



N° D'ORDRE :

UNIVERSITE DE M'SILA
FACULTE DES MATHÉMATIQUES ET DE L'INFORMATIQUE
DEPARTEMENT DE L'INFORMATIQUE

MEMOIRE DE FIN D'ETUDE
PRESENTE POUR L'OBTENTION DU DIPLOME DE MASTER
DOMAINE : MATHÉMATIQUES ET INFORMATIQUE
FILIERE : INFORMATIQUE
SPECIALITE : SYSTEM D'INFORMATION AVANCE
PAR : ATOUI Oussama

SUJET

**DESIGN OF AN INTERACTIVE AUGMENTED
REALITY PRESENTATION SOFTWARE BASED ON
HAND GESTURE RECOGNITION**

**SOUTENU PUBLIQUEMENT LE : 10/06/2014 DEVANT LE JURY
COMPOSE DE :**

MOKHTARI RABAH

PRESIDENT

UNIVERSITE DE M'SILA

MEFTAH LAKEHAL

RAPPORTEUR

UNIVERSITE DE M'SILA

KAMEL MOHAMED

EXAMINATEUR

UNIVERSITE DE M'SILA

PROMOTION : 2013 /2014

شكر وعرفان

أولا أحمد الله على أن وفقني لإستكمال بحثي فهو وحده صاحب الفضل والمنة الحقيقية وكل من سواه ممن كان سببا في إعائتي في بحثي فهو كذلك منة ونعمة من الله توجب له الشكر الدائم، فأحمده سبحانه حمدا يليق به لا أحصي به ثناء عليه بل هو كما أثنى على نفسه وفوق ما يثني عليه عبده

ولكن هذا لا يمنعني من تقديم خالص الإمتنان لكثير من الناس ممن ساهم في إعائتي عن طريق مباشر أو غير مباشر فلا يشكر الله من لا يشكر الناس وأخص بالذكر والديّ العزيزين اللذين لم يدخرا جهدا في إعائتي في كل أعباء حياتي وليس فقط في هذا البحث

كما أشكر كلا من زوجتي وإخوتي وأصحابي وأساتذتي اللذين تسابق فضلهم ليديركني حتى وجدت كثير المشقة في ترتيبهم الأول فالأول ثم لم أستطع فكتبت أولهم قدوما على خلدي ثم الذي يلوه إلى آخرهم

هذا وأصلي على نبينا محمد وعلى آله وأصحابه ونسائه وإخوانه إلى يوم الدين والحمد لله

رب العالمين

TABLE OF CONTENTS

شكر و تعارف	ii
Table of Contents	iii
Introduction	7
1 Overview	7
2 Motivation	7
3 System Description	8
4 Thesis Organization	8
CHAPTER I Interactive Augmented Reality: State of the art	9
1 Introduction	9
2 Augmented reality	9
2.1 How Does Augmented Reality Differ from Virtual Reality?	11
2.2 Milgram Reality-Virtuality Continuum	13
2.3 Application Domains for Augmented Reality	14
2.3.1 Medical	15
2.3.2 Entertainment	16
2.3.3 Military	18
2.3.4 Engineering Design	19
2.3.5 Robotics and Telerobotics	20
2.3.6 Manufacturing, Maintenance and Repair	21
2.3.7 Consumer Applications	23
2.4 Performance Goals	23
3 Human Computer Interaction (HCI)	25
3.1 Human-Computer Interaction: Definition, Terminology	25
3.2 Overview on HCI	26
3.2.1 Existing HCI Technologies	26
3.2.2 Recent Advances in HCI	28
3.2.2.1 Intelligent and Adaptive HCI	28
3.2.2.2 Ubiquitous Computing and Ambient Intelligence	29
3.3 HCI Systems Architecture	30
3.3.1 Unimodal HCI Systems	30
3.3.1.1 Visual-Based HCI	31

3.3.1.2	Audio-Based HCI	31
3.3.1.3	Sensor-Based HCI	32
3.3.2	Multimodal HCI Systems	33
4	User-Augmented Reality interaction.....	34
5	Latest news about augmented reality apps.....	35
5.1	Google Glass.....	35
5.1.2	Some features that will be offered by Google Glass.....	37
5.2	Leap Motion	39
5.2.1	How Leap Motion (Should) Work.....	39
5.2.2	The Potential Applications of Leap Motion.....	39
5.2.2.1	Medicine.....	40
5.2.2.2	Entertainment	40
5.2.2.3	3D Modelling and Graphics	41
6	Conclusion.....	41
CHAPTER 2	computer presentation: STATE OF THE ART	43
1	Introduction.....	43
2	Presentation software: Definition and Terminology.....	45
2.1	Definition.....	45
2.2	Terminology.....	45
2.2.1	Presentation:	45
2.2.2	Slide:	46
2.2.3	Slide Show:	46
3	History	47
4	Presentation program architectures.....	51
5	Most appropriate presentation programs.....	51
5.1	Best presentation software characteristics.....	51
5.1.1	Features.....	51
5.1.2	Compatibility	51
5.1.3	Ease of Use	51
5.1.4	Help and Support.....	52
5.2	Presentation Software similar To PowerPoint	52
5.2.1	Prezi.....	52
5.2.2	GoAnimate	53

5.2.3	Google Docs.....	53
5.2.4	Sliderocket.....	53
5.2.5	Zoho Show.....	54
5.2.6	Augmented Reality Presentation from RTT.....	54
6	Conclusion.....	56
CHAPTER 3 Techniques used to create ASPC.....		57
1	Introduction.....	57
2	Important technical terms : overview.....	57
2.1	Image processing.....	57
2.2	Computer vision.....	57
2.3	Gesture recognition.....	58
2.3.2	Vision based hand gesture taxonomies and representations.....	60
2.3.3	Vision based hand gesture recognition techniques.....	64
2.3.3.2	Detection.....	65
2.3.3.3	Tracking.....	65
2.3.3.4	Recognition.....	65
3	Techniques used to develop ASPC.....	65
3.2	Pre-Processing and Convex-Hull Method.....	66
3.2.1	Preliminary Processing.....	67
3.2.1.1	Skin color detection using HSV color space.....	67
3.2.1.2	Morphology Processing.....	69
3.2.1.3	Contour Finding.....	70
3.2.2	Convex Hull.....	75
3.2.2.1	Convex Polygon.....	75
3.2.2.2	Definition of Convex Hull.....	76
3.2.2.3	Three-Coin Algorithm.....	77
3.2.3	Convexity Defect.....	83
3.3	Hand click detection (HCI part).....	88
3.3.1	Centroid.....	88
3.3.1.2	Calculate the centroid based on moment of inertia.....	89
3.3.2	Hand click detection.....	90
3.4	Virtual interface display (AR part).....	91
3.4.1	Virtual buttons.....	91

3.4.2	Virtual slide show	91
3.4.3	Virtual white board	91
4	Conclusion	92
CHAPTER 4	implementation and test.....	93
1	Introduction.....	93
2	Development environment.....	93
2.1	Operating system	93
2.2	Microsoft Visual Studio 2012	93
2.3	OpenCV.....	93
2.3.1	Installing and Configuring with Visual Studio 2012.....	94
2.3.1.1	Set up Environmental Variable.....	94
2.3.1.2	Configure Visual Studio	97
3	Hardware Architecture.....	100
3.1	Simple web camera	100
3.2	Computer.....	100
3.2.1	Screen Display	100
4	ASPC components	100
4.1	General ASPC aspect	100
4.2	Interaction with slides.....	101
5	Conclusion	103
	Global conclusion	104
	Bibliography	106

INTRODUCTION

1 Overview

Many systems today are too difficult to use because of complex user interfaces. This is partially due to a lack of competence in designing user interfaces, many engineers suffer from. A more important reason is that with the growing computational power of modern systems, devices and applications become more complex and integrate more features. Soft- and hardware that was only available to a small amount of specialists a few decades ago, is now a well-integrated part of everyone's daily life. Good user interface design is therefore no longer an option but a hard requirement for developing highly usable applications.

Augmented Reality (AR) research aims at developing new human computer interfaces. Instead of showing information on isolated displays, it puts data right where it belongs: into the real world. AR thereby blurs the distinction between the real world and the user interface and combines them in a natural way allowing the creation of simple and intuitive user interfaces even for complex applications.

In the future, computers won't be just computers. Many daily devices will be integrated with computers, like cars, television, household appliances, etc. Since so many types of machines will come up, we need various Human Computer Interactions (HCIs) to cope with different type of machines. The hand gesture will be the most popular one. It is the most intuitive way for human to communicate with their bodies.

With the communication by hand, people will no longer need to use computer devices in a certain location. For example, if a TV is integrated with gesture recognition function, people can manipulate it everywhere in their house. A remote control will be needless. The gap between human and machines will be closer. It increases the freedom of usability.

Presentations are one of important domains which need a deep interaction between the presenter and information (slides). In addition, the use of switch-based devices (section 3.2.1) constitutes a major obstacle to the achievement of passing information to people in many situations, such as academic conference or business.

2 Motivation

As previous section described (As described above), we believe (it can be believable that) the use of presentation technology (section 2.1) is very important to reach the audience. Despite the

fact that the technology used currently handles a lot of difficulties, there is a significant problem facing the area of the presentation which is the dispersing the audience between the monitor and the presenter. AR and HCI is a good trend to solve this problem.

Along this thesis, combining the presenter to the slides (section 2.2) is the major work. And showing them on only one screen using a virtual display of slides and buttons (AR part), where the user can interact with them through a real time hand gesture recognition (HCI part).

3 System Description

Our system and its computation are based on the contour of hand and forearm.

We are performing the hand recognition according to the program needs. The original input data will be the RGB image captured by a web camera. We perform some preliminary processing to generate a binary image which provides enough information of hand contour. The binary image will be used to calculate the contour and the convex hull of the contour to extract the hand input which is one type of click used to click the virtual buttons and drawing on the screen.

The computation cost of our system is low since only computation geometry algorithms are performed; and there's no need to apply classifier and no pre-trained data base is needed.

4 Thesis Organization

The thesis is consisted of four chapters with a general introduction and global conclusion. In chapter 1 and 2, state of art for related works will be reviewed. Chapter 3 will discuss the methods to obtain the binary image and calculate the contour, convex hull, hand click and virtual presentation. Chapter 4 focuses on the tools used to develop the project and shows results. In global conclusion, is a discussion about future works to perform and develop this kind of systems.

CHAPTER I

INTERACTIVE AUGMENTED REALITY: STATE OF THE ART

1 Introduction

For years, the computer was something mystical locked up in the basements of large companies and universities. Then came home computers, desktops and video consoles, bringing both entertainment and an increase in productivity to users. In the last twelve years, the Internet has become the largest source of information on the planet, linking together computer users from all over the world. Then came the surge in mobile computing; laptops, personal digital assistants and especially mobile phones have become common- place in todays Information Society. Computers are getting smaller and smaller and users want access to information and communication everywhere they go.

Users would like to have direct access to information and communication, where ever they might be. Instead of having increasingly smaller and increasingly more difficult to use displays, information and interfaces should be superimposed directly onto the vision of users. This is the idea behind Augmented Reality (AR). In this chapter, the current state of AR will be shown, with a number of real - world applications in use today. The focus will be on the interaction between the user and the AR and the types of interfaces users will see in the near future. These new ways of interaction are sometimes called “Post-WIMP”: AR isn't concerned about Windows, Icons, Menus and Pointers, but in new ways for users to interact.

2 Augmented reality

An augmented reality system is a system that creates a view of a real scene by incorporating computer-generated virtual objects, including those with full three-dimensional properties, into the scene. As the user of such a system moves about the real scene, the virtual objects appear as if they actually exist in the scene. Ideally, the virtual objects should interact with the user and real objects in the scene in a natural manner. The application domains described in Section 2.3 reveal that augmented reality can take on a number of different forms. In all the applications that are discussed later, augmented reality enhances the user's performance in and perception of the world.



Figure 1.1 shows a view of what the user might see while using an augmented reality

The ultimate goal is to create a system such that the user cannot tell the difference between the real world and the virtual augmentation of it. To the user of this ultimate system it would appear that he is working in a single real environment.

System in the medical domain. It depicts the merging and correct registration of data from a pre-operative imaging study onto the patient's head. Providing this view in the operating theater would enhance the surgeon's performance and possibly eliminate the need for any other calibration fixtures during the procedure. Consider another application scenario as depicted in Figure 1.2. The city planners want to visualize what the landscape will look like when a new footbridge is built. They would go to the proposed site (Figure 1.2.a) and (donning an augmented reality display device) see the area with the new bridge merged into their view of the landscape (Figure 1.2.b). If adjustments were needed they could be performed directly on the visual model of the new bridge.



(a) (b)
Figure 1.2 Preview of a new footbridge using an augmented reality display (Fraunhofer Institute for Computer Graphics 1997)

2.1 How Does Augmented Reality Differ from Virtual Reality?

Virtual reality is a technology that encompasses a broad spectrum of ideas. It defines an umbrella under which many researchers and companies express their work. The phrase was originated by Jaron Lanier, the founder of VPL Research — one of the original companies selling virtual reality systems. The term was defined as “a computer-generated, interactive, three-dimensional environment in which a person is immersed.” [1]. There are three key points in this definition. First, this virtual environment is a computer-generated three-dimensional scene that requires high performance computer graphics to provide an adequate level of realism. The second point is that the virtual world is interactive. A user requires real-time response from the system to interact with it in an effective manner. The last point is that the user is immersed in this virtual environment. One of the identifying marks of a virtual reality system is the head-mounted display often worn by users. These displays block out all the external world and present to the wearer a view that is under the complete control of the computer. The user is completely immersed in an artificial world and becomes divorced from the real environment. For this immersion to appear realistic the virtual reality system must accurately sense how the user is moving and determine what effect that will have on the scene being rendered in the head mounted display.

In contrast, an augmented reality system generates a composite view for the user. The user sees an image that is the combination of the real scene being viewed and a computer-generated virtual scene. The virtual scene augments the real environment with additional information. There are still the requirements that as the user moves through the workspace the effect on the computer generated world

must be determined. Once that is done the operation of combining the virtual scene with the real scene must be performed.

The discussion above highlights the general similarities and differences between virtual reality and augmented reality systems. A very visible difference between these two types of systems is the immersiveness of the system. Virtual reality strives for a totally immersive environment. The visual, and in some systems aural and proprioceptive, senses are under control of the system. This requires the virtual reality system to model its artificial world completely. Depending on the desired level of verisimilitude, this is a very complex simulation problem. In contrast, an augmented reality system augments the real world scene and attempts to maintain the user's sense of being in the real world. The rationale behind this is twofold. First, real environments contain a wealth of information much of which is impossible to model and simulate by computer. Secondly, if the end goal is to enhance the performance of a real-world task, it will be performed most naturally if the user still feels immersed in the task environment. To maintain the user's immersion in the real world an augmented reality system merges the virtual images with a view of the real scene to create the augmented display. This merging requires a mechanism to combine the real and virtual that is not present in other virtual reality work. Developing the technology for merging the real and virtual image streams either optically or in video is an active research topic [2] that is not addressed in this thesis.

Requiring a means to visually merge the views of the real and virtual worlds is not the major hurdle brought on by the inclusion of the real world in an augmented reality system. The nature of the visual interface between computer and user is different. Both virtual reality and augmented reality systems give the user a sense of immersion in the virtual environment by ensuring that the user receives a consistent set of sensory inputs. The primary performance goal for a virtual reality system is to present visual stimuli that are consistent with the changes in body position sensed by the user. This requires that motions or changes made by the user will result in the appropriate changes in the perceived virtual world. Because the user is looking at a virtual world there is no natural connection between the user's internal proprioceptive coordinate system and the virtual world coordinate system. A connection must be artificially created [3]. Any inconsistency the user perceives results from a misregistration between the coordinate system the user maintains internally to describe body position and the coordinate system that describes the graphics system's viewpoint in the virtual scene. Errors are perceived here as conflicts between the visual system and the kinesthetic or proprioceptive systems. The phenomenon of visual capture gives the vision system a stronger influence in our perception [4]. This stronger influence allows a user of a virtual reality system to accept or adjust to the visual stimulus by overriding the discrepancies occurring with input from the other sensory systems.

Contrast this to the primary performance goal for an augmented reality system which is to render views of virtual objects that are consistent with the user's view of the real environment containing the objects. In this case, errors of misregistration are between two visual stimuli which we are trying to fuse to see as one scene. The observer is more sensitive to these errors [3]. Any inconsistency, which manifests itself as a difference between two visual stimuli, i.e. the virtual and real images, derives from a misregistration between the coordinate system describing the user's viewpoint in the real scene and the graphics system's viewpoint in the virtual scene. This imposes a tougher registration requirement on augmented reality systems compared to virtual reality systems.

2.2 Milgram Reality-Virtuality Continuum

Milgram [5] describes a taxonomy that identifies how augmented reality and virtual reality are related. He defines the Reality-Virtuality continuum shown as Figure 1.3. The real

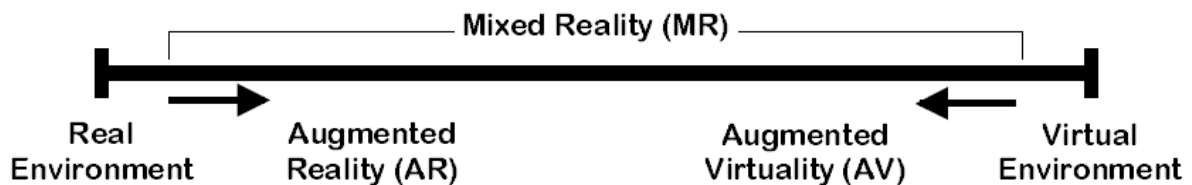


Figure 1.3 Milgram's Reality-Virtuality Continuum

world and a totally virtual environment are at the two ends of this continuum with the middle region called Mixed Reality. Augmented reality lies near the real-world end of the spectrum with the predominate perception being the real-world augmented by computer generated data. Augmented virtuality is a term created by Milgram to identify systems that are mostly synthetic with some real world imagery added — such as texture mapping video onto virtual objects. An example of this technology is live video texture-mapped onto the graphics image of an avatar — a computer-generated virtual character that is a stand-in for the user's presence within a virtual environment [6].

Milgram further defines a taxonomy for the Mixed Reality displays. The three axes he suggests for categorizing these systems are: Reproduction Fidelity, Extent of Presence Metaphor and Extent of World Knowledge. Reproduction Fidelity relates to the quality of the computer generated imagery ranging from simple wireframe approximations to complete photorealistic renderings. The real-time constraint on augmented reality systems forces them to be toward the low end on the Reproduction Fidelity spectrum. The current graphics hardware capabilities can not produce real-time photorealistic renderings of the virtual scene. Milgram also places augmented reality systems

on the low end of the Extent of Presence Metaphor. This axis measures the level of immersion of the user within the displayed scene. This categorization is closely related to the display technology used by the system. There are several classes of displays used in augmented reality. Each of these gives a different sense of immersion in the virtual environment presented to the user. In an augmented reality system, some display technologies utilize the user's direct view of the real world. Immersion in that environment comes from the user simply having his eyes open. It is contrasted to systems where the merged view is presented to the user on a separate monitor for what is sometimes called a "Window on the World" [7] view.

The third, and final, dimension that Milgram uses to categorize Mixed Reality displays is Extent of World Knowledge. Augmented reality does not simply mean the superimposition of a graphic object over a real world scene. This is technically an easy task. To do a realistic merging of virtual objects into a real scene, knowledge about the world is needed. One difficulty in augmenting reality, as defined here, is the need to maintain accurate registration of the virtual objects with the real world image. This requires detailed knowledge of the relationship between the frames of reference for the real world, the camera viewing it and the user. To properly compute the visual interactions between real and virtual objects, data about their locations and orientations in three-dimensional space are needed. Correct dynamic interactions with virtual objects require knowledge of the dynamic characteristics of both the real and virtual objects in the augmented environment. In some domains models for the real and virtual worlds are well known, which makes the task of augmenting reality easier or might lead the system designer to use a completely virtual environment.

2.3 Application Domains for Augmented Reality

Only recently have the capabilities of real-time video image processing, computer graphic systems and new display and haptic technologies converged to make possible the creation of an augmented environment. In this environment, images of three-dimensional virtual objects are correctly registered with the view of the 3D environment surrounding the user and the user can interact naturally with the virtual objects. Researchers working with augmented reality systems have proposed them as solutions in many domains. The literature discusses application areas ranging from entertainment to military training. Many of the domains, such as medical [8], are also considered domains for virtual reality systems. This section highlights some of the proposed applications for augmented reality

2.3.1 Medical

Because imaging technology is so pervasive throughout the medical field, it is not surprising that this domain is viewed as one of the more important for augmented reality systems. Most of the medical applications deal with image-guided surgery. Pre-operative imaging studies, such as CT or MRI scans, of the patient provide the surgeon with the necessary view of the internal anatomy. From these images the surgery is planned. The surgeon visualizes the path through the anatomy to the affected area where, for example, a tumor must be removed, by first creating a 3D model from the multiple views and slices in the preoperative study. Through their extensive training, surgeons become very adept at mentally creating the three-dimensional visualization that is needed to render a diagnosis. Some newer systems do have the ability to create 3D volume visualizations from the imaging study. Figure 1.4 shows how augmented reality can be applied so that the surgical team sees the CT or MRI data correctly registered on the patient in the operating theater while the procedure is progressing. Being able to accurately register the images at this point will enhance the performance of the surgical team and may eliminate the need for the painful and cumbersome stereotactic frames [9] currently used for registration. Descriptions of other work in the area of image-guided surgery using augmented reality can be found in [10].

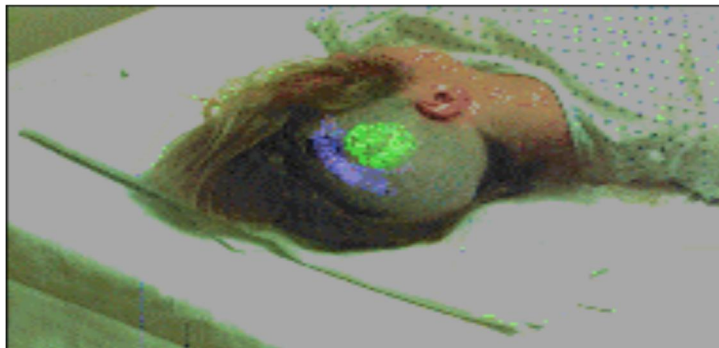


Figure 1.4 Image guided surgical procedure (Ettinger, Grimson et al. 1998)

Another application for augmented reality in the medical domain is in ultrasound imaging (State, Chen et al. 1994). Using an optical see-through display the physician can view a volumetric rendered image of the fetus overlaid on the abdomen of the pregnant woman. The image appears as if it were inside the abdomen and is correctly rendered as the user moves. Figure 1.5.a shows an image from the system along with a second application in ultrasound imaging seen in Figure 1.5.b. In this second application, the augmented reality image helps the surgeon guide a biopsy needle to

the site of a suspected tumor during a mock breast biopsy procedure. The V-shaped object in the left part of the image is used for registering the ultrasound image with the view of the real scene that the surgeon is seeing.

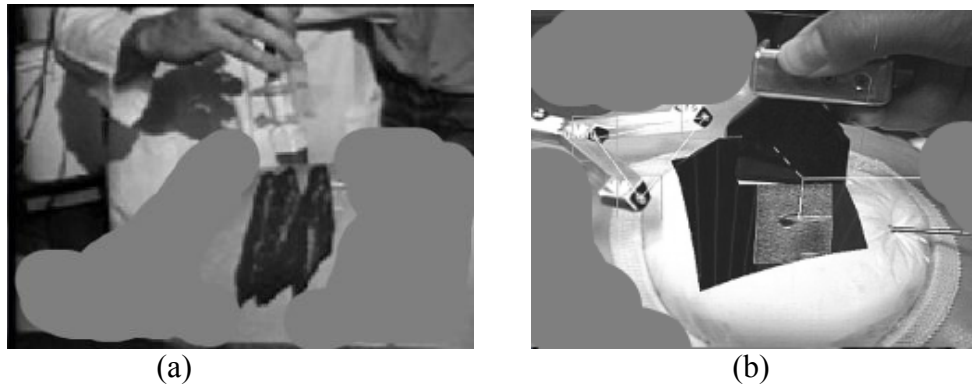


Figure 1.5 Ultrasound imaging using augmented reality displays (a (UNC - Chapel Hill 1995); b (UNC - Chapel Hill 1997))

Finally, a future application is the extension of the virtual reality based craniofacial surgery visualization and simulation systems [11] by adding an augmented reality display. These systems currently allow the surgeon to experiment, within the safety of a virtual environment, with different treatment approaches for the reconstructive work to be done. The model that the surgeon works on is a 3D volume visualization derived from a pre-operative CT or MR study. Augmented reality would allow the surgeon to see the final results directly on the patient rather than only with the volume visualization.

2.3.2 Entertainment

A simple form of augmented reality has been in use in the entertainment and news business for quite some time. When you watch an evening weather report the weather reporter often stands in front of changing weather maps. In the studio the reporter is actually standing in front of a blue or green screen. This real image is augmented with computer generated maps using a technique called chroma-keying. It is also possible to create a virtual studio environment so that the actors appear to be positioned in a studio with computer generated decorating [7] an example of which is shown in Figure 9. The work with virtual studios goes beyond the simple chroma-keying of an actor over a flat two-dimensional background image. It extends the traditional keying to define the background in three-dimensions and key the live action as a three-dimensional image also. Balcisoy and Thalmann (1997) present a virtual studio populated by virtual humans that interact

with human actors. The interactions are very carefully scripted and the human actor does not have the ability to feel a virtual actor. With tracking of the human, appropriate visual interactions can be accomplished. This system uses calibrated cameras and careful measurements of the location of objects in the real studio.

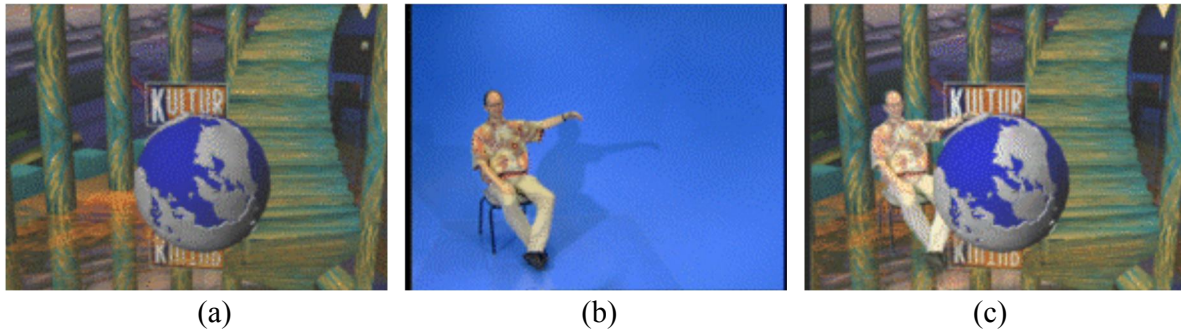


Figure 1.6 Virtual set technology. (a) Virtual set background, (b) live action, (c) combined video (Gibbs 1995).

Movie special effects make use of digital compositing to create illusions. Strictly speaking, with current technology, this is not considered augmented reality because it is not generated in real-time. Most special effects are created off-line, frame by frame with a substantial amount of user interaction and computer graphics system rendering. But some work is progressing in computer analysis of the live action images to determine the camera parameters and use this to drive the generation of the virtual graphics objects to be merged.

