GESTURE RECOGNITION TECHNOLOGY

Technical Seminar Report submitted

in partial fulfillment of the requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

In

INFORMATION TECHNOLOGY

Submitted

By

V.AKHIL

HT.NO:20D41A1258

DEPARTMENT OF INFORMATION TECHNOLOGY

SRI INDU COLLEGE OF ENGINEERING AND TECHNOLOGY

(An Autonomous Institution under UGC, Accredited by NBA, Affiliated to JNTUH)

Sheriguda, Ibrahimpatnam

(2020-2024)

SRI INDU COLLEGE OF ENGINEERING AND TECHNOLOGY

(An Autonomous Institution under UGC, Accredited by NBA, Affiliated to JNTUH)

DEPARTMENT OF INFORMATION TECHNOLOGY

CERTIFICATE

Certified that the Technical Seminar entitled "**GESTURE RECOGNITION TECHNOLOGY"** is a Bonafide work carriedout by **V.AKHIL(20D41A1258) in** partial fulfillment for the award of **Bachelor of Technology** in **SICET**,Hyderabad for the academic year **2023-2024**. The project has been approved as it satisfies academic requirements in respect of the work prescribed for the **IV YEAR, I-SEMESTER** of **B. TECH** course.

STAFF MEMBER IN-CHARGE

HOD/IT

Dr.P.Epsiba

ABSTRACT

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

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

CONTENTS

Introduction

A child being sensed by a simple gesture recognition algorithm detecting handlocation and movement

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

Gesture recognition can be seen as a way for computers to begin to understand human body language, thus building a richer bridge between machines and humans than primitive **text user interfaces** or even **GUIs** (graphical user interfaces), which still limit the majority of input to keyboard and mouse.

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

Interface with computers using gestures of the human body, typically hand movements. In gesture recognition technology, a camera reads the movements of the human body and communicates the data to a computer that uses the gestures as input to control devicesor applications. For example, a person clapping his hands together in front of a camera can produce the sound of cymbals being crashed together when the gesture is fed through a computer.

One way gesture recognition is being used is to help the physically impaired to interact with computers, such as interpreting sign language. The technology also has the potential to change the way users interact with computers by eliminating input devices suchas joysticks, mice and keyboards and allowing the unencumbered body to give signals to the computer through gestures such as finger pointing.

Unlike haptic interfaces, gesture recognition does not require the user to wear any special equipment or attach any devices to the body. The gestures of the body are read by a camera instead of sensors attached to a device such as a data glove.

In addition to hand and body movement, gesture recognition technology also can be used -to read facial and speech expressions (i.e., lip reading), and eye movements.

The literature includes ongoing work in the computer vision field on capturing gesturesor more general human pose and movements by cameras connected to a computer.

Gesture recognition and pen computing:

In some literature, the term gesture recognition has been used to refer more narrowly to non-text-input handwriting symbols, such as inking on a graphics tablet, multi-touch gestures, and mouse gesture recognition. This is computer interaction through the drawing of symbols with a pointing device cursor (see discussion at Pen computing).

Gesture Only Interfaces

The gestural equivalent of direct manipulation interfaces is those which use gesture alone. These can range from interfaces that recognize a few symbolic gestures to those that implement fully fledged sign language interpretation. Similarly interfaces may recognize static hand poses, or dynamic hand motion, or a combination of both. In all cases each gesture has an unambiguous semantic meaning associated with it that can be used in the interface. In this section we will first briefly review the technology used to capture gesture input, then describe examples from symbolic and sign language recognition. Finally we summarize the lessons learned from these interfaces and provide some recommendations for designing gesture only applications.

