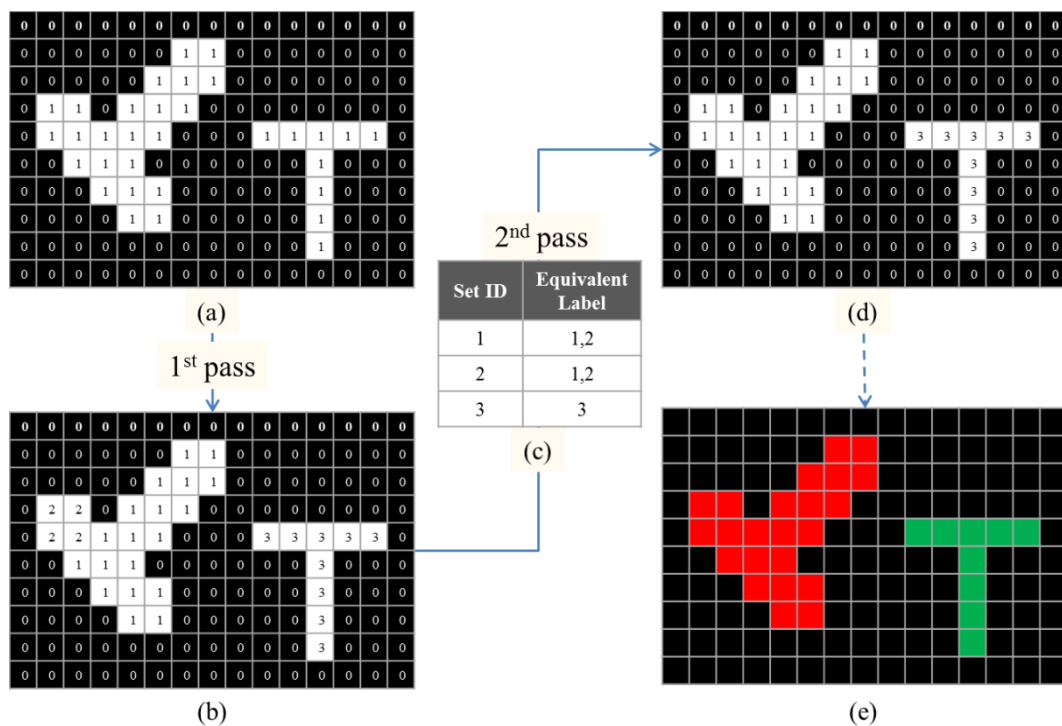


Fig. 56. Connected-component labeling operator: (a) Row data of detected image converts into binary image. After the first pass, the following labels are generated as (b). The label equivalence relationships are given in (c) and then (d) merging the equivalent labels with the label value that is the smallest for a given connected region. Final result (e) in color to clearly see two different regions that have been found

The connected-component labeling not only determines the number of user, but meanwhile denotes each region with its unique label. Hence, for *Mark separation*, each labeled region is used as the mask to isolate the projected light-marks. The operation $I(x, y) \cap F$, where $I(x, y)$ denotes the row data and F denotes the labeling

masks, as illustrated in Fig. 57, yields the mark extraction onto different images. This process is essential to eliminate the disturbance that might happen in multi-user interaction.

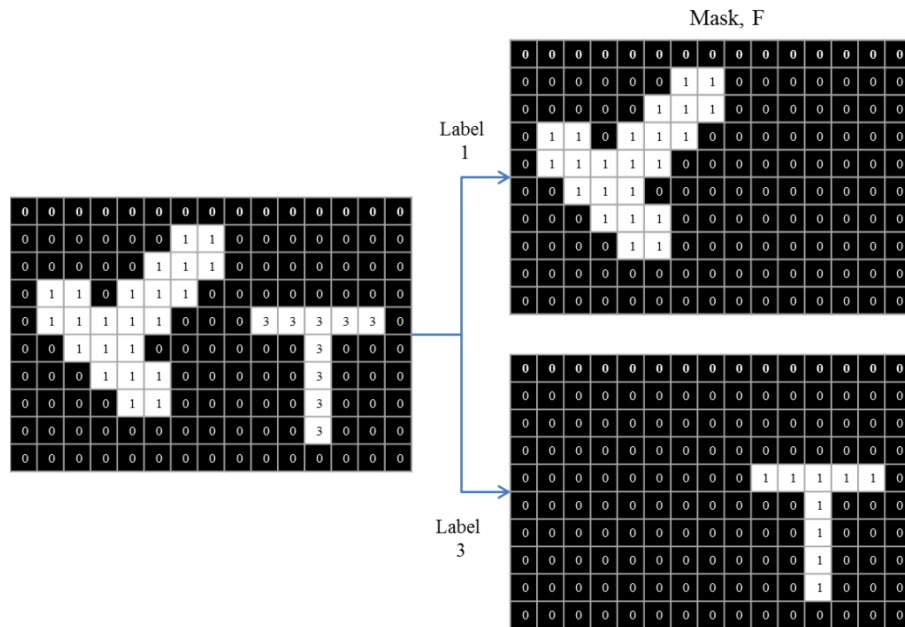


Fig. 57. Unique label on each connected regions are utilized to separate different marks to different binary images.

By isolating the detected marks, the 2D coordinate tracking is simplified to a single mark tracking which is similar to that in sequential-lighting method. Therefore, the full search method is applied to extract the 2D coordinates (x and y) by finding the maximum accumulation of the search window, as shown in Fig. 58.

However, due to the effect of the extra freedom in tilt, the depth value (z) can no more be identified by the adaptive window method. For non-normal incidence of a light pen, the projected light mark would distort and results in depth error.

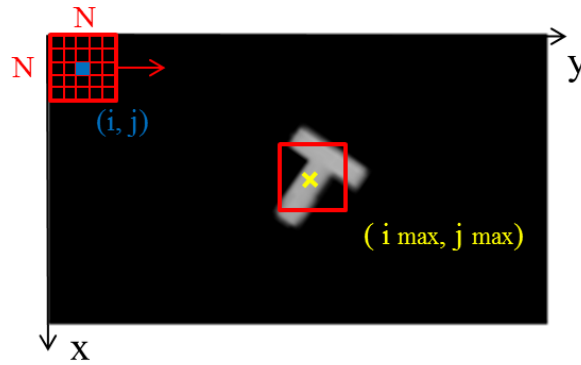


Fig. 58. 2D coordinate of a T mark is acquired by the full search method, where the origin of the maximum accumulation in search window (i_{max} and j_{max}) is assigned to the 2D coordinate (x and y).

Depth and Orientation tracking: The orientation angles (θ and ϕ) and the short/long branches are defined as Fig. 59. In our proposed system, the motion of tilt is limited to the short branch of a multi-T mark. Hence the projected long branch experiences a more serious image distortion for non-normal incidence than the short branch. Therefore, the depth value (z) should be identified by the length of the short branch instead of a window as that in the adaptive window method. By applying the rotation filter and branch filter, the rotation angle (ϕ) and the short/long branches can be defined. Following by the image rotation process and branch ratio method, the depth value (z) and the angle of tilt (θ) can be acquired as well.

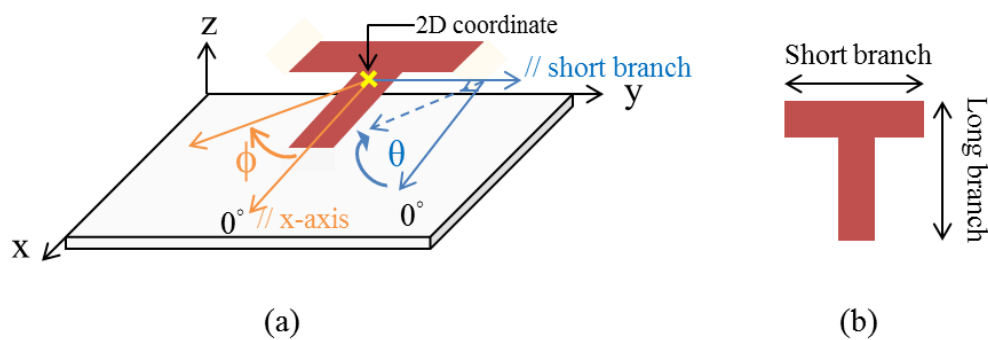


Fig. 59. (a) Orientation angles about z-axis is called the rotation angle (ϕ), and the tilt angle (θ) is rotated about the axis parallel to the short branch. (b) A normal

direction of a multi-T mark with the definition of short and long branches.

Rotation filter is utilized to acquire the orientation angle about z-axis (ϕ). Considering the feasibility in hardware implementation, a look up table for rotation angle from 0° to 170° , as shown in Fig. 60, with approximate ratio was established instead of calculating the precise rotation angle. By knowing the 2D coordinate of a multi-T mark, the maximum convolution of the mark and the rotation filter reveals the highest similarity of the mark to the filter. Therefore, the accordant rotation filter could be defined. However, without knowing the orientation of the short branch, the precise rotation angle (ϕ) cannot be identified. For a rotation filter, it actually indicate two angles: θ and $\theta+180^\circ$ due to the same orientation of their long branches. Hence, it is essential to define the orientation of a short branch by the following branch filter method.

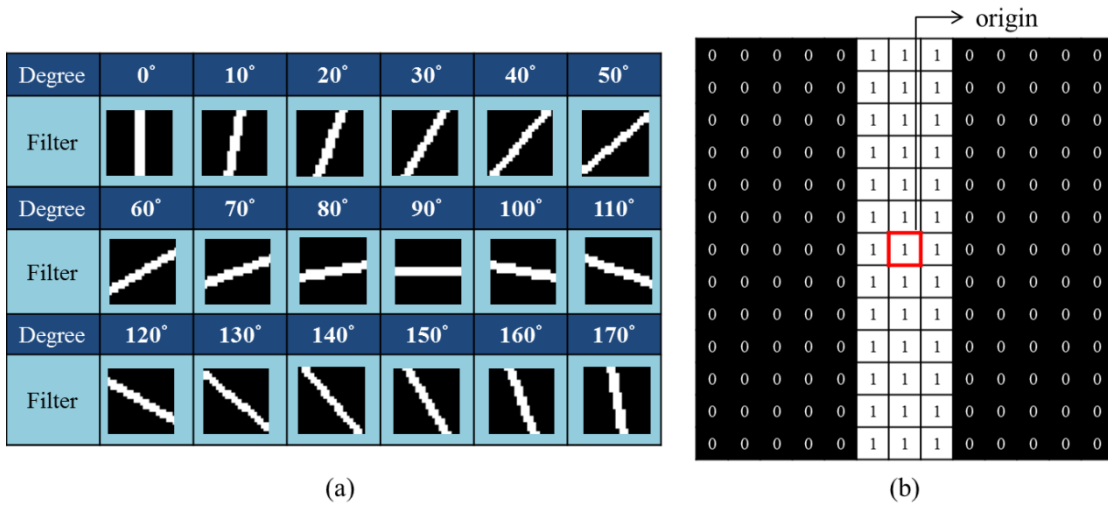


Fig. 60. (a) Rotation filters from 0 to 170 degree with 10 degree's step where a normal T with a long branch parallel to x-axis is defined to 0 degree. (b) Matrix of the rotation filter at 0 degree.