Princeton Electronic Billboard has developed an augmented reality system that allows broadcasters to insert advertisements into specific areas of the broadcast (National Association of Broadcasters 1994). For example, while broadcasting a baseball game this system places an advertisement in the image so that it appears on the outfield wall of the stadium. The electronic billboard requires calibration to the stadium by taking images from typical camera angles and zoom settings in order to build a map of the stadium including the locations in the images where advertisements will be inserted. By using pre-specified reference points in the stadium, the system automatically determines the camera angle being used and referring to the pre-defined stadium map inserts the advertisement into the correct place. A French company, Symah Vision, has also developed a similar application. Another application in sports broadcasting is Fox network's FoxTrax system [3] for tracking the path of a hockey puck during a game. The path of the puck is overlaid on the image of the hockey rink as a blue streak. The streak changed color based on the

speed of the puck. This system requires a detailed calibration process for each television camera and the ice ring itself.

Augmented reality can be applied to enhance games that people play. A system developed for pocket billiards players uses a head-mounted display and wearable computer to analyze the layout of the table and suggest possible shots for the player to take. The trajectory of the shot is displayed as augmented graphics over the image of the pool table. Or consider a futuristic game of paintball where players wear augmented reality headsets. The image that the players see is not only of the game area and their real opponents but virtual players are also playing along with them.

2.3.3 Military

The military has been using displays in cockpits that present information to the pilot on the windshield of the cockpit or the visor of their flight helmet. This is a form of augmented reality display. SIMNET, a distributed war games simulation system, is also embracing augmented reality technology. By equipping military personnel with helmet mounted visor displays or a special purpose rangefinder [12] the activities of other units participating in the exercise are seen. While looking at the horizon, for example, the display-equipped soldier sees a helicopter rising above the tree line. Actually, another participant is flying this helicopter in simulation. In wartime, the display of the real battlefield scene could be augmented with annotation information or highlighting to emphasize hidden enemy units.

The University of Rochester is participating in the Video Surveillance and Monitoring (VSAM) project funded by the Defense Advance Research Projects Agency (DARPA). Figure 1.7 shows the scenario for using augmented reality in this project. Aerial reconnaissance units fly overhead and generate reference marks for registration. Autonomous ground units with video surveillance equipment monitor sections of the same area. Aerial views augmented by the information from the surveillance units are generated for strategic command and control. Another scenario equips ground level warfighter or urban crime/terrorist fighters with special displays. Aerial reconnaissance units identify suspect objects and transmit the location of these objects to the ground units. The suspect objects may be hidden from the view of the ground forces but will appear in the augmented view on their displays.

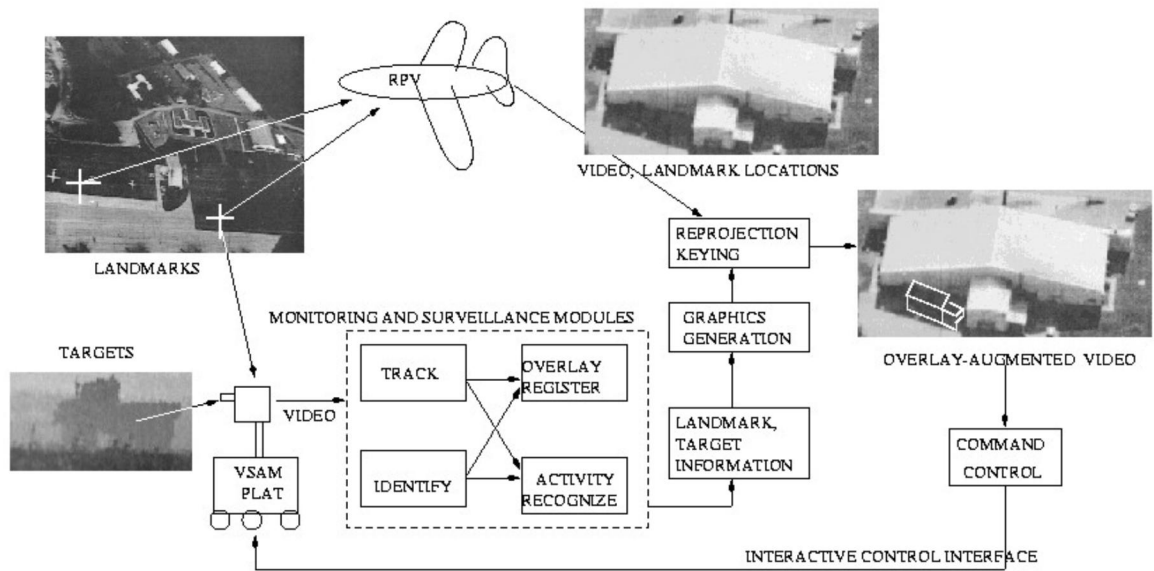


Figure 1.7 Video Surveillance and Monitoring (VSAM) scenario

2.3.4 Engineering Design

Imagine that a group of designers is working on the model of a complex device for their clients. The designers and clients want to do a joint design review even though they are physically separated. If each of them had a conference room that was equipped with an augmented reality display this could be accomplished. The physical prototype that the designers have mocked up is imaged and displayed in the client's conference room in 3D. The clients walk around the display looking at different aspects of it. To hold discussions the client points at the prototype to highlight sections and this is reflected on the real model in the augmented display that the designers are using (Figure 1.8). Or perhaps in an earlier stage of the design, before a prototype is built, the view in each conference room is augmented with a computer generated image of the current design built from the CAD files describing it. This allows real time interaction with elements of the design so that either side can make adjustments and changes that are reflected in the view seen by both groups. A technique for interactively obtaining a model for 3D objects called 3D stenciling that takes advantage of an augmented reality display is being investigated in our department by Kyros Kutulakos.

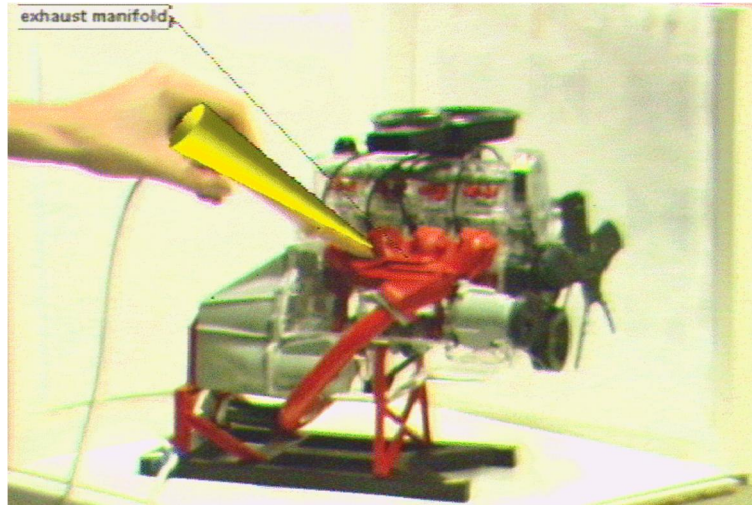


Figure 1.8 Engineering design using an augmented reality display (Breen 1995)

2.3.5 Robotics and Telerobotics

In the domain of robotics and telerobotics an augmented display can assist the user of the system. Croby and Nafis (1994) describe an augmented reality telemanipulation system for nuclear reactor inspection operations. In a telerobotic system, the operator uses a visual image of the remote workspace to guide the robot. Annotation of the view would still be useful just as it is when the scene is in front of the operator. There is an added potential benefit. Since often the view of the remote scene is monoscopic, augmentation with wireframe drawings of structures in the view facilitates visualization of the remote 3D geometry. If the operator is attempting a motion he first practices it on a virtual robot that he sees as an augmentation to the real scene (Figure 1.9). The operator can decide to proceed with the motion after seeing the results. The robot executes the motion pattern directly which in a telerobotics application eliminates the oscillations often present due to long communication delays to the remote site.

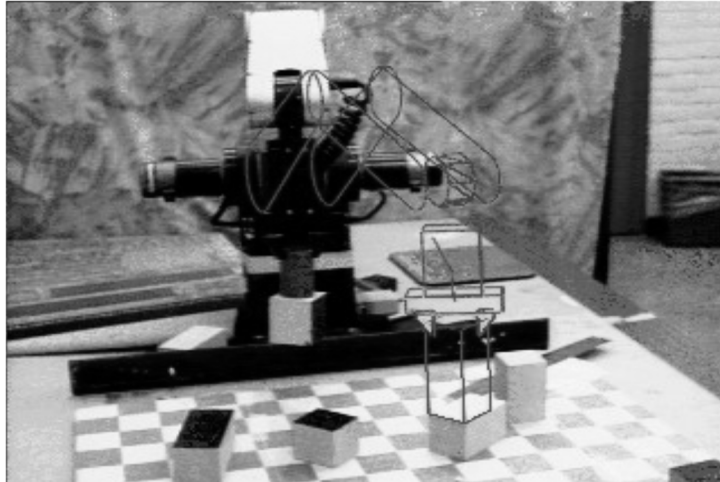


Figure 1.9 Augmented reality in robotics (Rastogi, Milgram et al. 1995)

2.3.6 Manufacturing, Maintenance and Repair

When a maintenance technician approaches a new or unfamiliar piece of equipment instead of opening several repair manuals he simply puts on an augmented reality display. In this display an augmented reality system shows the image of the equipment augmented with annotations and information pertinent to the repair. For example, the location of fasteners and attachment hardware that must be removed are highlighted. As part of the next step an inside view of the machine highlights the boards that need to be replaced (Feiner, MacIntyre et al. 1993b; Uenohara and Kanade 1995). Figure 1.10 shows an example of this. The military has developed a wireless vest worn by personnel that is attached to an optical see-through display (Urban 1995). The wireless connection allows the soldier to access repair manuals and images of the equipment. Future versions might register those images on the live scene and provide animation to show the procedures that must be performed.

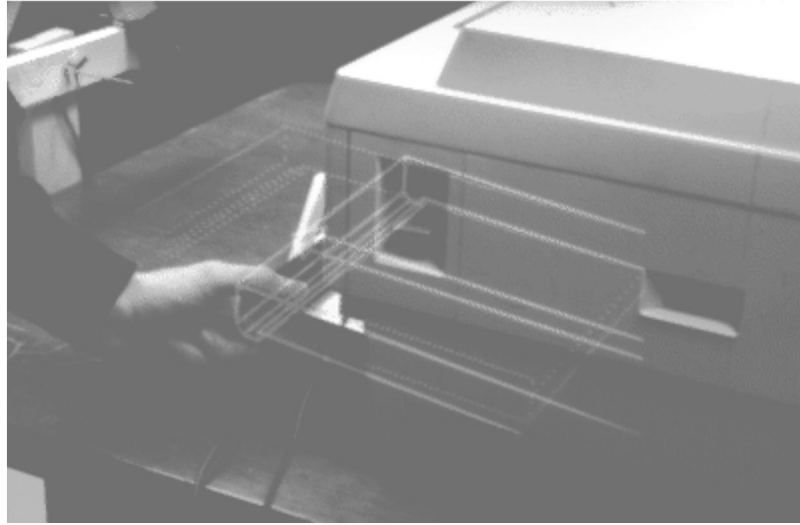


Figure 1.10 Equipment maintenance application for augmented reality (Feiner, MacIntyre et al. 1995)

Aircraft manufacturers are particularly active in incorporating augmented reality systems into their manufacturing and maintenance activities. Boeing researchers developed an augmented reality display to replace the large work frames used for making wiring harnesses for their aircraft. Using this experimental system (Figure 1.11.b), the technicians are guided by the augmented display that shows the routing of the cables on a generic frame used for all harnesses. The augmented display allows a single fixture to be used for making the multiple harnesses. Research undertaken by the University of Southern California in conjunction with McDonnell-Douglas is using an augmented reality system to guide technicians through maintenance and repair sequences Figure 1.11.a.

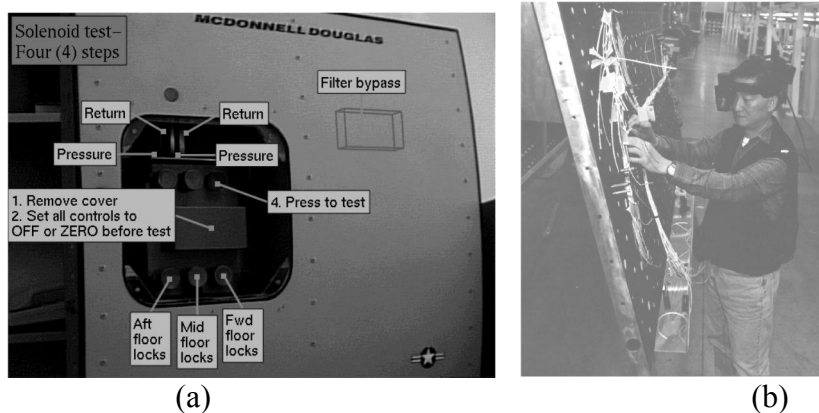


Figure 1.11 Aircraft manufacturing use of augmented reality, (a) (Neumann and Cho 1997); (b) (Dr. David Mizell, The Boeing Company)

2.3.7 Consumer Applications

Virtual reality systems are already used for consumer design. There are programs available for home computers that assist the homeowner with remodeling projects, such as building a new deck. These computer graphics programs allow you to graphically build the new deck attached to the back of a general model of a house. It is conceivable that a future system would allow you to make a video tape of your house shot from various viewpoints in your backyard. The video is input to your computer so that the design program could show the new deck in its finished form attached to your house. Or use a tape of your current kitchen and the augmented reality program replaces your current kitchen cabinetry with virtual images of the new kitchen that you are designing.

Applications in the fashion and beauty industry that would benefit from an augmented reality system can also be imagined. If the dress store does not have a particular style dress in your size an appropriate sized dress could be used to augment the image of you. As you look in the three sided mirror you see an image of the new dress on your body. You view changes in hem length, shoulder styles or other particulars of the design before you place the order. When you head into some high-tech beauty shops today you see what a new hair style will look like on a digitized image of yourself. But with an advanced augmented reality system you see the new styling as you move. If the dynamics of hair are included in the description of the virtual object you also see the motion of your hair as your head moves.

2.4 Performance Goals

The application domains and projects described so far in this chapter motivate the desire to develop interactive augmented reality systems. These current and future applications help to shape the requirements specification for an interactive augmented reality system. Some applications will not demand an “ultimate” augmented reality system.

An augmented reality system with a high degree of verisimilitude and utility will possess the following ideal characteristics:

- *Constrained cost to allow for broader usage.* To allow augmented reality systems to be deployed in a wide range of applications the cost for the system should be constrained. This leads to a goal of using inexpensive consumer grade video cameras, personal computer processing and lower resolution display technology.
- *Perfect static registration of virtual objects in the augmented view.* When a virtual object has been placed at a location in the real scene it should appear to the user to remain at that same

position in 3D space unless an object has interacted with it. If the system can not meet this goal for all static viewpoints, it will not be possible to meet the following more difficult dynamic registration requirement.

➤ *Perfect dynamic registration of virtual objects in the augmented view.* Visual updates should be performed at a rate of, at least 15 Hz, and preferably, 30 Hz. Perhaps more importantly, latencies should be minimized. If changes in the rendering of the virtual objects lag behind the user action triggering them the virtual objects will appear to “swim” around in three-dimensional space.

➤ *Perfect registration of visual and haptic scenes.* This can be phrased as WYSIWYF or “What you see is what you feel.” The visual image of a virtual object should match with its haptic counterpart. The user should feel the surface of a virtual object at the same time and in the same place that the augmented view shows the contact.

➤ *Virtual and real objects are visually indistinguishable.* In addition to photorealistic rendering of the virtual objects—the usual consideration for computer graphics applications—there are additional requirements specific to interactive augmented reality applications. Visual occlusions between virtual and real objects must occur correctly. This is not only for virtual objects occluding real ones, but also for the more difficult case of real objects occluding virtual ones. Lighting in the augmented view must be matched between the real and virtual worlds.

➤ *Virtual objects exhibit standard dynamic behavior.* When the user interacts with a virtual object it should move with the same dynamic behavior that an equivalent real object would exhibit. This includes correctly rebounding from collisions between virtual objects or between virtual and real objects. To accommodate this characteristic, the system’s internal representation of objects should help to compute the graphics rendering of virtual objects and the interactions and dynamics of all objects.

➤ *The user has unconstrained motion within the workspace.* The system should allow movement without constraints or limitations. It would be ideal to have no mechanical limitations, blind spots or motion constraints.

➤ *Minimal a priori calibration or run-time setup is required.* To determine the location of the viewer many augmented reality systems require calibration of the video camera viewing the scene. This calibration process is tedious to perform and will often limit operation to a single focal length. Lenses on standard consumer grade video cameras can not be zoomed in or out because they do not provide feedback of zoom position. During start-up of the system the user should not have to

perform extensive setup such as measurement of the locations of fiducials, or complicated procedures for placing objects into the scene.

3 Human Computer Interaction (HCI)

Utilizing computers had always begged the question of interfacing. The methods by which human has been interacting with computers has travelled a long way. The journey still continues and new designs of technologies and systems appear more and more every day and the research in this area has been growing very fast in the last few decades.

The growth in Human-Computer Interaction (HCI) field has not only been in quality of interaction, it has also experienced different branching in its history. Instead of designing regular interfaces, the different research branches have had different focus on the concepts of multimodality rather than unimodality, intelligent adaptive interfaces rather than command/action based ones, and finally active rather than passive interfaces.

This section intends to provide an overview on the state of the art of HCI systems and cover most important branches as mentioned above. In the next section, basic definitions and terminology of HCI are given. Then an overview of existing technologies and also recent advances in the field is provided. This is followed up by a description on the different architectures of HCI designs. The final sections pertain to description on some of the applications of HCI and future directions in the field.

3.1 Human-Computer Interaction: Definition, Terminology

Sometimes called as Man-Machine Interaction or Interfacing, concept of Human-Computer Interaction/Interfacing (HCI) was automatically represented with the emerging of computer, or more generally machine, itself. The reason, in fact, is clear: most sophisticated machines are worthless unless they can be used properly by men. This basic argument simply presents the main terms that should be considered in the design of HCI: functionality and usability .

Why a system is actually designed can ultimately be defined by what the system can do i.e. how the functions of a system can help towards the achievement of the purpose of the system. Functionality of a system is defined by the set of actions or services that it provides to its users. However, the value of functionality is visible only when it becomes possible to be efficiently utilised by the user . Usability of a system with a certain functionality is the range and degree by which the system can be used efficiently and adequately to accomplish certain goals for certain

users. The actual effectiveness of a system is achieved when there is a proper balance between the functionality and usability of a system .

Having these concepts in mind and considering that the terms computer, machine and system are often used interchangeably in this context, HCI is a design that should produce a fit between the user, the machine and the required services in order to achieve a certain performance both in quality and optimality of the services . Determining what makes a certain HCI design good is mostly subjective and context dependant. For example, an aircraft part-designing tool should provide high precisions in view and design of the parts while a graphics editing software may not need such a precision. The available technology could also affect how different types of HCI are designed for the same purpose. One example is using commands, menus, graphical user interfaces (GUI), or virtual reality to access functionalities of any given computer. In the next section, a more detailed overview of existing methods and devices used to interact with computers and the recent advances in the field is presented.

3.2 Overview on HCI

The advances made in last decade in HCI have almost made it impossible to realize which concept is fiction and which is and can be real. The thrust in research and the constant twists in marketing cause the new technology to become available to everyone in no time. However, not all existing technologies are accessible and/or affordable by public. In the first part of this section, an overview of the technology that more or less is available to and used by public is presented. In the second part, an outlook of the direction to which HCI research is heading has been drawn.

3.2.1 Existing HCI Technologies

HCI design should consider many aspects of human behaviours and needs to be useful. The complexity of the degree of the involvement of a human in interaction with a machine is sometimes invisible compared to the simplicity of the interaction method itself. The existing interfaces differ in the degree of complexity both because of degree of functionality/usability and the financial and economical aspect of the machine in market. For instance, an electrical kettle need not to be sophisticated in interface since its only functionality is to heat the water and it would not be cost-effective to have an interface more than a thermostatic on and off switch. On the other hand, a simple website that may be limited in functionality should be complex enough in usability to attract and keep customers .

Therefore, in design of HCI, the degree of activity that involves a user with a machine should be thoroughly thought. The user activity has three different levels: physical , cognitive , and affective . The physical aspect determines the mechanics of interaction between human and computer while the cognitive aspect deals with ways that users can understand the system and interact with it. The affective aspect is a more recent issue and it tries not only to make the interaction a pleasurable experience for the user but also to affect the user in a way that make user continue to use the machine by changing attitudes and emotions toward the user .

The focus of this paper is mostly on the advances in physical aspect of interaction and to show how different methods of interaction can be combined (Multi-Modal Interaction) and how each method can be improved in performance (Intelligent Interaction) to provide a better and easier interface for the user. The existing physical technologies for HCI basically can be categorized by the relative human sense that the device is designed for. These devices are basically relying on three human senses: vision, audition, and touch.

Input devices that rely on vision are the most used kind and are commonly either switch-based or pointing devices. The switch-based devices are any kind of interface that uses buttons and switches like a keyboard. The pointing devices examples are mice, joysticks, touch screen panels, graphic tablets, trackballs, and pen-based input. Joysticks are the ones that have both switches and pointing abilities. The output devices can be any kind of visual display or printing device.

The devices that rely on audition are more advance devices that usually need some kind of speech recognition. These devices aim to facilitate the interaction as much as possible and therefore, are much more difficult to build. Output auditory devices are however easier to create. Nowadays, all kind of non-speech and speech signals and messages are produced by machines as output signals. Beeps, alarms, and turn-by-turn navigation commands of a GPS device are simple examples.

The most difficult and costly devices to build are haptic devices. These kinds of interfaces generate sensations to the skin and muscles through touch, weight and relative rigidity. Haptic devices are generally made for virtual reality or disability assistive applications. The recent methods and technologies in HCI are now trying to combine former methods of interaction together and with other advancing technologies such as networking and animation. These new advances can be categorized in three sections: wearable devices, wireless devices, and virtual devices. The technology is improving so fast that even the borders between these new technologies are fading away and they are getting mixed together. Few examples of these devices are: GPS navigation

systems, military super-soldier enhancing devices (e.g. thermal vision, tracking other soldier movements using GPS, and environmental scanning), radio frequency identification (RFID) products, personal digital assistants (PDA), and virtual tour for real estate business. Some of these new devices upgraded and integrated previous methods of interaction. As an illustration in case, there is the solution to keyboarding that has been offered by Compaq's iPAQ which is called Canesta keyboard as shown in figure 1.12. This is a virtual keyboard that is made by projecting a QWERTY like pattern on a solid surface using a red light. Then device tries to track user's finger movement while typing on the surface with a motion sensor and send the keystrokes back to the device .



Figure 1.12 Canesta virtual keyboard

3.2.2 Recent Advances in HCI

In following sections, recent directions and advances of research in HCI, namely intelligent and adaptive interfaces and ubiquitous computing, are presented. These interfaces involve different levels of user activity: physical, cognitive, and affection.

3.2.2.1 Intelligent and Adaptive HCI

Although the devices used by majority of public are still some kind of plain command/action setups using not very sophisticated physical apparatus, the flow of research is directed to design of intelligent and adaptive interfaces. The exact theoretical definition of the concept of intelligence or being smart is not known or at least not publicly agreeable. However, one can define these concepts by the apparent growth and improvement in functionality and usability of new devices in market.

As mentioned before, it is economically and technologically crucial to make HCI designs that provide easier, more pleasurable and satisfying experience for the users. To realize this goal, the interfaces are getting more natural to use every day. Evolution of interfaces in note-taking tools is a good example. First there were typewriters, then keyboards and now touch screen tablet PCs that

you can write on using your own handwriting and they recognize it change it to text and if not already made, tools that transcript whatever you say automatically so you do not need to write at all.

One important factor in new generation of interfaces is to differentiate between using intelligence in the making of the interface (Intelligent HCI) or in the way that the interface interacts with users (Adaptive HCI). Intelligent HCI designs are interfaces that incorporate at least some kind of intelligence in perception from and/or response to users. A few examples are speech enabled interfaces that use natural language to interact with user and devices that visually track user's movements or gaze and respond accordingly.

Adaptive HCI designs, on the other hand, may not use intelligence in the creation of interface but use it in the way they continue to interact with users . An adaptive HCI might be a website using regular GUI for selling various products. This website would be adaptive -to some extent- if it has the ability to recognize the user and keeps a memory of his searches and purchases and intelligently search, find, and suggest products on sale that it thinks user might need. Most of these kinds of adaptation are the ones that deal with cognitive and affective levels of user activity. Another example that uses both intelligent and adaptive interface is a PDA or a tablet PC that has the handwriting recognition ability and it can adapt to the handwriting of the logged in user so to improve its performance by remembering the corrections that the user made to the recognized text.

Finally, another factor to be considered about intelligent interfaces is that most non-intelligent HCI design are passive in nature i.e. they only respond whenever invoked by user while ultimate intelligent and adaptive interfaces tend to be active interfaces. The example is smart billboards or advertisements that present themselves according to users' taste. In the next section, combination of different methods of HCI and how it could help towards making intelligent adaptive natural interfaces is discussed.

3.2.2.2 Ubiquitous Computing and Ambient Intelligence

The latest research in HCI field is unmistakably ubiquitous computing (UbiComp). The term which often used interchangeably by ambient intelligence and pervasive computing, refers to the ultimate methods of human-computer interaction that is the deletion of a desktop and embedding of the computer in the environment so that it becomes invisible to humans while surrounding them everywhere hence the term ambient.

The idea of ubiquitous computing was first introduced by Mark Weiser during his tenure as chief technologist at Computer Science Lab in Xerox PARC in 1998. His idea was to embed computers everywhere in the environment and everyday objects so that people could interact with many computers at the same time while they are invisible to them and wirelessly communicating with each other .

UbiComp has also been named the Third Wave of computing. The First Wave was the mainframe era, many people one computer. Then it was the Second Wave, one person one computer which was called PC era and now UbiComp introduces many computers one person era. Figure 1.13 shows the major trends in computing.

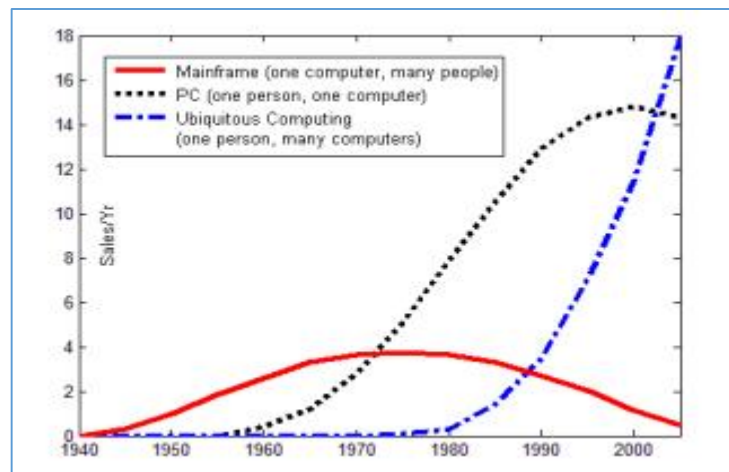


Figure 1.13 Major trends in computing

3.3 HCI Systems Architecture

Most important factor of a HCI design is its configuration. In fact, any given interface is generally defined by the number and diversity of inputs and outputs it provides. Architecture of a HCI system shows what these inputs and outputs are and how they work together.

Following sections explain different configurations and designs upon which an interface is based.

3.3.1 Unimodal HCI Systems

As mentioned earlier, an interface mainly relies on number and diversity of its inputs and outputs which are communication channels that enable users to interact with computer via this interface. Each of the different independent single channels is called a modality . A system that is based on only one modality is called *unimodal*. Based on the nature of different modalities, they can be divided into three categories:

- Visual-Based
- Audio-Based
- Sensor-Based

The next sections describe each category and provide examples and references to each modality.