Tracking Technologies

Gesture-only interfaces with syntax of many gestures typically require precise hand pose tracking. A common technique is to instrument the hand with a glove which is equipped with a number of sensors which provide information about hand position, orientation, andflex of the fingers. The first commercially available hand tracker, the Data glove, is described in Zimmerman, Lanier, Blanchard, Bryson and Harvill (1987), and illustratedin the video by Zacharey, G. (1987). This uses thin fiber optic cables running down the back of each hand, each with a small crack in it. Light is shone down the cable so when the fingers are bent light leaks out through the cracks. Measuring light loss gives an accurate reading of hand pose. The Data glove could measure each joint bend to an accuracy of 5 to 10 degrees (Wise et. al. 1990), but not the sideways movement of the fingers (finger abduction). However, the Cyber Glove developed by Kramer (Kramer 89) uses strain gauges placed between the fingers to measure abduction as well as more accurate bend sensing (Figure XX). Since the development of the Data glove and Cyber glove many other gloves based input devices have appeared as described by Sturman and Zeltzer (1994).

Natural Gesture Only Interfaces

At the simplest level, effective gesture interfaces can be developed which respond to natural gestures, especially dynamic hand motion. An early example is the Theramin, an electronic musical instrument from the 1920’s. This responds to hand position using two proximity sensors, one vertical, the other horizontal. Proximity to the vertical sensor controls the music pitch, to the horizontal one, loudness. What is amazing is that music can be made with orthogonal control of the two prime dimensions, using a control systemthat provides no fixed reference points, such as frets or mechanical feedback. The hands work in extremely subtle ways to articulate steps in what is actually a continuous control space. The Theramin is successful because there is a direct mapping of hand motion to continuous feedback, enabling the user to quickly build a mental model of how to use the device.

Gesture Based Interaction

Figure XX: The Cyber Glove

The Cyber Glove captures the position and movement of the fingers and wrist. It has up to22 sensors, including three bend sensors (including the distal joints) on each finger, four abduction sensors, plus sensors measuring thumb crossover, palm arch, wrist flexion and wrist abduction. (Photo: Virtual Technologies, Inc.)

Once hand pose data has been captured by the gloves, gestures can be recognized using a number of different techniques. Neural network approaches or statistical template matching is commonly used to identify static hand poses, often achieving accuracy rates of better than 95% (Väänänen and Böhm 1993). Time dependent neural networks may also be used for dynamic gesture recognition **[REF]**, although a more common approach is to use Hidden Markov Models. With this technique Kobayashi is able to achieve an accuracy of **XX**% (Kobayashi et. al. 1997), similar results have been reported by **XXXX** and **XXXX**. Hidden Markov Models may also be used to interactively segment out glove input into individual gestures for recognition and perform online learning of new gestures(Lee 1996). In these cases gestures are typically recognized using pre-trained templates; however gloves can also be used to identify natural or untrained gestures. Wexelblat uses a top down and bottom up approach to recognize natural gestural features such as finger curvature and hand orientation, and temporal integration to produce frames describing complete gestures (Wexelblat 1995). These frames can then be passed to higher level functions for further interpretation.

Although instrumented gloves provide very accurate results they are expensive and encumbering. Computer vision techniques can also be used for gesture recognition overcoming some of these limitations. A good review of vision based gesture recognition is provided by Palovic et. al. (1995). In general, vision based systems are more natural to use that glove interfaces, and are capable of excellent hand and body tracking, but do not provide the same accuracy in pose determination. How ever for many applications this may not be important. Sturman and Zeltzer point out the following limitations for image based visual tracking of the hands (Sturman and Zeltzer 1994):

