POLITECNICO DI MILANO

Master in Computing System Engineering

Gesture Based Interaction: A survey

Relatore : Prof. Franca Garzotto
Correlatore : Ing. Matteo Valoriani

From
Bhavesh Kumar Singh 709924

Academic Year 2011/2012

TABLE OF CONTENTS

1 Introduction

"A gesture is a motion of the body that contains information. It is a form of non-verbal communication in which visible bodily actions communicate particular messages, either in place of speech or together and in parallel with spoken words. Gestures include movement of the hands, face, or other parts of the body. Gestures differ from physical non-verbal communication that does not communicate specific messages, such as purely expressive displays, proxemics, or displays of joint attention [1]. Gestures allow individuals to communicate a variety of feelings and thoughts, from contempt and hostility to approval and affection, often together with body language in addition to words when they speak [2].

We have a broad range of gestures that are almost universal, including pointing at objects, touching or moving objects, changing object shape, activating objects such as controls, or handing objects to others.The word gesture is used for many different phenomena involving human movement, especially of the hands and arms. Only some of these are interactive or communicative. The questions of how gestures are acquired and come to be recognized as meaningful by particular individuals in the course of their development (ontogeny of gesture and its recognition), and conventionalized, elaborated, or lost within particular cultures (evolution of gesture) are large and deep issues[3].

In computer interfaces, two types of gestures are distinguished[4]:

- **Offline gestures:** Those gestures that are processed after the user interaction with the object. i.e they occur when certain action has already happened.An example is the gesture to activate a menu.
- **Online gestures:** Direct manipulation gestures. They are used to scale or rotate a tangible object.

Psychological/linguistic studies of human gesture use classifications relevant for interaction, language evolution, and language acquisition, e.g. by hearing or deaf children, have all been undertaken [3]. Different cultures may differ in their use of the various types of gesture. Some symbolic gestures such as anger signs (e.g. the OK gesture with thumb and index fnger forming a circle) can have radically different interpretations in other cultures, or no set interpretation depending on the culture of the recipient (e.g. crossing fngers as a sign of wishing for luck, or the Chinese anger signs for some numbers such as 6, 7, 8). Tilting the head back (Greece) or nodding the head (Bulgarian) are used symbolically for `no', but would certainly not be interpreted that way in many other cultures [3]. Cultures also differ in their types and scope of movement in (class 2) expressive gestures: Consider, for example, the differences of rhythm, prosody, hand motions, eye contact, and facial expressions accompanying speech between British, Italian, Japanese, and French speakers [3]. Within cultures, differences between different individuals' uses of gestures can be regional, restricted to particular There are mainly three methods to recognize and detect gestures[3]:

- **Model-based** approaches at a kinematics model into the scene observed by sensors,
- **Gesture recognition based** on classifiers use learning algorithms to label gestures and
- **Heuristic-based** methods which directly search for hints related to a gesture.

Depending on the context of the overall robotic control system, all of these may be of use. The model-based approach is followed by researchers where the goal is to develop algorithms that geometrically model a gesture maintained by the robot into the current scene observed by stereo vision systems and a time-of-flight depth sensor[3].

2 Touchless Gestural Interaction

Tochless gestural interaction enables user to explore and control multimedia information space and/or digital devices through body movements and gesture, without the burden of a physical contact with technology.Inteaction is siad to be touchless if it can take place without mechanical contact between the human and any part of the artifical system (see figure 1).

Figure 1: The ipoint explorer, a touchless information system

Touchless interaction can be multi modal.Touchless gestural interaction in-the-small has received marginal interest by the HCI community.When using touchless interaction technologies,effort and time can be saved [15].

2.1 Types of Touchless gestural interaction

Here we defne two types of touchless interaction**: Direct** and **Indirect manupulation.** Manupulation refers to the specification of object properties.An object much be selected before it can be manipulated[15].

2.1.1 Direct manupulation: Resembles what in engineering called control, where a manipulated variable influence a controlled variable, and the controlled variable is fed back to provide information that can be used to readjust the manipulated variable.Directness is often only based on a representation of reality that can be manipulated. For instance: Dragging of a document icon onto a printer icon to produce a hardcopy of the document which that icon represented or user is turning a page with a gesture[19].

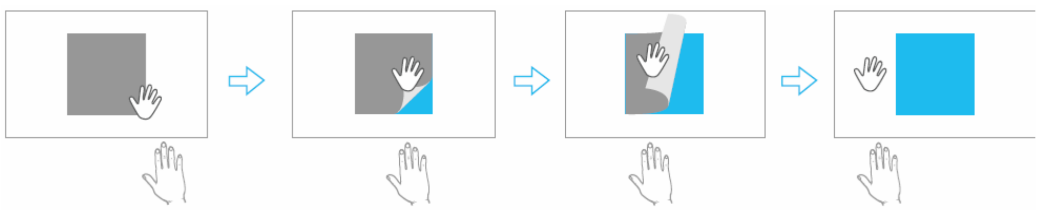


Figure 2: **Page pulling up and turning over as the user moves hand horizontally**

2.1.2 Indirect manupulation: in principle comprises any type of interactive behavior that can be interpreted as a speech act in the sense of searle. As an assertion, declaration, question and so on amongst which commands are just one example.because speech acts are minimal units of communication, indirect interaction could be propely termed communication based.

The manipulation paradigm lets user users manipulate objects directly using **Hand Gestures** in an application specific way.For example, virtual reality systems use thisapprocah by presenting systhesized pseudo-physical objects to the users.The sign language recognition lets users issue command with hand hand gesture.several Advantage of using free-hand gesture for interaction:

- ***Natural Interaction***: Gestures are a natural form of interaction and easy to use.
- ***Terse and Powerful:*** A single gesture can be used to specify both a command and its parameters
- ***Direct Interaction***: The hand becomes input device eliminates the need for intermediate transducers. The user can interact with surrounding machinery by simple designation and appropriate gestures.

Hand movements can be of three types [15]:
Semiotic: is used to convey meaning.
Ergotic: is used to change the state of physical environment.
Epistemic: is performed to feel the environment and thus to gain knowledge. Example of epistemic movements is the head motion performed to look around objects.

2.2 Example of Hand Gesture: A system that interprets gestures and translate them into a sequence of commands must have a way of segmenting the continuous stream of captured motion into discrete lexical entities. Hand gesture input in the Real world:

2.2.1 Computer-aided presentation:

Charade uses hand gestures to control computer-aided presentation [17]. Example, Data glove (see figure 2). Gestures are recognized by an algorithm that runs in real time and leaves sufficient CPU power to run the presentation software on the same machine. Model is based on the notions of active zone and gestural command. The active zone is a 2D area. Gestures are interpreted only when the users hand points at the active zone.

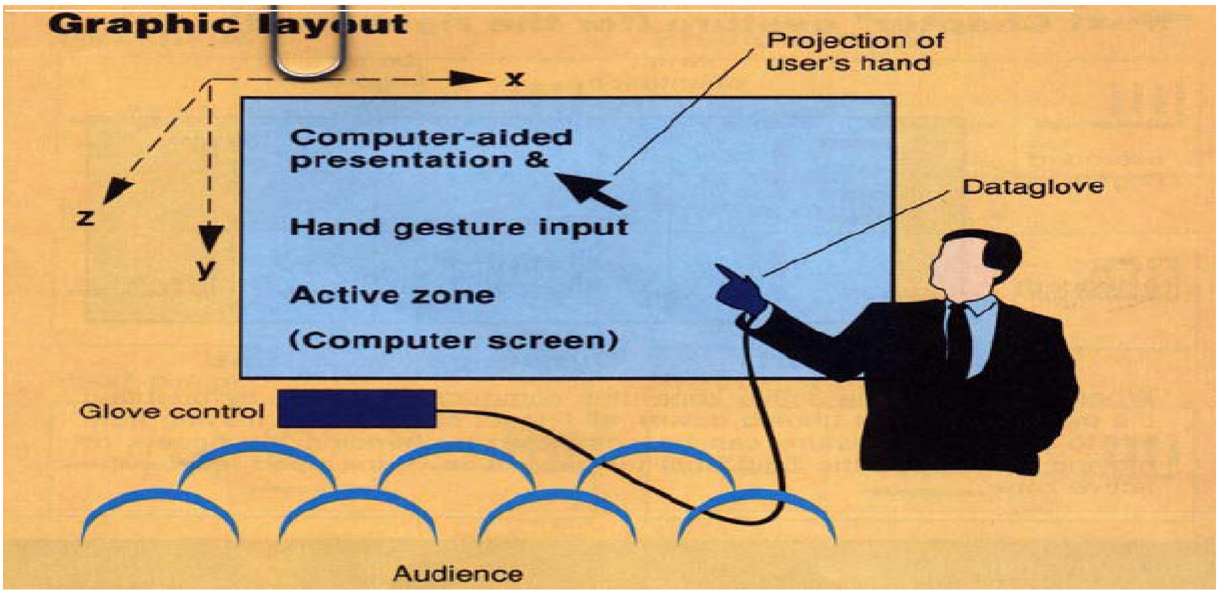


Figure 3: The Charade Interface

Each gestural command is described by a start position, dynamic phase and an end position. The Recognition of a command involves the following three steps:

Detection of the intention: since gestures are interpreted only when the hand is pointing at the active zone, the user can perform various gestures to communicate in the real world. It is also possible to use several active zones to command several systems.

Segmentation of gesture: start and end positions are defined by the wrist orientation (palm up, down, right or left) and finger position.Between 30 to 80 hand positions can be recognized efficiently.

Classification: Gestures are classified according to their start position and dynamic phase. The dynamic phase interprets the movement of the wrist and fingers to distinguish among commands.

2.2.2 Medical systems:

Hand Gesture of touchless interfaces are specially usuful in healthcare environments.control of home devices and appliances for people with physical handicaps and/or elderly users withimpaired mobolity. It explore large complex data volumes and manipulate high-quality images through intuitive actions benefit from 3D interaction, rather than constrained traditional 2D methods.

Hand gesture in medical systems and assistive technologies. Gesture can be used to control the distribution of resourses in hospitals, interact with medical instrumentatio, control visualization displays and help handicapped users as part of their rehabilitation therapy.For example, Face MOUSe satisfied the *come as you are* requirent, where surgeons control the motion of a laparoscope by making appropriate facial gesture without hand or foot switches or voice input[6].

Figure 4: Surgeon using Gestix to browse medical images.

2.2.3 Entertainment:

In computer-vision based, hand gesture-controlled games, the system must respond quickly to user gesture.With the advent of the Nintendo Wii, Sony Move,and Microsoft Kinect, game developers and researchers are now faced with the challenge of creating compelling interface techniques and gameplay mechanics that make use of these technologies[7]. With the advent of the motion controller on every major console and PC gaming platform, the video game industry has made it possible for literally anyone to build and explore 3D spatial interfaces for a variety of different applications. These devices, specifically, the Nintendo Wii Remote, Microsoft Kinect, and Playstation Move, provide user tracking devices that are both inexpensive and accurate.

2.2.3.1 Wiimote [8][9]

The Nintendo Wii controller (Wiimote) is one of the most popular remote human- computer interaction devices. Users can interact with a computer system via gesture recognition or pointing by exploiting the built-in accelerometer and the ingìfrared (IR) camera tracker. The wiimote is essentially a wireless device that employs the satndard Bluetooth HID (Human Interface Device) protocol to communicate, any bluettoth host can recognize it as a standard input device.Wiimote presents three axes of acceleration data in no particular frame of reference, with intermittent optical sensing. (The Wii MotionPlus gyroscope attachment can add three axes of orientation change, or angular velocity.) Although this means the Wiimote's spatial data doesn't directly map to a real-world positions, the device can be employed effectively under constrained use.

Figure 5: Wiimote

The Wii Remote or Wiimote interact with a sensor bar by using accelerometers infrared LEDs and triangulation.