3.3.1.1 Visual-Based HCI

The visual based human computer interaction is probably the most widespread area in HCI research. Considering the extent of applications and variety of open problems and approaches, researchers tried to tackle different aspects of human responses which can be recognized as a visual signal. Some of the main research areas in this section are as follow:

- Facial Expression Analysis
- Body Movement Tracking (Large-scale)
- Gesture Recognition
- Gaze Detection (Eyes Movement Tracking)

While the goal of each area differs due to applications, a general conception of each area can be concluded. Facial expression analysis generally deals with recognition of emotions visually . Body movement tracking and gesture recognition are usually the main focus of this area and can have different purposes but they are mostly used for direct interaction of human and computer in a command and action scenario. Gaze detection is mostly an indirect form of interaction between user and machine which is mostly used for better understanding of user's attention, intent or focus in context-sensitive situations. The exception is eye tracking systems for helping disabilities in which eye tracking plays a main role in command and action scenario, e.g. pointer movement, blinking for clicking. It is notable that some researchers tried to assist or even replace other types of interactions (audio-, sensor-based) with visual approaches. For example, lip reading or lip movement tracking is known to be used as an influential aid for speech recognition error correction.

3.3.1.2 Audio-Based HCI

The audio-based interaction between a computer and a human is another important area of HCI systems. This area deals with information acquired by different audio signals. While the nature of audio signals may not be as variable as visual signals but the information gathered from audio signals can be more trustable, helpful, and in some cases unique providers of information. Research areas in this section can be divided to the following parts:

- Speech Recognition

- Speaker Recognition
- Auditory Emotion Analysis
- Human-Made Noise/Sign Detections (Gasp, Sigh, Laugh, Cry, etc.)

Historically, speech recognition and speaker recognition have been the main focus of researchers. Recent endeavors to integrate human emotions in intelligent human computer interaction initiated the efforts in analysis of emotions in audio signals . Other than the tone and pitch of speech data, typical human auditory signs such as sigh, gasp, and etc helped emotion analysis for designing more intelligent HCI system.

3.3.1.3 Sensor-Based HCI

This section is a combination of variety of areas with a wide range of applications. The commonality of these different areas is that at least one physical sensor is used between user and machine to provide the interaction. These sensors as shown below can be very primitive or very sophisticated.

- Pen-Based Interaction
- Mouse & Keyboard
- Joysticks
- Motion Tracking Sensors and Digitizers
- Haptic Sensors
- Pressure Sensors
- Taste/Smell Sensors

Some of these sensors have been around for a while and some of them are very new technologies. Pen-Based sensors are specifically of interest in mobile devices and are related to pen gesture and handwriting recognition areas. Keyboards, mice and joysticks are already discussed in section 3.2.1.

Motion tracking sensors/digitizers are state-of-the-art technology which revolutionized movie, animation, art, and video-game industry. They come in the form of wearable cloth or joint sensors and made computers much more able to interact with reality and human able to create their world virtually. Figure 1.14 depicts such a device. Haptic and pressure sensors are of special interest for applications in robotics and virtual reality . New humanoid robots include hundreds of haptic sensors that make the robots sensitive and aware to touch. These types of sensors are also used in

medical surgery application . A few research works are also done on area of taste and smell sensors; however, they are not as popular as other areas.



Figure 1.14 Wearable motion capture cloth for making of video games (Taken from Operation Sports)

3.3.2 Multimodal HCI Systems

The term multimodal refers to combination of multiple modalities. In MMHCI systems, these modalities mostly refer to the ways that the system responds to the inputs, i.e. communication channels. The definition of these channels is inherited from human types of communication which are basically his senses: Sight, Hearing, Touch, Smell, and Taste. The possibilities for interaction with a machine include but are not limited to these types.

Therefore, a multimodal interface acts as a facilitator of human-computer interaction via two or more modes of input that go beyond the traditional keyboard and mouse. The exact number of supported input modes, their types and the way in which they work together may vary widely from one multimodal system to another. Multimodal interfaces incorporate different combinations of speech, gesture, gaze, facial expressions and other non-conventional modes of input. One of the most commonly supported combinations of input methods is that of gesture and speech.

Although an ideal multimodal HCI system should contain a combination of single modalities that interact correlatively, the practical boundaries and open problems in each modality oppose limitations on the fusion of different modalities. In spite of all progress made in MMHCI, in most of existing multimodal systems, the modalities are still treated separately and only at the end, results of different modalities are combined together.

The reason is that the open problems in each area are yet to be perfected meaning that there is still work to be done to acquire a reliable tool for each sub-area. Moreover, roles of different modalities and their share in interplay are not scientifically known. Yet, people convey multimodal communicative signals in a complementary and redundant manner. Therefore, in order to accomplish a human-like multimodal analysis of multiple input signals acquired by different sensors, the signals cannot be considered mutually independently and cannot be combined in a context-free manner at the end of the intended analysis but, on the contrary, the input data should be processed in a joint feature space and according to a context-dependent model. In practice, however, besides the problems of context sensing and developing context dependent models for combining multisensory information, one should cope with the size of the required joint feature space. Problems include large dimensionality, differing feature formats, and time-alignment.

An interesting aspect of multimodality is the collaboration of different modalities to assist the recognitions. For example, lip movement tracking (visual-based) can help speech recognition methods (audio-based) and speech recognition methods (audio-based) can assist command acquisition in gesture recognition (visual-based). The next section explains one of application of intelligent multimodal systems which is related by our thesis's topic.

4 User-Augmented Reality interaction

According to what we seen in previous sections, we can summarize the following means of user interaction:

- Motion sensors, changes in 6 dimensional space
- Vision focus
- Markers and props (gun, flashlight)
- Audible interaction and feedback
- Additional hand held computer, stylus
- Gloves, hand movement

However, there is a near limitless amount of different ways a user could interact with an AR system. Normal ways of human computer interaction, like WIMP (Windows, Icons, Menus, and Pointers) can be supplemented with anything a user's senses can take as input and with any action a user could come up with. A goal of AR is coming up with interaction that feels as natural as possible to the user, which is often particularly complex. Tan et. al. have done research in AR interaction by using different types of cards in engineering tasks. A special help card could be

placed or held next to another card and a small animation would show help information about the card (Tangible Bubble Help). For a short piece of information, the system automatically showed the name of an object once it enters the users workspace area, defined to be within arm's length (Tangible Tool tips). It appeared to be a very natural way for users, in this case engineers, to quickly gain extra information without having to resort to extra interaction with the system. There are tests underway within a large automobile company to use these results in trial applications.

Interaction between AR and users is such a complex subject, that it is linked to a broader but strongly related subject, called NUI: Natural User Interfaces. This paradigm is based on the assumption that users should interact with the virtual world just as they would with the real world (speech, handwriting, 3D projection). At the last Human Interaction symposium in 1999, there were a number of projects presented that focused on this topic:

Digital-Desk projects, about the use of a real desk and the adding of virtual tools to annotate papers on each other's desks, the use of retrieving computer files by using real world objects and the use of real playing chips on a virtual playing board for casino games. It appeared that the NUI/AR- enabled games gave the players a higher chance in winning in comparison with using conventional interfaces (command line, mouse + GUI, touchscreen + GUI). Familiarity with interfaces clearly helps users in this project. Their conclusion was that NUI/AR have many advantages over traditional interaction styles and Virtual Reality, with an emphasis on non-verbal communication (motor movements).

5 Latest news about augmented reality apps

Augmented reality is still something more people talk about than actually use on a day to day basis. However, that's beginning to change as organizations find new and clever ways to make it easy to use, compelling and rewarding[13].

We will give some examples of latest augmented reality apps:

5.1 Google Glass [14]

Google Glass is a headset that you wear like a pair of eyeglasses. Google has even announced that it's popping prescription lenses into some models. The headset has a small prism-like screen tucked into the upper corner of the frame that keeps you constantly plugged in to your e-mail, calls and other notifications so you don't have to miss a beat Figure 1.15.



Figure 1.15 An experimental model for Google Glass, Glass Explorers

The idea behind Glass is that bringing the technology closer will actually make it easier to disengage from it. Rather than having your head bent over a screen in your hand, you can look up. Rather than flicking through a list of notifications or e-mails to see if you missed anything important, you can make that decision immediately and get on with your day. Not that you can abandon your phone altogether. Glass is designed to provide a second screen for your device, so you need to hook it up to your phone to get the notifications and other information.

It's basically like wearing a heavier pair of glasses with a small screen that hangs just out of your direct line of vision.

The idea is to have that little notification bar from your phone in a place where you don't have to strain to see it; so you can actually look where you're going while you're out and about and stay plugged in to your inbox.

There are different ways to operate Glass. The device has a touchpad on the side (the part that goes over your ear) that you can tap or swipe for navigation. You can use voice commands for Glass by adding the phrase "Okay, Glass" to the start of whatever you tell it to do launch an app, take a picture, start a call, etc. Users can also wake up Glass by looking up.

Glass has its own store where developers can publish apps that take advantage of the device's unique design. These tend to offer quick bursts of information and seem most useful when you're doing something that requires your hands, such as cooking. You can get step-by-step instructions for a recipe from Glass, for instance, rather than soiling your cookbook with hands that are coated in sticky dough. In Figure 1.16, one person takes a picture of his kitchen appliances and shares it to the Fancy app. He then sees 20 "matching" items like storage drawers, shutters, rugs, and

wallpaper. From there, he can "Fancy" the item and save it to his collection, or buy it on the spot with a tap and a swipe

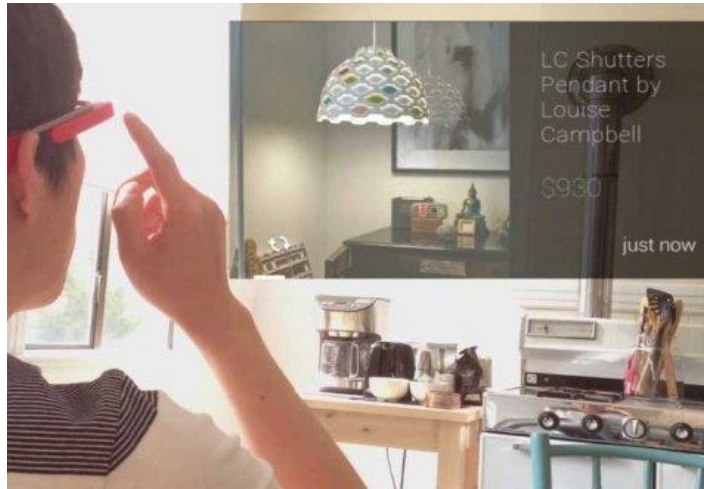


Figure 1.16 one person takes a picture of his kitchen appliances and shares it to the Fancy app

Until now, Google Glass isn't commercially available. But maybe you've seen people walking around with them; Those people are participating in Google's beta program, which the company calls "Glass Explorers." You have to apply to become one of those explorers, and it comes at a steep price: \$1,500. That's expensive, but remember that the devices out now are prototypes, aimed at developers and others who may have cool ideas about how to use the technology in the future.

The number of explorers has grown over the past year, though. It does appear that Google is setting the stage for a full commercial rollout this year

5.1.2 Some features that will be offered by Google Glass

➤ **GolfSight** It's a golf GPS rangefinder made specifically for Glass. It gives you accurate pin distance, course data, and scoring information, all without having to dig through your pockets.

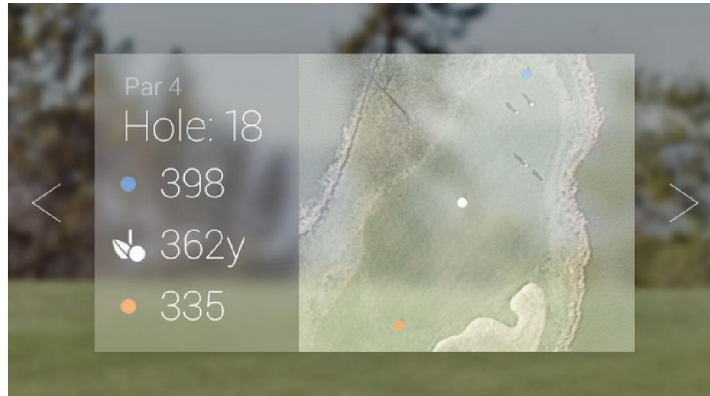


Figure 1.17 golf GPS rangefinder in Google Glass

➤ **Navigation** :Whether you love to cycle or run, Glass makes it easy to stay on course with turn-by-turn navigation.

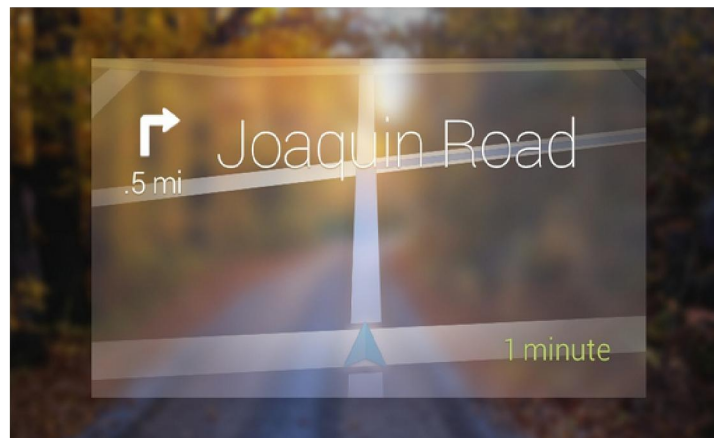


Figure 1.18 Navigation using Google Glass

➤ **LynxFit**: It's a personal fitness coach that helps keep you on track with your health goals. Subscribe to a workout challenge, track your activity, and receive real time coaching through Glass.

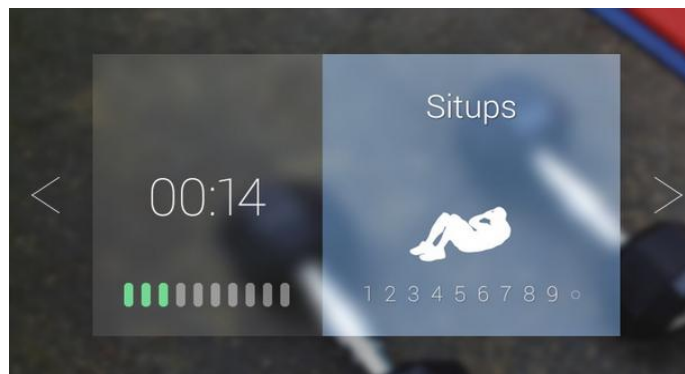


Figure 1.19 A personal fitness coach using Google Glass

5.2 Leap Motion[15]

Launched in July 24, 2013 with the promise of changing the way people interact with their computers, Leap Motion has already reached the 1 million app downloads (only in 3 weeks ago after launching) milestone through the app store that accompanies the device

5.2.1 How Leap Motion (Should) Work

The small rectangular device, no bigger than a computer mouse, can be easily plugged in the USB port of your device and after you download a small piece of software your Leap Motion device is fully functional. According to the company's specifications it can detect the position of your hand and fingers with a precision of 0.01mm within an eight cubic foot volume of space. To translate this, it means every finger twitching can be detected above the device within a space the size of a 33-inch screen in a square box.

The Leap can track any movement, pinch, wave or any other motion you can make with your fingers, hands or any other small devices such as pens. These are then translated into actual computer commands through the device's software and the apps that you can download from their Air Space app store. In this way you can control your computer, play games, draw, create 3D graphics, turn the pages of a digital newspaper or scroll through your digital photo albums.



Figure 1.20 Leap Motion hand detection

5.2.2 The Potential Applications of Leap Motion

With such a wide range of issues surrounding the device, it might be a surprise for many that Leap Motion has reached the 1 million mark in their Air Space app downloads. Although it benefited from a great hype raising the interest of thousands of early adopters, the potential of The Leap is more important than any other publicity and even then the initial drawbacks and it is certainly the one thing that will determine the long term success of this piece of hardware.

5.2.2.1 Medicine

Imagine Leap Motion integrated into hi-tech medical computers through which doctors can perform highly complicated and very delicate operations across continents, with the device calibrated in such a way that it can actually remove the normal shake that even surgeons have to a certain degree and eliminate any possible errors related to this.

Helped by augmented reality, doctors would be able to see and control everything without being in the same room as the patient. This can also be used to drive remote cameras through very sensitive areas of the body without touching or putting any mechanical pressure on the patient.



Figure 1.21 Leap Motion in Medicine

5.2.2.2 Entertainment

Although this is already the first application of the technology, it is still only at the beginning in terms of its potential. Coupled with augmented reality, The Leap could help control your character in an AR game that happens around you. Controlling radio devices (e.g. cars, planes, boats) without the need of physical controls can give you more accuracy and freedom.



Figure 1.22 Drawing using Leap Motion

5.2.2.3 3D Modelling and Graphics

Leap Motion 3D ModellingIt is the only device that gives full control of a 3D object on all three axes. Architects and designers can really benefit from this by getting the freedom they have always wanted when it comes to 3D modelling and 3D graphics. They can turn, scale and fully transform objects without having to use a physical 2D controller (e.g. mouse, keyboard, trackpad), thus saving precious time and allowing more creativity.



Figure 1.23 Modeling and drawing using leap Motion

6 Conclusion

Although promising, Augmented Reality is not there yet. There are a large number of technical challenges that need to be solved (pattern and 3D object recognition, optical character recognition, image processing, speech recognition as input, Registration, alignment issues for providing a usable output, next to a decreasing of the size of the systems), but especially the input with the user is an open book. Should there be a default set of interfaces for users to communicate with, or should this depend on the application at hand? Simply put, the field of AR has only in the past years begun experimenting with a large variety of interfaces with the user.

What if all these problems are solved and Ubiquitous Computing becomes a reality? What if we always could be connected with a virtual world, able to control computers with a wave of our hand, or by thought, or without even by thinking at all? The possibilities are endless, yet there could be a large number of risks and social problems connected to this Pandora's box. Privacy, security are but a few of the topics concerned. Having computers assist us with our daily work could even cause unrest among workers: If all you have to do is follow orders, aren't you just being used as a robot?

Still, having AR augmented to our lives should make the use of computers a lot easier for users. Instead of having mobile phones, desktops, laptops, PDA's and wristwatches you could make do

with a single, consistent way of controlling the components of computers we know today using a natural interface.

In this chapter, we have looked at what Augmented Reality is, or should be, we have seen a number of Augmented Reality research projects showing the various forms possible, we have looked at the different ways humans and Augmented Reality systems interact with each other. We have also seen a large number of limitations and problems. Still, AR seems to be the only direction in which the next generation of Human-Computer Interaction can evolve. There are many scientists working on solving separate issues with AR, in the not so- distant future AR will be reality.

CHAPTER 2

COMPUTER PRESENTATION:

STATE OF THE ART

1 Introduction

Presentations are one of the most powerful communication mediums because the spoken word allows us to make very deep and human connections with others. It is so effective to pass information to people in many situations, such as an academic conference or business.

Social work researchers and educators need to effectively communicate information in a variety of settings. A successful presentation requires the content to be clear and engaging.

Presentation technology are tools used to assist in conveying information during a presentation. When a speaker is verbally addressing an audience it is often necessary to use supplementary equipment and media to clarify the point. If the audience is large or the speaker is soft-spoken, a public address system may be employed.

At a basic level, visual content can be provided by drawing directly onto a blackboard or whiteboard during the presentation. At a more advanced level, flip charts, slide projectors, and overhead projectors are more suitable for displaying prepared content. The use of prepared material often results in a neater, more accessible, and better conveyed presentation.

Laser pointers or even a simple stick are often used to highlight a particular point of interest within a slide or image. Many manufacturers produce devices that provide remote control over electrical presentation equipment, thus allowing the speaker to move around the stage freely and to activate systems when they are required.

Occasionally still images do not provide enough information to convey the message clearly. For example, the presenter may attempt to describe a complex mechanical movement, in which case animation or video may provide the clearest method of delivery.

Computer presentation combined with digital projectors are frequently used because they provide a visually pleasing presentation and combine multiple media into a single device [16].

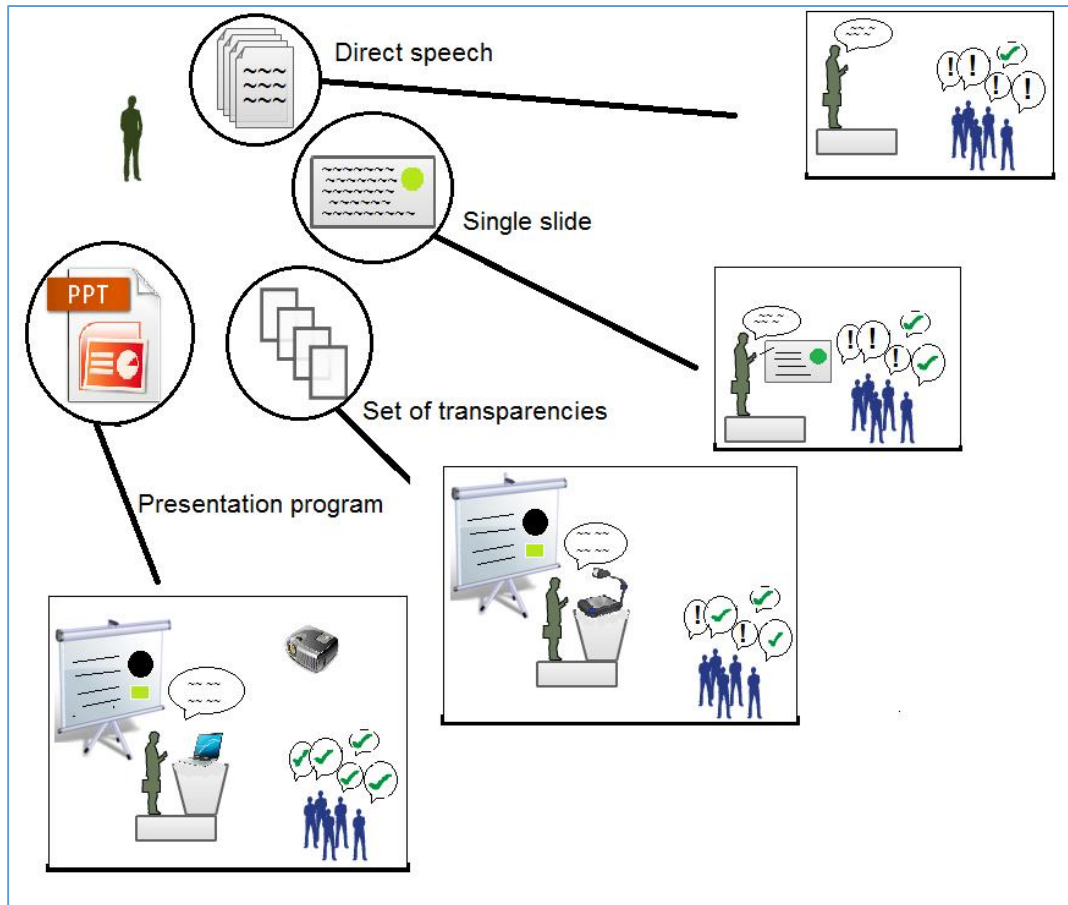


Figure 2.1 The impact of Computer Presentation to reach audiences compared with other means

Since my work is based on designing and construction of an augmented reality presentation software and as we talked before (chapter 1) about the state of interactive Augmented Reality, we need to speak about the state of presentation software before starting describe my practical part of my thesis (Chapter 3 and 4)

This chapter intends to provide an overview on the state of the art of computer presentation (or more precisely, it about presentation software) and cover most important branches. In the next sections, basic definitions and terminology of presentation programs are given. Then an overview of existing technologies besides presentation program, recent advances in the field and also a brief history is provided then a description on the different architectures of presentation programs. This is followed up by a branch of most appropriate presentation programs. The conclusion intends to an overview on my practical part of my thesis.

2 Presentation software: Definition and Terminology

2.1 Definition

Presentation software (sometimes called "presentation graphics") is a category of application program used to create sequences of words and pictures that tell a story or help support a speech or public presentation of information. Presentation software can be divided into business presentation software and more general multimedia authoring tools, with some products having characteristics of both. Business presentation software emphasizes ease- and quickness-of-learning and use. Multimedia authoring software enables you to create a more sophisticated presentation that includes audio and video sequences. Business presentation software usually enables you to include images and sometimes audio and video developed with other tools.

Some very popular presentation software, such as Microsoft's Powerpoint and Lotus's Freelance Graphics, are sold stand-alone or can come as part of office-oriented suites or packages of software. Other popular products include Adobe Persuasion, Astound, Asymetrix Compel, Corel Presentations, and Harvard Graphics. Among the most popular multimedia authoring tools are Macromedia Director and Asymetrix's Multimedia Toolbook. These authoring tools also include presentation capability as well. Most if not all of these products come in both PC and Mac versions.

Recently, a new presentation tool has arrived: your Web browser and the tools for creating Web pages, such as Microsoft's FrontPage and Adobe's PageMill. The ubiquity of these tools and the browser as a playback device make this a popular approach, especially when a presentation can combine HTML pages on the hard disk with links to outside sites (if you have a live Internet connection).[17]

2.2 Terminology

There are many terms related to the presentation programs need to be explained, in this section we will define some of them.

2.2.1 Presentation:

A presentation is the process of presenting a topic to an audience. It is typically a demonstration, lecture, or speech meant to inform, persuade, or build good will. The term can also be used for a formal or ritualized introduction or offering, as with the presentation of a debutante.

A study done by Wharton School of Business showed that the use of visuals reduced meeting times by 28 percent. Another study found that audiences believe presenters who use visuals are more professional and credible than presenters who merely speak. Other research indicates that

meetings and presentations reinforced with visuals help participants reach decisions and consensus more quickly.



Figure 2.2 A speaker giving a presentation using a projector

A presentation program, such as Microsoft PowerPoint, Apple Keynote, OpenOffice.org Impress or Prezi, is often used to generate the presentation content. Modern internet-based presentation software, such as the presentation application in Google Docs and SlideRocket also allow presentations to be developed collaboratively by geographically disparate collaborators. Presentation viewers can be used to combine content from different presentation programs into one presentation. [18]

2.2.2 Slide:

A slide is a single page of a presentation. Collectively, a group of slides may be known as a slide deck, hi slider or slide show. Historically, a slide was created on a transparency and viewed with an overhead projector. In today's digital age, a slide most commonly refers to a single page developed using a presentation program such as Microsoft PowerPoint or Apple Keynote. It is also possible to create them with a document markup language, for instance with the LaTeX class Beamer.[19]

2.2.3 Slide Show:

A slide show is a presentation of a series of still images on a projection screen or electronic display device, typically in a prearranged sequence. Each image is usually displayed for at least a few seconds, and sometimes for several minutes, before it is replaced by the next image. The changes may be automatic and at regular intervals or they may be manually controlled by a presenter or the viewer. Slide shows originally consisted of a series of individual photographic

slides projected onto a screen with a slide projector. When referring to the video or computer-based visual equivalent, in which the slides are not individual physical objects, the term is often written as one word, *slideshow*.



Figure 2.3 A slide show

A slide show may be a presentation of images purely for their own visual interest or artistic value, sometimes unaccompanied by description or text, or it may be used to clarify or reinforce information, ideas, comments, solutions or suggestions which are presented verbally. Slide shows are sometimes still conducted by a presenter using an apparatus such as a carousel slide projector or an overhead projector, but now the use of an electronic video display device and a computer running presentation software is typical [20].

3 History

The slide shows had their beginnings in the 1600s, when hand-painted images on glass were first projected onto a wall with a "magic lantern". By the late 1700s, showmen were using magic lanterns to thrill audiences with seemingly supernatural apparitions in a popular form of entertainment called a phantasmagoria. Sunlight, candles and oil lamps were the only available light sources. The development of new, much brighter artificial light sources opened up a world of practical applications for image projection. In the early 1800s, a series of hand-painted glass "lantern slides" was sometimes projected to illustrate story-telling or a lecture. Widespread and varied uses for amusement and education evolved throughout the century. By 1900, photographic images on glass had replaced hand-painted images, but the black-and-white photographs were sometimes hand-colored with transparent dyes. The production of lantern slides had become a

considerable industry, with dimensions standardized at 3.25 inches high by 4 inches wide in the US and 3.25 inches square in the UK and much of Europe.