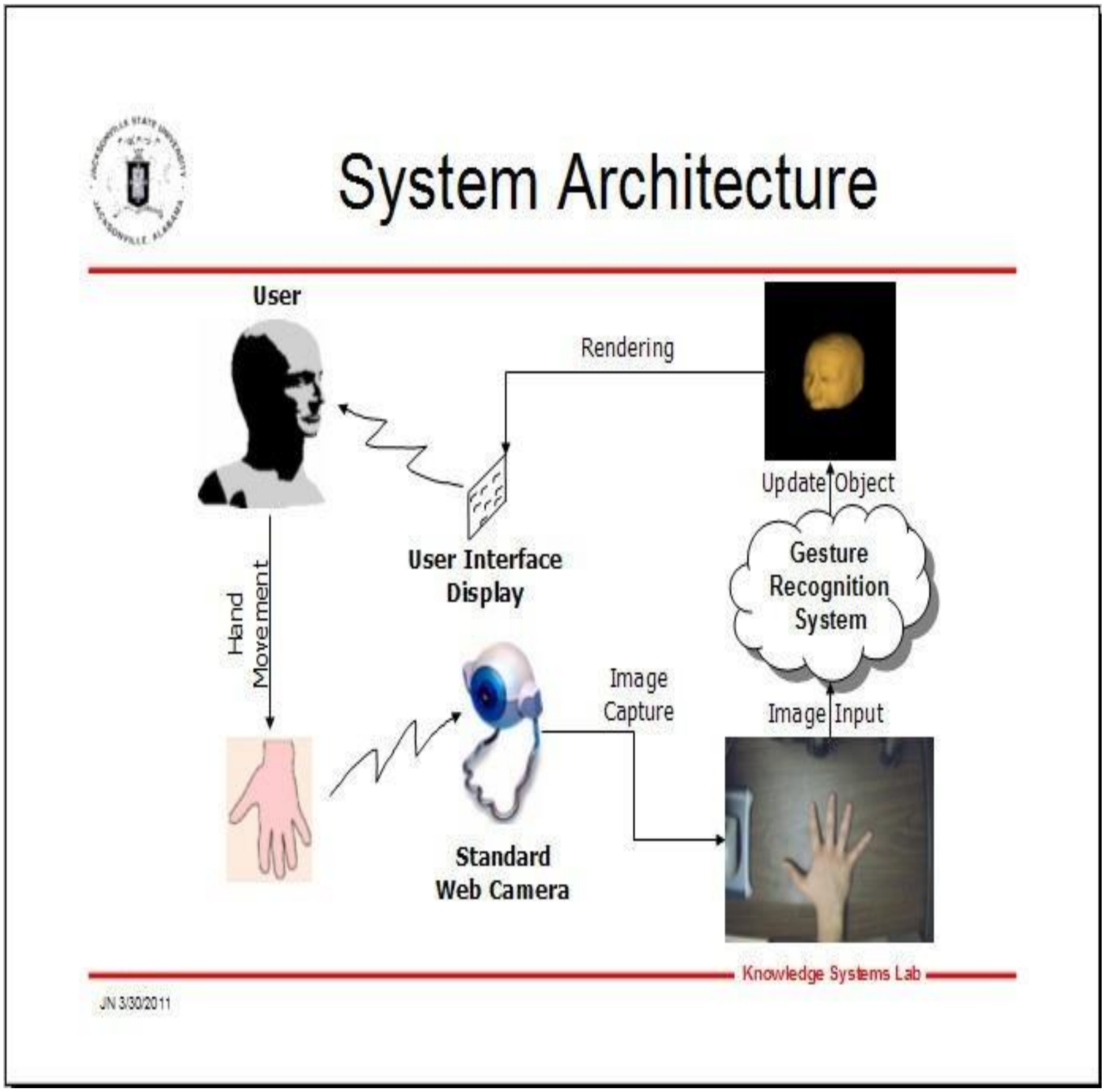
- The resolution of video cameras is too low to both resolve the fingers easily and coverthe field of view encompassed by broad hand motions.
- The 30- or 60- frame-per-second conventional video technology is insufficient to capture rapid hand motion.
- Fingers are difficult to track as they occlude each other and are occluded by the hand.

There are two different approaches to vision based gesture recognition; model based techniques which try to create a three-dimensional model of the users hand and use this for recognition, and image based techniques which calculate recognition features directly from the hand image. Rehg and Kanade (1994) describe a vision- based approach that uses stereo camera to create a cylindrical model of the hand. They use finger tips and joint links as features to align the cylindrical components of the model. Etoh, Tomono and Kishino **(**1991) report similar work, while Lee and Kunii use kinematic constraints to improve the model matching and recognize 16 gestures with XX% accuracy (1993). Image based methods typically segment flesh tones from the background images to find hands and then try and extract features such as fingertips, hand edges, or gross hand geometry for use in gesture recognition. Using only a coarse description of hand shape and a hidden markov model, Starner and Pentland are able to recognize 42 American Sign Language gestures with 99% accuracy (1995). In contrast, Martin and Crowley calculate the principle components of gestural images and use these to search the gesture space to match the target gestures (1997).

