Innovative ICT Solutions
for the Societal Challenges

Behind the scenes of Virtual Reality: face, gesture, gait and posture recognition, 360 degree video and computer graphics

Agata Manolova

amanolova@tu-sofia.bg/ http://telilab.tu-sofia.bg/

Technical University of Sofia
Faculty of Telecommunications
Bulgaria

24.05.2017, Valencia

UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Overview

- Defining virtual reality
- Four Key Elements of Virtual Reality Experience
- A brief History of VR
- Interface to the Virtual World-Input
- Interface to the Virtual World-Output
- Representation of the Virtual World
- Interacting with the Virtual World
- The Virtual Reality Experience

Virtual Reality Definition

- What is virtual reality?
  - Virtual –
    - being in essence or effect, but not in fact
  - Reality –
    - the state or quality of being real. Something that exists independently of ideas concerning it. Something that constitutes a real or actual thing as distinguished from something that is merely apparent."
  - What was the first VR?

Define VR

- Virtual reality is a high-end user-computer interface that involves real-time simulation and interactions through multiple sensorial channels. These sensorial modalities are visual, auditory, tactile, smell, and taste.

Virtual Reality

- What is “Virtual Reality”?
  - Brooks (1999) defines it as: “[an] experience .. in which the user is effectively immersed in a responsive virtual world” …
  - Sherman and Craig (2003) defines it as a medium composed of interactive computer simulations that sense the participant’s position and actions and replace or augment the feedback to one or more senses, giving the feeling of being mentally immersed or present in the simulation (a virtual world)
  - Immersive, Semi-immersive, Non-immersive

Virtual Reality

- Four Key Elements in Experiencing Virtual Reality
  - A virtual world an imaginary space, often (but not necessarily) realized through a medium (rendering pipeline, display, etc.)
  - Immersion (physical and mental) having a sense of “presence” within an environment; this can be purely a mental state, or can be accomplished through physical means.
    - Mental Immersion: a state of being deeply engaged, with suspension of disbelief.
    - Physical Immersion: bodily/physically entering into a projected area
  - Sensory feedback: visual/aural/haptic feedback to a participant
  - Interactivity: in a virtual reality experience, participants are able to move around and change their viewpoint, generally through movements of their head.

History of Virtual Reality

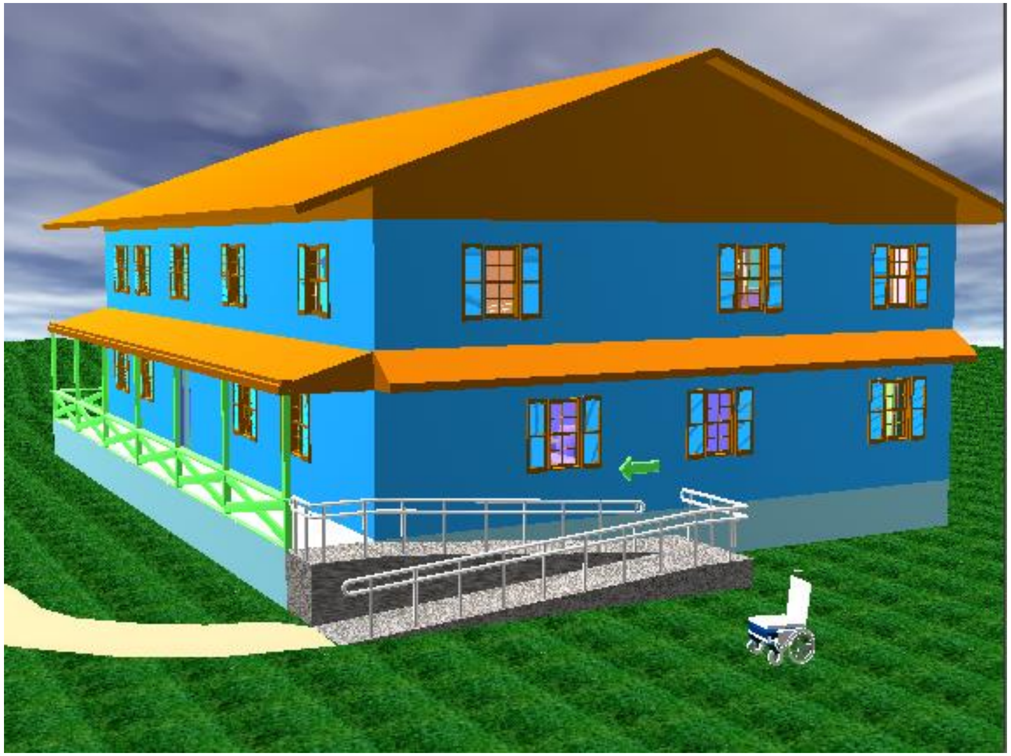
- The Ultimate Display – Ivan Sutherland, 1965 - Describe a virtual world provided by the computer
- Head Mounted Display (HMD) – Ivan Sutherland 1968 - Users are presented with left and right views of a computer generated 3D scene. The user's head movements caused corresponding spatial changes in the images.
- Evans & Sutherland Corp. manufacture Image Generators for pilot training starting in 1968