"Magic lantern shows" also served as a form of home entertainment and were especially popular with children. They continued to have a place among commercial public amusements even after the coming of projected "moving pictures". Between films, early movie theaters often featured "illustrated songs", which were community sing-alongs with the lyrics and illustrations provided by a series of projected lantern slides. Theaters also used their lanterns to project advertising slides and messages such as "Ladies, kindly remove your hats".

After 35 mm Kodachrome color film was introduced in 1936, a new standard 2x2 inch (5x5 cm) miniature lantern slide format was created to better suit the very small transparencies the film produced. In advertising, the antique "magic lantern" terminology was streamlined, so that the framed pieces of film were simply "slides" and the lantern used to project them was a "slide projector". The old-fashioned magic lantern show became an up-to-date "slide show".

Home slide shows were a relatively common phenomenon in middle-class American homes during the 1950s and 1960s. If there was an enthusiast in the family, any visit from relatives or the arrival of a new batch of Kodachrome slides from the film processing service provided an excuse to bring out the entire collection of 35 mm slides, set up the slide projector and the screen, turn out the lights, then test the endurance of the assembled audience with a marathon of old vacation photos and pictures taken at weddings, birthdays and other family events, all accompanied by live commentary.

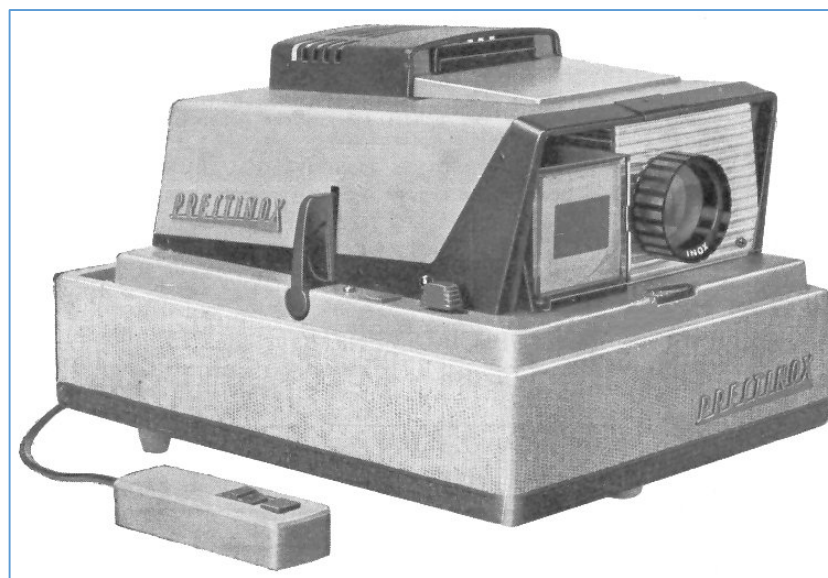


Figure 2.4 A 1960 slide projector

An image on 35 mm film mounted in a 2x2 inch (5x5 cm) metal, card or plastic frame is still by far the most common photographic slide format [20].

Early presentation graphics software ran on computer workstations, such as those manufactured by Trollman, Genigraphics, Autographix, and Dicomed. It became quite easy to make last-minute changes compared to traditional typesetting and pasteup. It was also a lot easier to produce a large number of slides in a small amount of time. However, these workstations also required skilled operators, and a single workstation represented an investment of \$50,000 to \$200,000 (in 1979 dollars).

In the mid-1980s, developments in the world of computers changed the way presentations were created. Inexpensive, specialized applications now made it possible for anyone with a PC to create professional-looking presentation graphics.

Originally, these programs were used to generate 35 mm slides, to be presented using a slide projector. As these programs became more common in the late 1980s several companies set up services that would accept the shows on diskette and create slides or print transparencies. In the 1990s dedicated LCD-based screens that could be placed on the projectors started to replace the transparencies, and by the late 1990s they had almost all been replaced by video projectors.

The first commercial computer software specifically intended for creating WYSIWYG presentations was developed at Hewlett Packard in 1979 and called BRUNO and later HP-Draw. The first software displaying a presentation on a personal computer screen was VCN ExecuVision, developed in 1982. This program allowed users to choose from a library of images to accompany the text of their presentation.



Figure 2.5 Brief History of Visual Aids

4 Presentation program architectures

It has three major functions: an editor that allows text to be inserted and formatted, a method for inserting and manipulating graphic images, and a slide-show system to display the content.

5 Most appropriate presentation programs

Before you choose any presentation application, you should know which features you need and which ones you don't. During our review process, we found several functions and attributes that are necessary for the best slideshow maker software.

5.1 Best presentation software characteristics [22]

5.1.1 Features

Any presentation application worth its salt will have the ability to add multimedia elements, from text and photos to audio and video. But the best ones also have built-in chart or diagram editors to help you better present data, word processing tools to organize your work and collaborate with others, password protection to keep your work safe and spell check to make sure you're text is polished. While nearly all applications offer slide transition effects and object animations, higher-quality products have more options, including 3D effects. And built-in video and image editors make tweaking your media a breeze. Other organizational tools to look for include decorative slide themes or templates, master slide layouts, layer functionality, alignment guides, slide organization tools and the ability to practice timing on your presentation.

5.1.2 Compatibility

Some presentation applications are more widely used than others, so if you choose a less-common product, you should make sure it's compatible with its more popular peers. This means it should be able to open or even save to file formats used by other applications, not just its native software format. You should also look for a product that allows you to save your work as a MOV file, a universal format that is read by QuickTime. The software should allow individual slides to be saved as image files (JPG, TIFF or PNG) or as PDF files. Having a web-ready compatible format is an added bonus.

5.1.3 Ease of Use

Though most presentation software is relatively straightforward and simple to use, you should still make sure your product is easy to navigate, with an intuitive software design. Since drag-and-

drop functionality is a key component to quick and easy slide creation, your slideshow presentation software should also have this feature.

5.1.4 Help and Support

Finally, there should be a wide range of self-help and technical support options available should you run into trouble. Self-help options include FAQs, manuals or help menus, troubleshooting guides and tutorials. Technical support options include live chat, email, phone and message boards.

If you find a presentation application with the features and tools listed here, you'll be able to create a slideshow presentation that will surely impress your friends, colleagues or bosses. The best presentation software applications offer many of the features listed here (figure 2.6)



Figure 2.6 2014 Best Presentation Software Comparisons and Reviews

5.2 Presentation Software similar To PowerPoint

In the last couple of years, many people have been searching for innovative ways to create and execute unique and memorable presentations. If you need to create a snazzy, animated presentation that captures attention, then you need business presentation software that's designed to do the job. Here are some examples of most appropriate presentation software.

5.2.1 Prezi

Prezi is about as different from PowerPoint as you can get. Rather than offering a slide-by-slide presentation, it's more of a visual and interactive mindmap, wherein you interact with different

elements on a virtual canvas. Convert your PowerPoint presentation slides into a dynamic user experience, or create a brand new ‘prezi’ from scratch.

Prezi is an online tool, so it’s available anywhere that you have an internet connection. There’s also the ability to work offline and to set security levels around who can see your animated presentations. Because Prezi is non-linear, it’s easy to jump from editing one part to another, although this can be difficult to get used to at first. Once you’re done, publish it on Prezi.com or embed it on your own website.

5.2.2 GoAnimate

In many cases, an informative, animated video presentation is the best way to get your message across. While not strictly business presentation software, GoAnimate provides you with the tools to create a standout animation so you can focus on the narrative. It’s a full suite of applications designed to help users create animated presentations for free.

GoAnimate is a great platform for presenting your idea, with a huge variety of pre-made themes, props and characters tailored for making professional business presentations. Rather than just queuing up images in slides, you can produce a professional-looking video from scratch in minutes – without drawing anything or operating a camera. Just drag and drop from the extensive content library and forget about recording or editing. Then download, share or publish your finished animated presentation to YouTube.

5.2.3 Google Docs

Almost everyone has a Google account these days, and therefore access to Google Docs! As a PowerPoint alternative, this may not be as powerful, but it is free.

Presentations are easily customized with text, images, videos, imported slides, and all that good stuff. It’s all done through the cloud, so it’s easy to collaborate on shared documents with other people. If you like, your presentation can be downloaded as a PowerPoint file as well as embedded or shared through a link.

5.2.4 Slidrocket

Now we’re talking high end presentation software. Slidrocket is a premium PowerPoint alternative that’s packed with features.

Here’s how it works: import a PowerPoint presentation or start from scratch with the theme library. Embed images, charts, videos, transitions, special effects – all of which can be stored,

searched for later and shared with others. Live data from real-time feeds online can also be embedded (slides automatically update with the latest information).

Extra capabilities include version control, privacy control – decide who can view, edit or share your slides – plus metrics to track who is viewing your slideshow, where they viewed it, who they've shared it with, which slides they looked at and for how long. Sliderocket presentations are also HTML5 compatible.

5.2.5 Zoho Show

Zoho Show's presentation software is a cloud-based option that supports the import of many file types (.ppt, .pptx, .pps, .ppsx, .odp, .sxi) and retains the look of the original document when viewed online. As a Powerpoint alternative, Zoho's clean and simple interface is easy to navigate.

Anyone creating business presentations will appreciate the custom flowcharts and diagrams, complete with dynamic shapes and connectors. Enhance your presentations with pictures and videos embedded straight from the web, Picasa and Flickr and add effects or touch-ups in seconds using the image editor. To make things more interesting, there's a wide gallery of animation and slide transition effects too. Plus Zoho enables universal style changes in a snap through the master slide feature, which rolls out changes to every single slide thereafter.

Version tracking is supported, as is real-time collaboration with other users. You get to choose who has permission to edit or view your presentation.

Ready to show off your final product? Send out a link to your business presentation, publish it to Zoho's Public Presentations section, or embed it on your website.

[<http://goanimate.com/video-maker-tips/6-best-business-presentation-software-and-powerpoint-alternatives>]

5.2.6 Augmented Reality Presentation from RTT[23]

For the Guangzhou, China launch of Dongfeng Nissan's new Teana, RTT developed a spectacular Augmented Reality (AR)-based presentation that offered the public a detailed, scaled product experience in a hitherto unknown form and quality. During the unveiling, a virtual image of the Teana was projected onto the stage and orchestrated interactively with AR in front of the public's eyes. This was made possible by tracking the presenter's on-stage movements in realtime and combining the information with scenes produced in RTT DeltaGen.

“A conventional product presentation is no longer enough to communicate all core messages and USPs of a product in an emotionally-charged, memorable way. That's why we joined forces

with RTT and used AR technology for the first time at a launch presentation. The results exceeded our wildest dreams,” said Jason Yang, Vice President Sales & Marketing, Dongfeng Nissan Passenger Vehicle Company.



Figure 2.7 Augmented Reality Presentation from RTT

6 Conclusion

In this chapter, it has been presented most issues related to the presentation technology, due to its importance to secure greater connection with audience. After this chapter, it appears how presentation is performing an important role to give the audience much understanding of the presenter topic.

For this reason, the software used in presentation must be more interactive with presenter and far be a reason to reduce his capacities. On the contrary, it must be flexible and harmonious with subtraction. For this, the combining of presenter with his/her presentation on the same screen using the interactive augmented reality technology contributes in a very big role in providing a successful presentation; one of these important roles is drawing the attention of audience in only one direction. It is the screen then, which hold all rather than dispersing the audience between the monitor and the presenter.

After all, it seems that the principal idea of this work becomes clear and isn't about providing the theories around the topic. So, putting it in use and getting some of its advantages are now what may be coming in the next chapters.

Finally, giving a name to the work ahead or product could be difficult. However, this is necessary do. Any product developer does it, because generating a name to the software product is very important. Therefore, this work is performed under the name ASPC that is the abbreviation of Augmented Slide Presenter Combiner. A further explanation and ore detail are provided ahead.

CHAPTER 3

TECHNIQUES USED TO CREATE ASPC

1 Introduction

Now we come to the most important phase that are interested in our project in the techniques used to create ASPC. In this chapter, we will talk about the most important stages that we experienced until we got ASPC a successful presentation program ready for use.

However, before we begin we need to explain some technical terms and give some important definitions.

2 Important technical terms : overview

In addition to the previews definitions in (Chapter1, Chapter2), we will define some technical terms, which belong to the project research and closely is linked to the AR and HCI technology.

2.1 Image processing

In imaging science, image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it.

Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images (producing the input image in the first place) is referred to as imaging[1].

Gonzalez, Rafael; Steve Eddins (2008). "4". *Digital Image Processing Using MATLAB* (2nd ed.). Mc Graw Hill. p. 163.

2.2 Computer vision

Computer vision is a field that includes methods for acquiring, processing, analyzing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions.[24][25][26][27] A theme in the development of this field has been to duplicate the abilities of human vision by electronically perceiving and understanding an image.[28] This image understanding can be seen

as the disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistics, and learning theory.[29] Computer vision has also been described as the enterprise of automating and integrating a wide range of processes and representations for vision perception.[30][31].

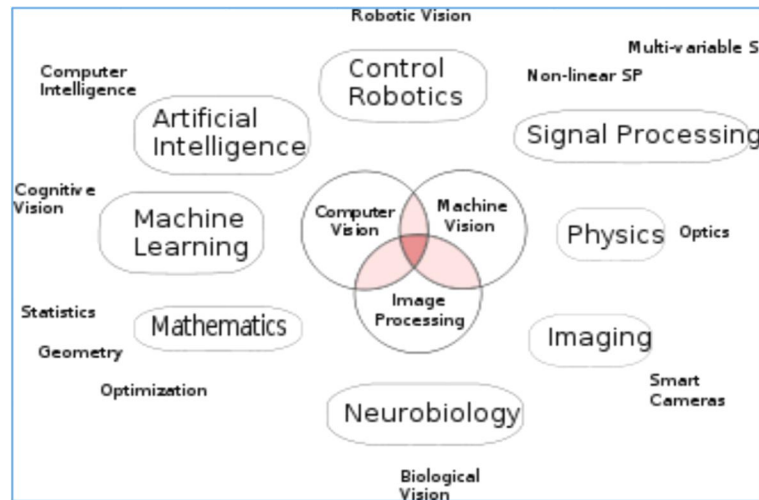


Figure 3.1 Relation between computer vision and various other fields

As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner. As a technological discipline, computer vision seeks to apply its theories and models to the construction of computer vision systems.

Sub-domains of computer vision include scene reconstruction, event detection, video tracking, object recognition, learning, indexing, motion estimation, and image restoration.

2.3 Gesture recognition

Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Current focuses in the field include emotion recognition from the face and hand gesture recognition. Many approaches have been made using cameras and computer vision algorithms to interpret sign language. However, the identification and recognition of posture, gait, proxemics, and human behaviors is also the subject of gesture recognition techniques.[32] Gesture recognition can be seen as a way for computers to

begin to understand human body language, thus building a richer bridge between machines and humans than primitive text user interfaces or even GUIs (graphical user interfaces), which still limit the majority of input to keyboard and mouse.



Figure 3.2 A child being sensed by a simple gesture recognition algorithm detecting hand location and movement

Gesture recognition enables humans to communicate with the machine (HMI) and interact naturally without any mechanical devices. Using the concept of gesture recognition, it is possible to point a finger at the computer screen so that the cursor will move accordingly. This could potentially make conventional input devices such as mouse, keyboards and even touch-screens redundant.

Gesture recognition can be conducted with techniques from computer vision and image processing.

The literature includes ongoing work in the computer vision field on capturing gestures or more general human pose and movements by cameras connected to a computer.[33][34][35][36]

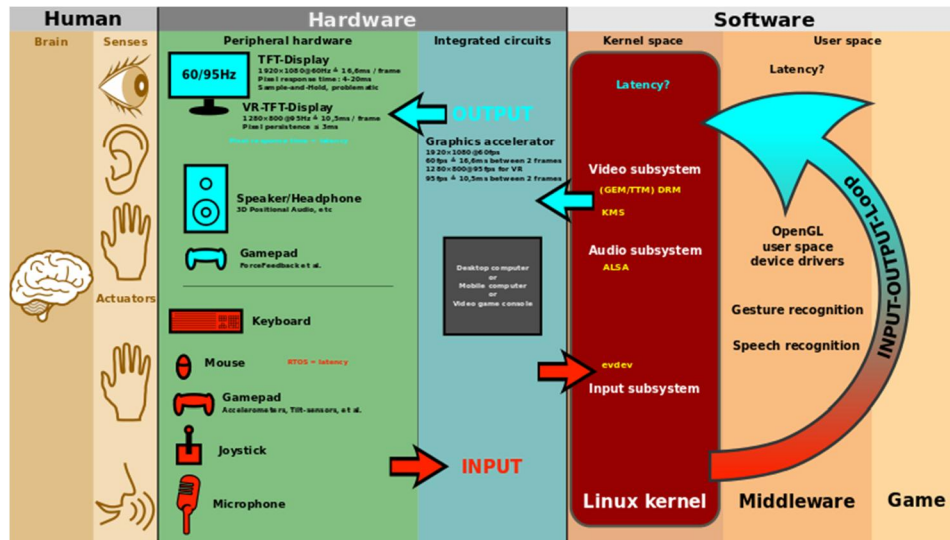


Figure 3.3 Gesture recognition is usually processed in middleware, the results are transmitted to the user applications.

Gesture recognition and pen computing: This computing not only going to reduce the hardware impact of the system but also it increases the range of usage of physical world object instead of digital object like keyboards, mice. Using this we can implement and can create a new thesis of creating of new hardware no requirement of monitors too. This idea may lead us to the creation of holographic display. The term gesture recognition has been used to refer more narrowly to non-text-input handwriting symbols, such as inking on a graphics tablet, multi-touch gestures, and mouse gesture recognition. This is computer interaction through the drawing of symbols with a pointing device cursor.[37][38][39]

2.3.2 Vision based hand gesture taxonomies and representations [40]

Psychological aspects of gestures based on hand gesture taxonomies and representations are also an important aspect of hand gesture recognition systems. Varying from person to person several hand gesture taxonomies have been proposed in the literature. Gesture acts a medium of communication for non vocal communication in conjunction with or without verbal communication is intended to express meaningful commands. These gestures may be articulated with any of the body parts or with combination of one or many of them. Gestures being major constituent of human communication may serve as an important means for human computer interaction too. Though the significance and meaning associated with different gestures differ very much with cultures having less or invariable or universal meaning for single gesture. For instance

different gestures are used for greeting at different geographical separations of the world. For example pointing an extended finger is a common gesture in USA & Europe but it is taken to be as a rude and offensive gesture in Asia. Hence the semantic interpretation of gestures depends strictly on given culture. Theoretically the literature classifies hand gestures into two type's static and dynamic gestures. Static hand gestures are defined as orientation and position of hand in the space during an amount of time without any movement and if a movement is there in the aforementioned time duration it is called dynamic gesture. Dynamic hand gestures include gestures involving body parts like waving of hand while static hand gestures include single formation without movement like jamming the thumb and forefinger to form the "ok" symbol is a static pose which represents static gesture.

Dynamic hand gestures done intentionally for communication are called conscious dynamic gestures, whereas unintentionally (unawareness) done gesture carried out during causal physical activity is known as unconscious dynamic gesture. Figure 3 shows the taxonomy of hand gesture categories. According to research (Hall 1973) 35 % of human communication consists of verbal communication and 65 % is non verbal gesture based communication. Gesture Ottenheimer(2005) can be categorized into five types i.e. emblems, affect displays, regulators, adaptors and illustrators.

Emblematic gestures also referred as emblem or quotable gestures are direct translation of short verbal communication like waving hand for good bye or nodding for assurance. The quotable gestures are specifically culture specific. Gestures conveying emotion or intensions are called affect displays.

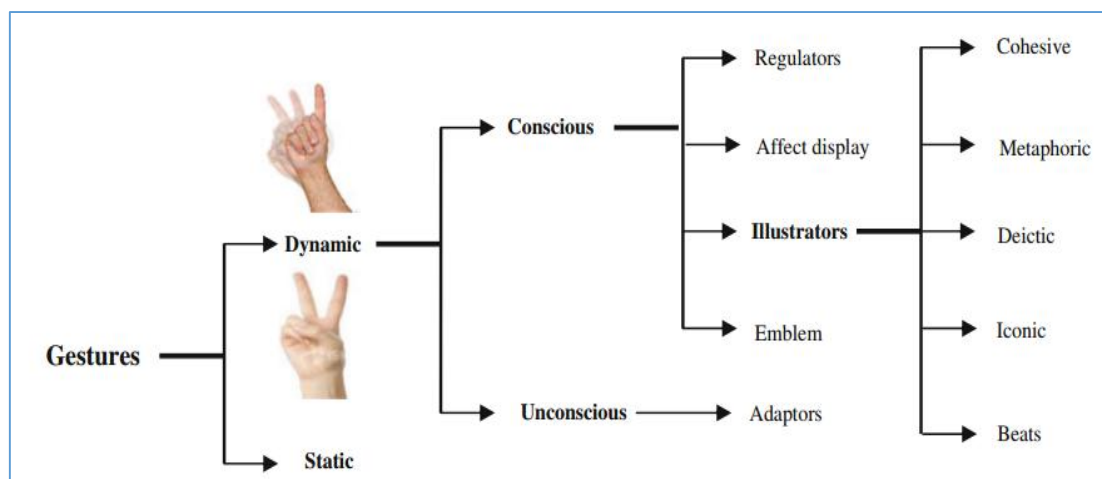


Figure 3.4 Vision based hand gesture taxonomies (Kanniche 2009)

The affect displays are generally universal less dependent on culture. Gestures controlling interaction are called regulators. Gestures like headshaking, quickly moving one's leg that enables the release of body tension are called adaptors. Adaptors are generally habit of communicators that are not used intentionally during a communication. Illustrator gestures emphasize the key point in speech to depict the communications pronouncing statements. Being emphasized by the communicators pronouncing statements these gestures are inherently dependent on communicators thought process and speech. These gesticulations could further be categorized into five sub category namely beats, deictic gestures, iconic gestures, metaphoric gestures and cohesive gestures (McNeill 1992).

- Beats are short and quick, rhythmic and after repetitive gestures.
- Concrete pointing to real location object or person and abstract pointing to abstract location or period of time are called deictic gestures.
- Hand movements depicting figural representation or actions for example moving hand upward with wiggling fingers to depict tree climbing are called iconic gestures.
- Abstractions are depicted by metaphoric gestures.
- Thematically related but temporally separated gestures are called cohesive gestures. The temporal separation of these thematically related gestures is due to interruption of current communicator by any other communicator.

Vision based Hand Gesture Representations: To abstract and model the human body parts motion several hand gesture representations and models have been proposed and implemented by the researchers. The two major categories of hand gesture representation are 3D model based methods and appearance based methods as depicted in Fig.3.5.

The 3D model based hand gesture recognition has different techniques for gesture representation namely 3D textured volumetric, 3D geometric model and 3D skeleton model. Appearance based hand gesture representation include color based model, silhouette geometry model, deformable gabarit model and motion based model.

The 3D model based hand gesture representation defines 3D spatial description of human hand for representation with temporal aspect being handled by automation. This automation divides the temporal characteristics of hand gesture into three phases (McNeill 1992) i.e. the preparation or prestroke phase, the nucleus or stroke phase and the retraction or poststroke phase in which every phase corresponds to one or many transitions of spatial states of the 3D human model. One or many

cameras focus on the real target and compute parameters spatially matching the real target and then follow its motion during the recognition process in 3D model. Thus the 3D model has an advantage that it updates the model parameters while checking the matches of transition in temporal model, leading to precise hand gesture recognition and representation, though making it computationally intensive with requirement of dedicated hardware. There are also many methods (Boulay 2007) that combine silhouette extraction with 3D model projection fitting by finding target self oriented. Generally three kinds of model are generally used.

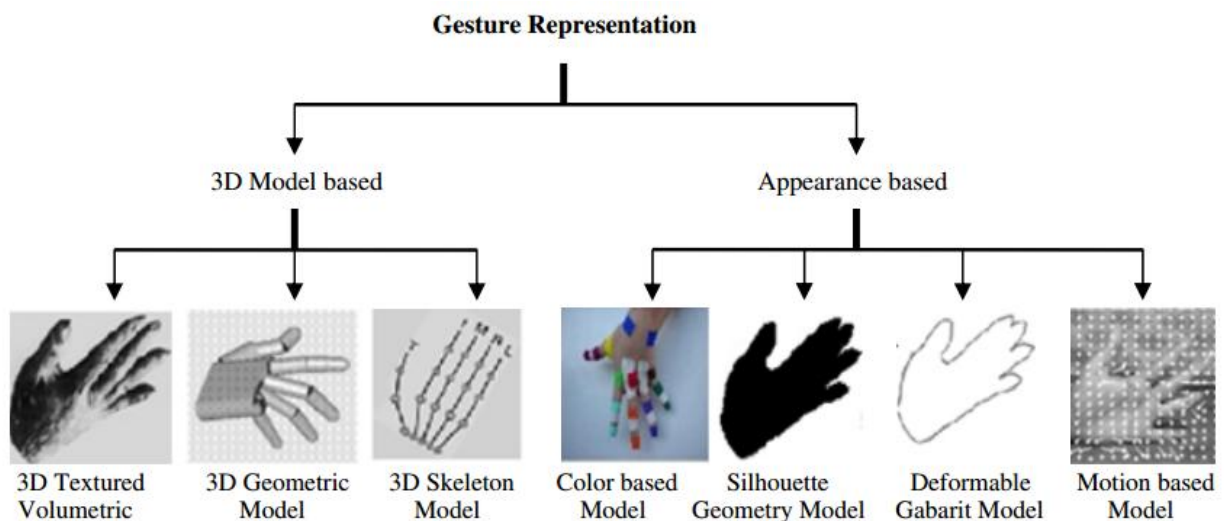


Figure 3.5 Vision based hand gesture representations
(Bourke et al. 2007)

- 3D textured kinematic/volumetric model contains very high details of human body skeleton and skin surface information.
- 3D geometric models are less precise than 3D textures kinematic/volumetric models with respect to skin information but contains essential skeleton information. Appearance based hand gesture representation methods are though broadly classified into two major subcategories i.e. 2D static model based methods and motion based methods, each sub category is having further variants. The commonly used 2D models include:
 - Color based model uses body markers to track the motion of body or body part. As Bretzner et al.(2002) proposed hand gesture recognition employing multi-scale color features, hierarchal models and particle filtering.

➤ Silhouette geometry based models include several geometric properties of the silhouette such as perimeter, convexity, surface, bounding box/ellipse, elongation, rectangularity, centroid and orientation. The geometric properties of the bounding box of the hand skin were used to recognize hand gestures (Birdal and Hassanpour 2008).

➤ Deformable gabarit based models: they are generally based on deformable active contours (i.e. snake parameterized with motion and their variants. Ju et al.(1997) used snakes for the analysis of gestures and actions in technical talks for video indexing.

➤ Motion based models are used for recognition of an object or its motion based on the motion of object in an image sequence. Local motion histogram was introduced by Luo et al.(2008) which uses an Adaboost framework for learning action models

2.3.3 Vision based hand gesture recognition techniques

Most of the complete hand interactive mechanisms that act as a building block for vision based hand gesture recognition system are comprised of three fundamental phases: detection, tracking and recognition. This section discusses some of the prominent vision based hand gesture recognition techniques used by most of the researchers by categorizing under the three verticals representing the three fundamental phases of detection tracking and recognition as shown in Fig.3.6.