Gesture types

In computer interfaces, two types of gestures are distinguished:

- Offline gestures: Those gestures that are processed after the user interaction withthe object. An example is the gesture to activate a menu.
- Online gestures: Direct manipulation gestures. They are used to scale or rotate atangible object.

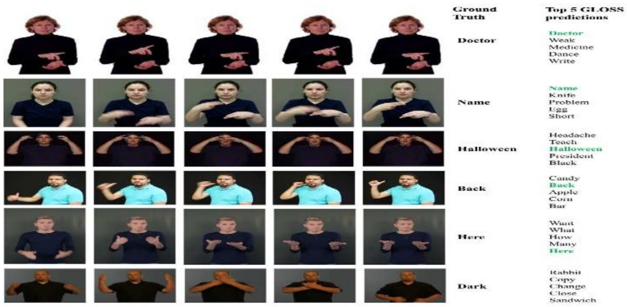


Here we can see that the user action is captured by a **camera** and the **image input** is fed into the **gesture recognition system** , in which it is processed and compared efficiently with the help of an **algorithm**. The virtual object or the **3-d model** is then updated accordingly and the user interfaces with machine with the help of a **user interface display.**

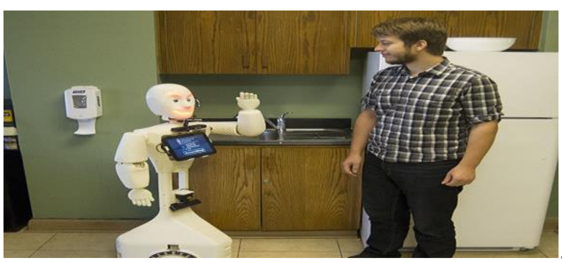
Uses

Gesture recognition is useful for processing information from humans which is notconveyed through speech or type. As well, there are various types of gestures which can be identified by computers.

- **Sign language recognition**. Just as speech recognition can transcribe speech to text, certain types of gesture recognition software can transcribe the symbols represented through sign language into text.



- **For socially assistive robotics**. By using proper sensors (accelerometers and gyros) worn on the body of a patient and by reading the values from thosesensors, robots can assist in patient rehabilitation. The best example can be stroke rehabilitation

.

- **Directional indication through pointing**. Pointing has a very specific purpose inour society, to reference an object or location based on its position relative to ourselves. The use of gesture recognition to determine where a person is pointing is useful for identifying the context of statements or instructions. This application is of particular interest in the field of robotics.
- **Control through facial gestures**. Controlling a computer through facial gestures is a useful application of gesture recognition for users who may not physically be able to use a mouse or keyboard. **Eye tracking** in particular may be of use for controlling cursor motion or focusing on elements
- **Alternative computer interfaces**. Foregoing the traditional keyboard and mouse setup to interact with a computer, strong gesture recognition could allow users to accomplish frequent or common tasks using hand or face gestures to a cameraof a display.

- **Immersive game technology**. Gestures can be used to control interactions withinvideo games to try and make the game player's experience more interactive or immersive

- **Virtual controllers**. For systems where the act of finding or acquiring a physical controller could require too much time, gestures can be used as an alternative control mechanism. Controlling secondary devices in a car, or controlling a television set are examples of such usage.
- **Affective computing**. In affective computing, gesture recognition is used in theprocess of identifying emotional expression through computer systems.
- **Remote control**. Through the use of gesture recognition, "remote control with thewave of a hand" of various devices is possible. The signal must not only indicate the desired response, but also which device to be controlled.

Input devices

The ability to track a person's movements and determine what gestures they may be performing can be achieved through various tools. Although there is a large amount of research done in image/video based gesture recognition, there is some variation within the tools and environments used between implementations.