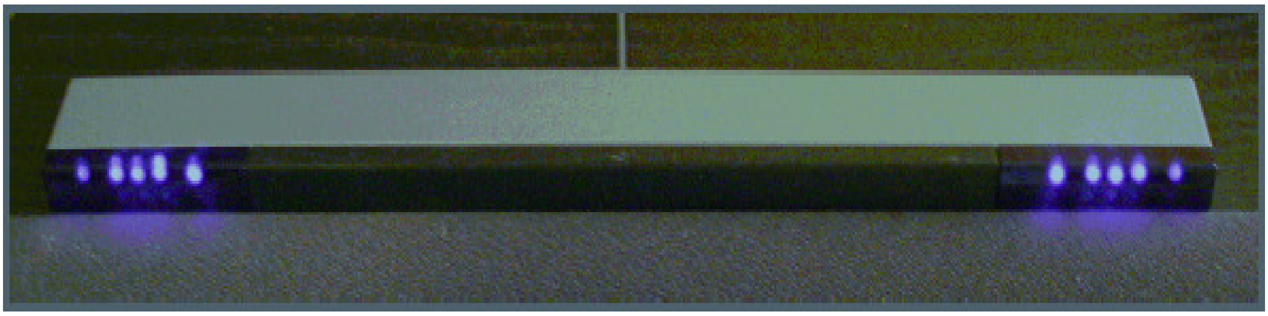


Figure 6: The Wiimote sensor bar has two groups of IR LEDs at fixed width

2.2.3.2 PlayStation[10]

The PlayStation Move system consists of a PlayStation Eye and one to four PlayStation Move motion controllers (see figure 7).

Figure 7: **The playstation move motion controller**

It combines the advantages of camera tracking and motion sensing with the reliability and versatility of buttons. The wireless controller is used with one hand, and it has several digital buttons on the front and a long-throw analog T button on the back. Internally, it has several MEMS inertial sensors, including a three-axis gyroscope and three-axis accelerometer. But the most distinctive feature of the controller is the 44mm-diameter sphere on the top which houses a RGB LED that applications can set to any color. The sphere Color can be varied to enhance interaction, but the primary purpose of the sphere is to enable reliable recovery of the controller 3D position using color tracking with the PlayStation Eye.

2.2.3.3 Microsoft Kinect [10][11]

Kinect is a motion sensing input device by Microsoft for the Xbox 360 video game console and Windows PCs. It aims to let users controls and interact with the Xbox 360 without the need to touch a game controller, through a natural user interface using body gestures and spoken commands. It is able to detect multiple bodies simultaneously and use theire movements and voices as input. The hardware for the Kinect (See figure 8) is comprised of a RGB camera, depth sensor, multi-array microphone, and motorized tilt system.

Figure 8: The Microsoft Kinect Sensor

RGB Camera is used to determine different features of the user and space by detecting RGB colors and is mainly used for facial recognition of the user.**Multi-array Microphone** is a set of four microphones that are able to isolate the voices of multiple users from the ambient noises in room. The **Depth sensor** (generally referred to as the 3D camera) is the most important portion of the Kinect which allows it to function.

It combines an infrared laser projector and a CMOS (complimentary metal-oxide semiconductor) sensor. The infrared projector casts out a myriad of infrared dots(See figure 9) that the CMOS sensor is able to see regardless of the lighting in the room.

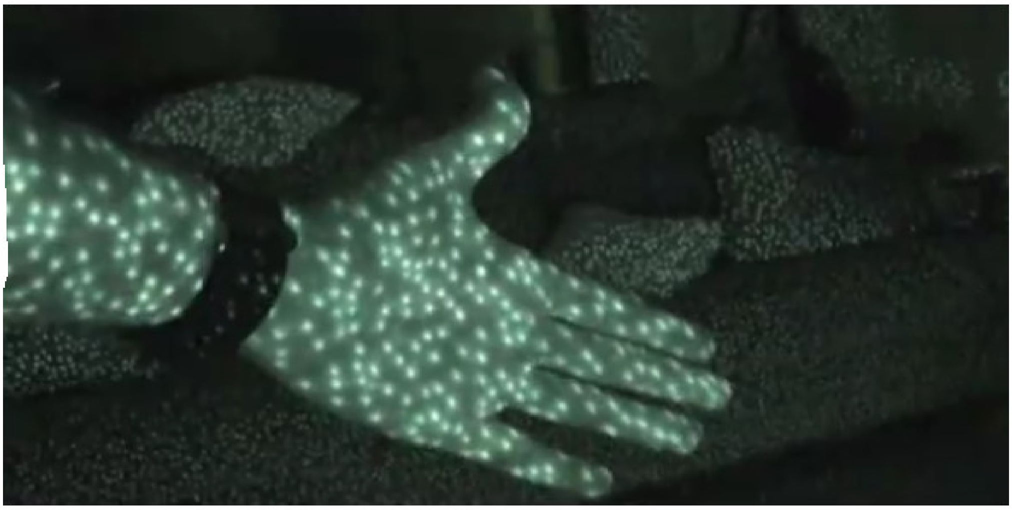


Figure 9: structured light pattern generated by the Kinect infrared laser projector.

To acquire 3D depth information, software component of the Kinect interprets the data from the CMOS sensor. Rays are cast out via the infrared projector in a pseudo-random array across a large area. The CMOS sensor is able to then read the depth of all of the pixels at 30 frames per second. camera is used to detect the location of the infrared dots. Depth calculations are performed in the scene using stereo triangulation.

The depth measurement requires that corresponding points in one image need to be found in the second image. Once those corresponding points are found, we can then find the disparity between the two images. If the images are rectified (along the same parallel axis), then, once we have the disparity, we can then use triangulation to calculate the depth of that point in the scene.Traditionally, stereoscopic triangulation requires two cameras, the Kinect is unique in that the depth sensor only has one camera to perform these calculations. The infrared projector is, in and of itself, a camera in the sense that it has an image to compare with the image taken from CMOS sensor camera. The projected speckles are semi-random in the fact that they are a generated pattern that the Kinect understands. Since the device knows where the speckles are located, it has an image which can be compared to find the focal points.

The CMOS sensor captures an offset image to detect differences in the scene where the disparity between dots can be analyzed and the depth can therefore be calculated. Assuming the images are rectified, the depth calculations are straightforward. The depth data is then interpreted and used in the system. To Visualize the depth information, a depth image can be generated by assigning a color coding to the data (see figure 10).

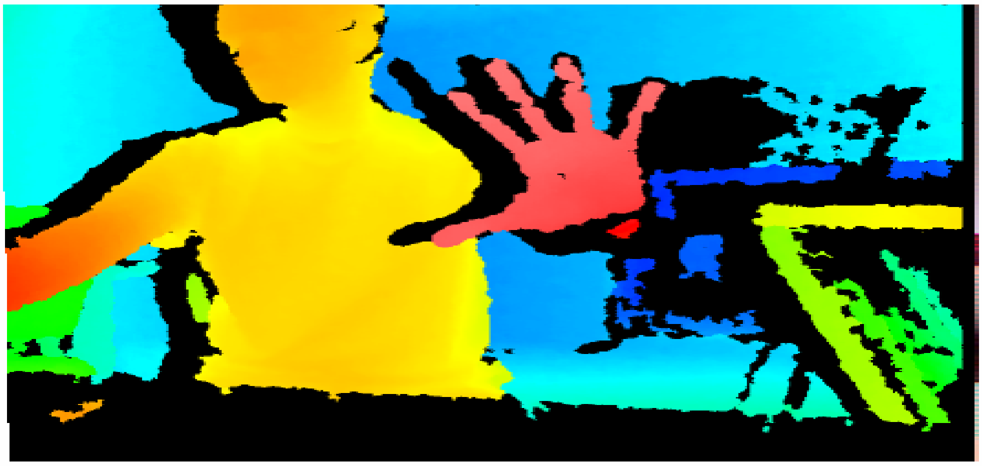


Figure 10: **A depth image from the Microsoft Kinect**

The Kinect software is able to track users' skeletons by combining the depth information and knowledge about human body kinematics. This knowledge was obtained by gathering and labeling data from special rigs that captured user motions in everyday life. The images and labels were used for training a machine learning algorithm to create probabilities and statistics about the human form and movement. The software goes through a series of steps to make sense of the input from the camera and have the users body be the controller. Beginning with the user stepping in front of the Kinect, a 3D surface is generated using the depth information, creating a point cloud of the user. The Kinect then creates a guess at the users skeleton. Next, using using the kinematic data,the Kinect makes an educated guess at determining the different parts of the body. A level of confidence is also assigned to each guess based on how confident the algorithm is about guessing the correct parts. Once this is done, the Kinect finds the most probable skeleton that would fit these body parts and their confidence levels assigned to them. This is performed real tine at 30 frames per second.

2.3 Functionality

A number of multimodal user interfaces afford interaction through the use of communicative free hand gesture [5]. Since hand gestures can be used for a number of different communicative purposes–e.g., pointing at an object to indicate reference, or tracing a path of motion–classification of hand gestures is an important problem [5]. One class of systems focuses on artificial gestures, such as waving, closed fist, or "thumbs up". These are not intended to correspond to the natural gestures that spontaneously arise during speech[5]. For such systems, the goal is to maximize ease and speed of recognition, rather than the naturalness of the user interface [5]. With such artificial gestures, gesture classes are distinguished purely on the basis of the dynamics of hand motion; however, mutual disambiguation with speech could be used to improve recognition[5]. There is, however, a growing set of user interfaces that attempt to allow users to communicate using more natural gestures. Here too, gesture classification has been taken to be primarily solve a problem for computer vision or pattern recognition applied to glove input devices. Mutual disambiguation has been applied to improve recognition by constraining the gesture recognition candidates based on a set of possible semantic frames. But the idea that gesture classes themselves are fundamentally multimodal entities – defined not only by the hand motion but also by the role of gesture within the linguistic context – has not yet been given full consideration[5]. We begin with a brief summary of the most frequently cited gesture taxonomy from the psychology literature; there has been some work on automatic classification for subsets of this taxonomy. Next, we present an empirical study of the ability of naive raters to classify gestures according to this taxonomy, evaluating the effect of removing either the visual or auditory modalities. Then we present a gesture classification system that uses only the linguistic context; no hand-movement information is used. Linguists have created a taxonomy of gesticulation, and gestures that naturally co-occur with speech are now typically divided into several classes: deictic,iconic, metaphoric and beat[5].

McNeil notes that these types should not be thought of as discrete, mutually exclusive bins, but rather as feature that may be present in varying degrees, possibly in combination. Thus identification of the extent to which each feature is present would be ultimate goal, rather than gesture classification. The following definitions are quoted and summarized from cassell.

2.3.1 Deictics Gestures: Gestures that evoke the speech referent: Deictics can refer to actual physical entities and locations, or to spaces that have previously been marked as relating to some idea or concept. These are the types of gestures most generally seen in HCI and are the gestures of pointing, or otherwise directing the listener's attention to specific events or objects in the environment. They are the gestures made when someone says "Put that there.

2.3.2 Iconic gestures: Gestures that depict the speech referent. Tthese gestures are used to convey information about the size, shape or orientation of the object of discourse. depict by the form of the gesture some features of the action or event being described.For example, when someone says The plane flew like this", while moving their hand through the air like the flight path of the aircraft.

2.3.3 Metaphoric gestures: are also representational, but the concept they represent has no physical form; instead the form of the gesture comes from a common metaphor." For example, a speaker might say, "it happened over and over again," while repeatedly tracing a circle.

2.3.4 Beat gestures: Gestures that relate to conversational process. Beat gestures are small baton-like movements that do not change in form with the content of the accompanying speech. They serve a pragmatic function, occurring with comments on one's own linguistic contribution, speech repairs and reported speech." Speakers that emphasize important points with a downward motion of the hand are utilizing beat gestures.

Ggesture types are defined in terms of the role they play in the discourse, rather than in terms of a specific hand trajectory or class of trajectories. Indeed, researchers have found that there is no canonical set of hand trajectories that define each gesture class. For example, Cassell states, "Deictics do not have to be pointing index fingers." For non-deictic gestures, it is even harder to characterize a "typical" set of hand shapes or trajectories; there are perhaps an infinite variety of possible iconic and metaphoric gestures. Clearly, some amount of linguistic evidence – prosodic, lexical, or semantic – is necessary to classify gestures [5].