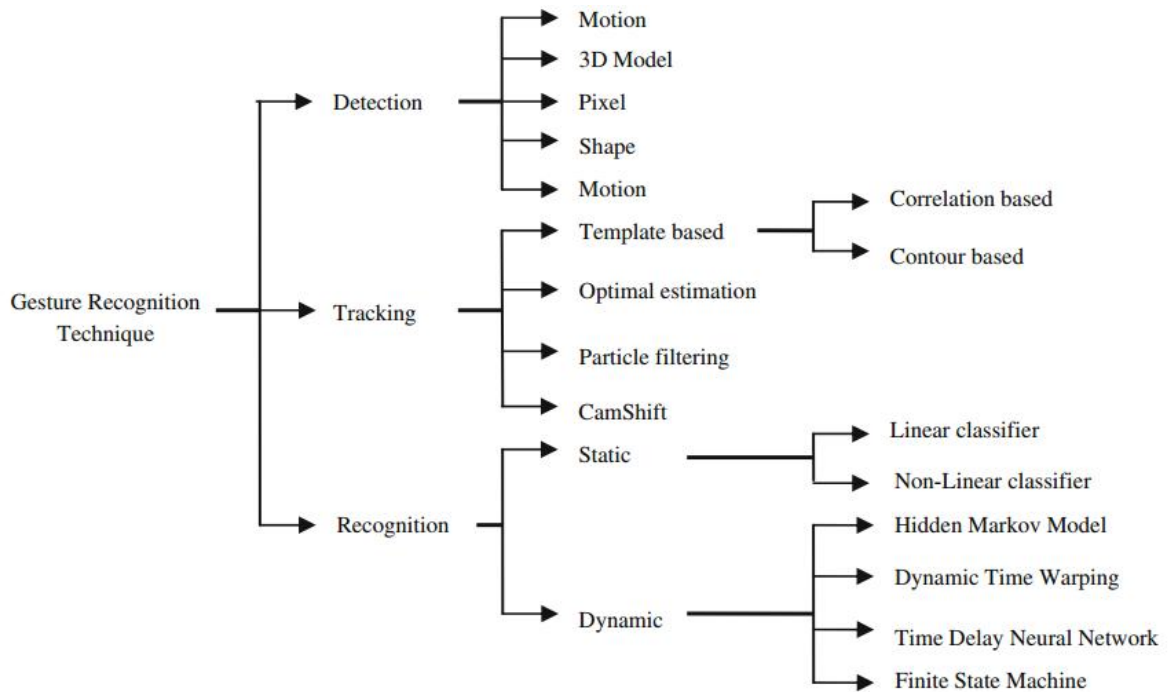


Figure 3.6 Vision based hand gesture recognition techniques

2.3.3.2 Detection

The primary step in hand gesture recognition systems is the detection of hands and the segmentation of the corresponding image regions. This segmentation is crucial because it isolates the task-relevant data from the image background, before passing them to the subsequent tracking and recognition stages. A large number of methods have been proposed in the literature that utilize a several types of visual features and, in many cases, their combination. Such features are skin color, shape, motion and anatomical models of hands. InCote et al.(2006), Zabulis et al.(2009).

2.3.3.3 Tracking

If the detection method is fast enough to operate at image acquisition frame rate, it can be used for tracking as well. However, tracking hands is notoriously difficult since they can move very fast and their appearance can change vastly within a few frames. Tracking can be defined as the frame-to-frame correspondence of the segmented hand regions or features towards understanding the observed hand movements. The importance of robust tracking is twofold. First, it provides the inter-frame linking of hand/finger appearances, giving rise to trajectories of features in time. These trajectories convey essential information regarding the gesture and might be used either in a raw form (e.g. in certain control applications like virtual drawing the tracked hand trajectory directly guides the drawing operation) or after further analysis (e.g. recognition of a certain type of hand gesture). Second, in model-based methods, tracking also provides a way to maintain estimates of model parameters variables and features that are not directly observable at a certain moment in time.

2.3.3.4 Recognition

The overall goal of hand gesture recognition is the interpretation of the semantics that the hand(s) location, posture, or gesture conveys. Vision based hand gesture recognition techniques can be further classified under static and dynamic gestures. To detect static gestures (i.e. postures), a general classifier or a template-matcher can be used.

3 Techniques used to develop ASPC

After these brief explanations, we can now display the most important stages and techniques used to develop ASPC. However, after the presentation of many information about hand gesture recognition we can determine the type of hand gesture recognition used for ASPC.

In my project, the hand presents the only way to interact with ASPC (HCI project part) by virtual buttons shown with the user (AR part), so that the user can click on them.

We find that the user needs a single input presented in hand click, so we need computer to detect ,in general, the hand in the video and ,in particular, detect the hand input (click).fig.3.7. Through that, the technique chosen is detection of hand shape .fig3.6 based on deformable representation .fig.3.5.

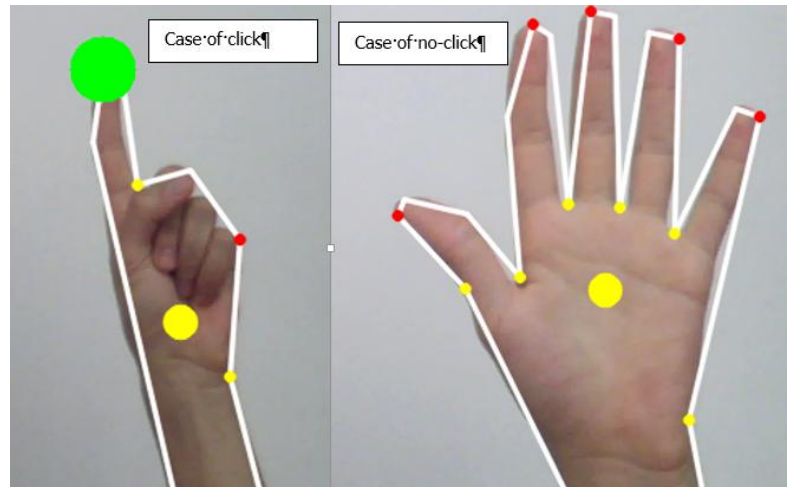


Figure 3.7 Clicking in ASPC by closing the hand and pointing index finger (arises the top of index finger)

According to this technique with some additions and modifications, We have developed ASPC as follows.

3.2 Pre-Processing and Convex-Hull Method

In this section, we will introduce how to extract the interested area from color image. We use a single web camera to capture a series of images. After transforming the image into HSV color space, we might be able to extract the interested area by defining the range of hue, saturation and value. A binary image will be produced and morphology processing will be performed. Erosion will eliminate the noises while the dilation will fill up the defects to smooth the contour of interested area.

Theo Pavlidi's algorithm [41] is used to obtain the contour sequence. Instead of using connected-component, we can determine the contour we are interested in by choosing the longest sequence. The convex-hull can be found by the Three-Coins algorithm [42][43], it's a set of points which can wrap the contour like a rubber band.

By comparing the convex-hull area and contour area, convexity defect area can be calculated. Convexity defect means the area, which is in the convex-hull area but not in the contour area.

Convexity defects help us to find the features of a contour. The depth of a defect area is the most important feature with the centroid of contour in this thesis. The hand (also the click) can be detected by a set of defection depths and using the centroid.

3.2.1 Preliminary Processing

The color model captured from a web camera is composed of RGB values, but it will be influenced by the light very easily, we must convert the RGB color space into another color space which is not sensitive to light variation. In addition to RGB, there are some other commonly used color spaces such as Normalized RGB [44], HSV [45], YCbCr [46], and so forth. In order to make the system achieve real-time processing and adapt to most of the environments, we choose the color model with simple converting formulas and low sensitivity. In this thesis, we choose HSV color space to extract human skin areas. The skin color areas will be represented in binary image. Hence we will perform morphology processing including erosion and dilation. Noises will be eliminated by erosion and the contour of skin color area will be smoothed by dilation.

3.2.1.1 Skin color detection using HSV color space

The HSV model comprises three components of a color: hue, saturation, and value. It can distinguish the value from the hue and saturation. A hue represents the gradation of a color within the optical spectrum, or visible spectrum of light. Saturation is the intensity of a specific hue, which is based on the purity of a color. The practicability of the HSV model can be referred to two main factors: 1) the value separated from the color image, and 2) both the hue and saturation related to human vision. Hence, for developing a human vision-based system, the HSV model is an ideal tool [45]. The chroma component (hue and saturation) and the luminance component (value) of the HSV color model are defined by the stereo picture as shown in Figure 3.8. All colors composed of three so-called primary colors (red, green, and blue) are inside the chromatic circles. The hue H is the angle of a vector with respect to the red axis. When $H = 0^\circ$, the color is red. The saturation S is the degree of a color that has not been diluted with white. It is proportional to the distance from the location to the center of a circle. The longer distances are away from the center of a chromatic circle, the more saturation the colors will be. The value V is measured by the line which is orthogonal to the chromatic circle and passes through the center of the circle. It tends to be white along this center line to the apex.

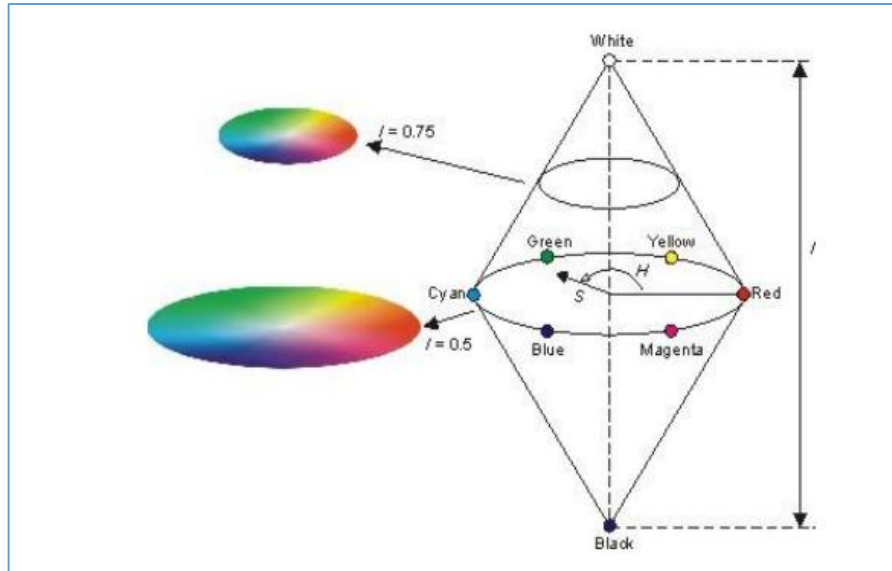


Figure 3.8 The stereo picture of the HSV color model.

The relationship between the HSV and RGB models is expressed below:

$$H = \begin{cases} \theta & \text{if } B \leq G \\ 360^\circ - \theta & \text{if } B > G \end{cases} \quad (3.1)$$

$$\text{where } \theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G) + (R-B)]}{\left[(R-G)^2 + (R-B)(G-B) \right]^{\frac{1}{2}}} \right\} \quad (3.2)$$

$$S = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)] \quad (3.3)$$

$$\text{and } V = \max(R, G, B) \quad (3.4)$$

We distinguish the skin color from non-skin color regions by setting upper and lower bound thresholds. In our experimental environment, we choose the H value from 2 to 39 and from 300 to 359; the S value between 0.1 and 0.9 for the range of skin colors. Some of skin color detection results are shown in Figure 3.9.

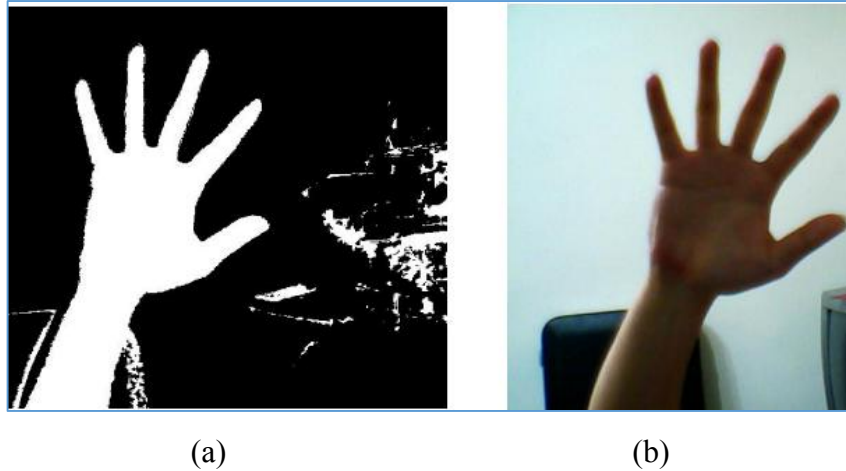


Figure 3.9 (a) Binary image (b) Original image

3.2.1.2 Morphology Processing

The erosion and dilation operations are both the conventional and popular methods used to clean up anomalies in the objects. The combination of these two operations can also achieve good performance. Therefore, a moving object region is morphologically eroded (once) then dilated (twice). Some examples of morphological operation are demonstrated in Figure 3.10

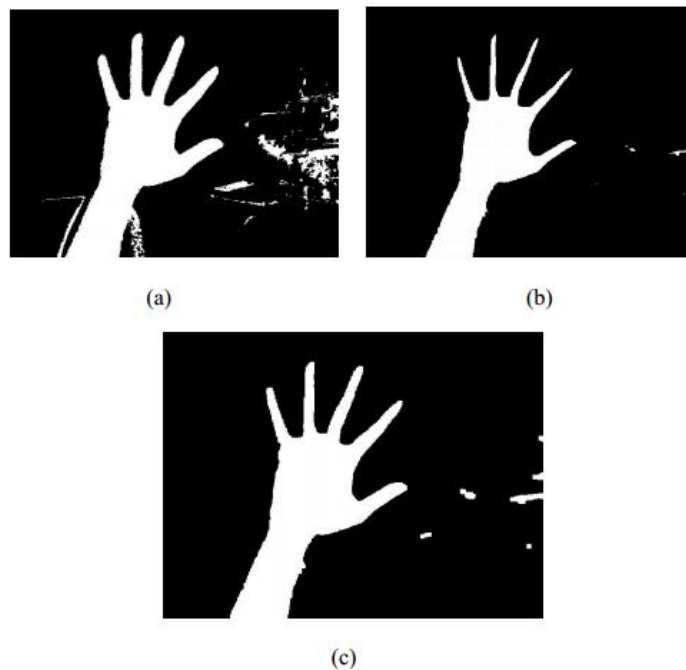


Figure 3.10 (a) input binary image (b) Erosion (c) Dilation

3.2.1.3 Contour Finding

After the previous processing, we should be able to obtain a binary image. The white pixels represent for skin color regions. The next step will be contour finding. A contour is a sequence of points which are the boundary pixels of a region. The contour of those regions will be found so that we can disregard those small areas and focus on the fore-arm area we are interested in. It can be done by comparing the length of their contour. The longest one is the one we are looking for. Unlike connected-component, contour finding doesn't need to trace the whole pixels in an area. Only boundary pixels will be visited. It brings not only lower computation cost but also comes with the contour information we need. In this thesis, we use Theo Pavlidis' Algorithm [40] to find contours. It works very well on 4-connected patterns which the fore-arm contours are always tending to be.

The algorithm begins with a start point. Locating a start point can be done by scanning each row of pixels from the bottom left corner. The scan of each row starts from the leftmost pixel proceeding to the right. When a white pixel is encountered, declare that pixel as start pixel. After deciding the start point, there are 3 pixels which we are interested in. That will be P1, P2, P3 shown in Figure 3.11; P1 is the upper-left pixel with respect to the start. P2 is the upper pixel with respect to the start. And P3 is the upper-right pixel with respect to the start.

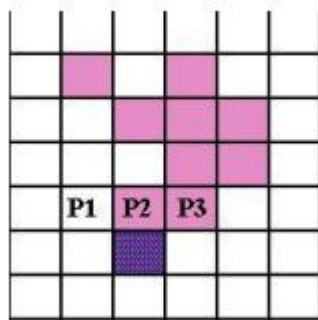


Figure 3.11 Position of p1, p2, p3 relative to start point

When locating the start pixel, there are four cases. The first thing to do is to check P1. If P1 is a white pixel, declare P1 to be your next start pixel and move one step forward followed by one step to your current left to land on P1 as shown in Figure 3.7. The next P1, P2 and P3 will be changed according to the location and change of orientation.

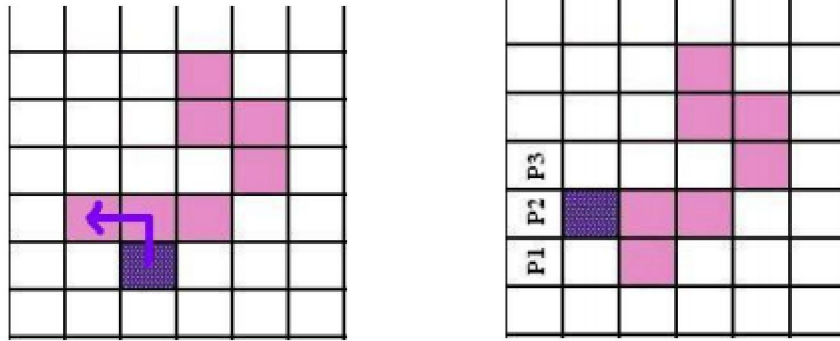


Figure 3.12 Move to p1 and change direction

Only if P1 is a black pixel, then we check P2. If P2 is a white pixel, we declare P2 to be the next start pixel and move one step forward to land on P2 as shown in Figure 3.13. The next P1, P2 and P3 will be changed according to the location and change of orientation.

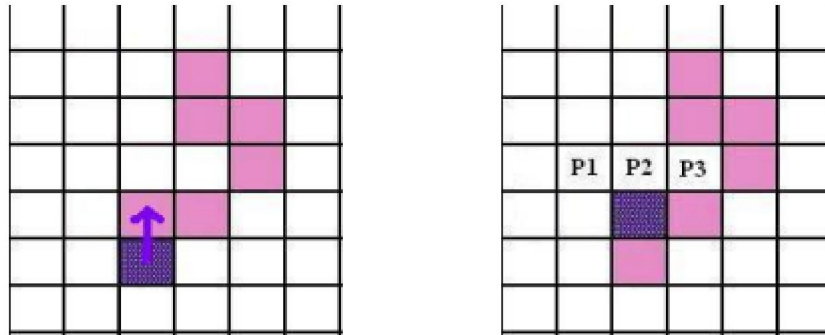


Figure 3.13 Move to p2

If P1 is black pixel and P2 as well, we check P3. If P3 is a white pixel, we declare P3 to be the next start pixel and move one step to your right followed by one step to your current left to land on P2 as shown in Figure 3.14. The next P1, P2 and P3 will be changed according to the location and change of orientation.

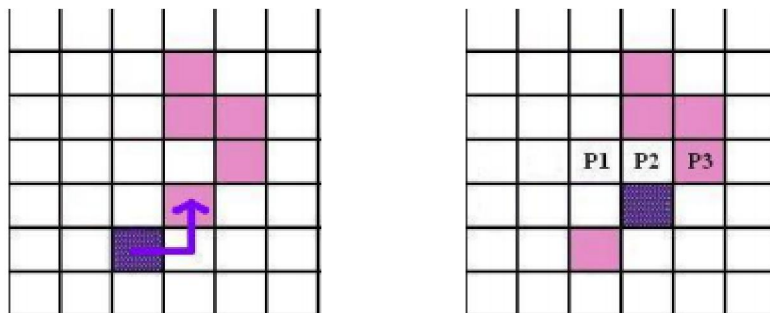


Figure 3.14 Move to P3

If none of P1, P2, P3 are white pixel, the orientation will be rotated 90 degrees clockwise while standing on the same pixel as shown in Figure 3.15. Afterwards we do the same procedure all over again.

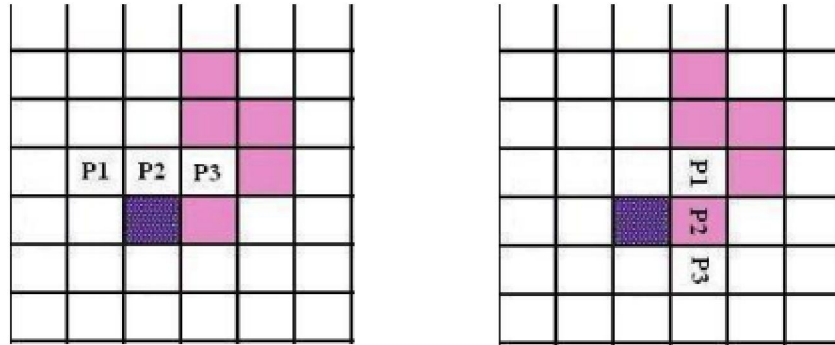


Figure 3.15 Rotate when p1, p2 and p3 are none

If the orientation has been rotated 3 times continuously without finding any white pixel in P1, P2 or P3, it means that we are locating on an isolated pixel. The pixel is not connected to any white pixels. This situation will cause the algorithm to terminate as shown in Figure 3.16

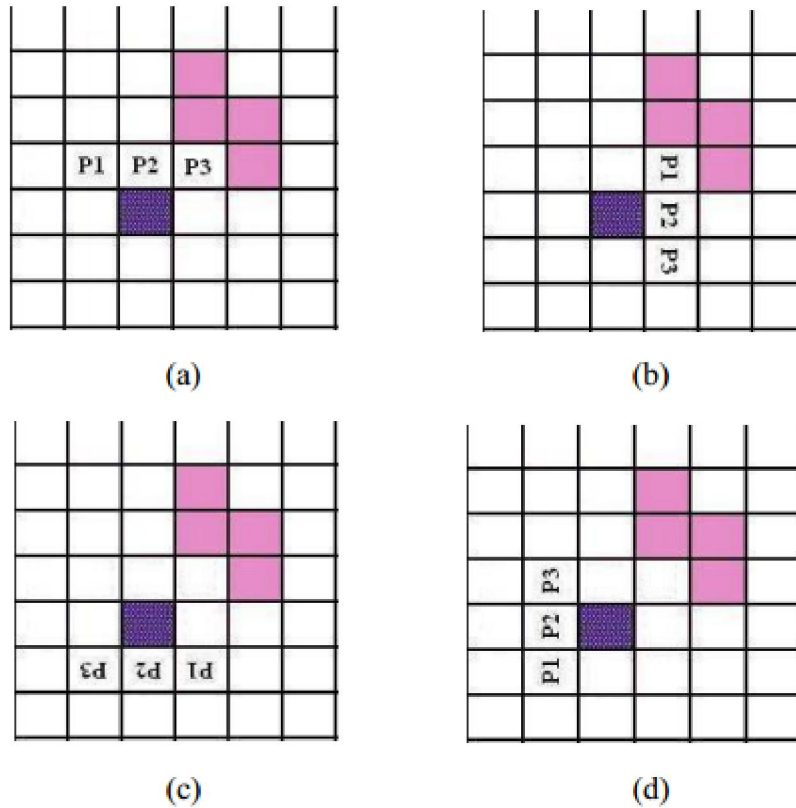


Figure 3.16 (a) p1, p2, p3 are none (b) first rotation (c) second rotation (d) third rotation

Formal Description of Pavlidis' Algorithm
<p>Input: A binary image containing connected group of white pixel</p> <p>Output: A sequence B (b_1, b_2, \dots, b_k) of boundary pixels as contour.</p>
<p>Begin</p> <ul style="list-style-type: none"> • Scan each pixel in rows from the bottom-left until you find a white pixel s • Insert s in B and set s to be the starting pixel • While (s haven't been visited twice) <ul style="list-style-type: none"> If pixel P1 is white <ul style="list-style-type: none"> ○ Insert P1 in B ○ Set s=P1 ○ Move one step forward followed by one step to your current left else if P2 is white <ul style="list-style-type: none"> ○ Insert P2 in B ○ Set s=P2 ○ Move one step forward else if P3 is white <ul style="list-style-type: none"> ○ Insert P3 in B ○ Set s=P3 ○ Move one step to the right, update your position and move one step to your current left else if s has been rotated 3 times <ul style="list-style-type: none"> ○ terminate the process and declare s as an isolated pixel else <ul style="list-style-type: none"> ○ rotate 90 degrees clockwise while standing on the current pixel p <p>End</p>

Table3.1 Pavlidis' Algorithm

When the search is finished, couple of contours could be found in the image. After doing some experiments and determining a background of a single color different from the color of the hand and depending on how the ASPC works (user must be on the left of screen) , we extracted the largest tow contours (Hand and head with a high probability). If the coordinate value of the largest contour center on the X axis is smaller than the one of the 2nd largest contour center, we choose the 2nd one as a hand. Else we choose the 1st one(as a hand). The algorithm used to extract the most largest tow contours is explained below, figure.3.18

```

for each list Of contours AS contour
  if contour > secondHighestContour
    if contour > highestContour
      secondHighestContour = highestContour
      highestContour = contour
    else
      secondHighestContour = contour

```

Table3.2 Extracting of largest tow contours

The results explained in figure.3.17.

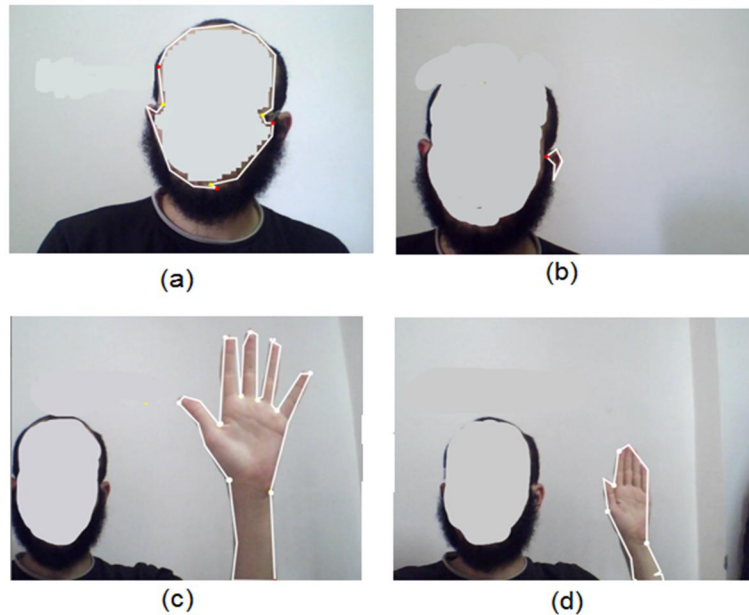


Figure 3.17 Hand contour tests

(a) face is the largest contour, (b) The right ear is the second largest contour (c) the right hand is the largest contour (d) the right hand is the second largest contour

3.2.2 Convex Hull

Finding a convex hull is a well-known problem in computation geometry. Let us imagine that there are several nails on the wall, a rubber band is used to surround those nails. Only the peripheries will touch the rubber band while the nails inside won't be able to affect the rubber band. The shape of the rubber band is probably how the convex hull of the nails will be looked like. It is obviously not difficult for a human to understand the shape of convex hull just in a glance. But as for machines, an algorithm is needed.

We calculate the convex hull of the fore-arm contour in order to find the desired information. Usually the arm part has a smooth contour and doesn't contain any important information. The hand part has more convex and concave contours and it usually contains the information we want. After the comparison of a fore-arm contour and its convex hull, we find out that the convexity defects are around the palm area. Hence we might be able to find the points which have the longest distance to the convex hull in each convexity defect. Those points are on the edge of hand click section since the convexity defects are around the palm area. Since the fore-arm is relatively smooth, so the neighbor defect usually won't create a point which has the longest distance to the convex hull on the arm contour. By these points, we can determine the hand click

The following section will introduce the convex polygon, and the 3-coin algorithm which is applied to find the convex hull of a set of points. After the convexity defects are found, we will discuss about the features we can find through it.

3.2.2.1 Convex Polygon

Before we begin to explain convex hull, the concept of convex polygon needs to be introduced first. A convex polygon is a polygon which does not contain any concave part. In Figure 3.18, the arrow is pointing to a concave part of a polygon. Although this character means "convex" in Chinese words, the character itself is not a convex polygon.