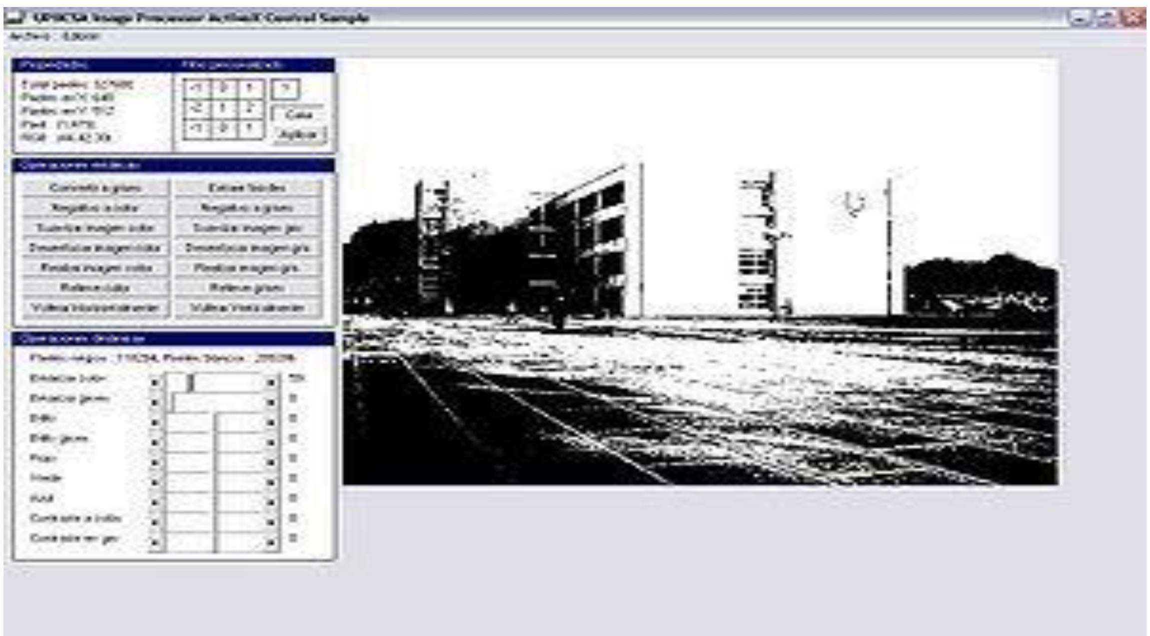
- **Depth-aware cameras**. Using specialized cameras such as **time-of-flight cameras**, one can generate a depth map of what is being seen through the camera at a short range, and use this data to approximate a 3d representation of what is being seen. These can be effective for detection of hand gestures due to their shortrange capabilities.
- **Stereo cameras**. Using two cameras whose relations to one another are known, a 3d representation can be approximated by the output of the cameras. To get the cameras' relations, one can use a positioning reference such as a lexian-stripe or **infrared** emitters. In combination with direct motion measurement **(6D-Vision)** gestures can directly be detected.
- **Controller-based gestures**. These controllers act as an extension of the body so that when gestures are performed, some of their motion can be conveniently captured by software. Mouse gestures are one such example, where the motion of the mouse is correlated to a symbol being drawn by a person's hand, as is the Wii Remote, which can study changes in acceleration over time to represent gestures.
- **Single camera**. A normal camera can be used for gesture recognition where the resources/environment would not be convenient for other forms of image-based recognition. Although not necessarily as effective as stereo or depth aware cameras, using a single camera allows a greater possibility of accessibility to a wider audience.