Branch Filter is used for identifying the orientation of a short branch in a multi-T mark and acquired the rotation angle (ϕ). The watershed of a branch filter is perpendicular to the according rotation filter, as shown in Fig. 61. By knowing the 2D

coordinate (x and y) and the rotation angle of a multi-T mark, the accordant branch filters, up and down, are applied for branch identification. If an accumulation result reveals a larger up filter, the short branch locates in the up filter. Hence rotation angle θ equals to the angle of accordant rotation filter. On the other hand, if the accumulation of a down filter is larger an up filter, the short branch locates in the down filter. Hence there should be a 180° modification, where the rotation angle is $\theta+180^\circ$.



















Degree	0°	10°	20°	30°	40°	50°
Filter						
Degree	60°	70°	80°	90°	100°	110°
Filter						
Degree	120°	130°	140°	150°	160°	170°
Filter						

Fig. 61. Branch filters from 0 to 170 degree is able to define the orientation of short branch in a multi-T mark.

Take a 30 degrees rotation filter as an example, as shown in Fig. 62. If 30° is obtained in a rotation filter, 30 degrees branch filters will be applied by accumulating the accordant up and down branch filters. If the up filter larger than the down filter, it reveals that the short branch locates in up filter area. Hence the rotation angle is 30° with is identical to the rotation filter. On the other hand, if the down filter is larger than the up filter, the capture T mark is like a mirroring image to the previous case. Hence there should be a 180° modification. The rotation angle is 210° .

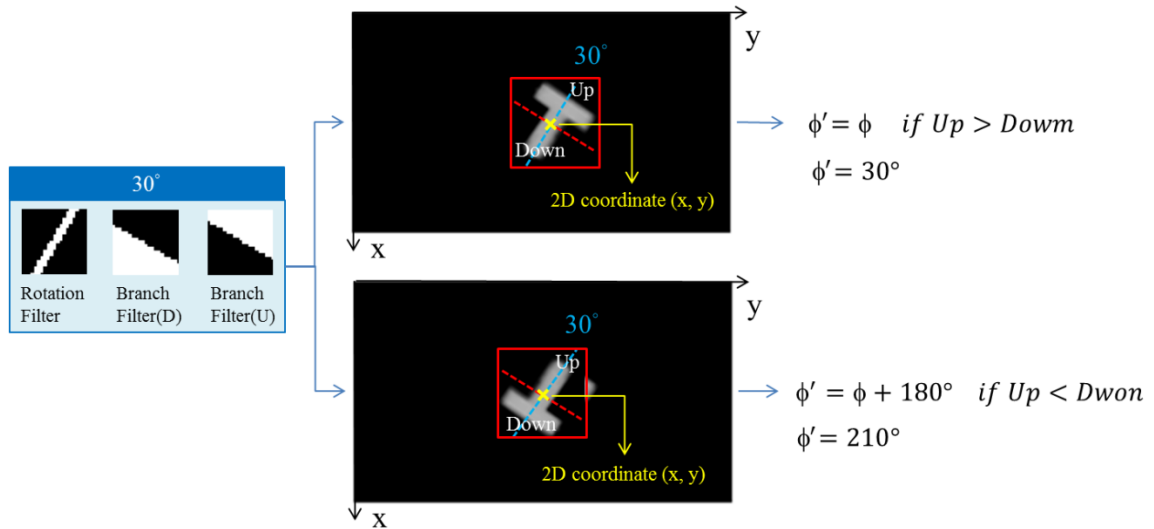


Fig. 62. Multi-T marks with rotation filter 30 degree indicate the rotation angles of (a) 30 degrees or (b) 210 degrees.

Image Rotation and Branch Ratio Method are applied to determine to depth value (z) and the tilt angle (θ). Image rotation with bilinear interpolation is processed to transform each separated multi-T mark into normal direction. By knowing the rotation angle ϕ by previous steps, the mark is rotated with $-\phi$ degree. Thus the long branch will be parallel to the x-axis. To calculate the length of the branches, scan lines accumulate the columns and rows from the image border respectively, as shown in Fig. 63. As long as the accumulation is above the threshold, the scan lines stop and the length of the short and long branches are equal to the remaining un-scanned columns and rows respectively. Other digital image processes such as skeleton can also be used to acquire the length of branch. However, in practical manufacturing of SoC (System-on-a-chip), a simplifier algorithm shall be chosen. Hence, rotation and scanning processes are provided.

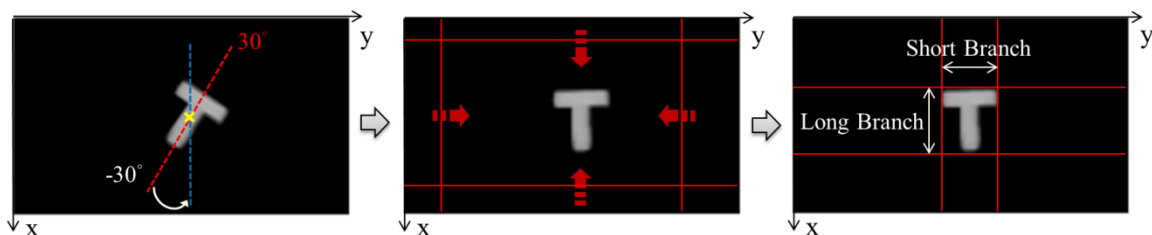


Fig. 63. Image rotation and length of braches acquisition.

Due to the light dispersion and the limitation in rotational movement, the length of the short branch is mainly affected by the height of the object. Therefore, a look-up table for depth value (z) according to the projected length of the short branch is built. The depth information can thus be acquired by knowing the length of the short branch.

Finally, the last axis information, the tilt angle (θ) is acquired by the branch ratio method. As illustrated in Fig. 64, for the tilt axis parallel to the short branch of T, the length of the projected short branch (S) barely changes according to the motion of tilt where

$$S_1 \sim S_2 \sim S_3$$

at the same height. On the contrary, there exists a sensible change on the length of the projected long branch (L). Hence, a look-up table for the ratio of branch is constructed. However, L increases with both direction of tilt. An additional light intensity factor has to be taken into account. In a positive tilt condition, the intensity at the end of the long-branch decreases. On the other hand, light decays at the head of the short-branch for a negative tilt. Therefore, a tilt angle (θ) can be obtained.

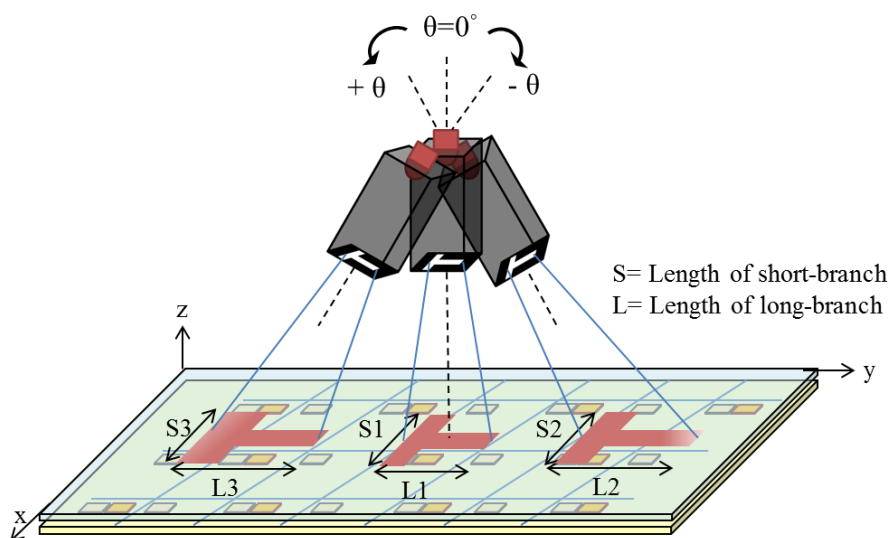


Fig. 64. Under different tilt angles, the branch ratio L/S changed accordingly. Hence $L1/S1 \neq L2/S2 \neq L3/S3$.

User identification: The third part of the multi-T-mark algorithm is user identification. Although the intensity of captured light might not be the same due to ambient light effect, the shape of the T-marks stays unchanged, which is a proper factor for user recognition. Hence, as long as we normalize the marks to the same size, where the size is related to the distance for the surface to the light pen, the ranking of the edge accumulation of marks will be the same order as that of the users by utilizing the ascendant numbers of blocking strips inside multi-T marks.

An edge filter is applied to extract the in-mark features. Even though the projection of the blocking strips might be blurred or distorted due to light dispersion especially for higher objects, the intensity discrepancy in strip regions will be retained and able to extract. Comparisons of various edge detectors have been made, as shown in Fig. 65. The Prewitt approximation performs a better result where the out-mark T can be kept with minimum distortion, due to the line factor of designed T mark which is match to the Prewitt operators. Meanwhile, the 1st derivative process in Prewitt edge filter maintains a better character of the in-mark blocks, which might be blurred due to light dispersion, than others which perform 2^{ed} derivative in the Laplacian of Gaussian filter or dual threshold in the Roberts and Canny filters [39].

Operator	Original Image	Prewitt	Zero-crossing
Image (z)			
Image (phi)			
Operator	Roberts	Laplacian of Gaussian	Canny
Image (z)			
Image (phi)			

Fig. 65. A comparison of various edge detectors with marks at different heights, which is a function of z , and orientation, which is a function of ϕ .

Before extracting the in-mark feature, the normalizing process, as in Eq. (8), is executed on each mark to tune the mark size equally.

$$T_{normalized} = T_{captured} \times \frac{K1}{S} \times \frac{K2}{L}$$

wehre $K1, K2 = \text{const.}$ $\left\{ \begin{array}{l} S = \text{the length of } \textit{short} \\ L = \textit{long} \text{ branch} \end{array} \right.$

Eq. (8)

Hence in accordance with the in-mark design, the user identification can be achieved by the sorting of the edge accumulation of extracted in-mark features, where we assume more in-mark strips, which results in larger accumulation result, for the latter no. of user, as illustrated in Fig. 66.