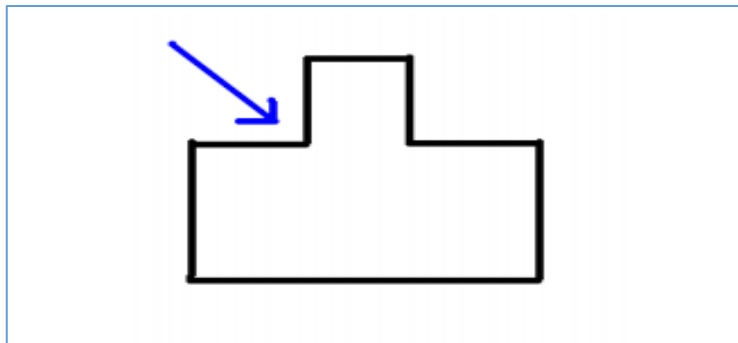


Figure 3.18 Not a convex polygon

In mathematics, there is a strictly definition of a convex polygon. Assume that we can choose any two points in a polygon (includes the boundaries and the area which the boundaries are covering) and connect the two points together within a straight line. If all the straight lines that any two points inside the polygon can form doesn't exceed the boundary of the polygon, we can say that the polygon is a convex polygon. Figure 3.19 shows that we can find two points (A and B) and connect them within a straight line. We can obviously see that part of the line (which is represented in red color) exceed the boundary, so we can say that the polygon is not a convex polygon.

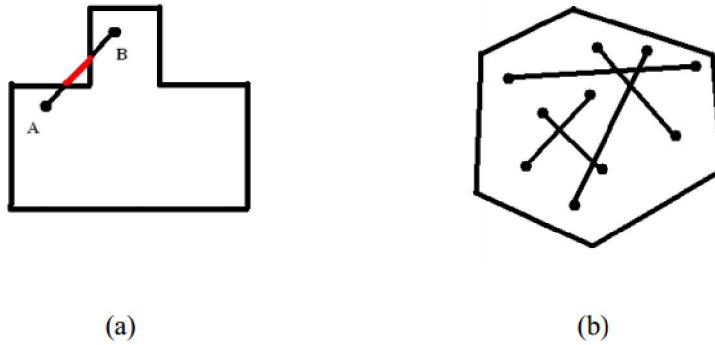


Figure 3.19 An example of a convex polygon, all the connected line of two points will never exceed the boundary. (a) Not a convex polygon (b) Convex polygon

3.2.2.2 Definition of Convex Hull

Once we know the definition of a convex polygon, we might be able to learn the concept of convex hull. For a given nonempty set of points in a plane, the convex hull of the set is the smallest convex polygon which covers all the points in the set. For example, in Figure 3.20 there are 10 points. The hexagon in the figure is the convex hull of the set. The six points which compose the hexagon are called “hull points”.

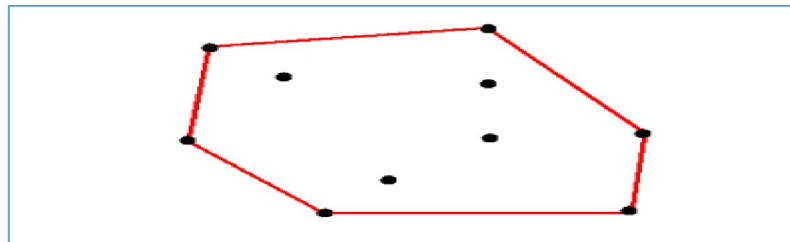


Figure 3.20 An example of convex hull

We know that a set S in a plane or in space is a convex polygon (or convex set) if and only if whenever points X and Y are in S , the line segment XY must be contained in S . The intersection of any collection of convex sets is also convex, as shown in Figure 3.21. For an arbitrary set W , the convex hull of W is the intersection of all convex sets containing W . The boundary of the convex hull is a polygon with all the vertices in W .

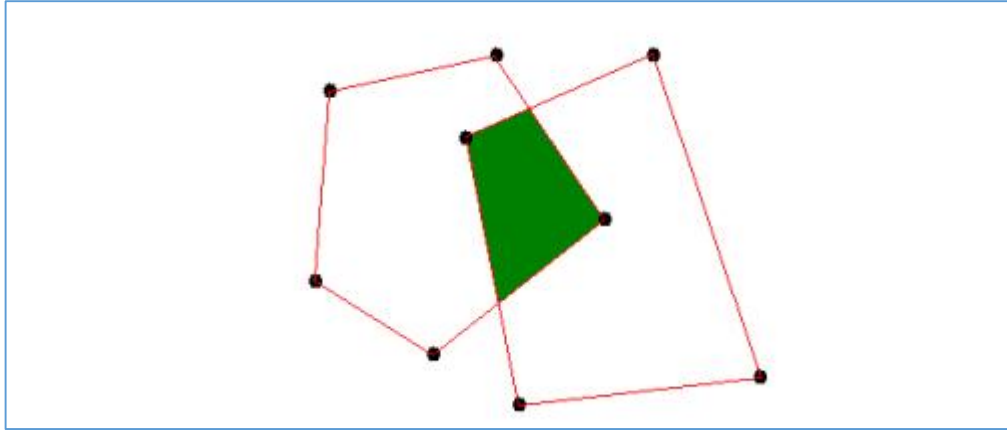


Figure 3.21 Any intersection of convex set is also a convex set.

3.2.2.3 Three-Coin Algorithm

The three-coin algorithm was developed by Graham and Sklansky [42,], trying to find the convex hull of a given set of points. To understand the algorithm, there's one thing we need to realize. Suppose there is a convex polygon, if we take a walk along the polygon's edge, we would only make right turns (or left turns). If it is not a convex polygon, a turn in an opposite direction will be expected.

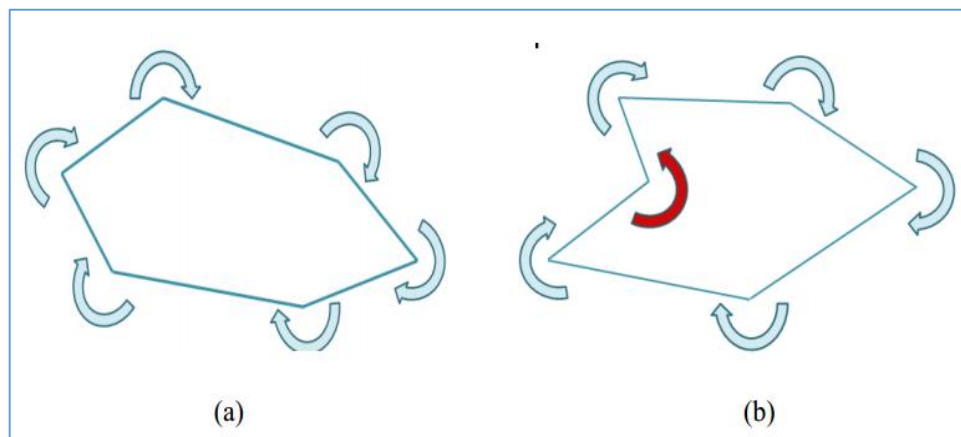


Figure 3.22 (a) convex polygon (b) Not a convex polygon

Graham's algorithm is slightly different than Sklansky's algorithm. In Graham's algorithm, the input will be a set of points in a plane. Sklansky's algorithm was designed for a simple polygon in a plane. Simple polygon means the polygon which doesn't have any intersected edges. Graham's algorithm needs to do a sort for all the points as show in figure 3.23. The sort gives number to each point, and the algorithm processes according to the order. Sklansky assumes the input to be a polygon. Once the starting point is decided, the algorithm follows the edges to process, which is just like following the orders. Since the fore-arm contour we found is a series of points, all of the points will be stored in a sequence structure, we obviously have the order of all the points. So we can skip the Graham's sorting. Both algorithms are the same except the sorting.

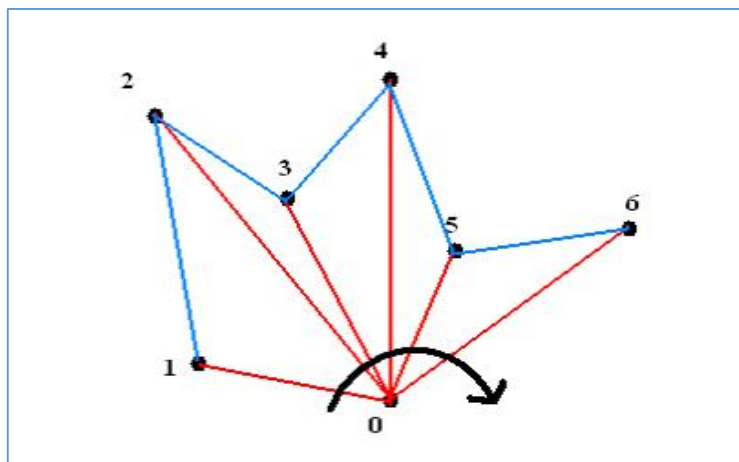


Figure 3.23 Graham's sorting

The fore-arm contour will be the input for the Sklansky's algorithm. The three-coin algorithm means that 3 points will be labeled each time. They will be represented in different colors (for example, black, red and grey). The following figures will illustrated the algorithm. Let's take the polygon in figure 3.24 for example.

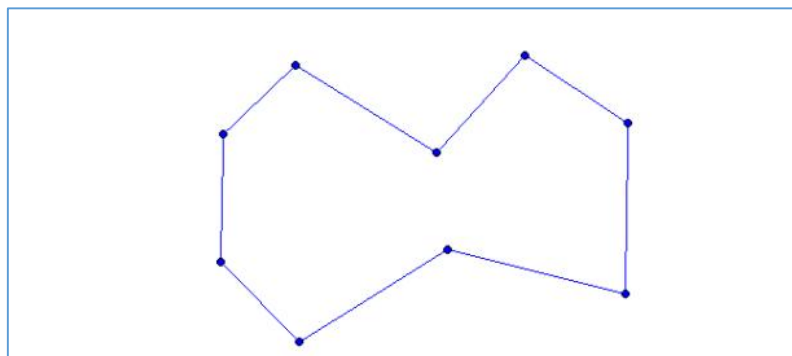


Figure 3.24 An example of input points

First of all, we choose a point to be the starting point, mark this point as a black coin. The starting point must be a convex vertex. We can choose the point which is the left most. The point after black point will be marked as a red coin. The point after the red coin will be marked as grey point. We also called them end coin, middle coin and front coin according to their order and regardless of their color.

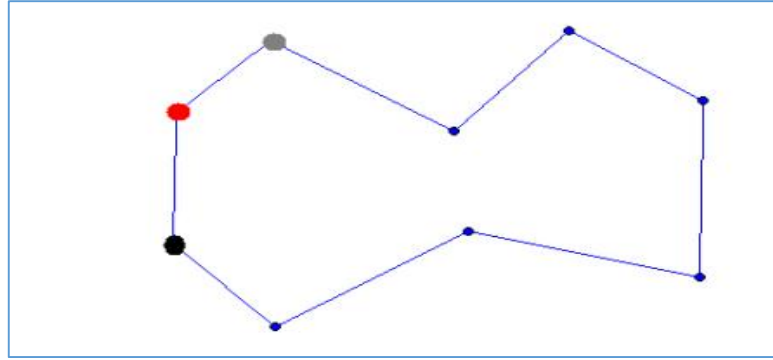


Figure 3.25 Place three coins

Let's check the path from the end coin (currently the black coin) to the middle coin (currently the red coin), then arrive the front coin (currently the grey coin). The path forms a right turn. Whenever encounter a right turn, move the coin at the end coin to the point next to the front point. Therefore, we have the black coin as front coin while the grey coin is the middle coin and red coin turns out to be the end coin.

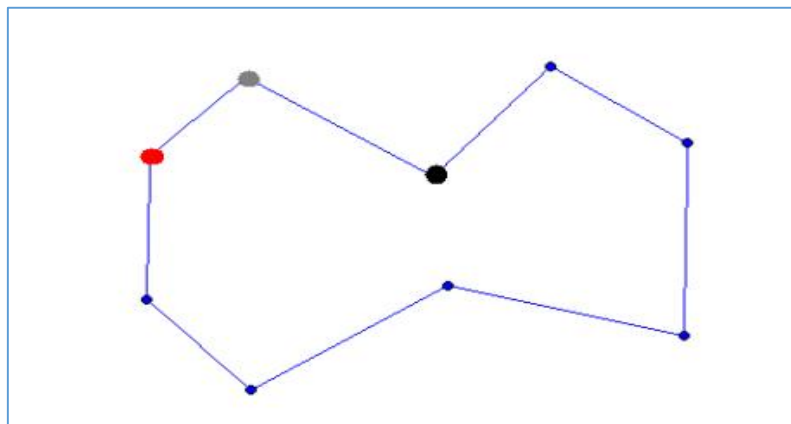


Figure 3.26 Move the end coin next to front coin

Once again, we check the path of the three coins. It forms a right turn again. So we move the end coin to the point next to the front coin. The front coin is now the red coin while the middle is the black point and the grey point become the end coin.

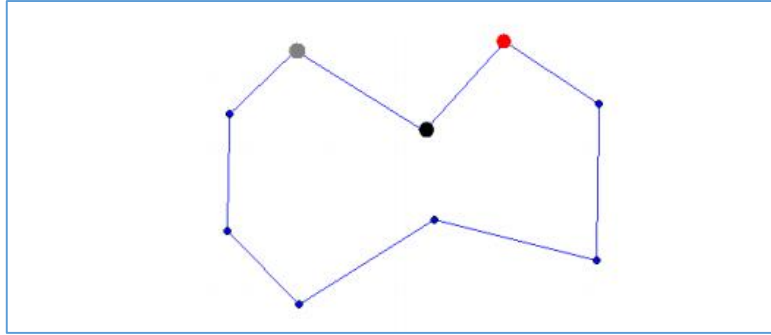


Figure 3.27 Move end coin next to front coin

We check the path of the three coins, this time it forms a left turn. Whenever we encounter a left turn, we should delete the point where the current middle coin stands on, and then move the middle coin to the point before the end point (currently the grey point). Right now, our front coin remains the same (the red coin). But the middle coin has been changed to the grey coin and the black coin is the end coin.

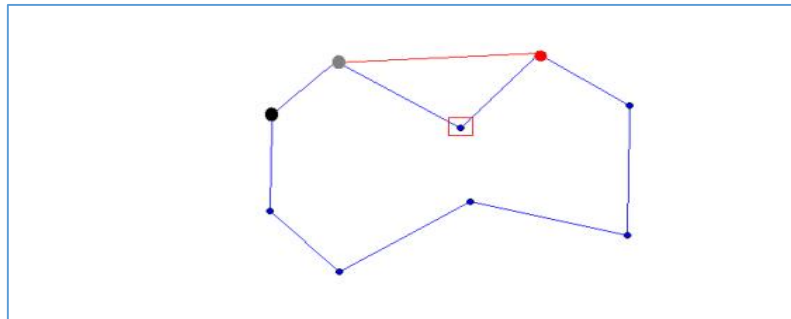


Figure 3.28 Remove middle point

Let's check the path again. It's a right turn this time. So we move the end coin to the point next to the front coin.

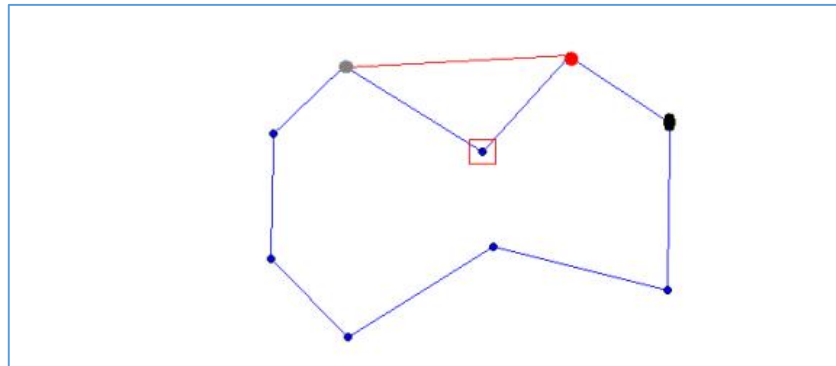


Figure 3.29 move end point next to front point

It forms a right turn again. So we move the end coin to the point next to the front coin. Previous step still forms a right turn, so we do the same procedure again. Still makes a right turn, so we do the same procedure.

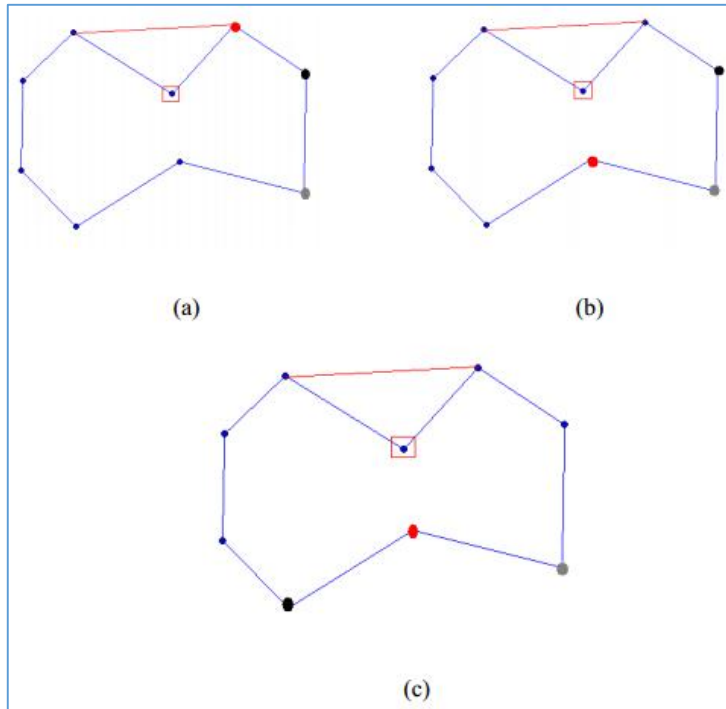


Figure 3.30 (a-c) move end point next to front point

Finally, the coins form a left turn. We delete the point where the middle coin stands on right now (currently the red coin). In addition, we move the middle coin to the point before the end coin (currently the grey coin).

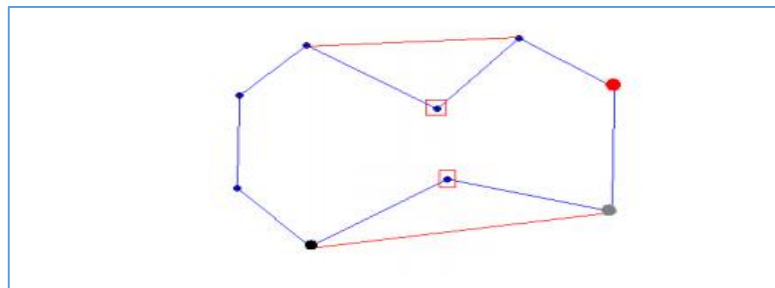


Figure 3.31 Remove middle point

It forms a right turn. As we move the end coin to the point next to the front coin, we find out that it comes back to the starting point. Therefore, the whole procedure will be terminated.

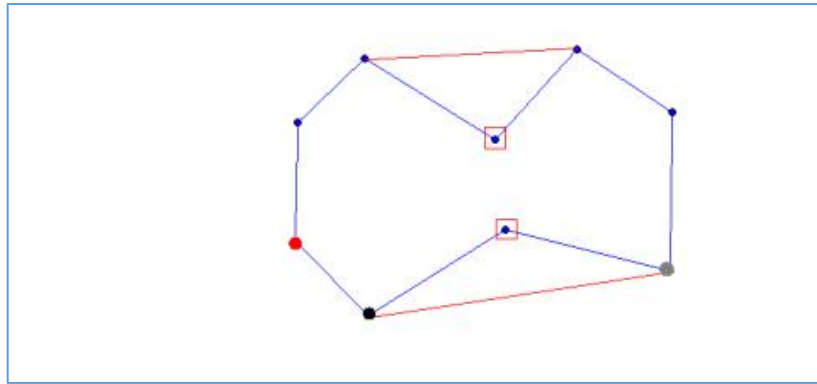


Figure 3.32 Process terminates when visit start point again and make a right turn

Link all the remained points. We got the convex hull of the set of points as shown in figure 3.33.

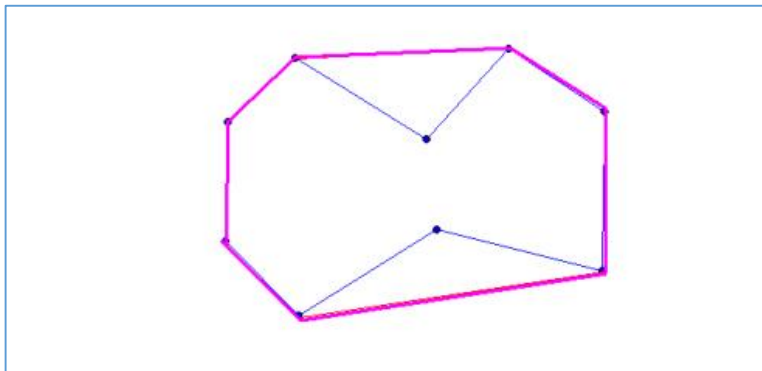


Figure 3.33 Convex hull of the input points

We can successfully generate the convex hull of the fore-arm contour by applying the three-coin algorithm as shown in figure 3.34. By observing the contours and its convex hull, we notice that the areas which the convex hull contains but is not included in the contour are called the convexity defects. Convexity defects give us useful information about the shape of a contour. In the case of fore-arm contour, we find out that the convexity defects are often being around the palm. We will discuss about convexity defect in the next section

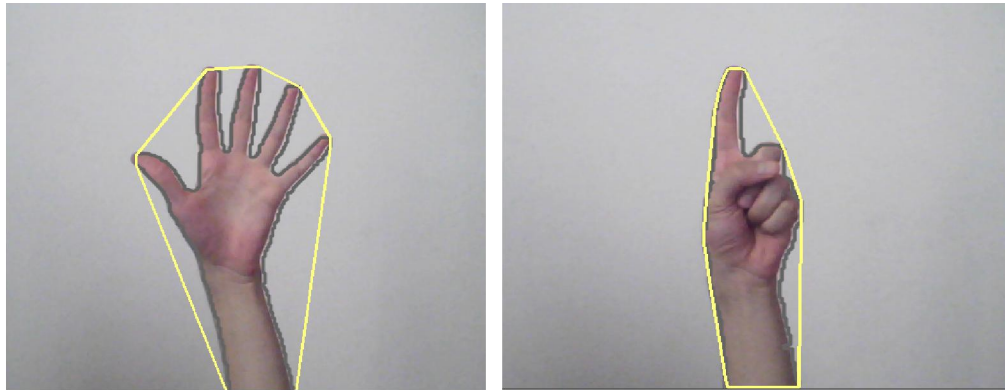


Figure 3.34 Convex Hull of hand contour

Formal Description of Three-Coin Algorithm	
Input: A set of N points which form a polygon in a plane.	
Output: A set of points which form the convex hull of the input.	
Begin	
●	Find the left most point as a starting point. Label it p0.
●	Label the remained N-1 points in a clockwise order.
●	Place three coins on p0, p1, p2 and label them "end coin", "middle coin", "front coin" respectively.
Do Until:	the "front coin" is on the starting point and they form a right turn
If	the 3 coins form a right turn (or the 3 coins lie on collinear vertices)
●	Take "end coin", put it on the point next to "front point".
●	"end coin" become "front coin", "front coin" become "middle coin", "middle coin" become "end coin".
Else	the coins form a left turn
●	Remove the point (and associated edges) that "middle coin" is on
●	Take "middle coin", put it on the point before "end point"
●	"middle coin" become "end coin", "end coin" become "middle coin"
End	
The remained points form a convex hull when connecting them in order.	

Table3.3 Three-Coin algorithm

3.2.3 Convexity Defect

We calculate the convex hull of the fore-arm contour in order to get the convexity defect of the contour. Convexity defect provides us very useful information to understand the shape of a contour. Many characteristic of complicated contours can be represented by convexity defects. Figure 3.35

illustrates the convexity defect of a star-like shape. The green lines represent the convex hull of the star-like shape. As you can see in the figure, the areas in yellow are contained in the convex hull. But they are not contained in the star. Those areas are so called convexity defects.



Figure 3.35 Yellow parts represent the convexity defects

The previous section tells us that the points which form the convex hull must be part of the contour. Our first step of searching a convexity defect will be finding the starting point of a convexity defect on the contour. The starting point of a convexity defect means a point on the contour which is also included in the convex hull points, but the next point on the contour is not included in the convex hull points. The figure 3.36 illustrates the starting point of a convexity defect. We search the contour clockwise. The red point is the first point which is included in the convex hull, but the next point is not included in the convex hull. We mark the point as a starting point.

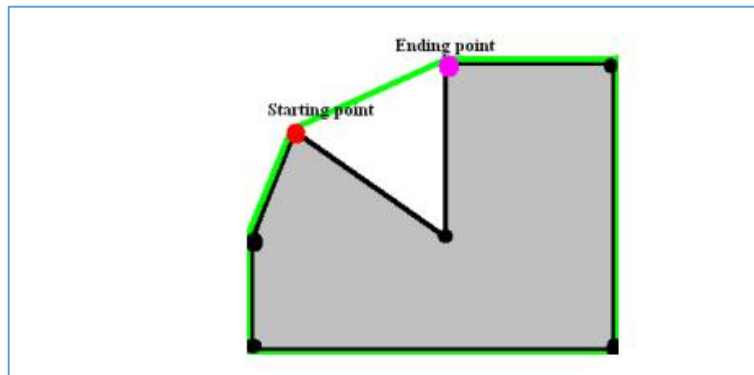


Figure 3.36 Starting and ending points of a convexity defect

Once we know about the starting point, the ending point will be similar. We define the ending point to be the point in the contour which is included in the convex hull points, but the point before it is not included in the convex hull points. As we can see in the figure 3.36, the purple point is the ending point of a convexity defect.

By connecting the starting point, ending point and the points on the contour between the starting point and the ending point, we get a convexity defect area as shown in figure 3.37.

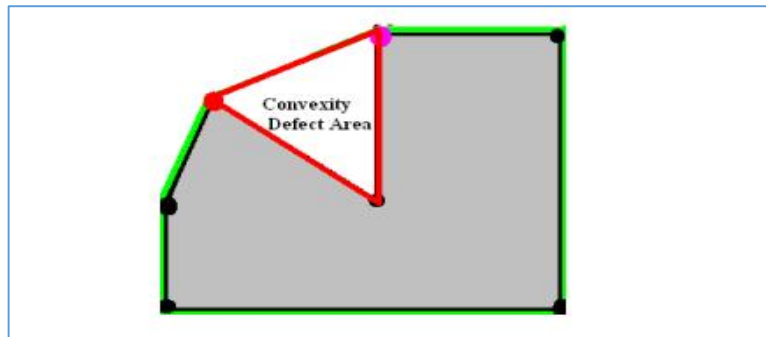


Figure 3.37 Convexity defect area

When all the points in the contour have been searched, we may find various convexity defects as shown in figure 3.38. Each convexity defect is composed by a starting point, ending point, and the points between them.