Proposed Approach

System Runtime:

- Current time – 41 ms for one image from camera

- Processing Capability on 1.6 Ghz Athlon:
    - 24 fps



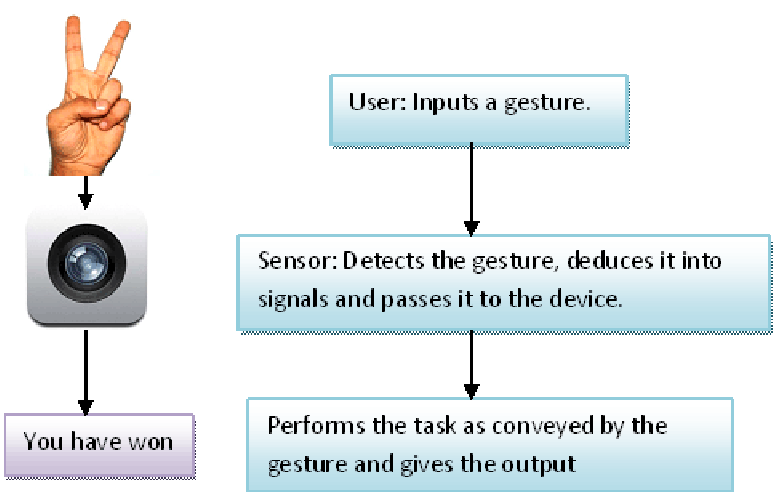
Monochrome black/white image

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

Image processing

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

- Euclidean geometry transformations such as enlargement, reduction, and rotation
- Color corrections such as brightness and contrast adjustments, color mapping,color balancing, quantization, or color translation to a different color space
- Digital compositing or optical compositing (combination of two or more images),which is used in film-making to make a "matte"
- Interpolation, demosaicing, and recovery of a full image from a raw image formatusing a Bayer filter pattern
- Image registration, the alignment of two or more images
- Image differencing and morphing
- Image recognition, for example, may extract the text from the image using opticalcharacter recognition or checkbox and bubble values using optical mark recognition
- Image segmentation
- High dynamic range imaging by combining multiple images
- Geometric hashing for 2-D object recognition with affine invariance



-

Challenges

There are many challenges associated with the accuracy and usefulness of gesture recognition software. For image-based gesture recognition there are limitations on the equipment used and image noise. Images or video may not be under consistent lighting, or in the same location. Items in the background or distinct features of the users maymake recognition more difficult.

The variety of implementations for image-based gesture recognition may also cause issue for viability of the technology to general usage. For example, an algorithm calibrated for one camera may not work for a different camera. The amount of background noise also causes tracking and recognition difficulties, especially when occlusions (partial and full) occur. Furthermore, the distance from the camera, and the camera's resolution and quality, also cause variations in recognition

In order to capture human gestures by visual sensors, robust computer vision methods arealso required, for example for hand tracking and hand posture recognition or for capturing movements of the head, facial expressions or gaze direction accuracy.

.

Gorilla arm

"Gorilla arm" was a side-effect that destroyed vertically-oriented touch-screens as amainstream input technology despite a promising start in the early 1980s.

Designers of touch-menu systems failed to notice that humans are not designed to hold their arms in front of their faces making small motions. After more than a very few selections, the arm begins to feel sore, cramped, and oversized—the operator looks like a gorilla while using the touch screen and feels like one afterwards. This is now considered a classic cautionary tale to human-factors designers; "Remember the gorilla arm!" is shorthand for "How is this going to fly in real use?".

Upcoming New Technologies:-

The Sixth Sense Device:-

Sixth sense is a wearable gestural interface device developed by Pranav Mistry, a PhD student in the Fluid Interfaces Group at the MIT Media Lab. It is similar to Telepointer, a neckworn projector/camera system developed by Media Lab student Steve Mann (which Mann originally referred to as "Synthetic Synesthesia of the Sixth Sense"). The Sixth sense prototype is comprised of a pocket projector, a mirror and a camera. The hardware components arecoupled in a pendant like mobile wearable device. Both the projector and the camera are connected to the mobile computing device in the user’s pocket. Theprojector projects visual information enabling surfaces, walls and physical objects around us to be used as interfaces; while the camera recognizes and tracks user's hand gestures and physical objects using computer-vision based techniques. The software program processes the video stream data capturedby the camera and tracks the locations of the colored markers (visual tracking fiducials) at the tip of the user’s fingers using simple computer-vision techniques. The movements and arrangements of these fiducials are interpreted into gestures that act as interaction instructions for the projected application interfaces. The maximum number of tracked fingers is only constrained by the number of unique fiducials, thus Sixth Sense also supports multi-touch And multi-user interaction.