3. Classification of Gesture

Gesture can be classified according to their function. To approach this problem,a classifcation of gesture for inferring intent and assisting in the understanding of human activity should closely relate gesture with limited categories of intent in situated human activity. The categories of the broad classifcation presented here thus correspond to and allow the attribution of limited kinds of intent to humans. This classifcation is developed as an aid for helping robots to achieve limited recognition of situated human gestural motion, so as to be able to respond appropriately if required, while these robots are working in an environment of ambient human activity (such as a home or office), in which, at times, the robots are also assisting or cooperating with the humans

[3]. Applications of this classification will require the mapping of physical aspects of gestural motion in interactional contexts to the five gestural classes (and their subtypes) suggested here.

Five Classes [3] [5] [6]

3.1 Irrelevant/Manipulative Gestures: These include irrelevant gestures, body/ manipulator motion, sideeffects of motor behaviour, and actions on objects. Broadly characterized, manipulation by a human is here understood as doing something to influence the non-animate environment or the human's relationship to it (such as position). Gestural motions in this class are manipulative actions (in this sense) and their side effects on body movement. These `gestures' are neither communicative nor socially interactive, but instances and effects of human motion. They may be salient, but are not movements that are primarily employed to communicate or engage a partner in interaction. Cases include, e.g. motion of the arms and hands when walking; tapping of the fingers; playing with a paper clip; brushing hair away from the face with the hand; scratching; grasping a cup in order to drink its contents.

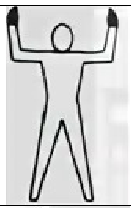
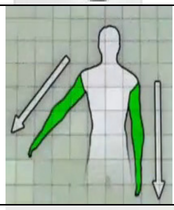
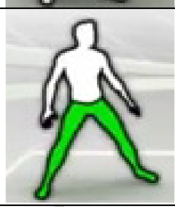
3.2 Side Effect of Expressive Behaviour*:* In communicating with others, motion of hands, arms and face (changes in their states) occur as part of the overall communicative behaviour, but without any specific interactive, communicative, symbolic, or referential roles. Example: persons talk excitedly raising and moving their hands in correlation with changes in voice prosody, rhythm, or emphasis of speech.

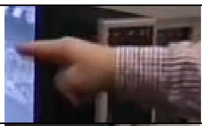
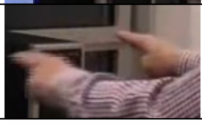
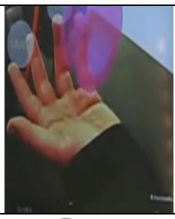
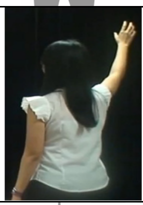
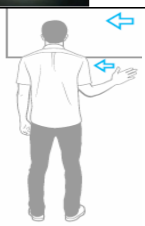
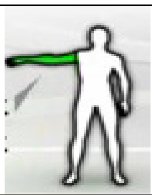
3.3 Symbolic Gestures: These are gestures that, within each culture, have come to have a Single meaning. An Emblem such as the "**OK**" gesture is one such example; however American Sign Language gestures also fall into this category.It is generally a member of a limited, circumscribed set of gestural motions that have specific, prescribed interpretations. Symbolic gesture interfaces are often used in immersive virtual environment where the user cannot see the real world to traditional input devices.

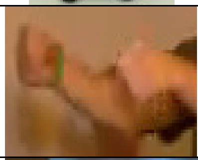
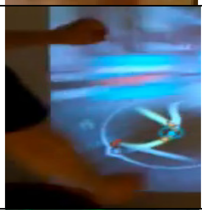
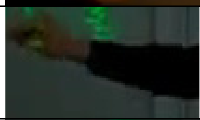
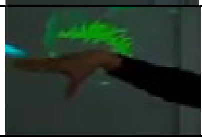
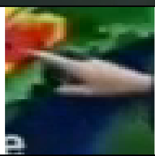
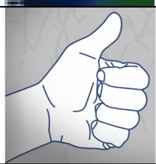
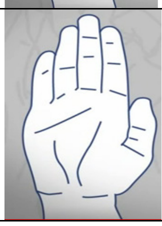
3.4 Interactional Gestures: These are gesture used to regulate interaction with a partner, i.e. used to initiate, maintain, invite, synchronize, organize or terminate a particular interactive, cooperative behaviour: nodding the head indicating that one is listening. The emphasis of this category is neither reference nor communication but on gestures as mediators for cooperative action.Interactional gestures thus concern regulating the form of interactions, including the possible regulation of communicative interactions but do not generally convey any of the content in communication. Interactional gestures are similar to class 1 manipulative gestures in the sense that they influence the environment, but in contrast to class 1, they influence the "animated environment" –doing something to influence human agents (or other agents) in the environment, but not by conveying symbolic or referential content.

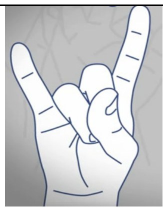
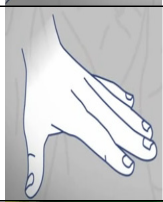
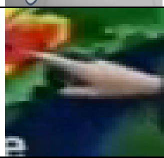
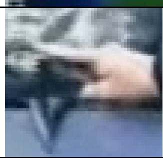
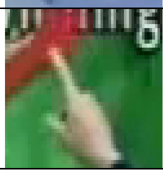
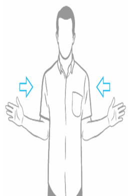
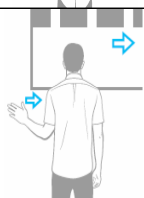
3.5 Referential/Pointing Gestures: These are used to refer to or to indicate objects (or loci) of interest –either physically present objects, persons, directions or locations the environment – by pointing (deixis – showing), or indication of locations in space being used as proxies to represent absent referents in discourse. Deictic gesture can involve a hand, finger, other directed motion, and/or eye gaze..

4. Interaction Task:

Interaction Task	Gesture	Progressive code
Start Application		1
Navigate		2
Change Value		3
Confirm		4
Pause		5
Pass & Shoot		6
Style Shoot		7
Block		8
Save		9

Scroll		10
Hand Apart 1		11
Pointing (Touchless)		12
Replace		13
Wave		14
Select		15
Scroll (Left)		16
Grab		17
Upward		18
Hit		19

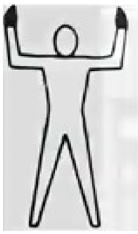
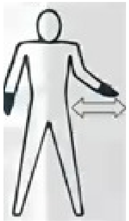
Swing Left		20
Swing Right		21
Acceleration		22
Steering		23
Control		24
Open		25
Touch		26
Thumbs up		27
Victory/ Peace		28
Ok		29
Stop		30

Texas		31
Come Here		32
Settle Down		33
Pointing		34
Drag		35
Track		36
Hand Apart(2)		37
Scroll Right		38
Tap		39
Upward		40

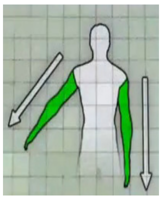
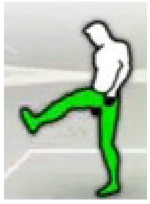
Hold (1)		41
Hold (2)		42

4.1 Gesture Classification Videos

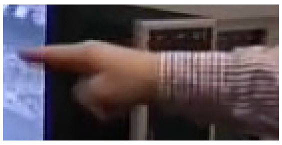
1. URL	http://www.youtube.com/watch?v=Qgd0rcv54Tg 
Name of Video	MOTIC-Don't touch the oven
Audio Yes/No	Yes
Progressive code	1.Start 2.Navigate 3. Change value 4.Confirm
Type of gesture	Interactional Gesture, Referential/Pointing Gestures
Time	1:29 minutes
Description	MOTIC(Motion-based Touchless interactive cooking system)has been built for being drmonstrated in mid April 2012 at the largest worldwide fair in the home furnishing sector and installed in Candy-Hoover show rroms worldwide.Motion-based touchless interaction empowers users to interact using movements and gestures, and without the burden of physical contact with technology. Movements and gestures: Start: Raising both hand up to Turning oven “ON” Vertical Browsing: Raising right arm up to raising temperature and left arm steady to confirm the temperature. Horizontal Browsing: Moving arms left/right to look at look at function.

Examples	 **Start**  **Navigate**  **Change value**  **Confirm**

2. URL	http://www.youtube.com/watch?v=uL_XsdqQ1BE&feature=related 
Name of Video	**Kinect sports: soccer Gameplay HD**
Audio Yes/No	Yes
Progressive code	**5**. Pause **6**. Pass & shoot **7**.style Shoot **8**.Block **9.** Save
Type of Gestures	Interactional Gesture, Referential/Pointing Gestures
Time	10:11 Minutes
Description	Kinect sports soccer allows user to select players against computer opponent to test your skill against computer controlled.Availability to change skill level to choose the computers opponent ability. Need to stand infront of sensor to play. **Start**:Stand as far back as you can within purple area. Then raise your both hand above your head. **Pause:** To pause the game, stretch your left arm out 45 degrees.Your arms needs to be Perfectly straight. **Pass and Shoot:** Kick in the direction you want to pass the ball and to score the goal. **Style Shoot:** Depending on height of the target. Use head to shoot. **Block:** step into the balls path to block the ball. **Save:** Get any part of your body to the target to save the shot.

Examples	 **Start**  **Pause**  **Pass & Shoot**  **Style Shoot**  **Block**  **Save**

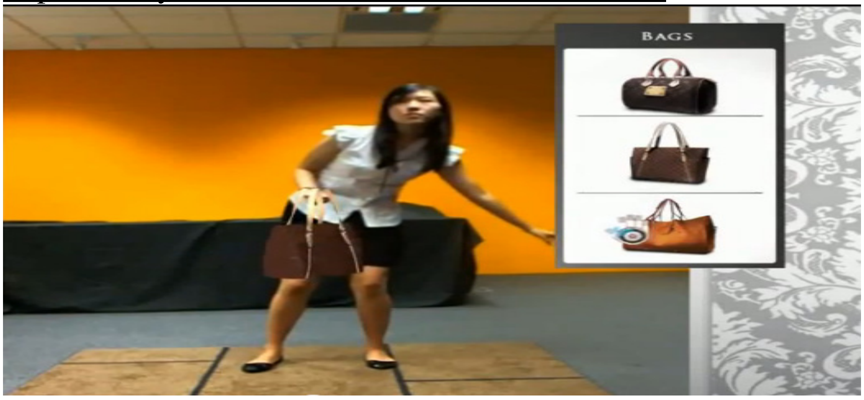
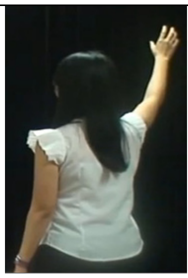
3. URL	http://www.youtube.com/watch?v=M-wLOfjVfVc 
Name of Video	Kinect control windows 7
Audio Yes/No	No
Progressive code	**10**.Scroll **11**.Hand Apart **12**. Pointing (Touchless)
Type of Gestures	Interactional Gesture, Referential/Pointing Gestures
Time	1:56 minutes
Description	Gesture-based Interaction techniques using Microsoft Kinect.Video showing Kinect to control Windows 7 desktop and other multitouch applications. **Scroll**: Using forefinger and middle finger together of right hand right hand to move the picture left-right and up-down on screen.Both hand horizontally in left and right direction to make a close view. **Hand Apart**: moving hands apart to enlarge the screen and reduce the screen by using the opposite gesture.You can rotate the screen by moving both hands. **Pointing**:Pointing on screen by using forefinger to write a letter or text on screen without touching the screen.Even you can use forefinger of both hands to draw with both hands on screen.