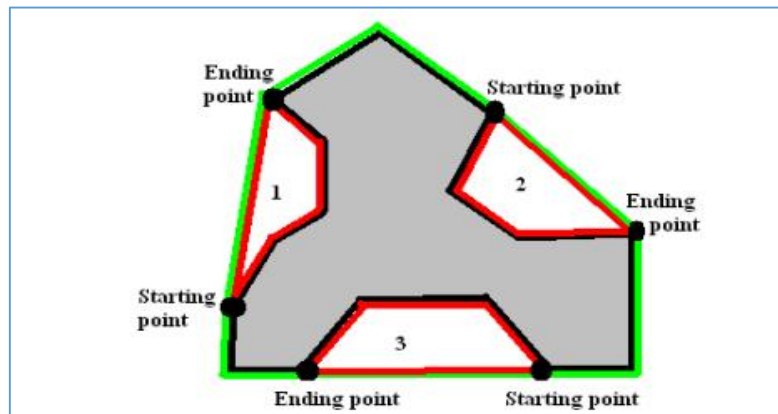


Figure 3.38 there are three convexity defects in the figure

Except the starting point and ending point, other useful information we can obtain from a convexity defect will be the depth of the defect and the depth point. The depth of the defect is the longest distance of all points in the defect to the convex hull edge of the defect. The point in the defect which has the longest distance to the convex hull edge of the defect will be the depth point.

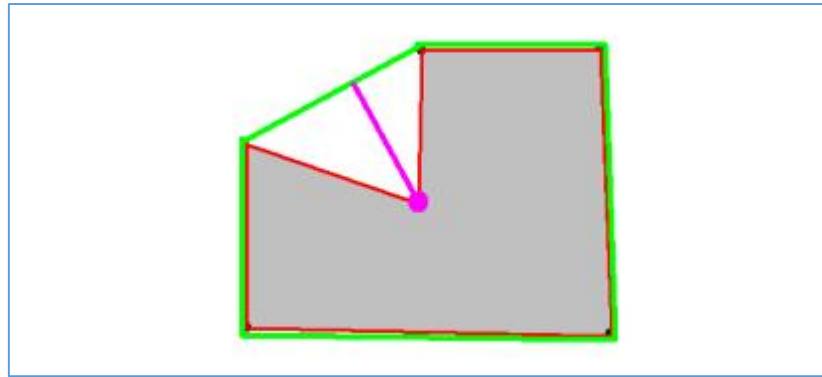


Figure 3.39 The depth point of a convexity defect

As shown in figure 3.39, there is only one convexity defect. The purple point is obviously the point with the longest distance to the defect's convex hull edge. And the length of the purple line is the depth of the purple point. Since the depth of the purple is the longest, it will also be the depth of the convexity defect. And the purple point is the depth point. Figure 3.40 shows all the depth points of a fore-arm contour.

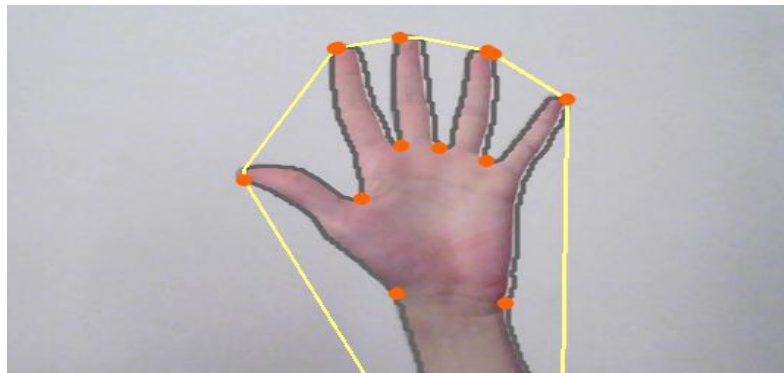


Figure 3.40 All depth points of a hand contour

The depth points are represented in yellow. Since there are 12 depth points in the figure, we know that there are 12 convexity defects in the fore-arm contour. Only six out of 12 convexity defects can be seen with human eyes while the rest of them are too small to be seen. Since the fingertips connect the convex hull edge, it would be easily to create a very small convexity defect if there is a concave contour. We should regardless such convexity defects and take the convexity defects with large area as what we want. The large convexity defects also have long depth. So we need to check the depth of all the convexity defects. If the depth is longer than a certain threshold, we draw the depth point of it. If it's not longer than a certain threshold, the defect will be ignored.

Figure 3.41 shows the result when the small convexity defects are ignored and only the depth points of large convexity defects will be shown.



Figure 3.41 Remove depth points which have small depth

An open hand gesture is a perfect example to show that the depth points of large convexity defects are around the hand palm. But it also works very well in other gestures. Figure 3.42 shows the distribution of depth points in different hand gestures.

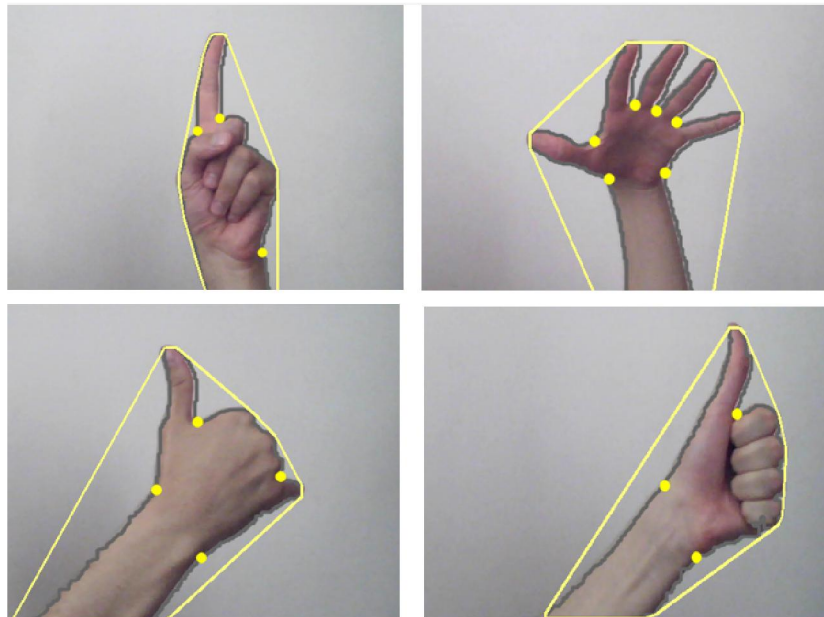


Figure 3.42 Examples of depth points of hand contour

We usually have a depth point in the wrist part of hand. That enables us to limit events (click and drawing) in hand. So even another object achieves the condition of clicking, the reaction of click can't happen until the program checks (comparison between depth points in each contour) if this action came from hand, for example the figure 3.43 shows the big difference between hand and head depth points. Figure 3.42 shows an example of head depth points.



Figure 3.43 Head depth points

3.3 Hand click detection (HCI part)

Based on previews techniques used to extract the hand from environment, we can now include other geometrics techniques to get a high estimation for hand click as shown in figure 3.7.

3.3.1 Centroid

In mathematics and physics, the centroid or geometric center of a two-dimensional region is, informally, the point at which a cardboard cut-out of the region could be perfectly balanced on the tip of a pencil, assuming uniform density and a uniform gravitational field. Formally, the centroid of a plane figure or two-dimensional shape is the arithmetic mean ("average") position of all the points in the shape. The definition extends to any object in n-dimensional space: its centroid is the mean position of all the points in all of the coordinate directions.

While in geometry the term barycenter is a synonym for "centroid", in physics "barycenter" may also mean the physical center of mass or the center of gravity, depending on the context. The center of mass (and center of gravity in a uniform gravitational field) is the arithmetic mean of all points weighted by the local density or specific weight. If a physical object has uniform density, then its center of mass is the same as the centroid of its shape.

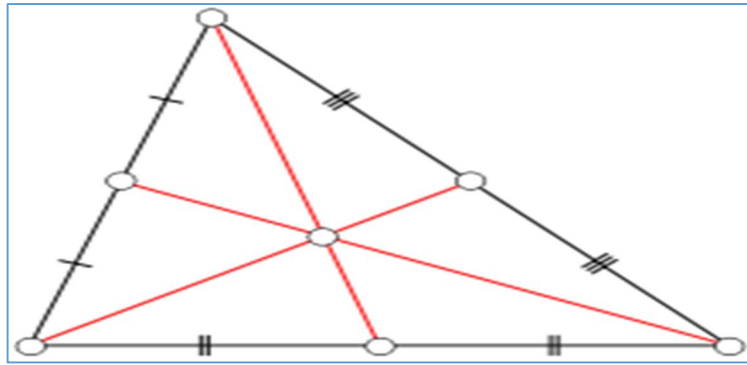


Figure 3.44 Centroid of a triangle

3.3.1.2 Calculate the centroid based on moment of inertia

The moment of a mass is a measure of its tendency to rotate about a point. Clearly, the greater the mass (and the greater the distance from the point), the greater will be the tendency to rotate.

We used Green's theorem to calculate the moment of inertia of the hand contour [46].

Finding of hand centroid using the moment of inertia is similar to centroid of complex shapes; we divide the complex shape into rectangles and find \bar{x} (the x-coordinate of the centroid) and \bar{y} (the y-coordinate of the centroid) by taking moments about the y- and x-coordinates respectively.

Because hand in binary image is thin plates with a uniform density, we can just calculate moments using the area [47]

The general formula of the equation is as shown below.

$$\bar{x} = \frac{\text{total moments in } x\text{-direction}}{\text{total area}} \quad (3.4)$$

$$\bar{y} = \frac{\text{total moments in } y\text{-direction}}{\text{total area}} \quad (3.5)$$

The result of calculation of centroid is as shown below.

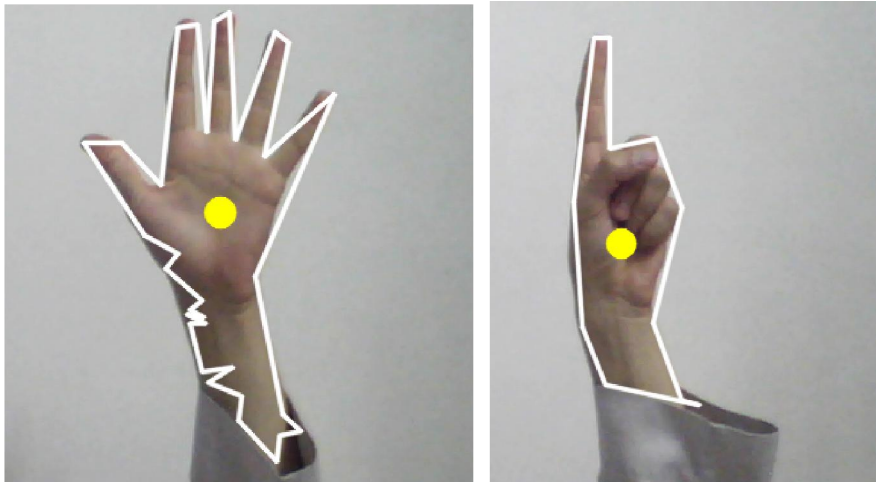


Figure 3.45 Centroid of hand

3.3.2 Hand click detection

Based on some properties of centroid we can detect the hand click with a high estimation.

As is well known, the centroid point of a complex shape will be closer to the simple constituent shape, which has the biggest area.

Because of this, the farthest point in the shape contour belongs to the smallest shape contour (assuming that shape is composed of two parts with low complexity).

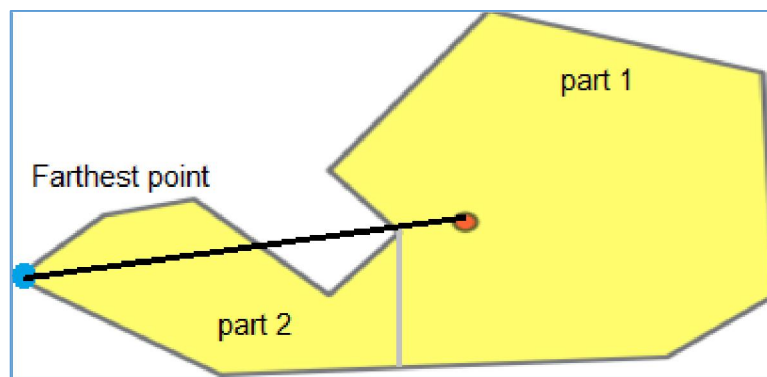


Figure 3.46 Centroid of complex shape

In our project, we can divide the hand contour into two parts, first presents in hand and second in forearm. The farthest point in hand contour from the centroid belongs to the forearm which presents always the smallest part in hand when hand is opened. But when hand points with index finger, the centroid will be closer to the forearm and the farthest point will be located directly on top of the index finger and this is what we are looking for.

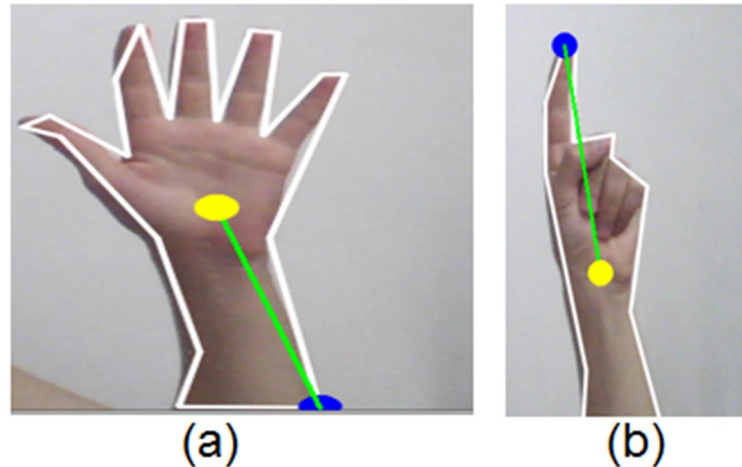


Figure 3.47 Farthest point in hand contour. (a) in opened hand (b) in pointed hand

In addition to the depth points calculated, we can detect the hand click easily.

3.4 Virtual interface display (AR part)

The displaying of virtual interface on screen is the last part from ASPC construction which presents the AR part, all of virtual components have the same idea.

The most hard work is finished by the detection of hand click and the rest base on simple principles in programing such as selecting parts on the screen, adding images and so on, so we will suffice by mentioning of different virtual components.

3.4.1 Virtual buttons

We have four virtual buttons used to interact with ASPC.

Next button for moving to the next slide.

Preview button for moving to the preview slide.

White board button for displaying of a virtual white board help presenter in explaining.

Marker button for drawing on the slide and white board.

3.4.2 Virtual slide show

It the place where the presenter display his sequence of slides and manage them using virtual button

3.4.3 Virtual white board

It is just a white image placed instead of slides where presenter can draws and writes

4 Conclusion

In this chapter, we focused on the technical part of the project and we explained the different steps used to develop ASPC.

In each step, we gave a depth information about theoretical algorithm that was the most is a geometrics rule. Moreover, how we use them in accordance with the requirements of the program.

CHAPTER 4

IMPLEMENTATION AND TEST

1 Introduction

In this final chapter, we will focus on ASPC components and give a brief overview on development environment and hardware architecture used to build ASPC.

2 Development environment

2.1 Operating system

We used the Microsoft Windows 7, it has the advantage of being the most used OS, it is easy to use, high efficiency and HCI.

2.2 Microsoft Visual Studio 2012

Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs for Microsoft Windows superfamily of operating systems, as well as web sites, web applications and web services. Visual Studio uses Microsoft software development platforms such as Windows API, Windows Forms, Windows Presentation Foundation, Windows Store and Microsoft Silverlight. It can produce both native code and managed code.

Microsoft Visual C++ is the platform which we used, it is a commercial (free version available), integrated development environment (IDE) product from Microsoft for the C, C++, and C++/CLI programming languages. It features tools for developing and debugging C++ code, especially code written for the Microsoft Windows API, the DirectX API, and the Microsoft .NET Framework.

Many applications require redistributable Visual C++ packages to function correctly. These packages are often installed independently of applications, allowing multiple applications to make use of the package while only having to install it once. These Visual C++ redistributable and runtime packages are mostly installed for standard libraries that many applications use.

2.3 OpenCV

OpenCV (Open Source Computer Vision) is a library of programming functions mainly aimed at real-time computer vision, developed by Intel Russia research center in Nizhny Novgorod, and now supported by Willow Garage and Itseez. It is free for use under the open source BSD license. The library is cross-platform. It focuses mainly on real-time image processing. If the library finds

Intel's Integrated Performance Primitives on the system, it will use these proprietary optimized routines to accelerate itself.

OpenCV's application areas include:

- 2D and 3D feature toolkits
- Egomotion estimation
- Facial recognition system
- Gesture recognition
- Human–computer interaction (HCI)
- Mobile robotics
- Motion understanding
- Object identification
- Segmentation and Recognition
- Stereopsis Stereo vision: depth perception from 2 cameras
- Structure from motion (SFM)
- Motion tracking
- Augmented reality

OpenCV is written in C++ and its primary interface is in C++, but it still retains a less comprehensive though extensive older C interface. There are now full interfaces in Python, Java and MATLAB/OCTAVE (as of version 2.5). The API for these interfaces can be found in the online documentation [48]. Wrappers in other languages such as C#, Ruby have been developed to encourage adoption by a wider audience.

2.3.1 Installing and Configuring with Visual Studio 2012

The most suitable way is the installation by using the pre-built libraries.

2.3.1.1 Set up Environmental Variable

- Right click on 'My Computer' and click 'Properties' in the drop down menu. Our System type is 32 bit operating system. So the system architecture is x86.
- Right click on 'My Computer' and click 'Properties' in the drop down menu.

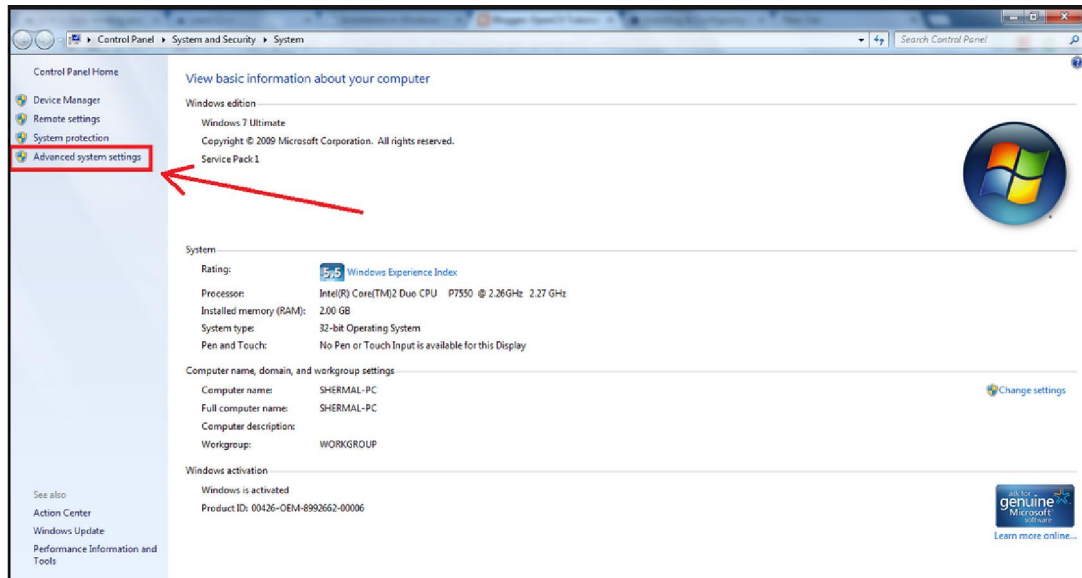


Figure 4.1 1st step in setting up Environmental Variable

- We click 'Advance System Settings' in the above window
- Then we click 'Environment Variables'

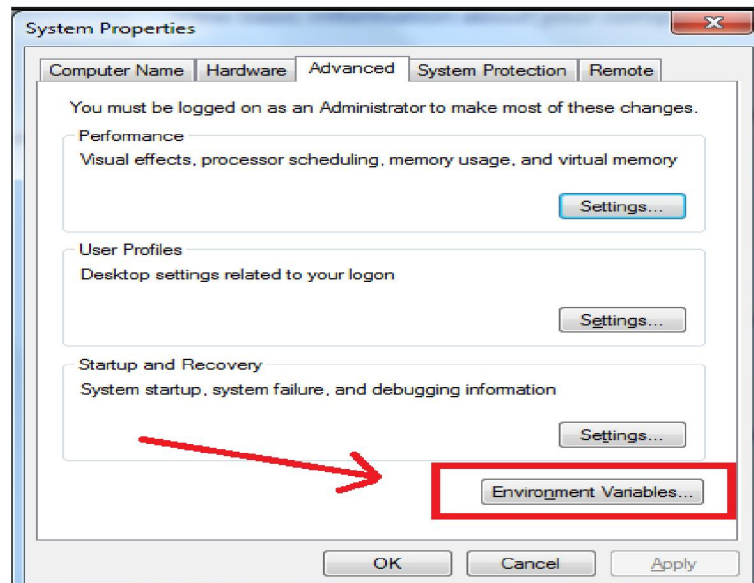


Figure 4.2 2nd step in setting up Environmental Variable

- We click 'New' button at the bottom of the window

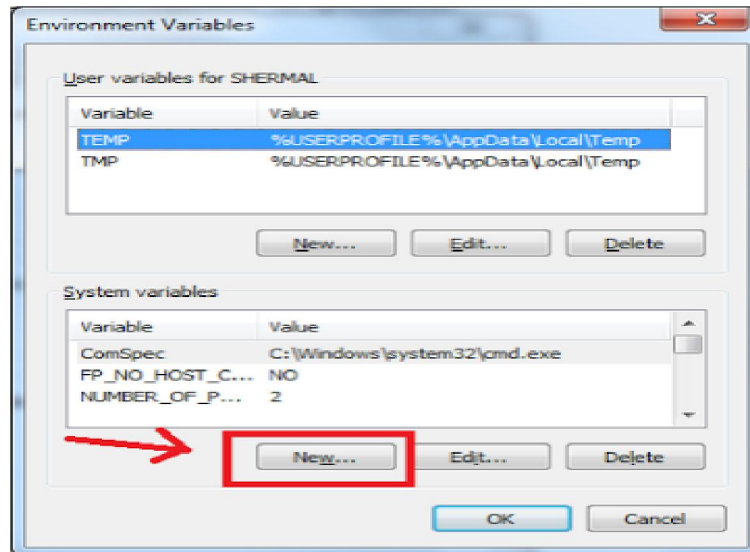


Figure 4.3 3rd step in setting up Environmental Variable

➤ We type OPENCV_DIR against Variable name:. And we type the location C:\opencv\build\ against the Variable value then we press ok:

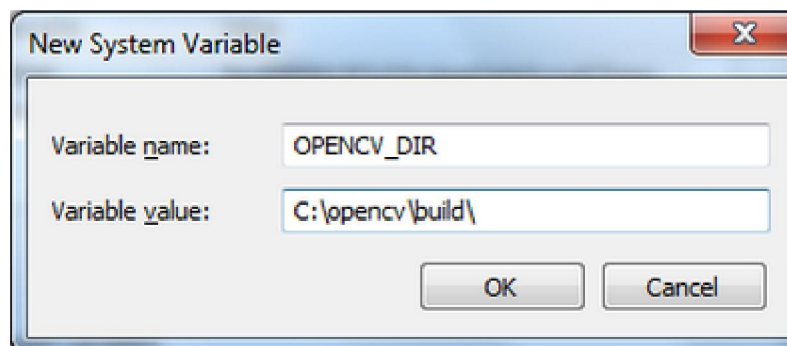


Figure 4.4 4th step in setting up Environmental Variable

Now we are going to edit a system variable. We click 'Path' or 'PATH' inside the list of System variables and then we click 'Edit' button at the bottom of the window Then we add a semicolon “;” to the end of the line and we add the following %OPENCV_DIR%\x86\vc11\bin after the semicolon. x86 is our system architecture and vc11 is the compiler type.

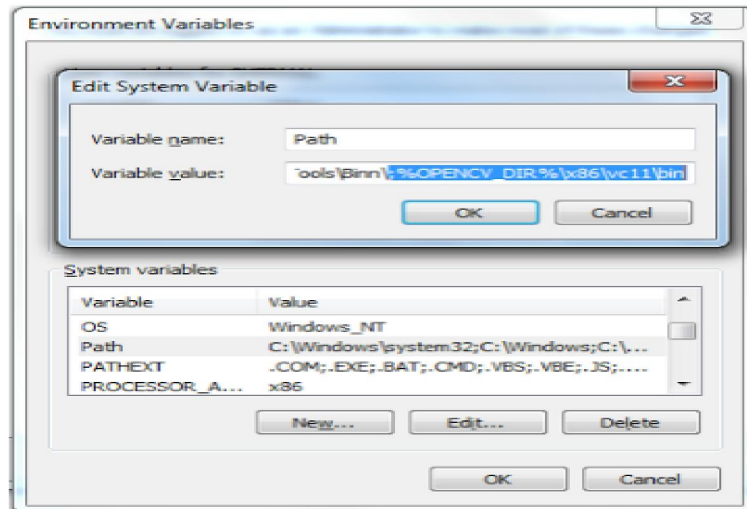


Figure 4.5 5th step in setting up Environmental Variable

2.3.1.2 Configure Visual Studio

- We start Microsoft Visual Studio
- We go to File>New Project...
- We do everything as shown in the below image and we click OK and then click Finish.

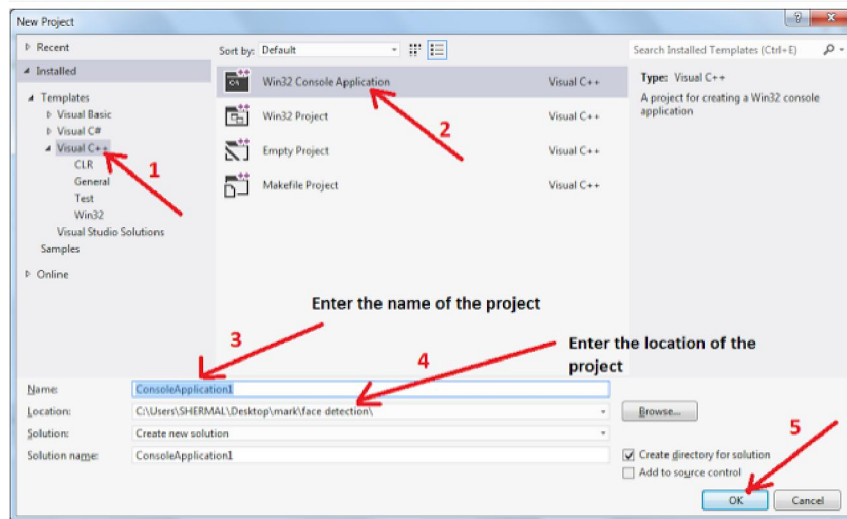


Figure 4.6 1st step in configuring visual studio

- We click properties in solution explorer of our project
- We choose “All Configurations” then we click the 3 places as pointed by next 3 boxes in the given order. Then we copy and paste `$(OPENCV_DIR)\include` against the “Additional Include Directories”.

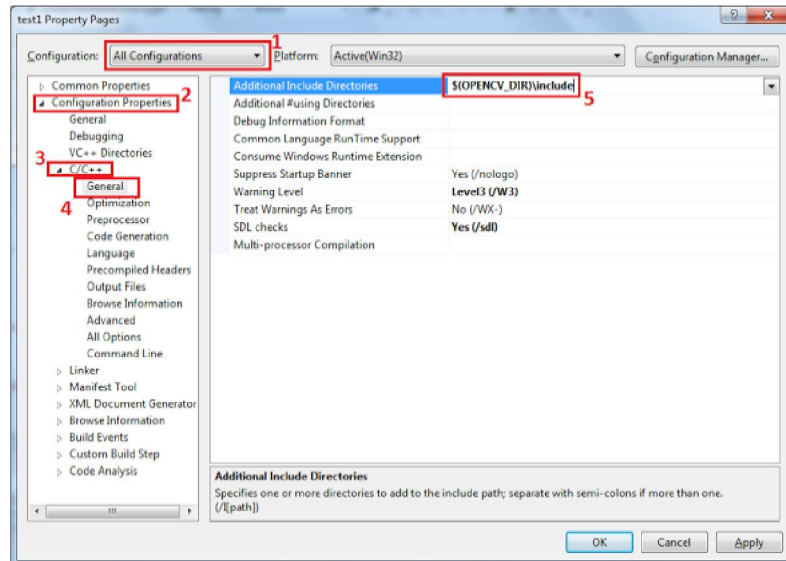


Figure 4.7 2nd step in configuring visual studio

We click the first 2 places as pointed by first 2 boxes in the given order. Then we copy and paste `$(OPENCV_DIR)\x86\vc11\lib` against the “Additional Library Directories”.