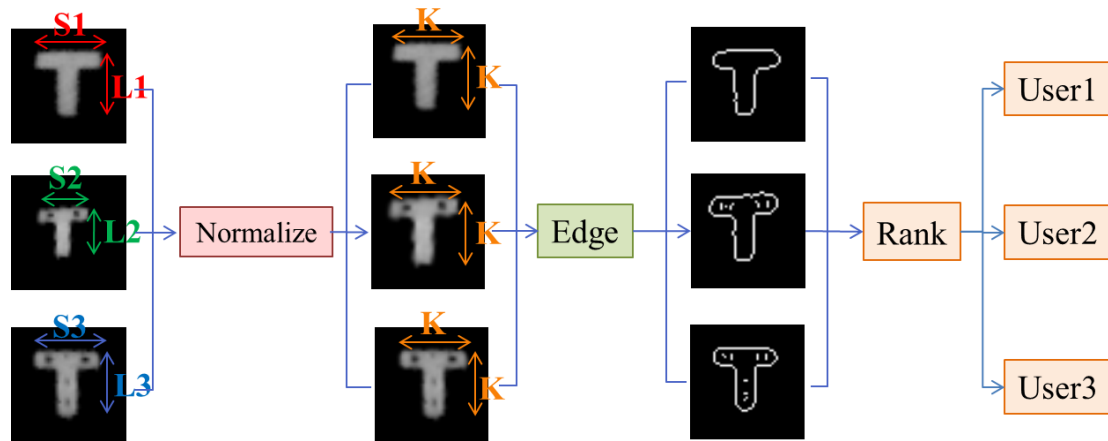
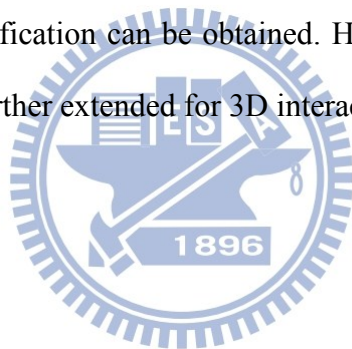


Fig. 66. User identification is achieved by normalizing (N) each multi-T mark to the same size and processing Prewitt edge filter (E) to extract in-mark feature. Finally the no. of user can be identified by ranking (R) the accumulation.

Accordingly, by using the Multi-T-mark algorithm, the 5-axis (x, y, z, θ , and ϕ) information with user identification can be obtained. Hence, more complex and user friendly interfaces can be further extended for 3D interaction applications.



Chapter 4 *Experiments and Results*

Experiments and Results

4.1 Experiment Setup

In order to prove the feasibility of the proposed algorithms, a prototype of a 3D interactive system was fabricated based on a 4-inch embedded optical sensor display, as shown in Fig. 67. The optical sensors were integrated with the TFT substrate to form a sensible LCD. The sensor resolution was set to be 68×120 , which was one fourth of the image resolution. The optical sensors were embedded in red and blue sub-pixels. The blue sensor was employed for easier verification of the proposed algorithms by utilizing visible light. However, since the blue sensor was affected by ambient light easily, the experiments had to operate in dark environment. The infrared light demonstrated real conditions in 3D interaction with lower ambient light affect. Meanwhile, image quality can be maintained by utilized infrared light as the input light source. The large bandwidth of red filter allows the infrared light passing thorough and being captured by the IR sensor. A FPGA-based control board drove the sensor array to operate at 30Hz and converted detected light into gray level images. A blue LED and an 850 nm IR LED with 25° divergence angle were used as input light sources with operating voltage at 3V and operating current at 0.15 A. Finally, the light marks were fabricated by the shadow mask preliminarily.

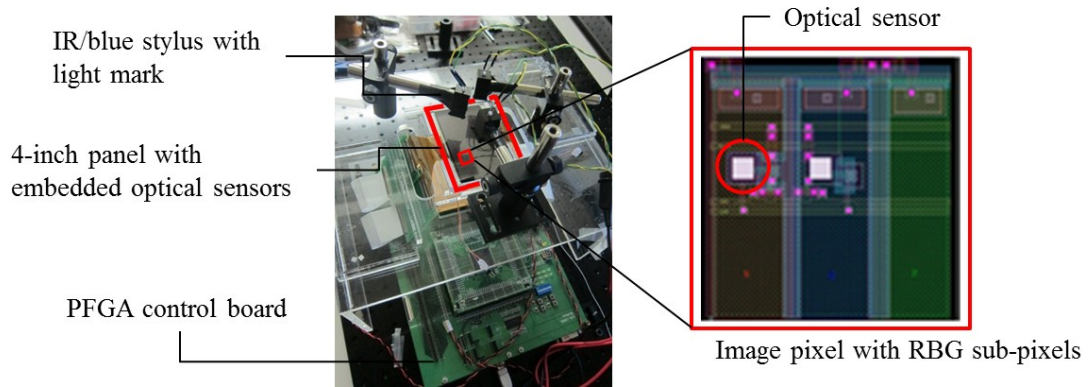


Fig. 67. Prototype of the 3D multi-interactive system with 4-inch sensible LCD notarizes the proposed algorithms.

4.2 Sequential-lighting Method

To demonstrate the feasibility of sequential illuminating method on our system, we modeled the situation that three users crossed over on their trajectories, as shown in Fig. 68. User 1 moved from the left to the right, user 2 moved from the top to the bottom, while user 3 stayed on the right side of the panel. To keep the experiment simple and clear, the heights of users were fixed as they were only moved horizontally. Without employing Sequential-lighting Method, three light marks were captured at the same time (sensor frame), as shown in Fig. 69. Hence, the overlapping of marks occurred and different users cannot be identified.

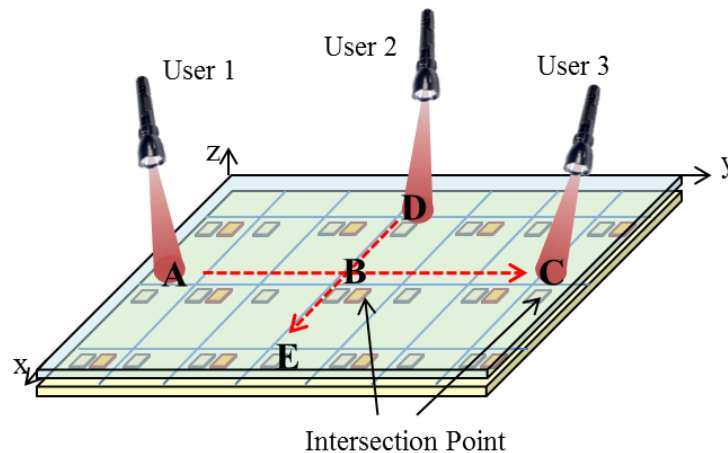


Fig. 68. Model of three users cross over: User 1 and 2 meet at position B and User 1 and 3 meet at position C where the overlapping might occur.

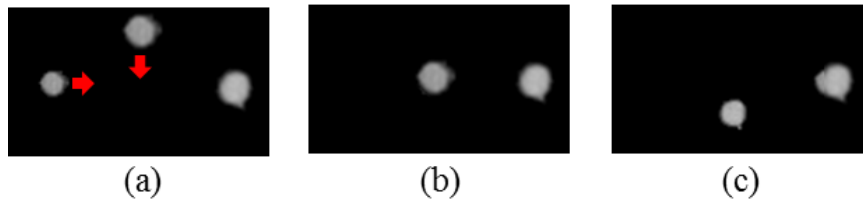


Fig. 69. (a) Without applying sequential-lighting method, three light marks are captured simultaneously. At (b) point B and (c) point C, the overlapping occurs. Different users cannot be identified.

Once Sequential-lighting method was applied, the light marks were captured by the optical sensor frame by frame, as shown in Fig. 70. It is obvious that there was only one mark being detected on each sensor frame. Therefore, mark overlapping no longer existed; that is, by proper synchronization, we can identify different users corresponding to accordant sensor frame. Moreover, the 3D coordinate (x , y , and z) of users were obtained with successful user recognition through Sequential-lighting algorithm.

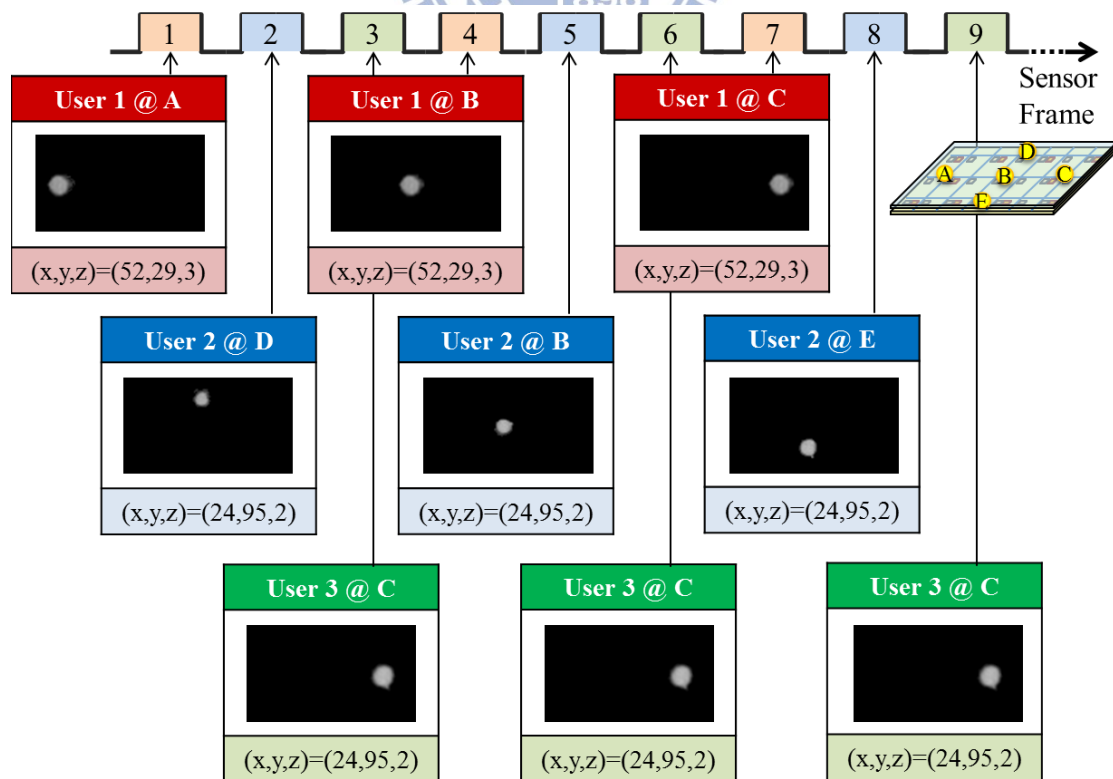


Fig. 70. Images captured by the optical sensors with 30 Hz sampling rate under

sequentially illuminating method. Each light-pen was set to be 10 Hz and synchronized with the optical sensors.