Example	 Scrolling by 2 fingers  Hand Apart  Pointing(Touchless)

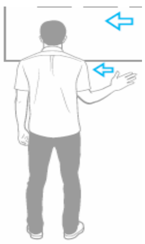
4.URL	http://www.youtube.com/watch?v=DoAqEKartsw 
Name of Video	Gesture based interaction for in situ visualization
Audio Yes/No	No
Progressive code	**12**.Pointing (Touchless) **13**. Replace
Type of Gestures	Pointing Gesture
Time	0:34 minutes
Description	In Simulation-Time, user can easily get information about the pump's state by pointing to one of the parts and grabbing the information in form of state indicators and graphical information. The goal is to support people with context aware information without distracting them from their tasks. The term Augmented Vision (AV) describes the insertion of additional digital content into the visual perception.In this CeBIT 2009 demonstrator, novel 3D hand recognition system serves as a natural interaction device applied on the example of a maintenance task. Recognizing gestures in real time, allowing an intuitive interaction with a complex machine. **Pointing:** In industrial pump, users pointing to one of the parts. **Replace:** Replacing or grabbing module from the pointed parts.
Example	

	 **Pointing**  **Replace**

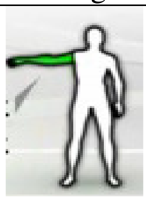
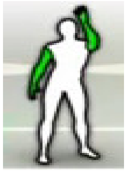
5. URL	http://www.youtube.com/watch?v=tLTTaC4izqI 
Name of Video	**EON Interact-Gesture Based Interaction for interactive 3D**
Audio Yes/No	No
Progressive code	**14**.Wave
Type of Gestures	Interactional Gesture
Time	1:05 minutes
Description	Ability to move and control virtual environment, just using your body and your gesture. EON Interact has support for gesture based interaction through Microsoft Kinect, enabling a more natural interaction method for users of 3D SBL (Simulation Based Learning) scenarios, that provides an improved 'hands-on' experience when using interactive 3D content for learning and product presentations. **Wave**: Hello with an open hand moving from left to right.
Example	 **Wave**

6. URL	http://www.youtube.com/watch?v=1dTVSUf0DxM 
Name of Video	EON Interactive Mirror for Virtual Dressing
Audio Yes/No	Yes
Progressive code	**15**.Select
Type of Gestures	Interactional Gesture
Time	1:10 minutes
Description	EON Interactive mirror is a **virtual try-on** systems where users can virtually try on clothes,bags, and accessories in real time. It enables the shopper to virtually try-on clothes, dresses, handbags and accessories using gesture based interaction. Changing from one dress to another is just a 'swipe' away and offers endless possibilities for mixing designs and accessories in a fun, quick and intuitive way. The shopper can snap a picture of current selections and share on Social Media to get instant feedback from friends.The virtual 3D clothing are based on actual designs from catalogue images, converted to 3D. **Select:** selection of bags and dresses.
Example	 **Select**

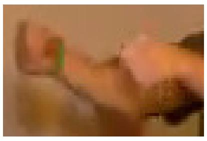
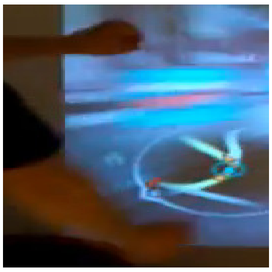
7.URL	http://www.youtube.com/watch?v=3Mddup_hTgA

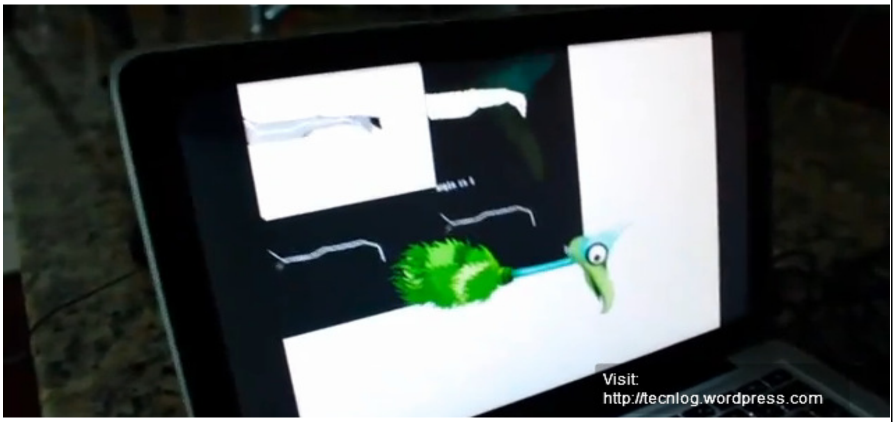
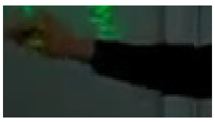
	
Name of Video	Ikea Gesture based Virtual catalog by TIC
Audio Yes/No	No
Progressive code	**16**.Scroll(Left)
Type of Gesture	Interactional Gesture
Time	0:36 minutes
Description	Gesture motion tracking based flip book. You can view and select all the available content via virtual catalog. **Scroll:** Swipe left to scroll or move content to the left.
Example	 **Scroll**

8. URL	http://www.youtube.com/watch?v=cchsdAjhHW0&feature=relmfu 
Name of Video	**Kinect - Kinect Sports: Table Tennis Gameplay HD**
Audio Yes/No	Yes
Progressive code	**1**.Start **17.**Grab **18**.Upward **19**.Hit **20**.Swing Left **21**.Swing Right
Type of Gesture	Interactional Gesture

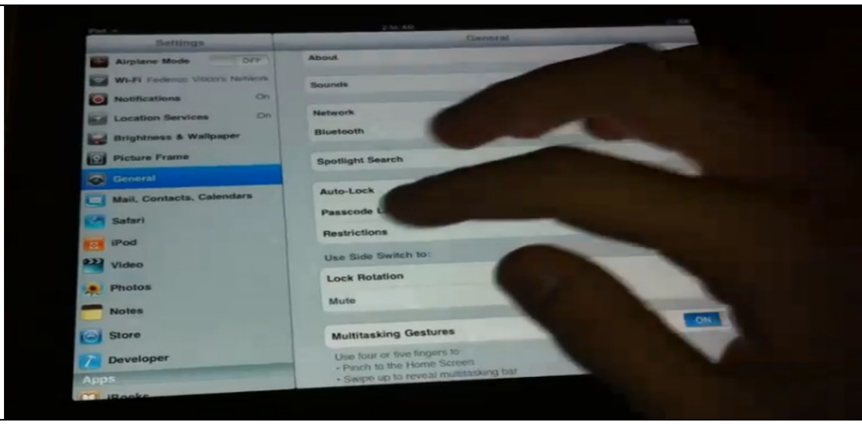
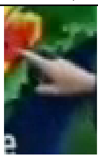
Time	8:49 minutes
Description	In Kinect sports you can choose which sports would you like to play,How would you like to play and with whom. In kinect sports there presents a sensor area and you must stand inside and infront of the sensor to play.Make sure that your play space is clear of people who are not playing. **Start:** Raise hand above head to play. **Grab:** used to reach right or left to grab a paddle with arm staright out. **Serve:** Toss the ball upwars with left or right arm then hit it with the paddle as it falls. **Direct:** swing left or right before you hit the ball to direct it.
Example	 **Strat**  **Grab**  **Upward**  **Hit**  **Swing Left**  **Swingh Right**

9. URL	http://www.youtube.com/watch?v=YhzJHq15OPo 
Name of Video	**Need for Speed Carbon played with Kinect/KinDrive**
Audio Yes/No	No
Progressive code	**22**.Acceleration **23**.Steering
Type of Gesture	Symbolic Gestures
Time	2:33 minutes
Description	Kin Driver is a two handed gestures. It's a dynamic Kinect tracker designed to be used with racing/driving games. Tracker essentially uses the incline between your hands and their distance to the rest of your body. **Acceleration**: both hands should be in parallel with same height. **Steering:** Move both hands left or right at the same time to drive left or right.

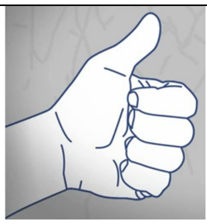
Example	 Acceleration  Steering

10. URL	http://www.youtube.com/watch?v=CeQwhujiWVk&feature=related 
Name of Video	**Kinect hack creates world's greatest shadow puppet**
Audio Yes/No	Yes
Progressive code	**24.**Control **25**.Open
Type of Gesture	Pointing Gestures
Time	2:35 minutes
Description	Export Connect with free net driver and open frame work doing computer vision with open cv.There are short of skeletonization code interacting with the points of elbow and wrist and mapping those points to movement of the bird. **Control:** used to control digital puppet. **Open:** When you open your hand. It sees the point of your hand and open mouth to speeches.
Example	 **Control**  **Open**

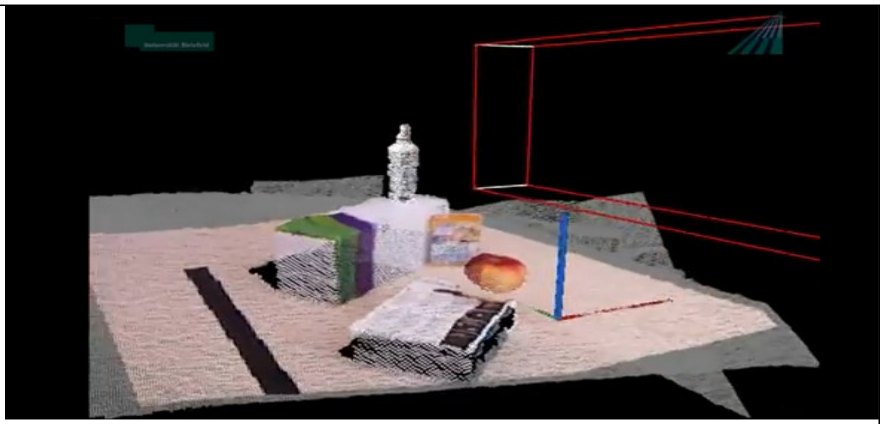
11. URL	http://www.youtube.com/watch?v=PJOFd_kCvoo

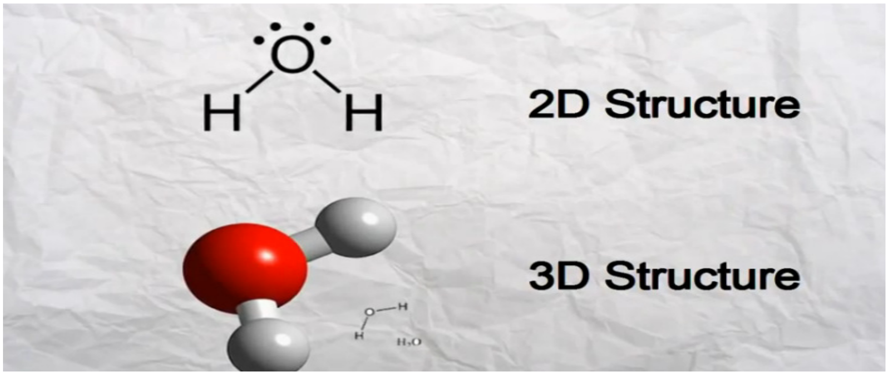
	
Name of Video	**iOS 4.3 Beta Multi-Touch Gestures for iPad**
Audio Yes/No	No
Progressive code	**26**.Touch
Type of Gesture	Interactional Gesture
Time	0:39 minutes
Description	A video showing the new multi-touch gestures in iOS 4.3 for iPad.You interact with your iPad using your fingers. Tap,Scroll,Pinch,and swipe your iPad with Multi-Touch gestures, and controls whats on your screen in a more fluid,natural, and intuitive way.
Example	 Touch

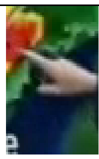
12. URL	http://www.youtube.com/watch?v=OWFPHW7BCCI 
Name of Video	**American Hand Gestures in Different Cultures - 7 Ways to Get Yourself in Trouble Abroad**
Audio Yes/No	Yes
Progressive code	**27**.Thumbs up **28**. Victory/Peace **29**. Ok **30**.Stop **31**.Texas **32**. Come Here **33**.Settle Down

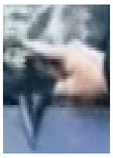
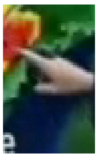
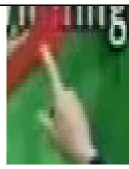
Type of Gesture	Symbolic Gestures
Time	2:04 minutes
Description	Hand Gestures does not have an identical meaning in every culture. It has different meaning in differenr culture. **Thumbs up:** In United States(U.S) the thumbs up sign means well done. But it dont use in Greece, Russia and west africa. **Victory/Peace**: index and middle fingers are up. Many people use this signs to denote victory and peace in america.But it dont use in Britain,Australia,Ireland, and New zealand. **Ok:** In U.S it shows everything is allright.But in Russia, Turkey, Brazil and mediterranean it means something very different.Something like you are homo sexual. In France and belgium its worth zero. **Stop:** Holding hand up, palm out and all five fingers are at attention in U.S it means stop. But in Greece it means saying someone to go staright to hell. **Texas:** In Norway it shows pride for Friendship.In entire Mediterranean telling their spouse being unfaithful. **Come Here:** In philippine curling your finger forward and motioning repeatdly its considered to be a gesture the feeding only uses on dog. Its Punishable with jail time if you used on a person. **Settle Down:** Placing both hand out, palms down,fingers out stretch to settle a crowd to tell people to wait is common place in U.S. but in Greece it means something different.
Example	 **Thumbs Up**  **Victory/Peace**  **OK**  **Stop**  **Texas**  **Come Here**  **Settle Down**