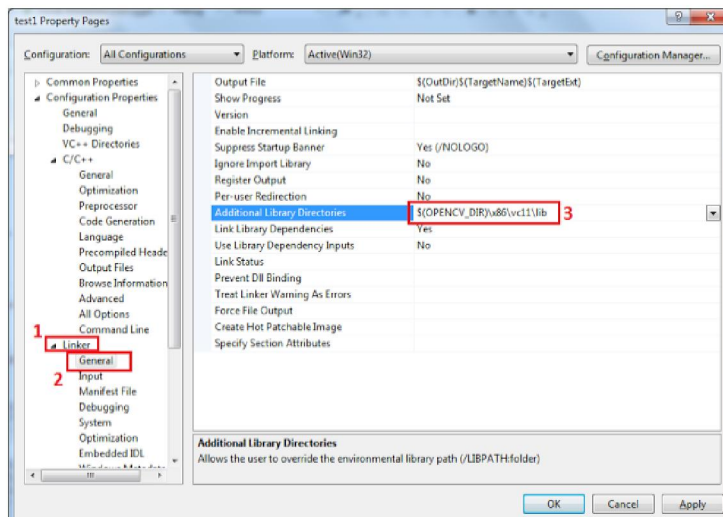


Figure 4.8 3^d step in configuring visual studio

➤ We click as illustrated below

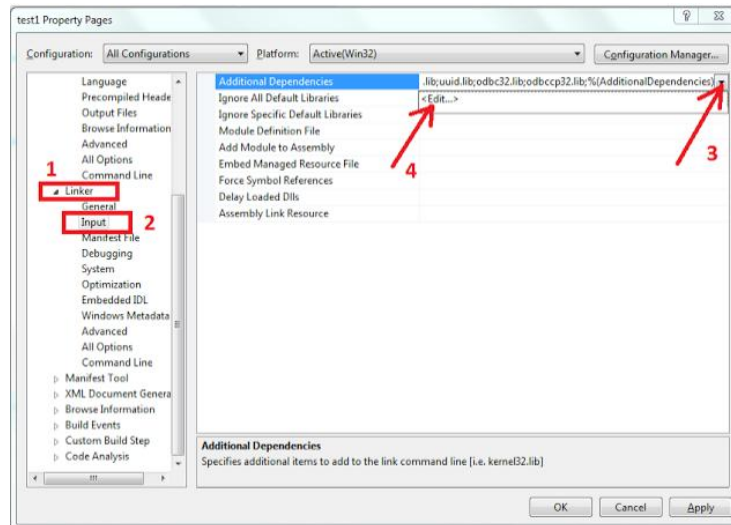


Figure 4.9 4th step in configuring visual studio

➤ After you we click the <Edit...> (4th arrow in the above image), dialog box will appear and we have to copy and paste following library file names.

opencv_calib3d245d.lib
 opencv_contrib245d.lib
 opencv_core245d.lib
 opencv_features2d245d.lib
 opencv_flann245d.lib
 opencv_gpu245d.lib
 opencv_haartraining_engined.lib
 opencv_highgui245d.lib
 opencv_imgproc245d.lib
 opencv_legacy245d.lib
 opencv_ml245d.lib
 opencv_nonfree245d.lib
 opencv_objdetect245d.lib
 opencv_photo245d.lib
 opencv_stitching245d.lib
 opencv_superres245d.lib
 opencv_ts245d.lib
 opencv_video245d.lib

opencv_videostab245d.lib

These are some of the file names, you can find in the “C:\opencv\build\x86\vc11” location.

We have done with all of the configurations of our IDE. Now we are ready to write any OpenCV application.

3 Hardware Architecture

3.1 Simple web camera

A webcam is a video camera that feeds or streams its image in real time to or through a computer to computer network. When "captured" by the computer, the video stream may be saved, viewed or sent on to other networks via systems such as the internet.

The webcam presents the live video input for ASPC.

3.2 Computer

The computation cost of our system is low since only computation geometry algorithms are performed; and there's no need to apply classifier and no pre-trained data base is needed. So we need a computer with medium capacities.

3.2.1 Screen Display

Since ASPC based on image segmentation, the plasma screen TV is better than using of slide projector.

4 ASPC components

ASPC is an interactive augmented reality presentation software based on hand gesture recognition. So its components are not similar to a normal presentation software. In ASPC the presenter is a one of its components and his hand is the input element.

ASPC does not have an editor to create the slides. Therefore, we can use PowerPoint to create slides and export it as JPEG images, and then we import it to ASPC.

In this section, we will focus on ASPC interfaces by illustrated figures.

4.1 General ASPC aspect

The interface of ASPC is video captured from web camera where the user must stays on the left of the screen and manipulates with his right hand.

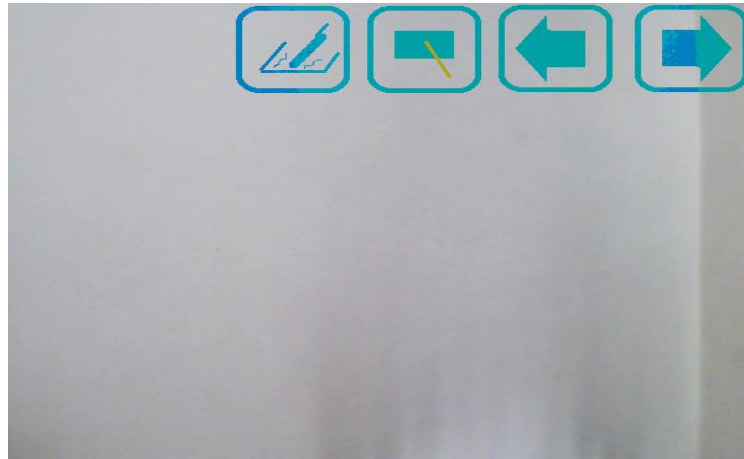


Figure 4.10 ASPC main interface view screen

4.2 Interaction with slides

- Presenter can point with his index finger and ASPC will show it as a laser pointer.

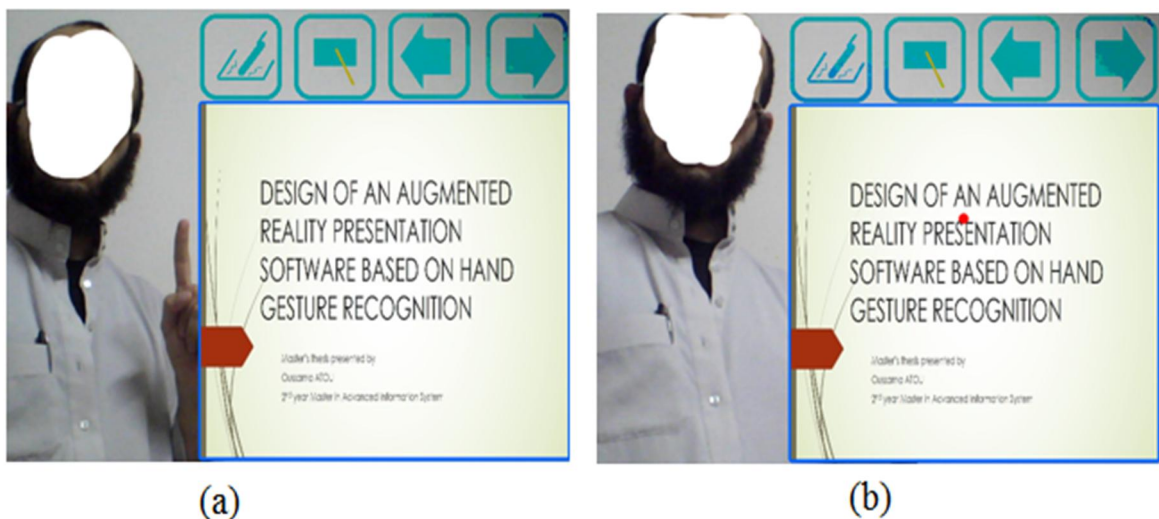


Figure 4.11 Laser pointer , ASPC does not shows the pointer only when user points with his index in the area of slide.

- The presenter must put his hand on the buttons areas then point with his index finger to perform a click. There are four virtual buttons, two to move forward and backward between slides, one for using the red marker and one for showing a white board.

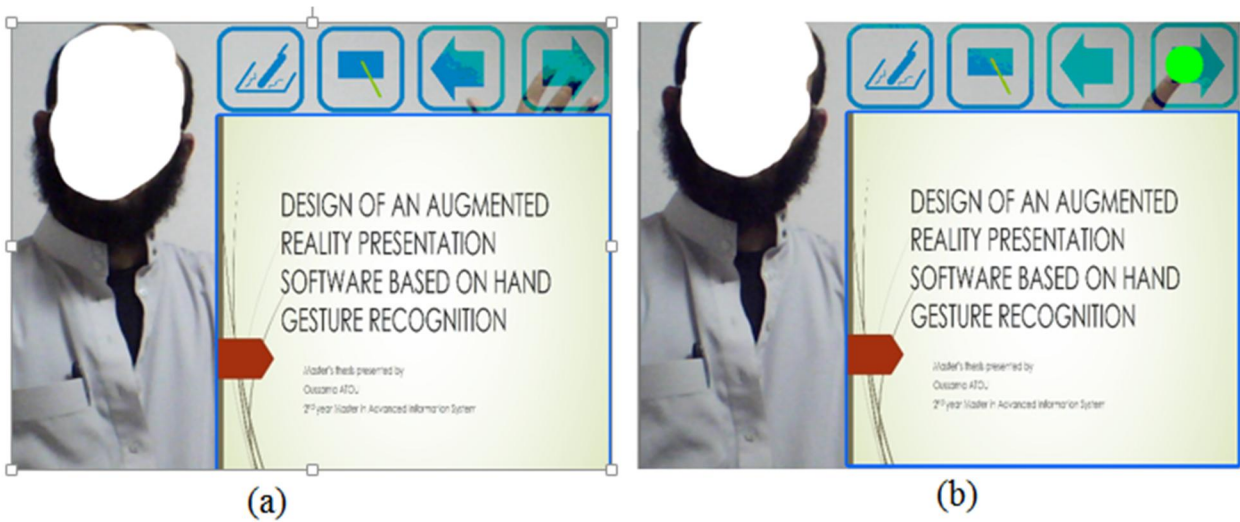


Figure 4.12 (a) Virtual buttons click in ASPC, (b) user clicked the next button.

- If the user clicks the marker button, he will be able to draw on the slide.

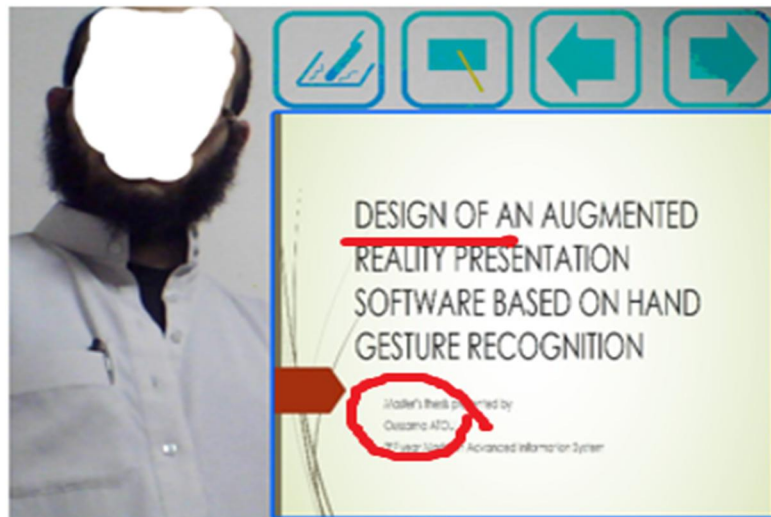


Figure 4.13 Presenter uses the virtual mark to draw on slides

- White board is a white slide where the user can draw and write on.

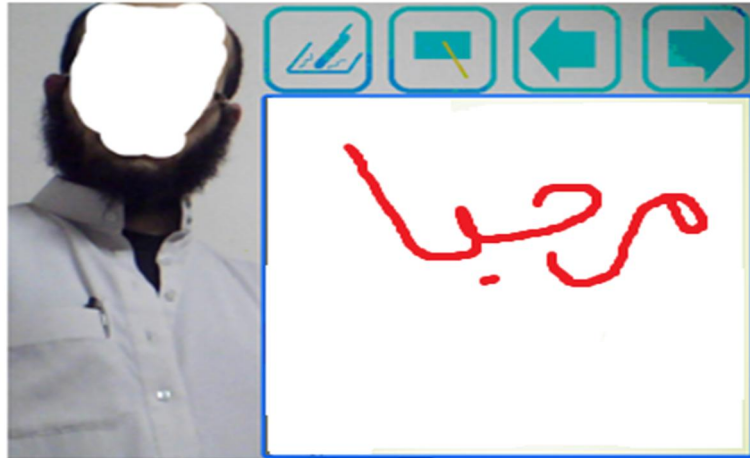


Figure 4.14 Using of white board

5 Conclusion

In this final chapter, we gave an overview on hardware architecture and development environment used to develop ASPC, as well as an illustrated presentation of ASPC components.

GLOBAL CONCLUSION

In this system, we can have an accurate hand click estimation based on a hand contour. Forearm can be included to the contour and the system has a good toleration to hand click despite the appearance of the user's face with his hand.

We obtained the color image from single web camera, and transform the color space into HSV color space. Our skin color region is defined under HSV color space. A binary image of hand can be obtained according to the skin color definition. Two morphological operations include erosion and dilation are performed. Erosion eliminates the noises while dilation smooths the boundary. When the usable image is generated, we apply Pavlidis' algorithm to search for the contours in the image. One or more contours can be found, we choose the right largest contour as our hand contour. When the hand contour has been chosen, we need to calculate the convex hull of it. Three-coin algorithm is applied to obtain the convex hull. We can compare the contour and its convex hull to find all the convexity defects.

The contour point inside one convexity defect which has the longest distance to the contour will be the depth point of the convexity defect. By observing the depth points in many different gestures, we notice that the depth points are tending to be around the hand palm. Minimum enclosing circle of all the depth points will be calculated with the centroid of hand to estimate the hand click.

Also, we used a simple and efficient methods to display virtual buttons, slideshow, white board, laser pointer where user can interact with them in real time and big flexibility which achieve the objective of creation of this program, which shows the advantages of natural interfaces compared to switch-based devices.

Finally, choosing a name for my product very important in programing world. Therefore, I chose ASPC as a name for my product, which is the abbreviation of Augmented Slide Presenter Combiner that is suitable for the main role of my program

1 Future work

Using stereo camera or depth information – We can use two camera from different viewpoint to locate those hand features more accurately. Since the depth camera is a common product nowadays, depth image can be obtained easily. Depth information can be included to know about exact condition of rotation or tilt; and even solve the problem of overlapped hands.

Combining with 3D Augmented Reality – 3D AR is getting more and more popular recently. User can use 3D objects for his presentation instead of using 2D slides. Using of 3D object can be so helpful in explaining of complicated information.

Creating of special editor for ASPC instead of using editors of other software such as PowerPoint and Prezi.

BIBLIOGRAPHY

- [1] Aukstakalnis, S. and D. Blatner (1992), *Silicon Mirage - The Art and Science of Virtual Reality*, Berkeley, CA: Peachpit Press.
- [2] Roland, J. P., R. L. Holloway and H. Fuchs (1994), "A comparison of optical and video see-through head-mounted displays," In *Proceedings SPIE Vol. 2351: Telemanipulator and Telepresence Technologies*, pages 293-307, 1994.
- [3] Azuma, R. (1995), *SIGGRAPH '95 Course Notes: A Survey of Augmented Reality*, Los Angeles: Association for Computing Machinery.
- [4] Welch, R. B. (1978), *Perceptual Modification: Adapting to Altered Sensory Environments*, New York: Academic Press.
- [5] Milgram, P. and F. Kishino (1994), "A Taxonomy of Mixed Reality Visual Displays," *IEICE Transactions on Information Systems*, E77-D(12):1321-1329.
- [6] Capin, T. K., H. Noser, D. Thalmann, I. S. Pandzic and N. M. Thalmann (1997), "Virtual Human Representation and Communication in VLNet," *IEEE Computer Graphics and Applications*, 17(2):42-53.
- [7] Feiner, S., B. MacIntyre, M. Haupt and E. Solomon (1993a), "Windows on the World: 2D Windows for 3D Augmented Reality," In *Proceedings of ACM Symposium on User Interface Software and Technology*, pages 145-155, 1993a.
- [8] Rosen, J. M., D. R. Laub, S. D. Pieper, A. M. Mecinski, H. Soltanian, M. A. McKenna, D. Chen, S. L. Delp, J. P. Loan and C. Basdogan (1996), "Virtual Reality and Medicine: From Training Systems to Performing Machines," In *Proceedings of the IEEE 1996 Virtual Reality Annual International Symposium*, pages 5-13, 1996.
- [9] Mellor, J. P. (1995a), *Enhanced Reality Visualization in a Surgical Environment*, Masters thesis, AI Lab, Massachusetts Institute of Technology.
- [10] Lorensen, W., H. Cline, C. Nafis, R. Kikinis, D. Altobelli and L. Gleason (1993), "Enhancing Reality in the Operating Room," In *Proceedings of the 1993 IEEE Visualization Conference*, pages 410-415, 1993.
- [11] Patel, V. V., M. W. Vannier, J. L. Marsh and L.-J. Lo (1996), "Assessing craniofacial surgical simulation," *IEEE Computer Graphics and Applications*, 16(1):46-54.

- [12] Urban, E. C. (1995), "The Information Warrior," IEEE Spectrum, 32(11):66-70. [24] Reinhard Klette (2014). Concise Computer Vision. Springer. ISBN 978-1-4471-6320-6.
- [25] Linda G. Shapiro and George C. Stockman (2001). Computer Vision. Prentice Hall. ISBN 0-13-030796-3.
- [26] Tim Morris (2004). Computer Vision and Image Processing. Palgrave Macmillan. ISBN 0-333-99451-5.
- [27] Bernd Jähne and Horst Haußecker (2000). Computer Vision and Applications, A Guide for Students and Practitioners. Academic Press. ISBN 0-13-085198-1.
- [28] Milan Sonka, Vaclav Hlavac and Roger Boyle (2008). Image Processing, Analysis, and Machine Vision. Thomson. ISBN 0-495-08252-X.
- [29] David A. Forsyth and Jean Ponce (2003). Computer Vision, A Modern Approach. Prentice Hall. ISBN 0-13-085198-1.
- [30] Dana H. Ballard and Christopher M. Brown (1982). Computer Vision. Prentice Hall. ISBN 0-13-165316-4.
- [31] Barghout, Lauren, and Jacob Sheynin. "Real-world scene perception and perceptual organization: Lessons from Computer Vision." Journal of Vision 13.9 (2013): 709-709.
- [33] Matthias Rehm, Nikolaus Bee, Elisabeth André, Wave Like an Egyptian - Accelerometer Based Gesture Recognition for Culture Specific Interactions, British Computer Society, 2007
- [34] Pavlovic, V., Sharma, R. & Huang, T. (1997), "Visual interpretation of hand gestures for human-computer interaction: A review", IEEE Trans. Pattern Analysis and Machine Intelligence., July, 1997. Vol. 19(7), pp. 677 -695.
- [35] R. Cipolla and A. Pentland, Computer Vision for Human-Machine Interaction, Cambridge University Press, 1998, ISBN 978-0-521-62253-0
- [36] Ying Wu and Thomas S. Huang, "Vision-Based Gesture Recognition: A Review", In: Gesture-Based Communication in Human-Computer Interaction, Volume 1739 of Springer Lecture Notes in Computer Science, pages 103-115, 1999, ISBN 978-3-540-66935-7,doi:10.1007/3-540-46616-9
- [37] Alejandro Jaimesa and Nicu Sebe, Multimodal human-computer interaction: A survey, Computer Vision and Image Understanding Volume 108, Issues 1-2, October–November 2007, Pages 116-134 Special Issue on Vision for Human-Computer Interaction,doi:10.1016/j.cviu.2006.10.019

- [38] Dopertchouk, Oleg; "Recognition of Handwriting Gestures", gamedev.net, January 9, 2004
- [39] Chen, Shijie; "Gesture Recognition Techniques in Handwriting Recognition Application", Frontiers in Handwriting Recognition p 142-147 November 2010
- [40] Vision based hand gesture recognition for human computer interaction: a survey Siddharth S. Rautaray· Anupam Agrawal November 2012.
- [41] T. Pavlidis, "Algorithms for Graphics and Image Processing", Computer Science Press, Rockville, Maryland, 1982.
- [42] R. L. Graham. "An efficient algorithm for determining the convex hull of a finite planar set", Information Processing Letters, 7:175–180, 1972.
- [43] Sklansky J., "Measuring Concavity on a Rectangular Mosaic", IEEE Transactions on Computing, 21, p1355, 1972.
- [44] M. Soriano, S. Huovinen, B. Martinkauppi, and M. Laaksonen, "Using the skin locus to cope with changing illumination conditions in color-based face tracking," in Proceedings of the IEEE Nordic Signal Processing Symposium, pp. 383-386, 2000.
- [45] R.C. Gonzalez and R.E. Woods, Digital Image Processing, 2nd Ed., Addison Wesley, Reading, Massachusetts, 1992.
- [46] S. L. Phung, A. Bouzerdoum, and D. Chai, "A novel skin color model in YCbCr color space and its application to human face detection," in Proceedings of IEEE International Conference on Image Processing, vol. 1, pp. 289-291, 2002.

Websites

- [13] Leap Motion: Drawbacks and Potential Applications
<http://augmentedtomorrow.com/drawbacks-applications-leap-motion/>
- [14] Augmented Reality Is Just Beginning To Change How We Interact With The Real World
<http://www.linkedin.com/today/post/article/20131121170712-48342529-augmented-reality-is-just-beginning-change-how-we-interact-with-the-real-world>
- [15] Directions magazine <http://www.directionsmag.com/podcasts/whats-new-in-augmented-reality/359794>
- [16] Presentation technology . http://en.wikipedia.org/wiki/Presentation_technology
- [17]presentation software <http://whatis.techtarget.com/definition/presentation-software-presentation-graphics>

- [18] Presentation <http://en.wikipedia.org/wiki/Presentation>
- [19] Presentation slide http://en.wikipedia.org/wiki/Presentation_slide
- [20] Slide show http://en.wikipedia.org/wiki/Slide_show
- [21] Web-based slideshow http://en.wikipedia.org/wiki/Web-based_slideshow
- [22] Presentation software <http://presentation-software-review.toptenreviews.com>
- [23] Augmented reality from RTT <http://www.rtt.ag/en/work/references/augmented-reality-presentation>
- [43] Green_theorem http://en.wikipedia.org/wiki/Green_theorem
- [44] Centroid area <http://www.intmath.com/applications-integration/5-centroid-area.php>
- [45] OpenCV C interface: <http://docs.opencv.org>

ASPC هو برنامج عرض يجمع بين المحاضر (العالم الحقيقي) وشريحة العرض التقديمي (العالم الافتراضي) باستخدام واحدة من أحدث التقنيات، والتي تعرف بزيادة الواقع، فعلى خلاف برامج العرض التقديمي الأخرى، يمكن للمستخدم التفاعل مع ASPC باستخدام إيماءات اليد، ومع الوقت سوف تصبح الوسيلة الأكثر إنتشاراً من وسائل تفاعل الإنسان مع الحاسوب الأخرى (HCI). مع ASPC، يمكن للمستخدم مشاهدة شرائح العرض والتفاعل معها كواجهة طبيعية، وهذا مفيد جداً للمحاضر لتسهيل وصول فكرته إلى الجمهور .

بالإضافة إلى ذلك، لاحظنا أن المستخدم يحتاج إلى بعض وسائل الدعم إضافة إلى عرض الشرائح، لذلك أضفنا سبورة بيضاء إقتراضية بيضاء حيث يمكن للمستخدم الرسم عليها مستعملاً الإشارة بالسبابة، وأضفنا مؤشر ليزر يظهر على الشرائح باستعمال سبابة المستخدم كما يمكنه الرسم وكتابة الملاحظات على الشرائح بطريقة سلسة وسهلة

الكلمات المفتاحية: زيادة الواقع ، تفاعل الإنسان مع الكمبيوتر ، إيماءات اليد ، رؤية الكمبيوتر

ABSTRACT

AUGMENTED SLIDE PRESENTER COMBINER (ASPC) IS A PRESENTATION SOFTWARE COMBINES BETWEEN THE PRESENTER (REAL WORLD) AND THE SLIDESHOW (VIRTUAL WORLD) USING A ONE OF LATEST TECHNOLOGY WHICH IS THE AUGMENTED REALITY. INSTEAD OF OTHER PRESENTATION SOFTWARE, USER CAN INTERACTS WITH ASPC USING HIS HAND GESTURES. HAND GESTURE RECOGNITION IS GOING TO BE THE MOST POPULAR TOOL USED FOR HUMAN COMPUTER INTERACTION (HCI).

WITH ASPC, USER CAN SHOWS SLIDES AND INTERACT WITH AS A NATURAL INTERFACE, AND THIS IS SO HELPFUL FOR THE PRESENTER TO REACH AUDIENCE.

IN ADDITION, WE HAVE NOTICED THAT THE USER NEEDS SOME AIDS RATHER THAN SLIDESHOW, SO WE ADDED TO ASPC A WHITE BOARD WHERE THE USER CAN DRAW WITH HIS INDEX FINGERTIP, A LASER POINTER RAISE WHEN USER POINTS WITH HIS INDEX FINGER BEHIND SLIDE AND ALSO USING THE RES MARKER HE CAN DRAW ON THE SLIDES EASILY

KEYWORD: AR, HCI, SLIDE, PRESENTATION, GESTURE RECOGNITION, HAND CLICK, PALM GEOMETRY, COMPUTER VISION.

RESUME

(ASPC) EST UN LOGICIEL DE PRESENTATION QUI COMBINE ENTRE LE PRESENTATEUR (MONDE REEL) ET LE DIAPORAMA (MONDE VIRTUEL) EN UTILISANT UN UNE DES DERNIERES TECHNOLOGIES, QUI EST LA REALITE AUGMENTEE. À LA PLACE D'UN AUTRE LOGICIEL DE PRESENTATION, L'UTILISATEUR PEUT INTERAGIT AVEC ASPC UTILISANT SES GESTES DE LA MAIN. AVEC ASPC, L'UTILISATEUR PEUT MONTRER DES DIAPOSITIVES ET D'INTERAGIR AVEC UNE INTERFACE NATURELLE, ET CELA EST SI UTILE POUR LE PRESENTATEUR D'ATTEINDRE PUBLIC.

EN OUTRE, NOUS AVONS REMARQUE QUE L'UTILISATEUR A BESOIN DE QUELQUES AIDES PLUTOT QUE DIAPORAMA, NOUS AJOUTER A ASPC UN TABLEAU BLANC OU L'UTILISATEUR PEUT DESSINER AVEC SON DOIGT D'INDEX, UN POINTEUR AUGMENTATION DE LASER LORSQUE LES POINTS DE L'UTILISATEUR AVEC SON INDEX DERRIERE DIAPOSITIVE ET AUSSI A L'AIDE LA RESOLUTION MARQUEUR, IL PEUT S'APPUYER SUR LES DIAPOSITIVES FACILEMENT

MOT-CLE : AR, HCL, RECONNAISSANCE DES GESTES DE LA MAIN, PRESENTATION, CLIC DE LA MAIN, LA VISION DE L'ORDINATEUR.