In our prototype, 30Hz sampling rate of the embedded sensor was fabricated. However, this might be insufficient for smooth trajectory tracking when the sequential illuminating method is applied and the sensor frame is divided by the number of users. A hardware improvement for a higher sampling rate should be processed. By enhancing the sampling rate, 60 Hz tracking can be generated for a single user, thus more robust and accurate trajectory tracking can be achieved.



4.3 Multi-ring Mark

The multi-ring mark system is utilized to achieve 3D virtual multi-touch interaction with user recognition enhancement under marks overlapping. Three ring marks were fabricated as shown in Fig. 71, where out-mark is employed for 3D coordinate (x, y, and z) acquisition and in-mark is used for user recognition.

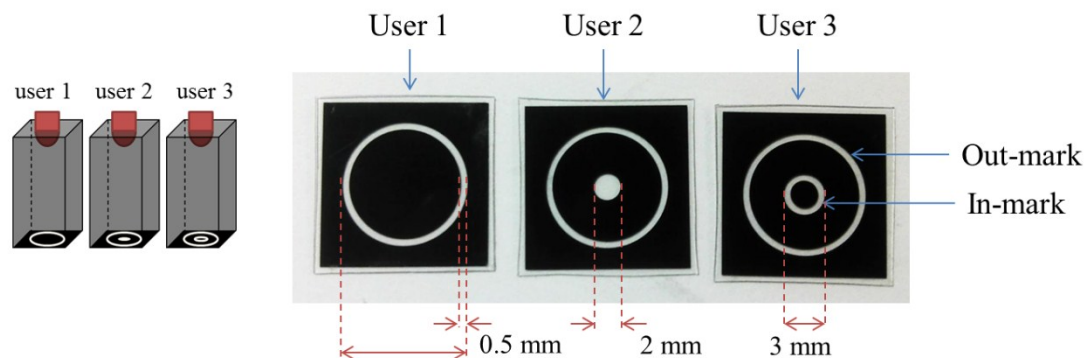


Fig. 71. Multi-ring mark includes two parts: one is out-mark for tracking 3D coordinate (x, y, and z), the other is in-mark for user identification.

To verify the proposed algorithm, three light pens, which mean three fingers or users, were placed randomly on the 4 inch display with embedded optical sensors. The working range of our prototype is from the surface to 50 mm above the display. First, we demonstrated the cases without light marks overlapping. As we randomly put the light pens on/above the panel, incident light was captured and then processed by Multi-ring mark algorithm.





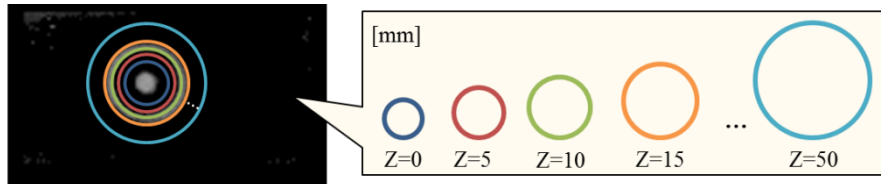
Capture Image	Ground Truth		Detected Result	Output Image
User 1 (○), 2 (⊙), 3 (⊗)	User	(x, y, z) [pixel, pixel, mm]	(x, y, z) [pixel, pixel, mm]	Detected user 1 (*), 2 (o), 3 (+)
	1	(21, 69, 30)	(21, 69, 30)	
	2	(38, 34, 10)	(39, 35, 10)	
	3	(39, 92, 0)	(39, 93, 0)	
	1	(27, 71, 15)	(27, 71, 15)	
	2	(43, 27, 25)	(44, 28, 25)	
	3	(31, 97, 15)	(31, 97, 15)	

Fig. 72. Experiment results of 3-axis detection and user recognition in Multi-ring mark system.

The results, as shown in Fig. 72, reveal well performance in 3D coordinate tracking and accurate user identification. In the 2D coordinate (x, and y) acquisition, the maximum error was only a few pixels difference from the actual position which was barely noticed by human perception. The height detection in z-axis was in accordance with the ground truth as well. Finally, the user recognition was achieved, which exhibited the viable method in user identification by accumulating the edges of the corresponding in-marks.

4.3.1 Resolution in Z-axis

In Multi-ring mark algorithm, the depth value is determined by the diameter of out-mark. Hence, a look-up table for out-mark was establish, as shown in Fig. 73, with the 5 mm interval. To determine the accuracy of height measurement, a light pen was positioned in the center and the four corners of the panel and moved vertically along the z-axis form 0 to 50mm in 5 mm steps, as shown in Fig. 74. The result, as shown in Fig. 75, reveals a linear response within the working region. The maximum error in depth detection is 5 mm.



z [mm]	0	5	10	15	20	25	30	35	40	45	50
Out-mark filter											
Diameter [pixel]	13	15	18	20	23	25	27	29	31	33	35

Fig. 73. Look-up table (LUT) for depth determination in multi-ring mark system

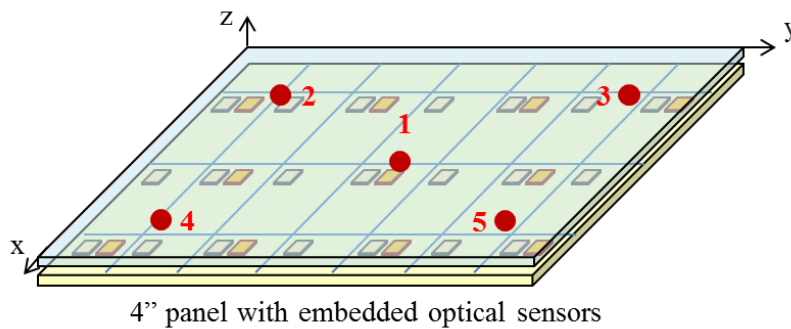


Fig. 74. 5 positions for height measurement.

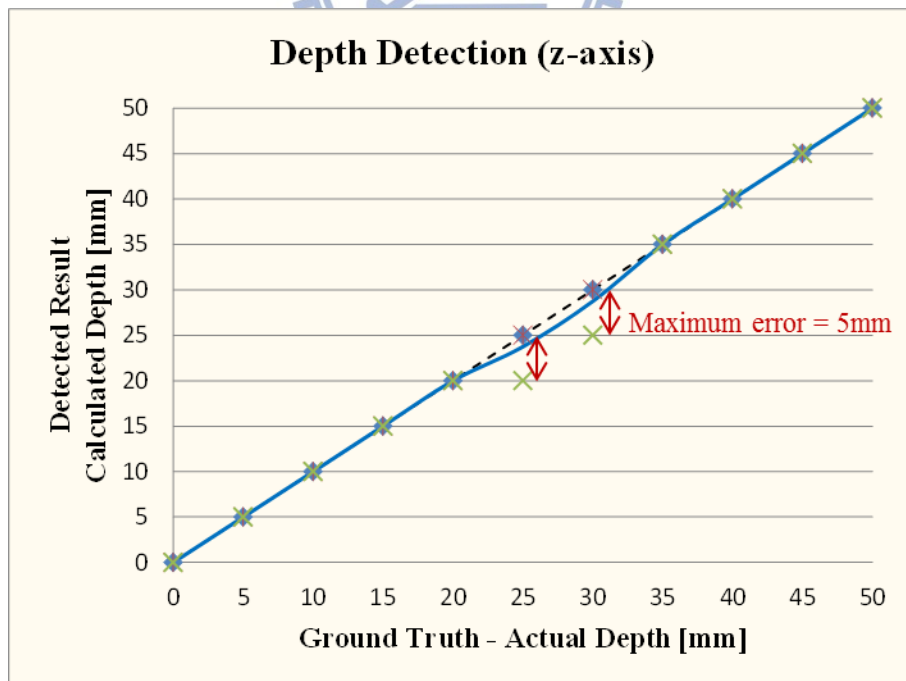


Fig. 75. Experiment results in depth detection with 5 mm step.

4.3.2 Overlapping with Normal Incident

To determine the ability of user recognition, we define the percentage of overlap as Eq. (9), where the overlap section of the line of centers is depicted in Fig. 76. For instance, when the projection light marks are just touch on their borders, l equals to 0 which means 0% of overlapping. While one of the out-marks is overlaying on the center of the other in-mark, there is 50% of overlapping. And 100% of overlapping happens when one of the marks is totally cover by another.

$$\% \text{ overlapping} \equiv \frac{\text{the overlap section of the line of centers}}{\min(\text{diameter of out mark})} = \frac{l}{D_{min}} \quad \text{Eq. (9)}$$

$$\text{Overlap section}(l) = L_c(\text{just touch}) - L_c'(\text{overlap})$$

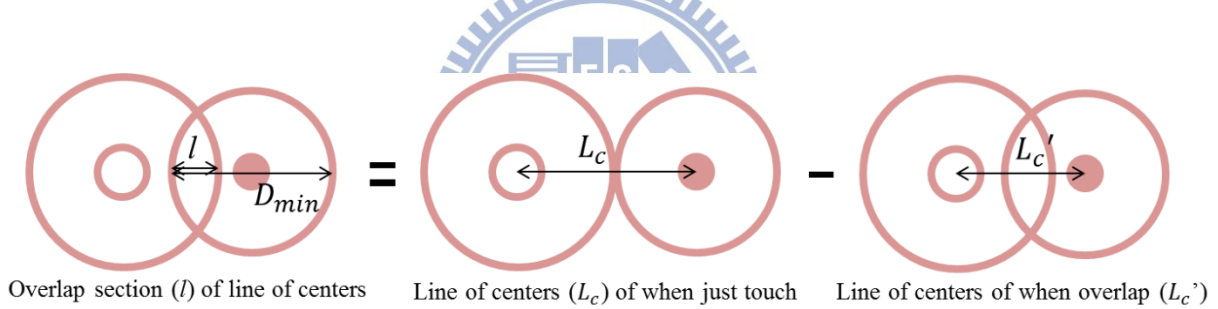


Fig. 76. Percentage of overlapping is defined as l / D , where l represents the length of overlap section and D_{min} is the minimum diameter of the projected light marks.

In normal incident cases, the percentage of overlapping increased when we raised the light pens due to light dispersion, as shown in Fig. 77. The maximum overlapping would happen when light pens were adjacent to each other at 50 mm in our proposed system.