13. URL	http://www.youtube.com/watch?v=WGFJ0o3JcDk

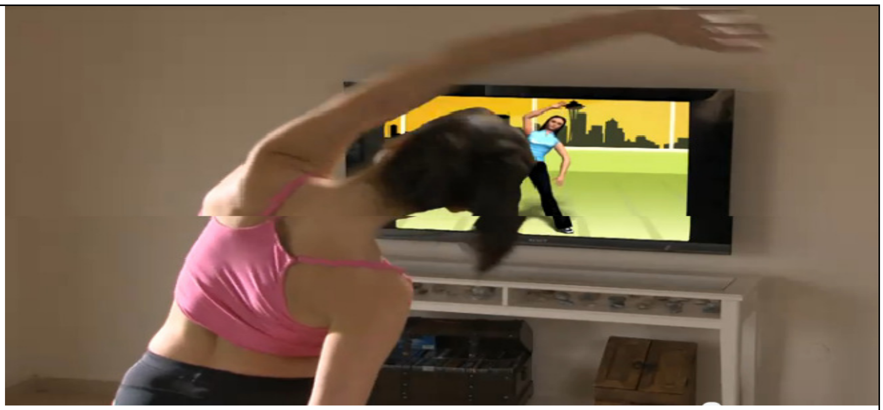
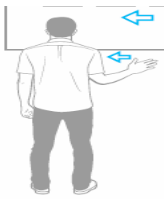
	
Name of Video	**Segmentation And Pointing Gesture Detection For Kinect Point Clouds**
Audio Yes/No	No
Progressive code	**12**.Pointing
Type of Gesture	Pointing Gesture
Time	1:45 minutes
Description	Using point gesures at colored point cloud. Objects are separated and pointed objects highlighted in white. Objectscan be fit in frame.deselect if the object is moved.
Example	 Ponting

14. URL	http://www.youtube.com/watch?v=m5ipCLIAIw4&feature=related 
Name of Video	**Augmented Chemistry Lab with gesture based keyboard**
Audio Yes/No	Yes
Progressive code	**34**.Pointing
Time	2:56 minutes
Description	A Vision based application for students to learn chemistry by designing and interacting with the virtual 3D molecular models with the help of a Tangible user interface. Augmented keyboard consists of fiduciary markers of all the elements of periodic table used in VSEPR

	theory(Valence,shell,electron,pair, and repulsion) with a rotating base marker.wearing a data glove user click a simple gesture of the thumb which turns on the led. **Pointing**: The users selects the element by using the gesture while pointing the specified elements by forefinger.
Example	 Pointing

15. URL	http://www.youtube.com/watch?v=SdbSxYQ2gAU 
Name of Video	**Gesture Tek-Weather**
Audio Yes/No	Yes
Progressive code	**35**.Drag **34**.Pointing **36**.Track
Type of Gesture	Referential/Pointing Gestures
Time	0:41 minutes
Description	weather services international has licensed the gesture stream technology to term their state of the art of weather presentation systems into interactive virtual sets. **Drag**: Dragging the snow data at different places. **Pointing**: pointing a particul location in circle. **Track**: Tracking a distance from pointed cirle to a define place to show the velocity of storm.
Example	 Drag  Pointing  Track

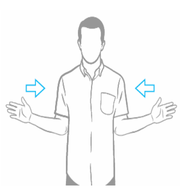
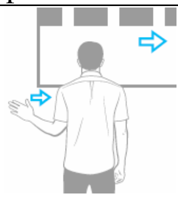
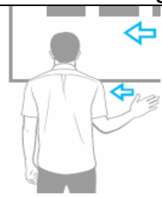
16. URL	http://www.youtube.com/watch?v=Mo7UzJhSZDQ

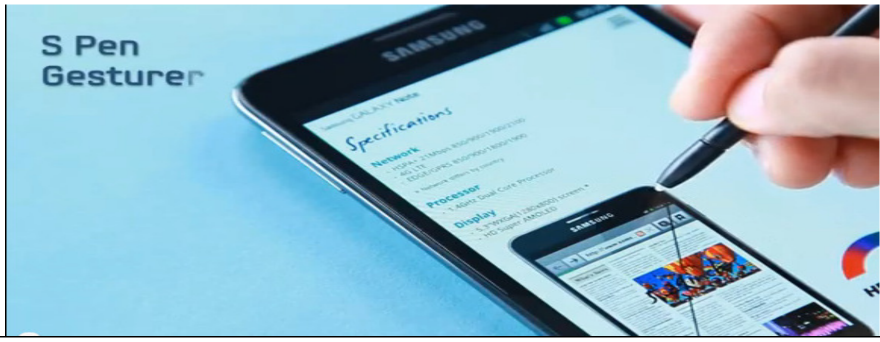
	
Name of Video	**Gesture & Motion Capture for TV - Extreme Reality**
Audio Yes/No	No
Progressive code	**14**.Wave **16**. Scroll(Left) **4**.Confirm
Type of Gesture	Side Effect of Expressive Behaviour, Referential/Pointing Gestures
Time	0:35 minutes
Description	Gesture Control Interface for TV Platforms, based on software only, using a single standard 2D camera. XTR3D Gesture Control Interface supports gestures up to 5 meters away from the screen, in real life environment and with multiple people in front of the TV. **Wave:** waving right hand right to left in front of TV displays program. **Scroll:** Moving right hand from right to the left in front of TV to see all the available programs. **Confirm:** Keeping constantly the right hand infront of menu to select the desire program.
Example	 **Wave**  **Scroll(Left)**  **Confirm**

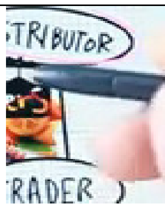
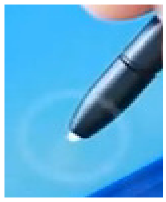
17. URL	http://www.youtube.com/watch?v=b6CT-YDChmE 

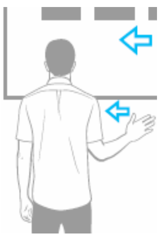
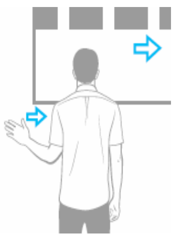
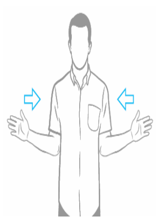
Name of Video	**Virtopsy - Potential use of gesture control in medicine using the Microsoft Kinect camera**
Audio Yes/No	Yes
Progressive code	
Type of gesture	Interaction gesture
Time	6:01 minutes
Description	This Video demonstrates, how the Microsoft Kinect can be used to control a PACS system, in this case OsiriX. The software is based on ofxKinect, libfreenect and open frameworks.Its touch- free control of medical viewers controlled by gesture and voice commands.wireless head sets allows to give voice command.There is three mode of access. You can change the mode with voice commands. Voice commands: **Change window**: recognized the voice command change window and go to new mode indicated by red energy on the kinect **Change window skeleton**: shows sekeleton bone window **Change move**: Recognize and indicated by red energy on the kinect. Now its in move mode, so you can use the hand gestures. With your right hand you can move the image with natural gesture. You can use your second hand to zoom in and out. **Change Stack navigation**: Recognize and indicated by red energy on the kinect, now you can use your right hand to move through data set up and down also by moving hand.you can use your second hand to fast the whole process. **Change next serious**: automatic goes from head scan to the double nose scan of same patient and again you can use your right hand to scroll through data set **Change Buttom**: To get buttom of the data set.
Example	Used Voice commands: Change window, Change window skeleton, Change move, Change next serious,and Change buttom.

18. URL	http://www.youtube.com/watch?v=MDUN01U--jE&feature=related 
Name of Video	**Toshiba AirSwing Gesture-based UI : DigInfo**
Audio Yes/No	Yes
Progressive code	**37**.Hands Apart **38**. scroll(Right) **16**. Scroll(Left)
Type of Gesture	Interaction gesture

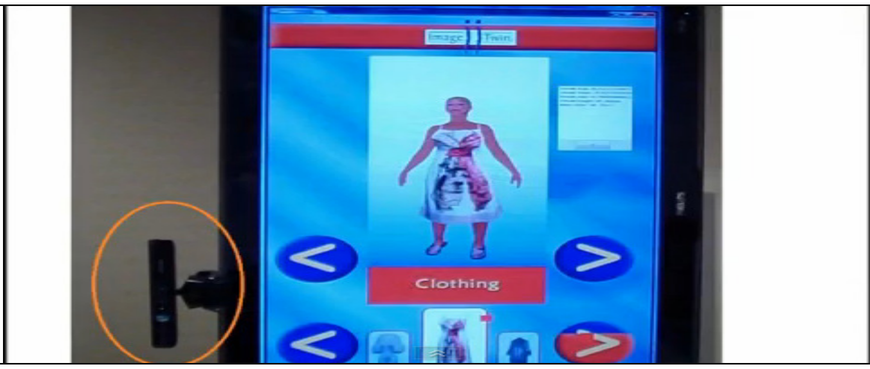
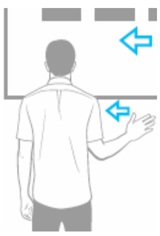
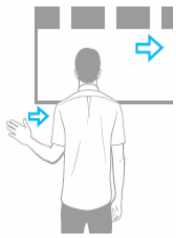
Time	2:14 minutes
Description	Air swing is designed to be operated just by shacking your hand.With air swing a many screen and semitransparent image of the users are showing on the display.The user can select content by shaking the hand super imposed buttoms of the menu screen. As the input device, air swing uses an ordinary web camera to obtain the pictures.The image processing is done by software and software uses only about 3% of the capacity of a 400MHZ ARM 11 CPU.Its uses a combination of hand movements. **Hands Apart:** If you move your hands apart, you can enlarge the screen and you can reduce the screen by using the opposite gesture. **Scroll:** Swipe right to scroll or move content to the right.Move your hand right to navigate through the photos and move your hand left to go back.
Example	   **Hands Apart** **Scroll Right** **Scroll Left**

19. URL	http://www.youtube.com/watch?v=pcxo_3hLeXA 
Name of Video	**[GALAXY Note] S Pen Use Case I: S Pen Gestures**
Audio Yes/No	Yes
Progressive code	**39**.Tap **40**.Upward **41**.Hold
Type of Gesture	Interaction and Pointing gestures
Time	0:26 minutes
Description	The Samsung Galaxy Note features an advanced input mechanism called the S-Pen.The GALAXY Note's advanced pen input functionality is combined with the full touch screen to create a new input experience. The incorporated digital S Pen can be used for accurate sketching and artwork, and instantly capturing ideas freely before they float away.In this videos demonstrate: **Tap**: Double tap the screen to bring up quick memo. **Upward:** Holding the pen button, you can drag upwards to call up the menu bar and drag from right to left to go back to the previous screen. **Hold**: Holding the pen button, tap and hold to capture any screen

Example	 Tap  Upward  Hold(1)

20. URL	http://www.youtube.com/watch?v=-qmmdGonQW4&feature=related 
Name of Video	**Multi Touch Music Wall**
Audio Yes/No	Yes
Progressive code	**16**.Scroll(Left) **38**.Scroll(Right) **42**.Hold(2) **37**.Hands Apart(2)
Type of Gesture	Interaction gesture
Time	0:34 minutes
Description	Worlds first multi touch interactive musical wall. This interactive wall allows multiple users to conduct virtual symphonies.Using gestures to create cacophony. **Scroll:** Moving hand right and left to navigate through the musical instruments. **Hold:** Holding hand to select a instrument. **Hands Apart:**moving hands apart to enlarge the instrument and reduce the instrument by using the opposite gesture.
Example	 Scroll Left  Scroll Right  Hold  Hands Apart

21. URL	http://www.youtube.com/watch?v=s0Fn6PyfJ0I&feature=related

	
Name of Video	**Kinect Virtual Fashion, the Future of Shopping at Home, Retail, and on Smartphone**
Audio Yes/No	Yes
Progressive code	**42**.Hold **16**.Scroll(Left) **38**.Scroll(Right) **4**.Confirm
Type of Gesture	Interaction gesture
Time	1:31 minutes
Description	Kinect virtual fashion using microsoft kinect.It capture your body size and shape with kinect.Personalize your avatar with a face photo.After applying your face you can select your hair style and color.You can try on clothes in 3D and get size and style advice.even you can scan product barcodes to access any content and see it on your smartphone. **Hold**: hold your hand to select a cloth.You can select more than one clothes but once at a time. **Scroll:** Move your hand left and right to move avatar upwards and backwards to see the fitting of different clothes. **Confirm:** confirm a selected cloth to transfer to smartphone and share with friend at social media to get feedback.
Example	 Hold(2)  scroll left  Scroll Right  Confirm

5. Gesture Recognition

Gestures and gesture recognition are terms increasingly encountered in discussions of human-computer interaction[12]. For many (if not most) people the term includes character recognition, the recognition of proof readers symbols, shorthand, and all of the types of interaction described in marking Interfaces. In fact every physical action involves a gesture of some sort in order to be articulated[12]. Furthermore, the nature of that gesture is generally an important component in establishing the quality of feel to the action. Nevertheless, what we want to isolate for discussion are interactions where the gesture is what is articulated and recognized, rather than a consequence of expressing something through a transducer[12].