Ivan Sutherland's
The Ultimate Display

"Don't think of that thing as a screen, think of it as a window, a window through which one looks into a virtual world. The challenge to computer graphics is to make that virtual world look real, sound real, move and respond to interaction in real time, and even feel real."

Key Element 1: Virtual World

Three-dimensional “virtual world” or “virtual environment”:



Real-time 3D-Computer Graphics

Goals:

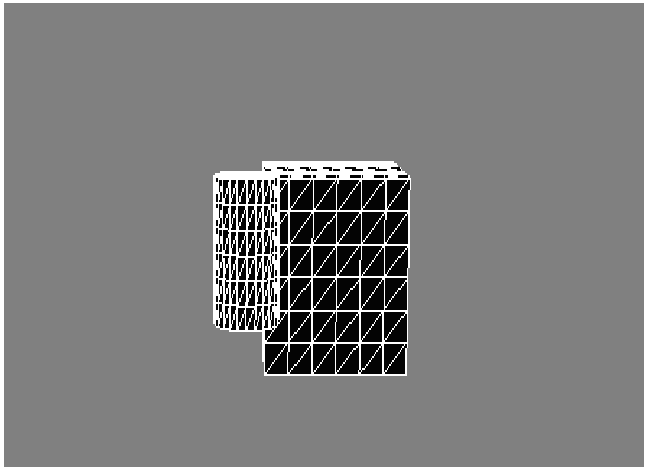
- Generating 3D worlds
- Visualize application domain specific data in >=3D
- Implement user interaction with the virtual world.
- Operate in real-time, minimize lag, maximize simulation rate.

Methods

- Computer graphics
  - Implement depth cues discussed earlier, e.g., mapping 3D world to 2D pictures to simulate user-eye-specific views in interactive rates (stereoscopy, immersion).
- CG methods differ in
  - primitives used and processed (e.g., vertex-, volume-, pixel-based, etc.).
  - lighting model and calculation used (local or global models).
  - direction of “lighting application”, shading (scanline shader, raytracer, etc.).
  - Mathematics (Trigonometry, linear algebra, analysis).

3D Graphics and Virtual World

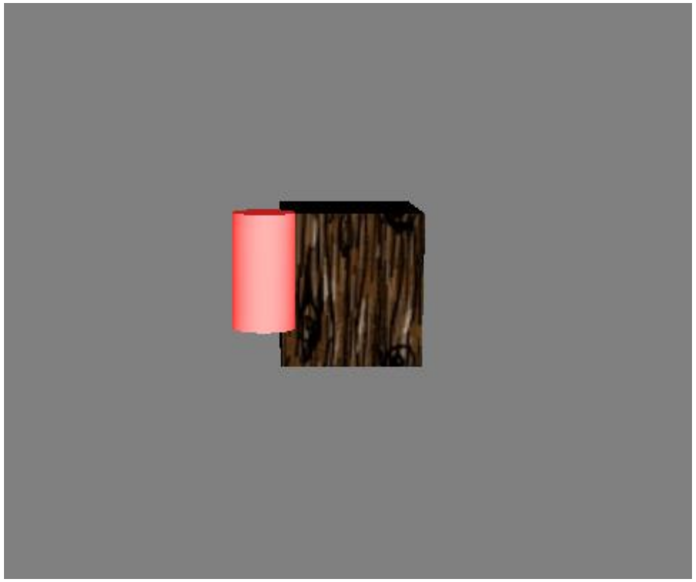
1) Model a 3D object: polygons, curves, colors, textures, transformations