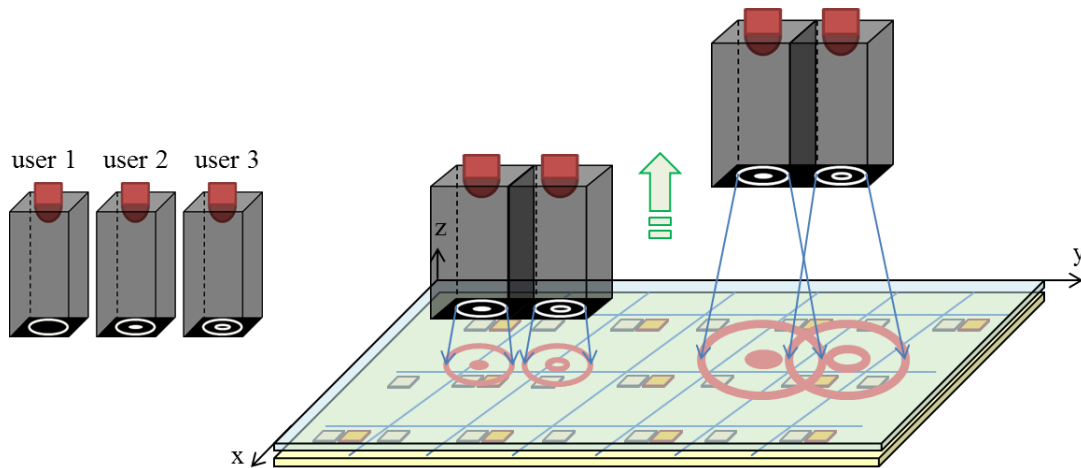


Fig. 77. In normal incident condition, the percentage of overlapping increase as the height of the light pens increase.

To demonstrate the viability of the proposed system during marks overlapping, users were positioned close together and raised in vertical movement from the surface to 50 mm above the panel. The results are shown in Fig. 78 to Fig. 80. For users lower than 20 mm, overlapping conditions would not occur due to the thickness of the light pen. For users greater than 20 mm, the marks started to overlap. It reached to maximum 40% of overlapping when user2 and user3 were placed at 50mm side by side. The results affirm the feasibility of our system in normal incident condition. Precise 2D coordinate (x, y) and correct depth (z) were obtained. Meanwhile, 3D information of each light mark was assign to accordance users accurately.

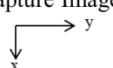





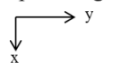














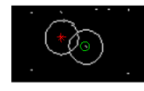
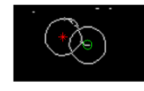

User 1 + 2	z = 0 mm	z = 5 mm	z = 10 mm	z = 15 mm	z = 20 mm
Capture Image 					
Output Result [pixel, pixel, mm]	x y z user 32 48 0 1 35 70 0 2	x y z user 27 54 5 1 32 75 5 2	x y z user 29 51 15 1 33 72 15 2	x y z user 28 50 15 1 34 73 15 2	x y z user 27 46 20 1 34 66 20 2
Output Image 					
Overlap	0	0	0	0	7.8%
z = 25 mm	z = 30 mm	z = 35 mm	z = 40 mm	z = 45 mm	z = 50 mm
					
x y z user 27 45 25 1 34 66 25 2	x y z user 29 45 30 1 36 67 30 2	x y z user 37 49 35 1 32 71 35 2	x y z user 29 45 40 1 37 66 40 2	x y z user 29 45 45 1 36 67 45 2	x y z user 29 48 50 1 34 71 50 2
					
11.5%	14.5%	22.2%	27.5%	30.0%	32.8%

Fig. 78. Overlap of user1 and user2 in vertical movement.

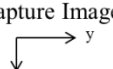





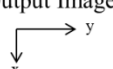

















User 1 + 3	z = 0 mm	z = 5 mm	z = 10 mm	z = 15 mm	z = 20 mm
Capture Image 					
Output Result [pixel, pixel, mm]	x y z user 37 40 0 1 41 60 0 2	x y z user 30 43 5 1 32 64 5 2	X y z user 36 45 10 1 37 66 10 2	x y z user 34 43 15 1 37 64 15 2	x y z user 29 46 20 1 30 66 20 2
Output Image 					
Overlap	0	0	0	0	12.9%
z = 25 mm	z = 30 mm	z = 35 mm	z = 40 mm	z = 45 mm	z = 50 mm
					
x y z user 31 43 25 1 33 64 25 2	x y z user 35 44 30 1 35 65 30 2	x y z user 30 39 40 1 32 61 40 2	x y z user 31 38 45 1 34 60 45 2	x y z user 31 38 45 1 34 60 45 2	x y z user 33 43 50 1 36 65 50 2
					
15.6%	22.2%	27.3%	28.7%	32.7%	36.6%

Fig. 79. Overlapping of user1 and user3 in vertical movement.

























User 2 + 3	z = 0 mm	z = 5 mm	z = 10 mm	z = 15 mm	z = 20 mm
Capture Image 					
Output Result [pixel, pixel, mm]	x y z user 33 37 58 1 33 58 0 2	x y z user 33 41 5 1 34 61 5 2	x y z user 37 42 10 1 38 62 10 2	X y z user 31 41 15 1 30 61 15 2	x y z user 32 41 20 1 30 61 20 2
Output Image 					
Overlap	0	0	0	0	12.6 %
z = 25 mm	z = 30 mm	z = 35 mm	z = 40 mm	z = 45 mm	z = 50 mm
					
x y z user 29 40 25 1 30 60 25 2	X y z user 33 39 30 1 32 60 30 2	x y z user 34 40 35 1 35 60 35 2	x y z user 35 42 40 1 36 62 40 2	x y z user 33 41 45 1 33 64 45 2	x y z user 33 39 50 1 35 60 50 2
					
19.9%	22.1%	30.9%	35.4%	36.4%	39.7%

Fig. 80. Overlapping of user2 and user3 in vertical movement.

4.3.3 Overlapping with Tilt Light-pen

In order to model all the cases from just touch to totally overlap, we tilted the light pens in order to achieve desired conditions, as shown in Fig. 81. The experiments for two and three marks overlapping were performed in turn.

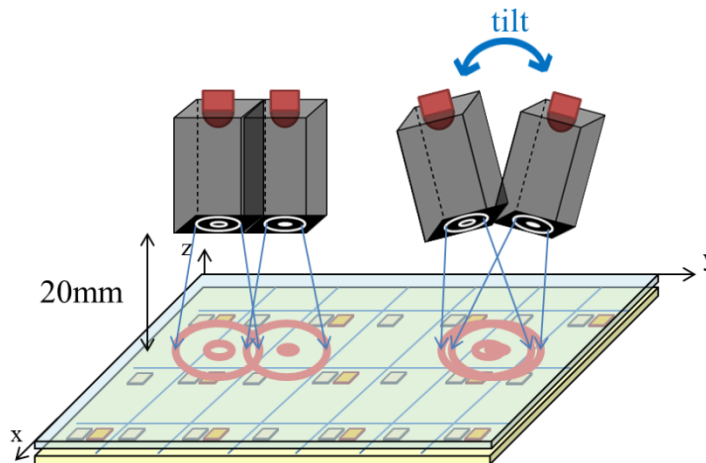


Fig. 81. Serious overlapping conditions can be attained by tilting the light pens.

First, two light pens, user1 and user2, were placed at 20 mm above the display. As the light pens were gradually tilted, light mark projections started to overlap from 0 to 100%. However, by the fact that the tilt of a light pen would cause light distortion in mark projection, the light pens were adjacent to each other to minimize the distortion. The results of 3D coordinate detection (x, y, and z) and user identification are shown in Fig. 82.

The result reveals that the proposed system enhanced the ability of user recognition. Accurate user identification with precision 2D coordinate and correct depth detection were obtained under conditions up to 40% of overlapping. However, for serious overlapping, such as in 50%, at least one of the in-marks was overlaid by the other out-mark. Due to the blocking of the characteristic mark for user recognition, the error user identification might happen. Furthermore, the depth detection was 5 mm higher than the ground truth for acute overlapping which was owing to the severe tilt of light pen which caused the distortion in the out-mark projection. For others, complex overlapping conditions resulted in redundant and mistaken points. Hence, incorrect number of users was found.

























Overlap %	0	10	20	30	40
Capture Image 					
Output Result [pixel, pixel, mm]	x y z user 38 68 20 1 36 48 20 2	x y z user 35 65 20 1 36 46 20 2	x y z user 31 64 20 1 30 46 20 2	x y z user 31 65 20 1 28 50 20 2	x y z user 32 67 20 1 29 53 20 2
Output Image 					
50	60	70	80	90	100
					
x y z user 30 59 25 1 35 69 25 2	x y z user 38 64 0 1 30 71 0 2 36 71 25 3 31 63 25 4	x y z user 35 68 25 1 30 62 25 2	x y z user 36 71 5 1 38 62 15 2 29 66 15 3 40 66 20 4 33 65 25 5	x y z user 34 61 25 1	x y z user 33 64 25 1
					

Fig. 82. Result of two marks overlapping experiment from 0 to 100%

Next, by knowing the ambit in two marks overlapping, from 0 to 40% of three users' overlap was testified. We positioned the light pens approximately at 30 mm above the display and tilted the pens to simulate 0 to 40% of overlapping.

As the result shown in Fig. 83, Multi-ring mark algorithm has succeeded in 3D coordinate (x, y, and z) tracking and user recognition for three points/users with 40% of overlapping. The experiment demonstrated precise calculation result in 2D coordinate (x, y) and accurate user identification. However, 5 mm depth detection error in z-axis was found since 10% overlap. The error might happen by occasion of the thinning process in the proposed algorithm. Though the human eye can easily tell the difference between the edges in the overlap area, it might be regard as the same line because of light dispersion in the image processing. Hence the boundary of out-mark might be suppress and result in error in depth.