If we remove ourselves from the world of computers and consider human-human interaction for a moment we quickly realize that we utilize a broad range of gesture in communication. The gestures that are used vary greatly among contexts and cultures.Gestures can exist in isolation or involve external objects. Free of any object, we wave, beckon, fend off, and to a greater or lesser degree (depending on training) make use of more formal sign languages. With respect to objects, we have a broad range of gestures that are almost universal, including pointing at objects, touching or moving objects, changing object shape, activating objects such as controls, or handing objects to others [12].

Gesture can take many forms,from using yourhand to target something on the screen,to specific,learned patterns of movement, to long stretcjes of continuous movement.Below are some examples of commonly used gestures[19].

Static Gestures:also called poses or postures, are one for which the user holds a certain still position until it is recognized.Example:

Figure 11: Example of Static Gestures

Continuous Gestures:track users as they move in front of the kinect.Example:

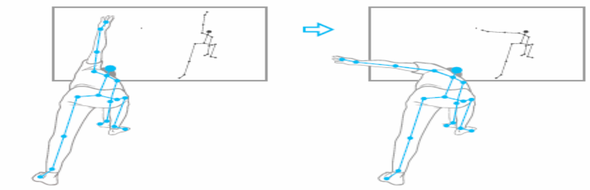


Figure 12:example of Continuous Gestures

Dynamic Gestures:give feedback to to user either during or after they do a defined movement.Example:

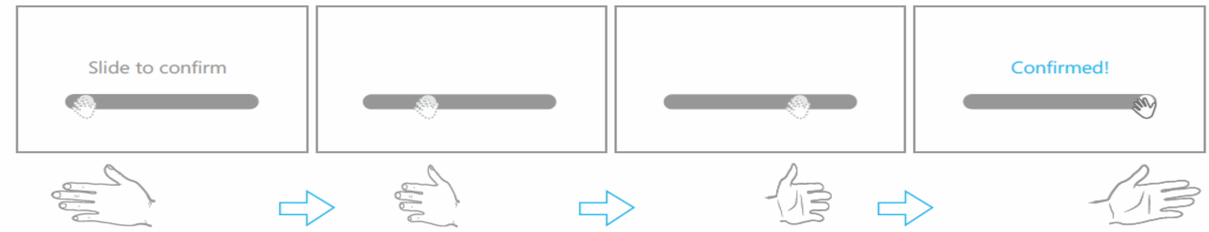


Figure 13: Example of Dynamic Gestures

<u>Symbolic Gesture Recognition</u>

Symbolic gesture interfaces are often used in immersive virtual environment where the user cannot see the real world to traditional input devices. In this setting there are typically a set of pretrained gestures used for navigation through the virtual environment and interaction with virtual objects.

As Baudel and Beaudouin-Lafon point out there are a number of advantages in using symbolic gestures for interaction, including[12]:

• Natural Interaction*:* Gestures are a natural form of interaction and easy to use.
• Terse and Powerful*:* A single gesture can be used to specify both a command and its parameters.
• Direct Interaction*:* The hand as input device eliminates the need for intermediate transducers.

<u>Sign Language Recognition [12]</u>

A obvious application for gesture interfaces is in the interpretation of formal sign language. In contrast with other gestures, sign language does not rely on other input modalities for interpretation and can be used to completely express syntactic and semantic information. Perhaps the most interesting use of sign language interfaces is in sign to speech translation.

5.1 Designing Gesture Interfaces

To design the gesture only interface following points must be kept in mind[12]:
Naturalness:
Are the following characteristics useful for controlling the application tasks?
• Pre-acquired sensorimotor skills
• Existing hand signs
• Absence of an intermediary device
• Task control maps well to hand actions (position and motion of hand).

Adaptability:
Are diverse modes of control used in the tasks?
Is it important to be able to switch between modes of control rapidly, and smoothly?

Coordination:
Do the tasks require the coordination of many degrees of freedom?

Fatigue:
Gestural communication involves more muscles than keyboard interaction or speech: the wrist, fingers, hand and arm all contribute to the expression of commands. Gestural commands must therefore be concise and quick to issue in order to minimize effort. In particular, the design of gestural commands must avoid gestures that require high precision over' a long period of time.As summarized in table 1[12].

Task Characteristics	**Hand Action Capabilities**	**Device Capabilities**
Degrees of Freedom	Degrees of Freedom	Degrees of Freedom
Task Constraints - Degrees of freedom - Physical constraints - Temporal constraints - External forces	Hand Constraints - range of motion - coupling - spatial interference - strength	Cross-coupling
Coordination	Coordination	Device Constraints
Resolution - spatial - temporal	Resolution - spatial - temporal	Fidelity
Speed	Speed	Resolution
Repeatability	Repeatability	Steadiness
Steadiness	Steadiness	Reliability
Endurance	Endurance	Mass
Expressiveness	Expressiveness	Comfort
Modality	Adaptability	Convenience
Task Analogy - comparison to existing methods - similarity to other tasks	Familiarity - similarity to existing skills - similarity to everyday motions	Sampling Rate Computation Required

Table 1: the Relationship between Task Characteristics, Hand Action and Device Capabilities.

5.1.1 Speech with Gestures Interfaces

Voice and gesture complement each other and when used together, creating an interface more powerful that either modality alone. Cohen shows how natural language interaction is suited for descriptive techniques, while gestural interaction is ideal for direct manipulation of objects. For example, unlike gestural or mouse input, voice is not tied to a spatial metaphor [12]. This means that voice can interact with objects regardless of degree of visual exposure, particularly valuable in a virtual environment where objects may be hidden inside each other or occluded by other objects [12].

Experiments in cognitive psychology have shown that a person's ability to perform multiple tasks is affected by whether these tasks use the same or different sensory modes.According to the multiple resource theory of attention the brain modularizes the processing of different types of information – when different tasks tap different resources much of the processing can go on in parallel [12]. So by adding speech input to the interface users should be able to perform gestural tasks at the same time as giving verbal commands with little cognitive interference. Some tasks are inherently graphical, others are verbal and yet others require both vocal and gestural input to be completed [12].As summarized in table 2[12].

	Direct Manipulation	Natural Language
Strengths	1. Intuitive 2. Consistent Look and Feel 3. Options Apparent 4. Fail Safe 5. "Direct Engagement" with object a. Point, act b. Feedback	1. Intuitive 2. Description, e.g., a. Quantification b. Negation c. Temporal d. Motivation/Cause 3. Context 4. Anaphora 5. Asynchronous
Weaknesses	1. Description 2. Anaphora 3. Operation on sets 4. Delayed actions difficult	1. Coverage is opaque 2. Overkill for short or frequent queries 3. Difficulty of establishing and navigating context; Spatial specification cumbersome 4. Anaphora problematic 5. Error prone (ambiguity, vagueness, incomplete)

Table 2: Complimentary Characteristics of Direct Manipulation and Natural Language Interfaces

5.1.2 Conversational Systems

Conversational interfaces are by definition interfaces that allow this type of relationship to exist between human and computer. Speech enabled interfaces are not necessarily conversational interfaces, because speech recognition is not conversational understandin.Similarly gestural input alone is not sufficient. Rather conversation interfaces require multimodal input and output controlled by an intelligent system that is a conversational expert. An important distinction between conversational and multimodal interfaces is the way in which the speech and gestural input is understood. In human conversation conversants actively collaborate together to achieve mutual understanding in a process known as "grounding". In order to achieve this the audio/visual contributions to conversation contain both propositional and interactional information. The propositional information consists of the conversational content, while the interactional information consists of cues that affect the conversational process. The gestural and multimodal interfaces described earlier try to achieve propositional understanding of the user's input.The Challenges in designing Conversational Interfaces[12]

Intuitiveness. Conversation is an intrinsically human skill that is learned over years of development and is practiced daily. Conversational interfaces provide an intuitive paradigm for interaction, since the user is not required to learn new skills.

Redundancy and Modality Switching: Embodied conversational interfaces support redundancy and complementarity between input modes. This allows the user and system to increase reliability by conveying information in more than one modality, and to increase expressiveness by using each modality for the type of expression it is most suited to. The Social Nature of the Interaction. Whether or not computers look human, people attribute to them human-like properties such as friendliness, or cooperativeness. An embodied conversational interface can take advantage of this and prompt the user to naturally engage the computer in human-like conversation.

Bandwidth: Combining conversational cues at an interface allows an increase in input and output bandwidth. The bandwidth limitations of current direct manipulation interfaces are made apparent by considering the interface model that the computer has of the user; a one eyed, one handed, deaf person!

These challenges are from an interface perspective, however there are also strong reasons for developing conversational agents from a communications perspective.

Interface Requirements[[12]
In attempting to develop a conversational interface there are a number of interface requirements. These can be divided into two groups; requirements on the interface of the conversational system, and requirements on the underlying architecture.In terms of underlying architectural requirements, Thorisson reviews the behavioral and psychological literature to arrive at the followin:
• Incremental Interpretation: Conversational understanding is not done "batch style", but occurs continuously as new input is received.
• Multiple Data Types: Conversational interfaces must understand and generate multiple input data types such as gesture, gaze, speech and prosody, and also internal representations such as spatial and relational.
• Seamlessness: Different data types are inserted seamlessly into the conversation, as are the cues that cause conversational mechanisms such as turn taking.
• Temporal Variability: Conversational input, out and understanding occurs over a variety of time scales ranging from less than a second in the case of recognizing intonation to many minutes for discourse understanding.
• Multi-Layered Input Analysis and Response Generation: In order to understand input on a variety of time scales and generate matching output, conversation require multiple layers the work at different time scales.
• High Level Understanding: Conversation involves more than understanding at the word or sentence level. Participants also take into account the conversational context and high level cues from the on-going discourse.

5.2 Common tasks in 3D user interfaces:

There are essentially four basic 3D interaction tasks that are found in most complex 3D applications. These tasks include navigation, selection, manipulation, and system control [10].

5.2.1 Navigation in 3D

The motor component of navigation is known as travel (e.g., viewpoint movement)[10]. There are several issues to consider when dealing with travel in 3D UIs. One such issue is the control of velocity and/or acceleration [10]. There are many methods for doing this, including gesture, speech controls, sliders, etc. Another issue is that of world rotation. In systems that are only partially spatially surrounding (e.g. a 4- walled CAVE, or a single screen), the user must be able to rotate the world or his view of the world in order to navigate. In fully surrounding systems (e.g. with an HMD or 6-sided CAVE) this is not necessary since the visuals completely surround the user [10]. Next, one must consider whether motion should be constrained in any way, for example by maintaining a constant height or by following the terrain. Finally, at the lowest level, the conditions of input must be considered - that is, when and how does motion begin and end (click to start/stop, press to start, release to stop, stop automatically at target location, etc.)? Four of the more common 3D travel techniques are gaze-directed steering, pointing, map-based travel, and "grabbing the air" [10].