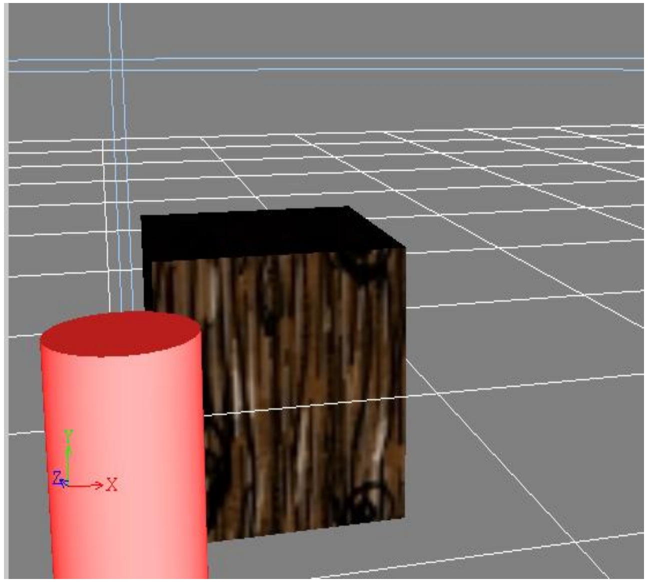


In computer graphics the most popular method for representing an object is the polygon mesh model.
A polygon mesh model consists of a structure of vertices, each vertex being a three-dimensional point. .

3D Graphics and Virtual World

2) Render the 3D objects: visibility, lights, camera, general illumination (shading, reflections, shininess, etc.)





3) Display the objects on a 2D screen: color, perspective, hardware (additional lights, shininess, camera position in 2nd picture)

3D Graphics and Virtual Wold

How do we move to a virtual environment? What else is needed?

Actions and animations, behaviors, interaction, immersion, audio, touch, gestures, smell, speech, stereo viewing, maybe more realism, physics

Nice intro video: https://www.youtube.com/watch?v=cvcAjgMUPUA

Key Element 2: Immersion

- **Mental Immersion**

  - sense of “presence”; state of being deeply engaged; suspension of disbelief; involvement
- **Physical Immersion**

  - bodily entering into a medium; synthetic stimulus of the body’s senses via the use of technology

Frequently embodied by an “avatar” – a virtual object used to represent a participant or physical object in a virtual world

Immersion:
Sense of “Presence”

The perception of being in a particular space or place.

- Physically immerse the participant in a computer-generated space.
- Provide computer-generated sensation to one or more of the human senses.

Key Element 3: Sensory Feedback

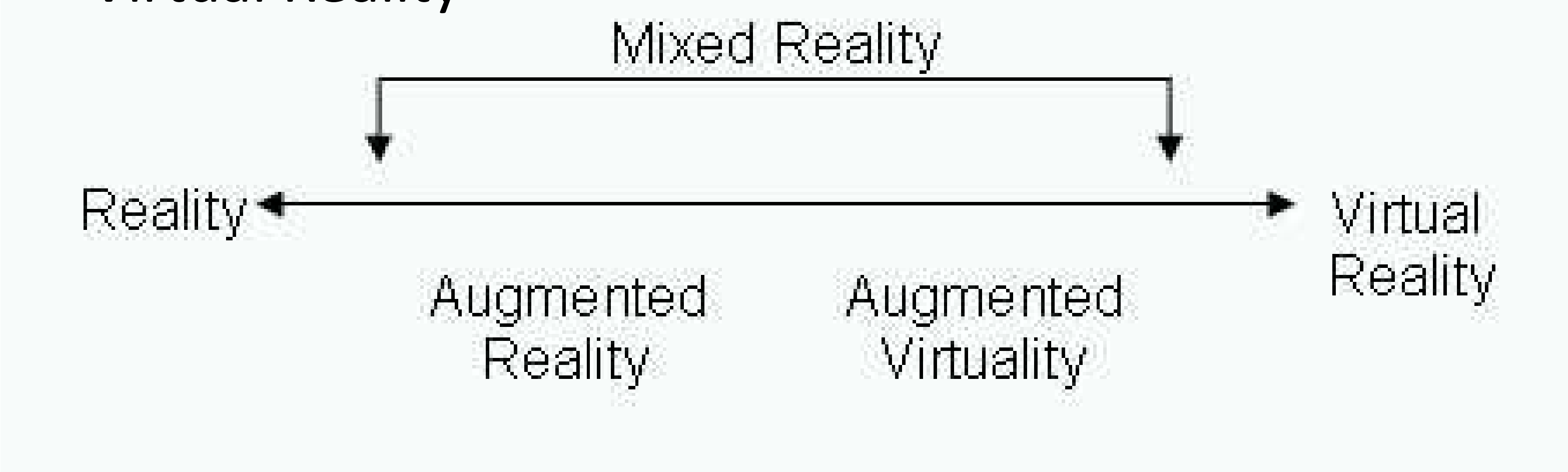
- visual
- aural
- haptic & tactile
- vestibular (balance: equilibrium, acceleration, and orientation gravity)
- olfactory, taste, magnetoreception