The Sixth Sense prototype implements several applications that demonstrate the usefulness, viability and flexibility of the system. The map application lets the user navigate a map displayed on a nearby surface using hand gestures, similar to gestures supported by Multi-Touch based systems, letting the user zoom in, zoom out or pan using intuitive hand movements. The drawingapplication lets the user draw on any surface by tracking the fingertip movements of the user’s index finger. Sixth Sense also recognizes user’s freehand gestures (postures). For example, the Sixth Sense system implement

Gestural camera

a gestural camera that takes photos of the scene the user is looking at by detecting the 'framing' gesture. The user can stop by any surface or wall and flick through the photos he/she has taken. Sixth sense also lets the user draw icons or symbols in the air using the movement of the index finger and recognizes those symbols as interaction instructions. For example, drawing a magnifying glass symbol takes the user to the map application or drawing an '@' symbol lets the user check his mail. The Sixth sense system also augments physical objects the user is interacting with by projecting more information about these objects projected on them. For example, a newspaper can showlive video news or dynamic information can be provided on a regular piece of paper. The gesture of drawing a circle on the user's wrist projects an analog watch.

Construction and Working: -

The Sixth sense prototype comprises a pocket projector, a mirror and a camera contained in a pendant like, wearable device. Both the projector and the camera are connected to amobile computing device in the user's pocket. The projector projects visual information enabling surfaces, walls and physical objects around us to be used as interfaces; while the camera recognizes and tracks user's hand gestures and physical objects using computer-vision based techniques. The software program processes the video stream data capturedby the camera and tracks the locations of the colored markers (visual tracking fiducials) at the tips of the user's fingers. The movements and arrangements ofthese fiducials are interpreted into gestures that act as interaction instructions for the projected application interfaces. Sixth sense supports multi-touch and multi-user interaction.

Example Applications: -

The Sixth sense prototype contains a number of demonstration applications.

- The map application lets the user navigate a map displayed on a nearby surfaceusing hand gestures to zoom and pan
- The drawing application lets the user draw on any surface by tracking thefingertip movements of the user’s index finger.
- Sixth sense also implements Augmented reality; projecting information ontoobjects the user interacts with.

The system recognizes a user's freehand gestures as well as icons/symbols drawn in the air with the index finger,

for example:

- A 'framing' gesture takes a picture of the scene. The user can stop by any surfaceor wall and flick through the photos he/she has taken.
- Drawing a magnifying glass symbol takes the user to the map application whilean ‘@’ symbol lets the user check his mail.
- The gesture of drawing a circle on the user’s wrist projects an analog watch.

Intel’s Gesture Technology:-

What's Next? Gesture Recognition Technology from Intel Labs allows you to interact with and control devices using simple hand gestures. Imagine a world where gestures like turning an "air knob" could turn up the volume on your TV or waving your hand would answer a phone that's in your pocket.

According to a , the target applications for AVX are interface technology to control gaming and entertainment. Intel expects that this forthcoming technology would reduce the need for specialized DSPs and GPUs. Smart computing is here.. Yes visibly smart. But my personal opinion would be that intel would make people more lazy by the launch of the next-generation gesture recognition technology. Its amazing to just thing about the world where we can control TV

Gesture Tek: -

Gesture Tek's Illuminate interactive multi-touch surface computing technology witha motion sensing gesture control interface lets users navigate interactive content on afloating panel, multimedia kiosk, multi touch surface screen, interactive table or interactive window. Surfaces can be configured with a multi-touch interface for multi- touch or multi-point interaction.

with no projector or hardware to be seen, the effect is unforgettable as GestureTek's dynamic interactive displays react to every point of your finger or wave of your hand, delivering a rich, interactive experience.

The hand tracking system lets you control multi-media in ways you never imagined, transforming an ordinary surface into an interactive multi-touch surface computing platform. Illuminate surfaces are available as interactive multi-touch display panels and windows, interactive kiosks and multi-touch tables. Multi-touch interactive surface displays come turnkey or can be customized to virtually any shape or size.