5.2.2 Selection

3D selection is the process of accessing one or more objects in a 3D virtual world. Note that selection and manipulation are intimately related, and that several of the techniques described here can also be used for manipulation. One of the most basic common issue for the implementation of selection techniques is how to indicate that the selection event should take place (e.g. you are touching the desired object, now you want to pick it up). This is usually done via a button press, gesture, or voice command, but it might also be done automatically if the system can infer the users intent. One also has to have efficient algorithms for object intersections for many of these techniques.The feedback given to the user regarding which object is about to be selected is also very important. Many of the techniques require an avatar (virtual representation) for the users hand. Finally, consider keeping a list of objects that are selectable, so that a selection technique does not have to test every object in the world, increasing efficiency. Four common selection techniques include the virtual hand, ray-casting, occlusion, and arm extension[10].

5.2.3 Manipulation

Manipulation is connected with selection, because an object must be selected before it can be manipulated. Thus, one important issue for any manipulation technique is how well it integrates with the chosen selection technique. Many techniques, as we have said, do both: e.g. simple virtual hand, ray-casting, and go-go. Another issue is that when an object is being manipulated, you should take care to disable the selection technique and the feedback you give the user for selection. If this is not done, then serious problems can occur if, for example, the user tries to release the currently selected object but the system also interprets this as trying to select a new object. Finally, thinking about what happens when the object is released is important. Does it remain at its last osition, possibly floating in space? Does it snap to a grid? Does it fall via gravity until it contacts something solid? The application requirements will determine this choice. Three common manipulation techniques include HOMER, Scaled-World Grab, and World-in-Miniature[10].

5.2.4 System Control

System control provides a mechanism for users to issue a command to either change the mode of interaction or the system state. In order to issue the command, the user has to select an item from a set. System control is a wide-ranging topic, and there are many different techniques to choose from such as the use of graphical menus, voice commands, gestures, and tool selectors. For the most part, these techniques are not difficult to implement, since they mostly involve selection.

5.3 Existing systems based on gesture interaction

Microsoft Touchless[14](www.officelabs.com/projects/touchless) is an open source Office Labs Grassroots prototype that can create a multitouch remote environment. The Touchless software development kit (SDK) lets developers create touchless multitouch-based applications using a webcam and visual tracking fiducial markers.

Sixth Sense[14] Following the same rationale, SixthSense is a wearable gestural interface that exploits natural hand gesture interactions with digital information displayed in the tangible world. At its core, the prototype is a pocket projector, a mirror, and a camera. The projector turns tangible surfaces—such as walls and physical objects—into screen displays by projecting visual information. Similarly to Microsoft Touchless, software processes the camera’s data and tracks colored markers on the tips of the user’s fingers using naive computer-vision techniques.

The PointScreen project[14] from the Fraunhofer IAIS (Institute for Intelligent Analysis and Information Systems) employs hand gesture recognition. PointScreen is a novel interface to manipulate digital artifacts touchlessly: the user navigates by pointing toward the screen. Instead of fiducial markers, PointScreen uses electric field proximity sensors. Employing noncontact sensors that measure a person’s interaction with the electric fields through body dynamics such as gestures and movement, it can efficiently process gestures for real-time navigation and interaction.

Data Gloves[14] Virtual reality wireless data gloves are valuable devices for providing touchless interaction through hand gestures.

CyberGlove II [14]from Imation is an instrumented glove system that provides up to 22 high-accuracy joint-angle measurements. The gloves use proprietary resistive bend-sensing technology that transforms hand and finger motions into real-time digital data. Because of their price tags, such devices aren’t suitable for low-cost prototyping projects.

Eye Writer [14] Another example of using the body as an interface is EyeWriter. By providing low-cost eye-tracking hardware, this project aims to let people suffering from amyotrophic lateral sclerosis (ALS) draw with their eyes. The software detects and tracks the position of the pupil from a glassmounted camera and uses a calibration sequence to map the coordinates to a computer screen or projection. The EyeWriter project software library was developed using OpenFrameworks.

Microsoft Project Natal [14] The idea of body-as-sensor is the main feature of Microsoft Project Natal, a "controller-free gaming and entertainment experience" for the Xbox 360 video game platform. It aims to let users control and interact with the Xbox 360 without touching a game controller, through body.

Wiimote[14] The Nintendo Wii controller (Wii Remote or Wiimote) is one of the most popular remote human-computer interaction devices. It's also one of the most sophisticated, providing a variety of multimodal I/O functionalities. The Wiimote is mostly advertised for its motion-sensing capabilities: users can interact with a computer system via gesture recognition or pointing by exploiting the built-in accelerometer and the infrared (IR) camera tracker. The most interesting part of the controller is its IR camera, which provides an image-processing engine that can track up to four moving objects and send their coordinates to a host, giving users fast and high-precision tracking at a very low cost. Specifications of the camera's hardware and software are confidential, but related websites contain much information.

The Razer Hydra[16], developed by Sixense, released in June 2011. It is a wireless motion and orientation detection game controller that uses a magnetic field wearable object to detect the absolute position and orientation of the controllers. The controllers themselves are similar to Wii controllers, but have six degrees of freedom and a precision of 1 mm and 1 degree. The use of magnetic fields eliminates the line of sight problem present in Kinect.

6. Conclusion

Based on the various classification dealt above it is true that the gestures recognition through digital interfaces and devices will enable us to reach to a higher level of proximity to mimic the human interactions. The gestures are most fundamental way of interacting with the physical environment. For human computer interaction instead of redrawing the physical interaction procedure and writing new protocols the researchers rather tried to mimic the existing behavioral and gestural patterns so as to be saved from visiting the unknown. However, some may even claim that since our way of interacting with the objects are working fine for us then why to reinvent the wheel. So whatever may be the motive the gesture based interactions enable us to interact with virtual objects. It also enables us to interact with the virtual avatars and roam around the virtual surroundings. Interacting with virtual objects is complex in nature as it involves multiple layers of overlapping interactions which we unknowingly process. There are offline gestures which are post action in nature i.e. they occur when certain action has already happened. These offline gestures are interesting in sense as they leave vital clue of change of position of rendered objects. Complement to offline gesture is online gesture that actually happens when the position or rendering of an object is changed by interacting with it. But there are several types of online gestures and some may not even interact with the virtual object but pass on sufficient clues. For example iconic gestures where the user may trace act of his doing by erratic hand movements. But what is interesting is most of the gestures are communicative in nature.

The communication may/ may not be expressed vocally. So because they are communicative in nature so there exist patterns in certain set of gestures as they happen at distinct time and each and every rhythm of movement is separated in time and thus has temporal pattern. Like Deictic is pointing to an object, in this case the pointing can be with index finger or with object held by the pointer. Here movement of the object/finger across directions then movement of the various joints of the body are all temporally separated and if the gesture can be recorded then such movements can be watched with clear distinction. Similarly there are metaphoric gesture the intensity and use of which particularly changes from culture to culture. In certain cultures metaphoric gestures try to gesturize every vocal word for example in

Italian culture the hand gestures almost complement vocal speech. Similar to metaphoric gesture there are beat gesture only the intensity of beat gestures are more and they may have to the expressed in words. Like for example if Italians would like to say “anything” then they may do so just with the hand movements where the dominant hand is raised to the waist level and rotated in anti-circular motion. Beat gestures just like metaphoric gestures are very much aligned with the culture.

Different researchers have classified gesture based on their utility like some gestures are classified as irrelevant gestures like motion of the arms and hands when walking although these gestures exits but are not relevant for the communication. Then there are gestures which are side effects of the expressive behavior like rythem of the speech, emphasis in speech etc. There are symbolic gestures like showing 2 fingers for number 2. The interactional gestures are online gestures that regulate the interaction with partner or object. The referential/pointing gestures are deictic gestures which relates to pinting to objects, persons, directions or locations.

Gesture recognition are increasingly used in Human computer interaction to recognize the gestures. There are many different types of interfaces which recognize gestures in combination with voice or alone. Then there are various commercial variants of the motion sensing capability mainly used in the digital gaming industry. Some of them are uses joysticks to map the gestures to the virtual domain. These joysticks can be wired or wireless. The only problem with wired controllers are that it limits the degrees of freedom, such problems are mitigated with wireless controllers but they have certain precondition like the hand should always be in front of the body for the camera to recognize the gestures. These preconditions also limit the movability of the gestures. Microsoft Kinect is one such type of systems which allows the player to move freely and interact with the game without holding any controller. This marked the entrance of pure motion sensing into gaming, and to date is the only one of its kind released to market. Designing a strong gesture set gets harder for every gesture. Gesture should be easy for users to learn and remember, and so that they are distinct enough from one another to avoid gesture collisions. Similar gesture uses the same gesture language. Research has shown that users can only remember up to 6 gestures.

6.1 New Gestural paradigms: A shift from grabbing the air

There are several limitations to the gesture recognition systems. Most of the gesture recognition systems recognize only one set of gestures and cannot recognize multiple gestures at once that are they fail to comprehend the gestures of the human body as a system where contraction in one set of muscles may result in expansion in another set of muscles. The Microsoft Kinect also has all the gesture recognition that is linear in nature so its game breadth is limited. However, there is immense commercial scope for systems that utilize mediums other than air. Basically moving away from the traditional approach of grabbing the air. For example, the earth is covered with 2/3 of water and swimming is most favorite sport all across the world but there are no motion sensor systems to exploit such environment. Moreover, swimming is a whole body motion with or without the voice.

6.2 Swimming

Swimming is movement through water using one's limbs, and usually without artificial apparatus. Human swimming typically consists of repeating a specific body motion or swimming stroke. There are many kinds of strokes, each defining a different swimming style or crawl. Most strokes involve rhythmic and coordinated movements of all major body parts

torso, arms, legs, hands, feet, and head. Breathing typically must be synchronized with the strokes, too. It is possible however to swim by moving only legs without arms or only arms without legs; such strokes may be used for special purposes, for training or exercise, or by amputees and paralytics. "Stroke" can also refer to a single completion of the sequence of body movements repeated while swimming in the given style. A style is also known as a stroke.
Several swimming styles are suitable for recreational swimming; many recreational swimmers prefer a style that keeps their head out of the water and has an underwater arm recovery. Breaststroke, side stroke, head up front crawl and dog paddle are the most common strokes utilized in recreational swimming. Although there are several styles of swimming: Front Crawl, Back stroke, Butterfly, Breast stroke, side stroke[20] but the basic ones are easy to be recognized. Some of the basic styles are:

6.2.1 Forward Crawl (also known as Freestyle)

Front crawl is the fastest swimming style in swimming. The front crawl, forward crawl, or freestyle is a swimming stroke usually regarded as the fastest of the four front primary strokes. The first position for front crawl or Freestyle is the streamline position that is to stay on the stomach with both arms stretched out to the front and both legs extended to the back.

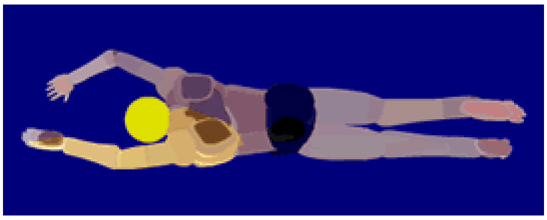


Figure 14: Freestyle swimming

Arm movement

The arm movement alternates from side to side. In other words, while one arm is pulling/pushing, the other arm is recovering. The arm strokes also provide most of the forward movement. The move can be separated into three parts: the downsweep, the insweep and the upsweep. From the initial position, the arm sinks slightly lower and the palm of the hand turns 45 degrees with the thumb side of the palm towards the bottom. This is called catching the water and prepares for the pull. The pull movement follows a semicircle, with the elbow higher than the hand, and the hand pointing towards the body center and downward. The semicircle ends in front of the chest at the beginning of the ribcage. The push pushes the palm backward through the water underneath the body at the beginning and at the side of the body at the end of the push. This pull and push is also known as the S-curve.

Sometime after the beginning of the pull, the other arm begins its recovery. The recovery moves the elbow in a semicircle in a vertical plane in the swimming direction. The lower arm and the hand are completely relaxed and hang down from the elbow close to the water surface and close to the swimmer's body. The beginning of the recovery looks similar to pulling the hand out of the back pocket of a pair of pants, with the small finger upwards. Further into the recovery phase, the hand movement has been compared to pulling up a center zip on a wetsuit. The recovering hand moves forward, with the fingers trailing downward, just above the surface of the water. In the middle of the recovery one shoulder is rotated forward into the air while the other is pointing backwards to avoid drag due to the large frontal area which at this specific time is not covered by the arm. To rotate the shoulder, some twist their torso while others also rotate everything down to their feet.