Key Element 4: Interactivity

- responsiveness to user actions
  - the ability to affect a virtual world
  - the ability to change one's viewpoint within a world
- collaborative environments
  - multiple users interacting within a virtual world and among themselves
- alternative realities
  - games, computer simulations of natural and artificial phenomenu, flight simulators

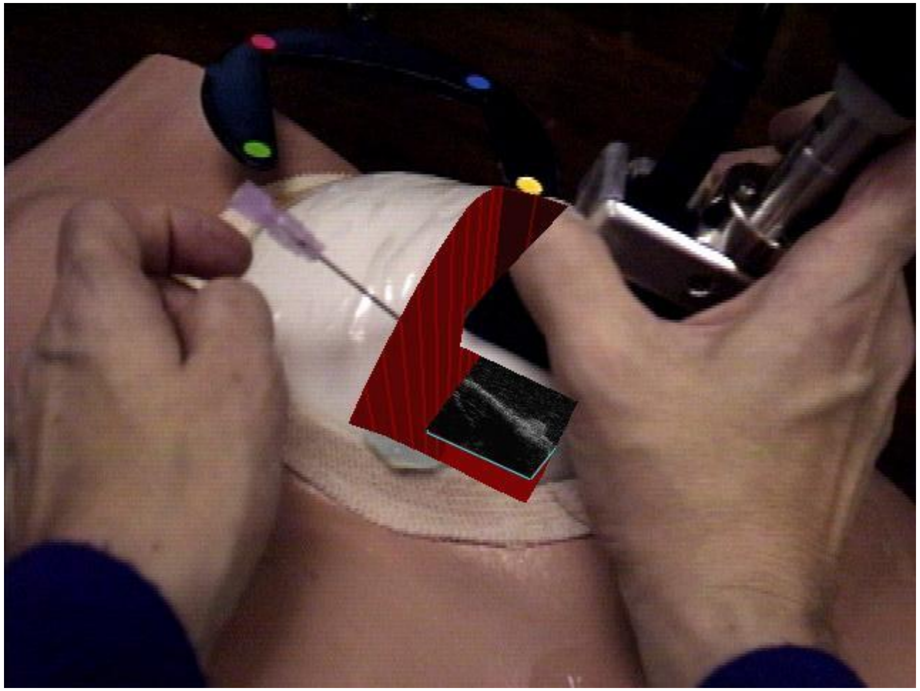
Virtual Environments

- Augmented Reality (Mixed Reality)
- Telepresence
- Artificial Reality
- Virtual Reality



Augmented Reality

- A combination of a real scene viewed by a user and a virtual scene generated by a computer that augments the scene with additional information.



Ultrasound Visualization Research at UNC – Chapel Hill

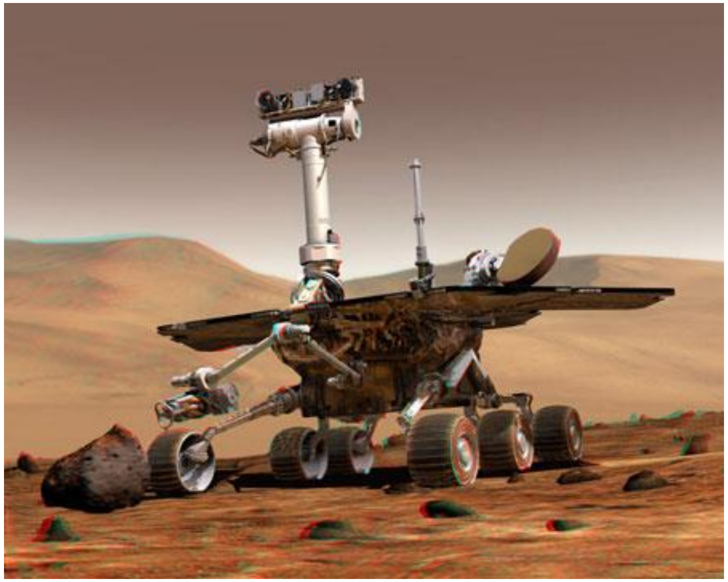
All Virtual Objects

All Real Objects

Telepresence

- The use of various technologies to produce the effect of placing the user in another location.