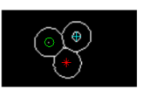

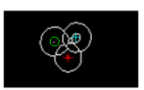


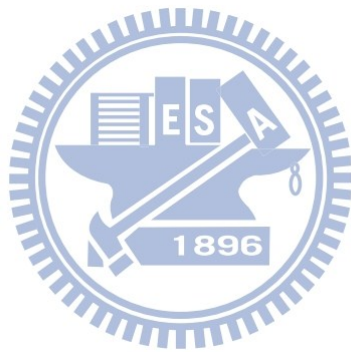
Overlap %	0	10	20	30	40																																																																																
Capture Image 																																																																																					
Output Result [pixel, pixel, mm]	<table border="1"> <thead> <tr> <th>x</th> <th>y</th> <th>z</th> <th>user</th> </tr> </thead> <tbody> <tr> <td>47</td> <td>58</td> <td>30</td> <td>1</td> </tr> <tr> <td>29</td> <td>43</td> <td>30</td> <td>2</td> </tr> <tr> <td>24</td> <td>67</td> <td>30</td> <td>3</td> </tr> </tbody> </table>	x	y	z	user	47	58	30	1	29	43	30	2	24	67	30	3	<table border="1"> <thead> <tr> <th>x</th> <th>y</th> <th>z</th> <th>user</th> </tr> </thead> <tbody> <tr> <td>44</td> <td>60</td> <td>25</td> <td>1</td> </tr> <tr> <td>29</td> <td>44</td> <td>25</td> <td>2</td> </tr> <tr> <td>24</td> <td>66</td> <td>30</td> <td>3</td> </tr> </tbody> </table>	x	y	z	user	44	60	25	1	29	44	25	2	24	66	30	3	<table border="1"> <thead> <tr> <th>x</th> <th>y</th> <th>z</th> <th>user</th> </tr> </thead> <tbody> <tr> <td>42</td> <td>58</td> <td>25</td> <td>1</td> </tr> <tr> <td>28</td> <td>46</td> <td>30</td> <td>2</td> </tr> <tr> <td>24</td> <td>65</td> <td>30</td> <td>3</td> </tr> </tbody> </table>	x	y	z	user	42	58	25	1	28	46	30	2	24	65	30	3	<table border="1"> <thead> <tr> <th>x</th> <th>y</th> <th>z</th> <th>user</th> </tr> </thead> <tbody> <tr> <td>40</td> <td>51</td> <td>25</td> <td>1</td> </tr> <tr> <td>25</td> <td>44</td> <td>30</td> <td>2</td> </tr> <tr> <td>27</td> <td>63</td> <td>30</td> <td>3</td> </tr> </tbody> </table>	x	y	z	user	40	51	25	1	25	44	30	2	27	63	30	3	<table border="1"> <thead> <tr> <th>x</th> <th>y</th> <th>z</th> <th>user</th> </tr> </thead> <tbody> <tr> <td>36</td> <td>54</td> <td>25</td> <td>1</td> </tr> <tr> <td>27</td> <td>67</td> <td>30</td> <td>2</td> </tr> <tr> <td>26</td> <td>59</td> <td>30</td> <td>3</td> </tr> </tbody> </table>	x	y	z	user	36	54	25	1	27	67	30	2	26	59	30	3
x	y	z	user																																																																																		
47	58	30	1																																																																																		
29	43	30	2																																																																																		
24	67	30	3																																																																																		
x	y	z	user																																																																																		
44	60	25	1																																																																																		
29	44	25	2																																																																																		
24	66	30	3																																																																																		
x	y	z	user																																																																																		
42	58	25	1																																																																																		
28	46	30	2																																																																																		
24	65	30	3																																																																																		
x	y	z	user																																																																																		
40	51	25	1																																																																																		
25	44	30	2																																																																																		
27	63	30	3																																																																																		
x	y	z	user																																																																																		
36	54	25	1																																																																																		
27	67	30	2																																																																																		
26	59	30	3																																																																																		
Output Image 																																																																																					

Fig. 83. Result of three marks' overlapping experiment.



4.4 Multi-T Mark

To prove the feasibility of Multi-T-mark algorithm, IR light pens with multi-T-marks, as shown in Fig. 84, were placed at different positions with different orientations. Due to the panel size limitation in our prototype, the maximum of three light pens, which represents three users, can interact with our prototype simultaneously. The definition of 5 axes (x , y , z , θ , and ϕ) is depicted in Fig. 85. We use θ to represent the tilt angle about the short branch of T-mark, and ϕ as the symbol for rotation angle about x -axis.

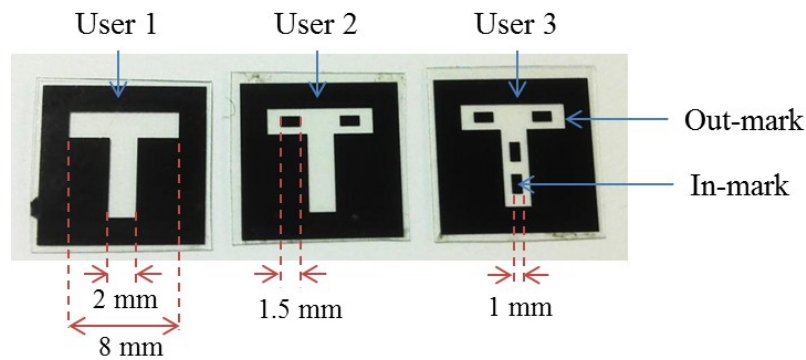


Fig. 84. Multi-T Marks are utilized to determine 5-axis information by out-mark T and meanwhile achieve user identification by in-mark, the inner blocking strips.

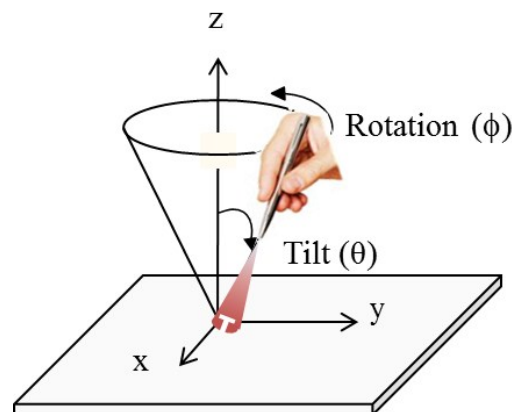


Fig. 85. Schematic of orientation coordinates where rotation angle (θ) and tilt angle (ϕ) are the angles rotate about the short bar of T and z -axis respectively.

4.4.1 5-axis Information and User Identification

The results in Fig. 86 demonstrate the high accuracy of 5-axis detection. Only some of the 2D (x, y) coordinate detections have the displacement of a few pixels, which is less than 1 mm thus can be ignored. The output image also evinces that robust user identification can be achieved by the in-mark patterns.

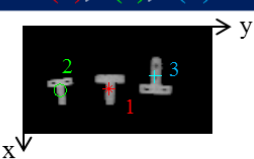
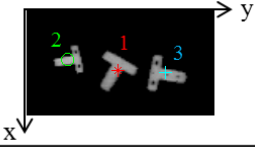
Ground Truth		Detected Results		Output Images	
User	(x, y, z, θ , ϕ) [pixel, pixel, cm, °, °]	(x, y, z, θ , ϕ) [pixel, pixel, cm, °, °]	Detected user 1(*), 2(o), 3(+)		
1	(39, 54, 2, 0, 0)	(40, 55, 2, 0, 0)			
2	(40, 23, 1,350, 20)	(41, 24, 1,350, 20)			
3	(31, 83, 2,180, 30)	(32, 84, 2,180, 30)			
1	(40, 55, 2, 30, 10)	(40, 59, 2, 30, 20)			
2	(32, 26, 1, 80, 10)	(33, 27, 1, 90, 10)			
3	(41, 89, 3,280, 10)	(41, 89, 3,270, 10)			

Fig. 86. Experiment results of 5-axis detection and 3-user recognition.

4.4.2 Resolution in Depth (z)

In the proposed system, the z-axis is determined by the length of a short branch. Since human depth perception (z) is not as sensitive as that of 2D coordinates (x,y), which suppresses the effect on hand vibration, we set the depth resolution to units of 1 cm. The look-up table for z-axis detection is shown in Fig. 87.

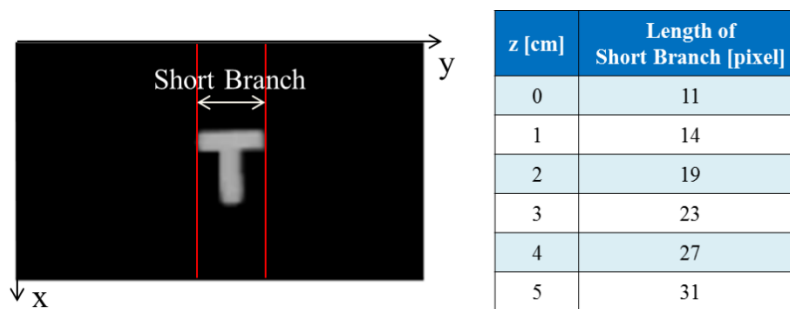


Fig. 87. Look-up table for depth value detection by using the length of the short branch.

To further check the resolution in the z axis, three objects with various T-marks were placed at different heights from 0 to 50 mm. To demonstrate the continuous transfer in real conditions, a 5 mm interval was set in the experiments. As show in Fig. 88, the average depth detection of the three T-marks is almost linear, and is close to the ground truth (actual depth) value with maximum ± 5 mm error. This trend conforms to human perception.

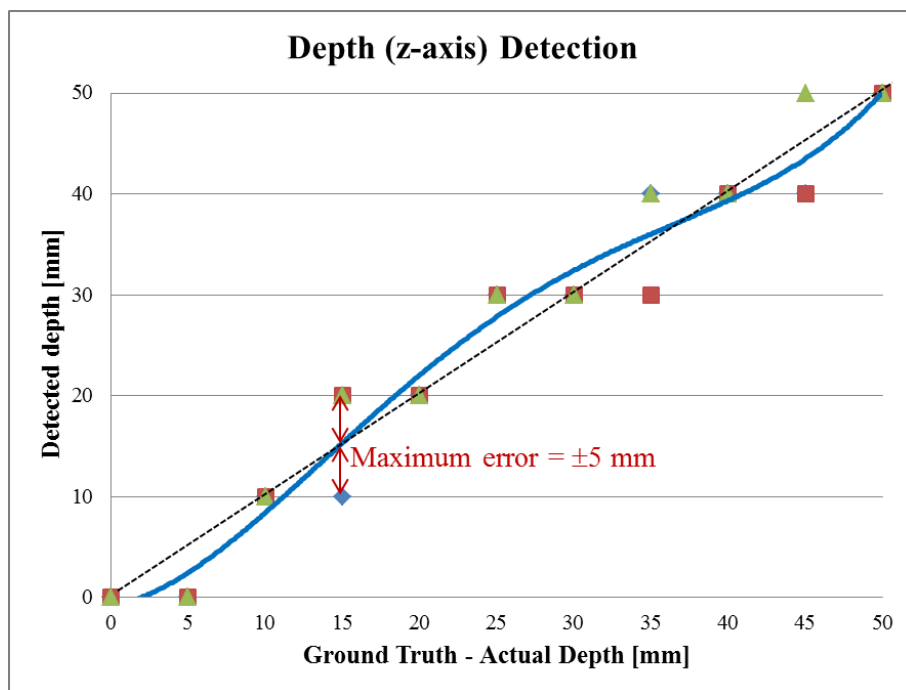


Fig. 88. Result of depth measurements.