A recreational variation of front crawl involves only one arm moving at any one time, whiles the other arm rests and is stretched out at the front. This style is called a "catch up" stroke and requires more strength for swimming. This is because the hand is beginning the pull from a stationary position rather than a dynamic one. This style is slower than the regular front crawl and is rarely used competitively: however, it is often used for training purposes by swimmers, as it increases the body's awareness of being streamlined in the water. Total Immersion is a similar technique.

Leg movement

The leg movement in freestyle is called the flutter kick [5]. The legs move alternately, with one leg kicking downward while the other leg moves upward. While the legs provide only a small part of the overall speed, they are important to stabilize the body position. This lack of balance is apparent when using a pull buoy to neutralize the leg action.

The leg in the initial position bends very slightly at the knees, and then kicks the lower leg and the foot downwards similar to kicking a football. The legs may be bent inward (or occasionally outward) slightly. After the kick, the straight leg moves back up. A frequent mistake of beginners is to bend the legs too much or to kick too much out of the water.

Ideally, there are 6 kicks per cycle, although it is also possible to use 8, 4, or even 2 kicks; Franziska van Almsick, for example, swam very successfully with 4 kicks per cycle. When one arm is pushed down, the opposite leg needs to do a downward kick also, to fix the body orientation, because this happens shortly after the body rotation. Alternatively, front crawl can also be swum with a butterfly (dolphin) kick, although this reduces the stability of the swimming position. A breaststroke (frog) kick with front crawl arms (the Trudgen) is awkward, because the breathing pattern for front crawl needs a rotation, and a breaststroke kick resists this rotation.

Breathing

Normally, the face is in the water during front crawl with eyes looking at the lower part of the wall in front of the pool, with the waterline between the brow line and the hairline. Breaths are taken through the mouth by turning the head to the side of a recovering arm at the beginning of the recovery, and breathing in the triangle between the upper arm, lower arm, and the waterline. The swimmer's forward movement will cause a bow wave with a trough in the water surface near the ears. After turning the head, a breath can be taken in this trough without the need to move the mouth above the average water surface. A thin film of water running down the head can be blown away just before the intake. The head is rotated back at the end of the recovery and points down and forward again when the recovered hand enters the water. The swimmer breathes out through mouth and nose until the next breath. Breathing out through the nose may help to prevent water from entering the nose. Swimmers with allergies exacerbated by time in the pool should not expect exhaling through the nose to completely prevent intranasal irritation.

Standard swimming calls for one breathe every third arm recovery or every 1.5 cycles, alternating the sides for breathing. Some swimmers instead take a breath every cycle, i.e., every second arm recovery, breathing always to the same side. Most competition swimmers will breathe every other stroke, or once a cycle, to a preferred side. However some swimmers can breathe comfortably to both sides. Sprinters will often breathe a predetermined number of times in an entire race. Elite sprinters will breathe once or even no times during a fifty meter race. For a one hundred meter race sprinters will often breathe every four strokes, once every two cycles, or will start with every four strokes and finish with every two strokes. In water polo, the head is often kept out of the water completely for better visibility and easier breathing, at the price of a much steeper body position and higher drag.

Body movement
The body rotates about its long axis with every arm stroke so that the shoulder of the recovering arm is higher than the shoulder of the pushing/pulling arm. This makes the recovery much easier and reduces the need to turn the head to breathe. As one shoulder is out of the water, it reduces drag, and as it falls it aids the arm catching the water; as the other shoulder rises it aids the arm at end of the push to leave the water.Side-to-side movement is kept to a minimum: one of the main functions of the leg kick is to maintain the line of the body.

6.2.2 Backstroke (also known as the back crawl)

The backstroke, also sometimes called the "back crawl", ""inverted seizure"" or the upside-down freestyle. This has the advantage of easy breathing, but the disadvantage of swimmers not being able to see where they are going. The swimming style is similar to an upside down front crawl. Both backstroke and front crawl are long-axis strokes. In Individual medley backstroke is the second style swum; in the team medley it is the first style swum. In the initial position, the swimmer lies flat on their back; arms stretched with extended fingertips, and legs extended backwards.

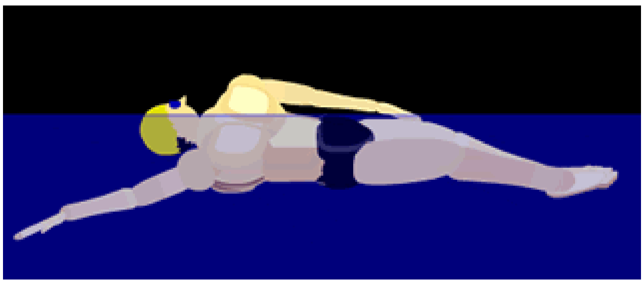


Figure 15: **Backstroke**

Arm movement
In backstroke, the arms contribute most of the forward movement. The arm stroke consists of two main parts: the power phase (consisting of three separate parts) and the recovery. The arms alternate so that always one arm is underwater while the other arm is recovering. One complete arm turn is considered one cycle. From the initial position, one arm sinks slightly under water and turns the palm outward starts the Catch phase (first part of the power phase). The hand enters downward (pinkie finger first) then pulling out at a 45 degree angle, catching the water.

During the power phase the hand follows a semi-circular path from the Catch to the side of the hip. The palm is always facing away from the swimming direction, while remaining straight as an extension of the arm, and the elbow always points downward towards the bottom of the pool. This is done so that both the arms and the elbow can push the maximum amount of water back in order to push the body forward. At the height of the shoulders, the upper and lower arms should have their maximum angle of about 90 degrees. This is called the Mid-Pull of the power phase.

The Mid-Pull phase consists of pushing the palm of the hand as far down as possible with the fingers pointing upward. Again, the goal is to push the body forward against the water. At the very end of the Mid-Pull, the palm flaps down for a last push forward down to a depth of 45 cm, creating the Finish of the Power phase. Besides pushing the body forward, this also helps with the rolling back to the other side as part of the body movement. During the power phase, the fingers of the hand can be slightly apart, as this will increase the resistance of the hand in the water due to turbulence.

To prepare for the recovery phase, the hand is rotated so that the palms point towards the legs and the thumb side points upwards. At the beginning of the recovery phase of the one arm, the other arm begins its power phase. The recovering arm is moved in a semicircle straight over the shoulders to the front. During this recovery, the palm rotates so that the small finger enters the water first, allowing for the least amount of resistance, and the palms point outward. After a short gliding phase, the cycle repeats with the preparation for the next power phase.

A variant is to move both arms synchronized and not alternating, similar to an upside down butterfly stroke. This is easier to coordinate, and the peak speed during the combined power phase is faster, yet the speed is much slower during the combined recovery. The average speed will usually be less than the average speed of the alternating stroke.

Another variant is the old style of swimming backstroke, where the arm movement formed a complete circle in a windmill type pattern. However, this style is not commonly used for competitive swimming, as a lot of energy is spent on pushing the body up and down instead of forward. Furthermore, the added strain on the shoulder is considered less than ideal and can lead to injuries.It is also possible to move only one arm at a time, where one arm moves through the power and recovery phases while the other arm rests. This is slow, but it is used frequently to teach students the movement, as they have to concentrate on only one arm.

Leg movement

The leg movement in backstroke is similar to the flutter kick in front crawl. They make a small contribution to the forward speed, yet are very significant for stabilizing the body.
The leg stroke is also alternating, with one leg sinking down straight to about 30 degrees out of the horizontal. From this position, the leg makes a fast kick upward, slightly bending the knee at the beginning and then stretching it again in the horizontal. However, there are also frequent variants with four or only two kicks per cycle. Usually, sprinters tend to use 6 kicks per cycle, whereas long distance swimmer may use less.

Body movement

Due to the asynchronous movement of the arms, there is a roll of the body around its own axis. This is normal and helps swimming effectively. The overall position of the body is straight in the horizontal to reduce drag. Beginners frequently let their posterior sink too low and increase drag. To avoid this, the upper legs have to be moved to the extreme down position at each kick even with a little help by the back and the foot tips have to be fixed in the extreme lower position and the head is held out of the water to act as a counter-weight.

6.3 Concluding the conclusion

Motion sensing, though important in gaming, will ultimately make its biggest impact in non-gaming applications. Though it has been effective in expanding the gaming demographic to include females and older users, it will never completely replace the controller based gaming of today. This is due to the inability of pure motion sensing to completely supplant legacy controllers with regard to games requiring peripherals, such as driving and shooting games. However I believe that in near future more immersive gesture recognition which may also encompass water based movements and techniques will come. This will also include gesture recognition using body as a system rather than standalone. Future work will focus on making the system more natural by utilizing the weak magnetic fields on which most of the biological life forms are based for movement and navigation than the current setup.

7. References

1) Kendon, Adam. (2004) Gesture: Visible Action as Utterance. Cambridge: Cambridge University Press. ISBN 0-521-83525-9
2) Xu J, Gannon PJ, Emmorey K, Smith JF, Braun AR. (2009). Symbolic gestures and spoken language are processed by a common neural system. Proc Natl Acad Sci U S A. 106:20664–20669. doi:10.1073/pnas.0909197106 PMID 19923436
3) A Methodological Approach relating the Classification of Gesture to Identification of Human Intent in the Context of Human-Robot Interaction, Chrystopher L. Nehaniv, Kerstin Dautenhahn,Jens Kubacki, Martin Haegele, Christopher Parlitz
4) Dietrich Kammer, Mandy Keck, Georg Freitag, Markus Wacker, Taxonomy and Overview of Multi-touch Frameworks: Architecture, Scope and Features
5) Visual and Linguistic Information in Gesture Classification, Jacob Eisenstein and Randall Davis
6) Vision-Based Hand-Gesture Applications Juan Pablo WACHS, MATHIAS KÖLSCH, HELMAN STERN, AND YAEL EDAN
7) WINGRAVE, C., WILLIAMSON, B., VARCHOLIK, P., ROSE, J., MILLER, A., CHARBONNEAU, E., BOTT, J., AND LAVIOLA, J. 2010. Wii remote and beyond: Using spatially convenient devices for 3d uis. IEEE Computer Graphics and Applications 30, 2, 71–85.
8) SHIRAI, A., GESLIN, E., AND RICHIR, S. 2007. Wiimedia: motion analysis methods and applications using a consumer video game controller. In Sandbox ’07: Proceedings of the 2007 ACM SIGGRAPH symposium on Video games, ACM, New York, NY, USA, 133–140.
9) SHIRATORI, T., AND HODGINS, J. K. 2008. Accelerometer-based user interfaces for the control of a physically simulated character. ACM Trans. Graph. 27, 5, 1–9
10) 3D Spatial Interaction: Applications for Art, Design, and Science Joseph J. LaViola Jr. Daniel F. Keefe
11) ""Project Natal" 101". Microsoft. June 1, 2009. Archived from the original on June 1, 2009. Retrieved June 2, 2009.
12) GESTURE BASED INTERACTION, Chapter 14, BillBuxton.
13) Johnston, M., Cohen, P., McGee, D., Oviatt, S., Pittman, J., Smith, I. (1997) Unification-based Multimodal Integration.
14) Human-Display Interaction Technology: Emerging Remote Interfaces for Pervasive Display Environments Andrea Bellucci, Alessio Malizia, Paloma Diaz, and Ignacio Aedo
15) Touchless Interaction-Novel Chances and Challenges René de la Barré, Paul Chojecki, Ulrich Leiner, Lothar Mühlbach, and Detlef Ruschin
16) Motion Sensing Technology, Team Bazinga Clare Chen, Grace Li, Peter Ngo, Connie Sun, June 2, 2011
17) Charade: Remote control of objects using free-hand gestures Thomas Baudel,Michael Beaudouin-Lafon
18) Designing Touch less gestural Interaction In-The-Small: Case Study
19) Human Interface Guidelines: Kinect for Windows v 1.5.0
20) Types of swimming styles. http://www.livestrong.com/article/430964-5-types-of-swimming-styles/