Gesture Tek's Illuminate 'point to control' and touch screen computing surfaces are popular in bars, nightclubs, retail stores, museums, science centers, real estate showrooms and corporate presentation centers - anywhere menu-based multi-media content is used for edutainment, entertainment, or to present advertising and corporate information.

References

1. ^ Matthias Rehm, Nikolaus Bee, Elisabeth André, Wave Like an Egyptian - Accelerometer Based Gesture Recognition for Culture Specific Interactions, British Computer Society, 2007
2. ^ Pavlovic, V., Sharma, R. & Huang, T. (1997), "Visual interpretation of handgestures for human-computer interaction: A review", IEEE Trans. Pattern Analysis and Machine Intelligence., July, 1997. Vol. 19(7), pp. 677 -695.
3. ^ R. Cipolla and A. Pentland, Computer Vision for Human-Machine Interaction, Cambridge University Press, 1998, ISBN 978-0521622530
4. ^ Ying Wu and Thomas S. Huang, "Vision-Based Gesture Recognition: A Review", In: Gesture-Based Communication in Human-Computer Interaction,Volume 1739 of Springer Lecture Notes in Computer Science, pages 103-115,1999, ISBN 978-3-540-66935-7, doi 10.1007/3-540-46616-9
5. ^ Alejandro Jaimesa and Nicu Sebe, Multimodal human–computer interaction: A survey, Computer Vision and Image Understanding Volume 108, Issues 1-2, October–November 2007, Pages 116-134 Special Issue onVision for Human-Computer Interaction, doi:10.1016/j.cviu.2006.10.019
6. ^ We consider online gestures, which can also be regarded as direct manipulations like scaling and rotating. In contrast, offline gestures are usually processed after the interaction is finished; e. g. a circle is drawn to activate a context menu.
7. ^ Thad Starner, Alex Pentland, Visual Recognition of American SignLanguage Using Hidden MarkovModels, Massachusetts Institute of Technology
8. ^ Kai Nickel, Rainer Stiefelhagen, Visual recognition of pointing gestures for human-robot interaction, Image and Vision Computing, vol 25, Issue 12,December 2007, pp 1875-1884
9. ^ Lars Bretzner and Tony Lindeberg "Use Your Hand as a 3-D Mouse ...", Proc. 5th European Conference on Computer Vision (H. Burkhardt and B. Neumann, eds.), vol. 1406 of Lecture Notes in Computer Science, (Freiburg, Germany), pp. 141--157, Springer Verlag, Berlin, June 1998.
10. ^ Matthew Turk and Mathias Kölsch, "Perceptual Interfaces", University of California, Santa Barbara UCSB Technical Report 2003-33

11. ^ M Porta "Vision-based user interfaces: methods and applications",International Journal of Human-Computer Studies, 57:11, 27-73, 2002. ^ Afshin Sepehri, Yaser Yacoob, Larry S. Davis "Employing the Hand as an Interface Device", Journal of Multimedia, vol 1, number 2, pages 18-29
12. "Telepointer: Hands-Free Completely Self Contained Wearable Visual Augmented Reality without Headwear and without any Infrastructural Reliance", IEEE International Symposium on Wearable Computing (ISWC00), pp. 177, 2000, Los Alamitos, CA, USA
13. "Cyborg: Digital Destiny and Human Possibility in the Age of the Wearable Computer", Steve Mann with Hal Niedzviecki, ISBN 0385658257 (Hardcover), Random House Inc, 304 pages, 2001.
14. ^ Intelligent Image Processing, John Wiley and Sons, 384pp, 02001NOV02, ISBN 0-471-40637-6
15. ^ sixth sense. Pranav Mistry.
16. ^ CNet News: MIT's 6th Sense device could trump Apple's multitouch
17. ^ New York Times - At TED, Virtual Worlds Collide With Reality
18. ^ http://www.ted.com/talks/pranav_mistry_the_thrilling_potential_of_sixthsense_technology.html TED Talks - Pranav Mistry: The thrilling potential ofSixth sense technology