4.4.3 Resolution in Orientation (θ , ϕ)

In tilt angle (θ) detection, we set the resolution to 10 degrees, and the detection range from -20° to 20° . The tilt angle range is limited by the panel size of our prototype and the serious mark distortion over 30° . By subtracting each branch of the T-mark, the rotation angle can be calculated by the branch ratio. The detection, as shown in Fig. 89, reveals a satisfaction result with a maximum error of 10° . An

obvious error occurs at $\pm 10^\circ$ which might due to indistinctive image distortion for small tilt of a light pen.

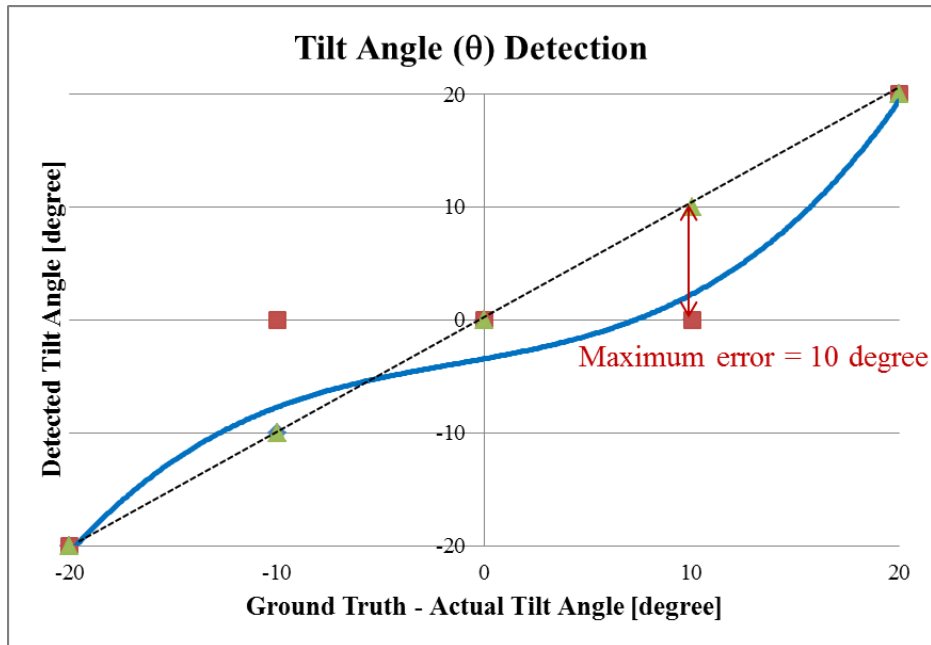


Fig. 89. Result of tilt angle measurements.

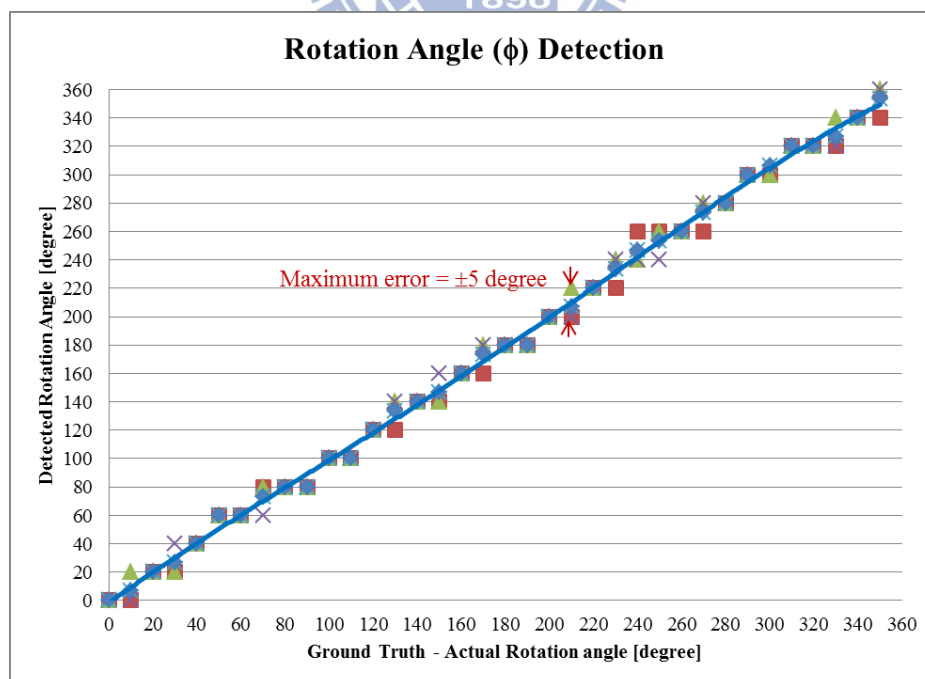
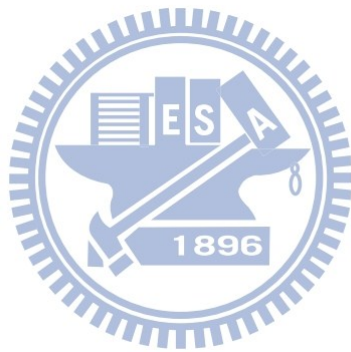


Fig. 90. Result of rotation angle measurements.

Meanwhile, the resolution of the tilt angle (ϕ) was verified. In the experiments, we calibrated the rotation angle to zero degrees as the vertical branch of T is parallel to x-axis. A full circle rotation was made with 5° intervals to verify the fidelity of continuous rotation about z-axis. The result, as shown in Fig. 90, reveals a highly match with the ground truth.



Chapter 5 *Conclusion and Future Work*

Conclusion and Future Work

5.1 Conclusion

Due to the increasing popularity in 3D displays and a widespread user interface of multi-touch technology, more and more researchers have devoted into the 3D multi-user/ multi-touch interaction development. To focus on near-field 3D interaction, the embedded optical sensor based structure is chosen in our proposed system. By virtue of its thin-form factor, camera free, and apposite working range, the optical sensor based structure is suitable for mobile display integration and creates a continuous working space from the surface to near distance.

We have successfully demonstrated that an embedded optical sensor panel in collaboration with the lighting systems can be used as a 3D input device. A brief summary and comparison of the proposed systems are shown in Fig. 91.

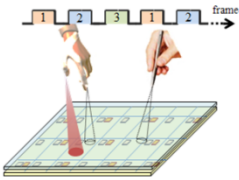
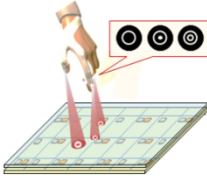
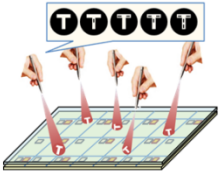
System	Sequential-lighting	Multi-ring Mark	Multi-T Mark
Diagram			
# of users	> 3	> 3	> 3
User identification	✓	✓	✓
Detected axes	(x, y, z)	(x, y, z)	(x, y, z, θ , ϕ)
Resolution/range in z	5 mm / 0 ~ 50 mm	5 mm / 0 ~ 50 mm	10 mm / 0 ~ 50 mm
Resolution/range in θ	×	×	10 degree / 0°~360°
Resolution/range in ϕ	×	×	10 degree / -20°~ 20°
Advantages	Integrable with thin-form panel Continuous Working Space Image Quality Maintenance		
	100% overlap	40% overlap	Orientation Information
Limitation	Optical Sensing Rate	More serious overlap	Overlap

Fig. 91. Summary of the proposed systems.

The experiments were executed on a 4 inch panel with in-cell optical sensors integrating on the TFT substrate. The sequential-lighting system is one of the most solid user interfaces for multi-user/ multi-touch interaction because it intrinsically eliminates the overlapping issue. However, a hardware improvement for a faster response in optical sensor needs to be achieved. Hence, two other systems, the multi-ring mark system and multi-T mark system, are proposed according to the circumstances in multi-finger and multi-user interaction.

In the multi-ring mark system, precise 3D coordinate information (x , y , and z) can be determined with resolution in pixel in x - and y -axis, 5 mm in z -axis. A continuous working space is constructed from a display surface to 50 mm above which is competent for 3D auto-stereoscopic display interaction or other 3D near distance interaction. The number of users can be further extended by the diversely designed of in-mark pattern. Meanwhile, user identification is applicable for not only two but also three user with up to 40% overlapping due to the feature of the proposed ring mark. Therefore, user identification ability has been improved. Thus, with 3D coordinate acquisition and user collaboration, more complex gestures or multi-touch can be achieved for 3D interaction.

On the other hand, the multi-T mark system is proposed where the out-mark T is utilized to determine the 5-axis information (x , y , z , θ , and ϕ) of users and the in-mark renders features for multi-user identification. Larger than three users collaboration can be easily achieved by modifying the number of the blocking strips. The multi-T mark system builds a high-freedom 3D interaction environment with 5 axes extraction from 0 to 50 mm with 10 mm step in depth sensing (z -axis). Additionally, the system provides orientation information where rotation angle ϕ from 0° to 360° in 10° step, and maximum $\pm 20^\circ$ of tilt angle θ in 10° steps.

5.2 Future Work

Following, we hope to further introduce the bare hand 3D multi-touch system. Considering the relatively low sensitivity of the optical sensor in existence, lighting systems are proposed which are endurable for ambient light effect. However, it is still inconvenient for hand holding equipment in some 3D interaction applications. Therefore, instead of illuminating the incident light from a light pen or a light glove, the system emits IR light from the back light and thus can be reflected by interaction objects such as fingers.

Based on the embedded optical sensor structure, an additional shielding layer is constructed, as shown in Fig. 92. Through properly design the pattern of shielding layer and the relative position to the optical sensors, the directional optical sensors which mean the optical sensors are able to capture light from different incident angles can be achieved. There are four sensors in a sensing group with a xy-sensor and three z-sensors. The xy-sensor is able to determine the 2D coordinate (x, y) of the objects. With no blocking pattern above the xy-sensor, the sensor can capture light from all direction. Hence, 2D coordinate (x, y) can be determined by the xy-sensor with maximum intensity. On the other hand, z-sensors can only detect light from specific directions. By knowing the relation between the incident angle and the depth value (z) , the depth information can therefore be acquired.

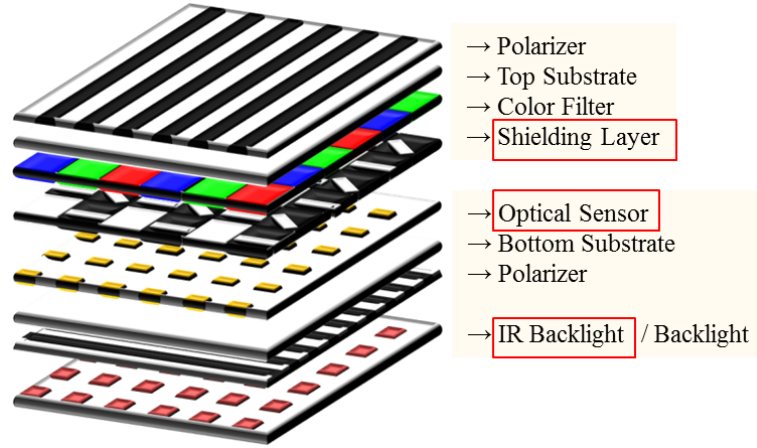


Fig. 92. Structure for the bare-hand 3D multi-touch system.