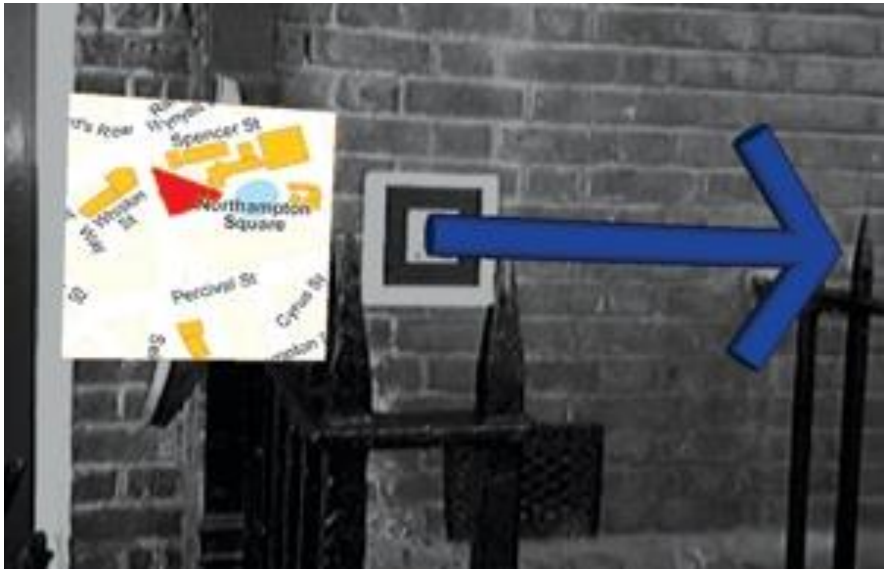
All Virtual Objects



All Real Objects

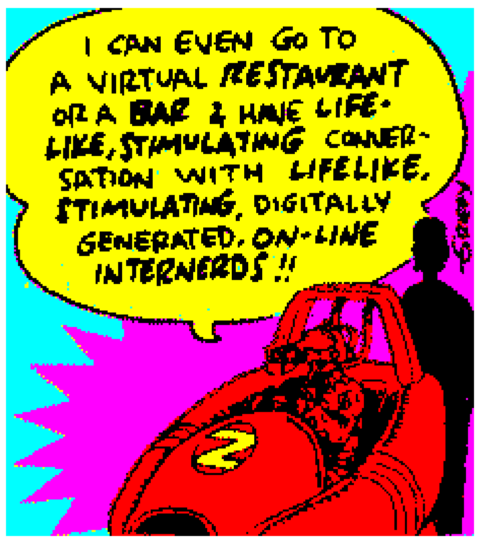
Mixed Reality

- Merging of the real and virtual worlds



Virtual Reality

- Ideal for VR is that everything you experience is computer-generated.



All Virtual Objects

All Real Objects

Some Terms in Virtual Reality

- Special effects on films are not VR if the actors are not aware of the virtual characters. These special effects are examples of computer animation.
- Cyberspace - A location that exits only in the minds of the participants, often as a result of technology that enables geographically distant people to interactively communicate.
- Avatar - A virtual object used to represent a participant or physical object in a virtual world. Adapted from Hindi, meaning the earthly embodiment of a deity.
- Many of the experts in the field consider the description Virtual Reality to be inappropriate. Virtual Environments (VE) is thought to be a more accurate description.