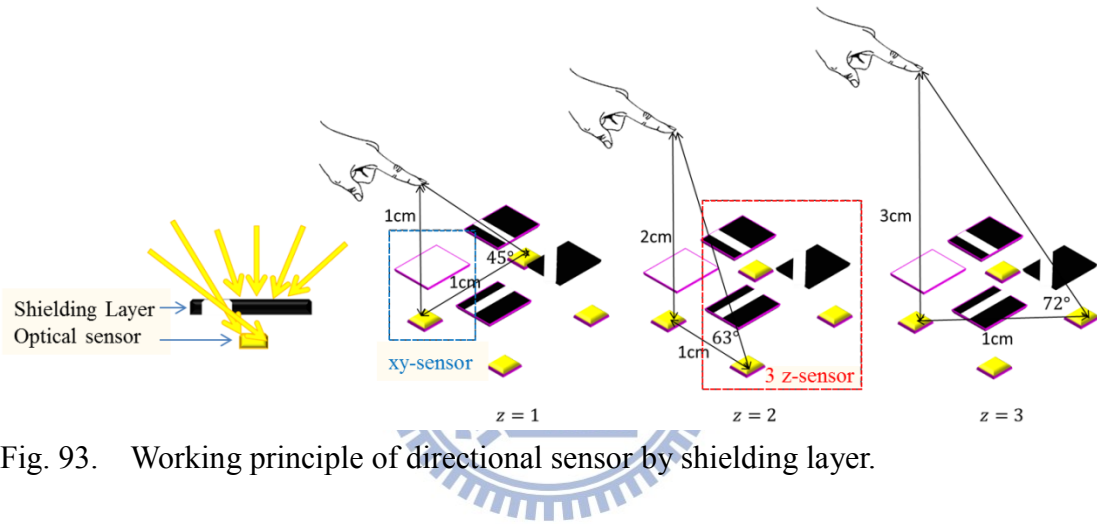


Fig. 93. Working principle of directional sensor by shielding layer.

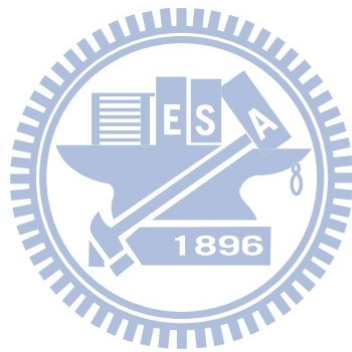
The working principle of shielding layer is illustrated in Fig. 93. By calculating the 2D coordinate from xy-sensor, z-sensors in the same sensing-group are activated to obtain the depth value. For instance, if an object is positioned around 1 cm above the panel, the maximum reflectance of light should be 45 degree, which is according to the structure design. Hence, the depth value can be calculated by the weighting of z-sensors of the sensing-group by Eq. (10).

$$Z = \frac{I_1 \times 1 + I_2 \times 2 + I_3 \times 3}{I_1 + I_2 + I_3}$$

where $I_n = \text{Intensity at } z - \text{sensor for } n \text{ cm}$

Eq. (10)

Additionally, multiple objects or fingers can be detected simultaneously. Due to the separation of four sensors in a group, reflected light from different objects can be captured by the accordant sensing-group. Hence a bare-hand 3D multi-touch interaction can be achieved.



Reference

- [1] C.L. Sholes, C. Glidden, and S.W. Soule, "Improvement in Type-writing Machines," US 79868, 1868.
- [2] N.R. Ball, J.N. Vardalas, *Ferranti-Packard: Pioneers in Canadian Electrical Manufacturing*. McGill Queens Univ Pr, 1994.
- [3] S. Saito, "History of Touch Interface," Creative Environment for Emerging Electronic Culture, 2010.
- [4] G. Largillier, "Developing the first commercial product that uses multi-touch technology," *Information Display*, pp. 14-18, 2007.
- [5] S. Knoedel, M. Hachet, "Multi-touch RST in 2D and 3D spaces: Studying the impact of directness on user performance," In *IEEE Symposium on 3D User Interfaces (3DUI)*, pp.75-78, 2011.
- [6] G. Walker, *Fundamentals of Touch Technologies and Applications*. Society for Information Display, 2012, Short Course S-3.
- [7] I. Maxwell, "An overview of optical touch technologies," *Information Display*, pp.26-30, 2007.
- [8] J. Schöning, P. Brandl, F. Daiber, F. Ehtler, O. Hilliges, J. Hook, "Multi-Touch Surfaces A Technical Guide," *Interactive Surfaces*, 2008.
- [9] J. Y. Han, "Low-cost multi-touch sensing throughfrustrated total internal reflection," In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pp. 115–118, 2005.
- [10] J. Y. Han, "Multi-touch interaction wall," *SIGGRAPH'06: ACM SIGGRAPH 2006 Emerging technologies*, pp. 25, 2006.
- [11] E. Costanza and J. Robinson, "A region adjacency tree approach to the detection and design of fiducials," *Vision, Video and Graphics (VVG)*, pp. 63–70, 2003.
- [12] NUI Group Authors, *Multi-Touch Technologies*, 1st ed. ch1-6, 2009
- [13] G. Walker, "Camera based optical touch technology," *Information Display*, pp. 30-34, 2011.
- [14] A. Abileah, W.d. Boer, T. Larsson, T. Baker, S. Robinson, R. Siegel, N. Fickenscher, "Integrated Optical Touch Panel in a 14.1" AMLCD", Planar Systems Inc.
- [15] S. Hodges, S. Izadi, A. Butler, A. Rrustemi, and B. Buxton, "ThinSight: versatile multi-touch sensing for thin form-factor displays," *Proceedings of the 20th annual ACM symposium on User interface software and technology (UIST' 07)*, pp. 259-268, 2007.
- [16] N. Marquardt, R. Jota, S. Greenberg, and J.A. Jorge, "The continuous interaction

space: interaction techniques unifying touch and gesture on and above a digital surface,” In *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction*, pp. 461-476, 2011.

[17] N. Holliman, *3D Display Systems*. Department of Computer Science, University of Durham, 2005.

[18] D. Valkov, F. Steinicke, G. Bruder, and K. Hinrichs, “2D touching of 3D stereoscopic objects,” In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, pp. 1353-1362, 2011.

[19] R. Ott, M. Gutiérrez, D. Thalmann, and F. Vexo, “Advanced virtual reality technologies for surveillance and security Applications,” In *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications (VRCIA '06)*, pp. 163-170, 2006.

[20] K.N. Tarchanidis, J.N. Lygouras, "Data glove with a force sensor," In *IEEE Transactions on Instrumentation and Measurement*, vol.52, no.3, pp. 984- 989, 2003.

[21] R. Ott, M. Gutiérrez, D. Thalmann, and F. Vexo, “Vr haptic interfaces for teleoperation: an evaluation study,” In *Proceedings of Intelligent Vehicles Symposium, 2005. IEEE*, pp. 789- 794, 2005.

[22] T. Schou and H.J. Gardner, “A Wii remote, a game engine, five sensor bars and a virtual reality theatre,” In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces (OZCHI '07)*, ACM, pp. 231-234, 2007.

[23] H.J. Luinge, P.H. Veltink, "Inclination measurement of human movement using a 3-D accelerometer with autocalibration," *Neural Systems and Rehabilitation Engineering, IEEE Transactions*, vol.12, no.1, pp.112-121, 2004.

[24] Microsoft Research, Resources:

<http://www.engadget.com/2011/02/25/microsoft-research-shows-off-next-generation-gesture-interfaces/>

[25] R. Raskar, H. Nii, B. deDecker, “Prakash: lighting aware motion capture using photosensing markers and multiplexed illuminators,” *ACM SIGGRAPH 2007*, ACM, New York, NY, USA, article 36.

[26] W.d. Boer, A. Abileah, P. Green, T. Larsson, S. Robinson, and T. Nguyen, “56.3: Active matrix LCD with integrated optical touch screen,” *SID Symposium Digest of Technical Papers*, vol. 34, pp. 1494- 1497, 2003.

[27] K. Yi, C. Choi, S. Suh, B. Yoo, J.J. Han, D. Park, and C. Kim, "45.2: Distinguished paper: Novel LCD display with a sensible backlight," *SID Symposium Digest of Technical Papers*, vol. 41, pp. 673-676, 2010.

[28] C. Brown, D. Montgomery, J.L. Castagner, H. Kato, and Y. Kanbayashi, "31.3: A system LCD with integrated 3-dimensional input device," *SID Symposium Digest of*

Technical Papers, vol. 41, pp. 453-456, 2010.

[29] H.U. Tung, "3D multi-interactive system," *Optics and Photonics Taiwan*, 2010.

[30] C.C. Chao, Guo-Zhen Wang, Yi-Pai Huang, "Multi-Mark Based 3D Interactive System for Multi-user," TDC 2012.

[31] Guo-Zhen Wang, Ming-Ching Ma, Shang-Yu Tung, Yi-Pai Huang, Hung-Wei Tseng, Jui-Chi Lo, Chung-Hong Kuo, "A Virtual Touched 3D Interactive Display with Embedded Optical Sensor Array for 5-axis Detection" ,SID, pp.737-740, 2011.

[32] L. Lipton, M. Ackerman, "Liquid crystal shutter system for stereoscopic and other applications," US Patent 4,967,268, 1990.

[33] H. Kim, "Interaction with hand gesture for a back-projection wall," *In Proceedings of Computer Graphics International*, pp.395-402, 2004.

[34] M. Fiala, "ARTag, a fiducial marker system using digital techniques", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005*, vol.2, pp. 590- 596, 2005.

[35] T. Matsumoto, L.O. Chua, T. Yokohama, "Image thinning with a cellular neural network," *Circuits and Systems, IEEE Transactions on* , vol.37, no.5, pp.638-640, 1990.

[36] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, 3rd ed. New Jersey: Prentice Hall, ch9, 2008.

[37] K. Suzuki, I. Horiba, N. Sugie, "Linear-time connected-component labeling based on sequential local operations," *Computer Vision and Image Understanding*, Vol.89, Issue 1, pp.1-23, 2003.

[38] L. Shapiro, G. Stockman. *Computer Vision*. Prentice Hall, 2002, pp. 69–73.

[39] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, 3rd ed. New Jersey: Prentice Hall, ch10, 2008.