A Generic Virtual Reality System

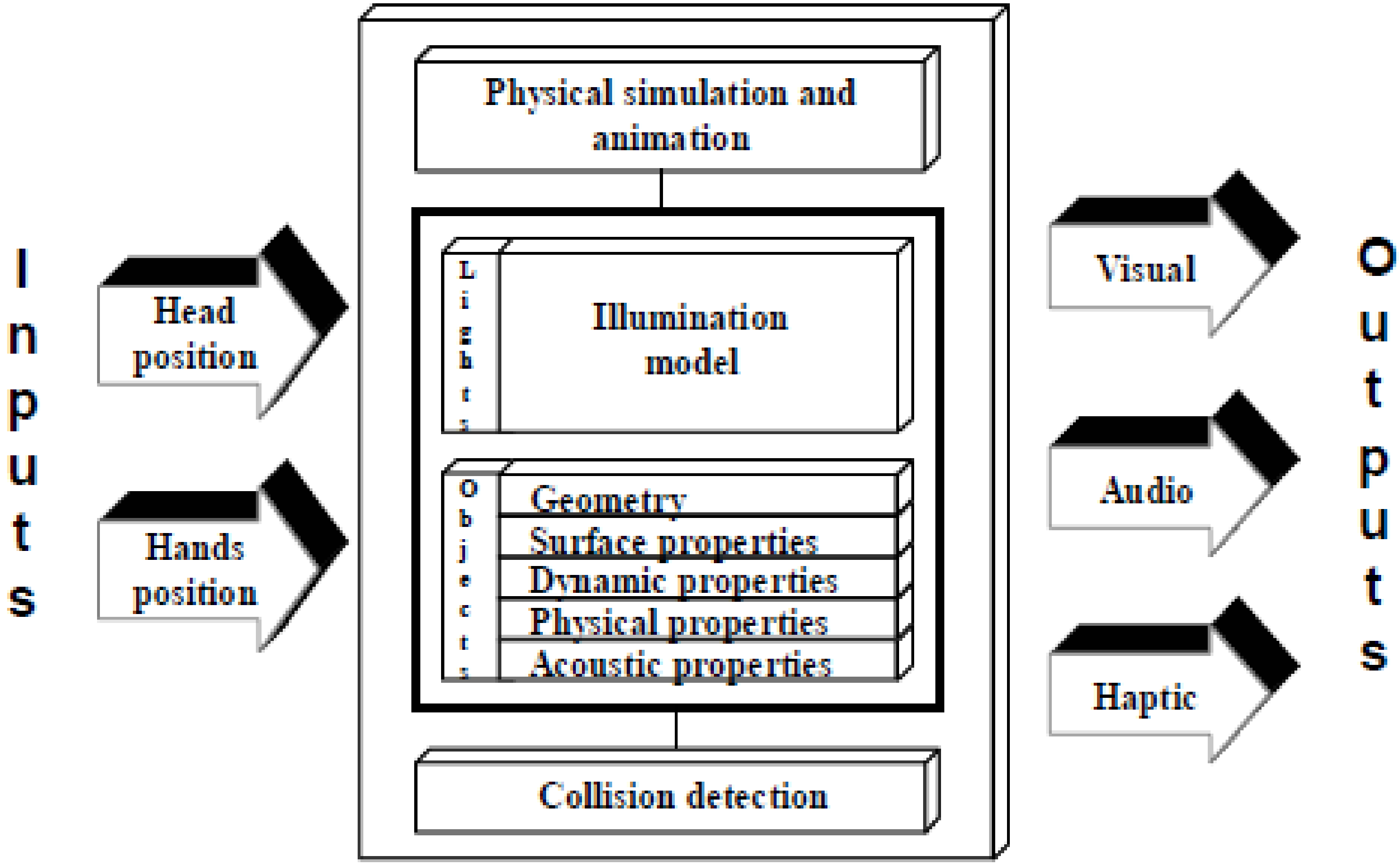


Figure: The inputs, processes and outputs in a generic VR system

Interface to the Virtual World – INPUT

- User monitoring
  - Monitoring of the participant’s actions in a VR experience.
- World monitoring
  - Information can be gathered and brought into the experience from sources not directly related to the participant.

User monitoring

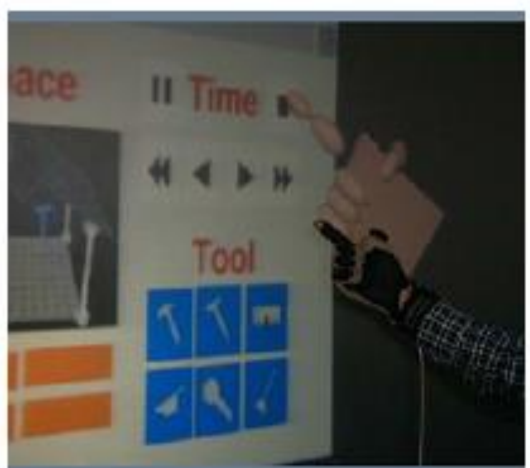
- User monitoring - user movements
- Position tracking
  - A position sensor is a device that reports its location and/or orientation to the computer.
  - The position sensor is the most important tracking device of any VR system
  - Tells the VR System where the users are located within a VR space.

User monitoring

- Body tracking
- Tracking the head
- Tracking the hand and fingers
- Tracking the eyes
- Tracking the torso
- Tracking the feet
- Tracking other body parts
- Indirect tracking

User monitoring

- The head is tracked in most VR systems -- to properly render and display the world.
- Tracking the hand and fingers is generally done to give the user a method of interacting with the world.

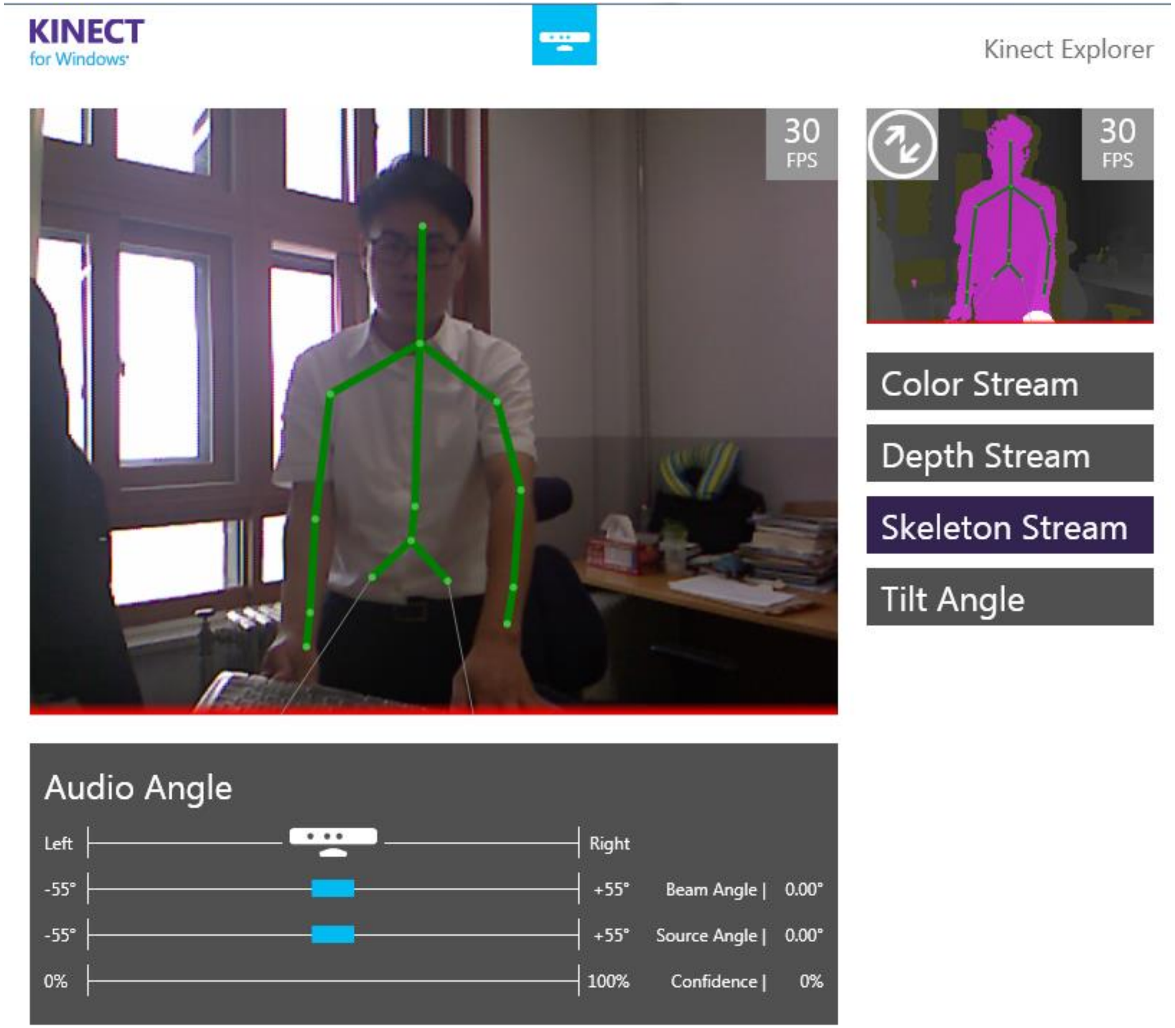


User monitoring

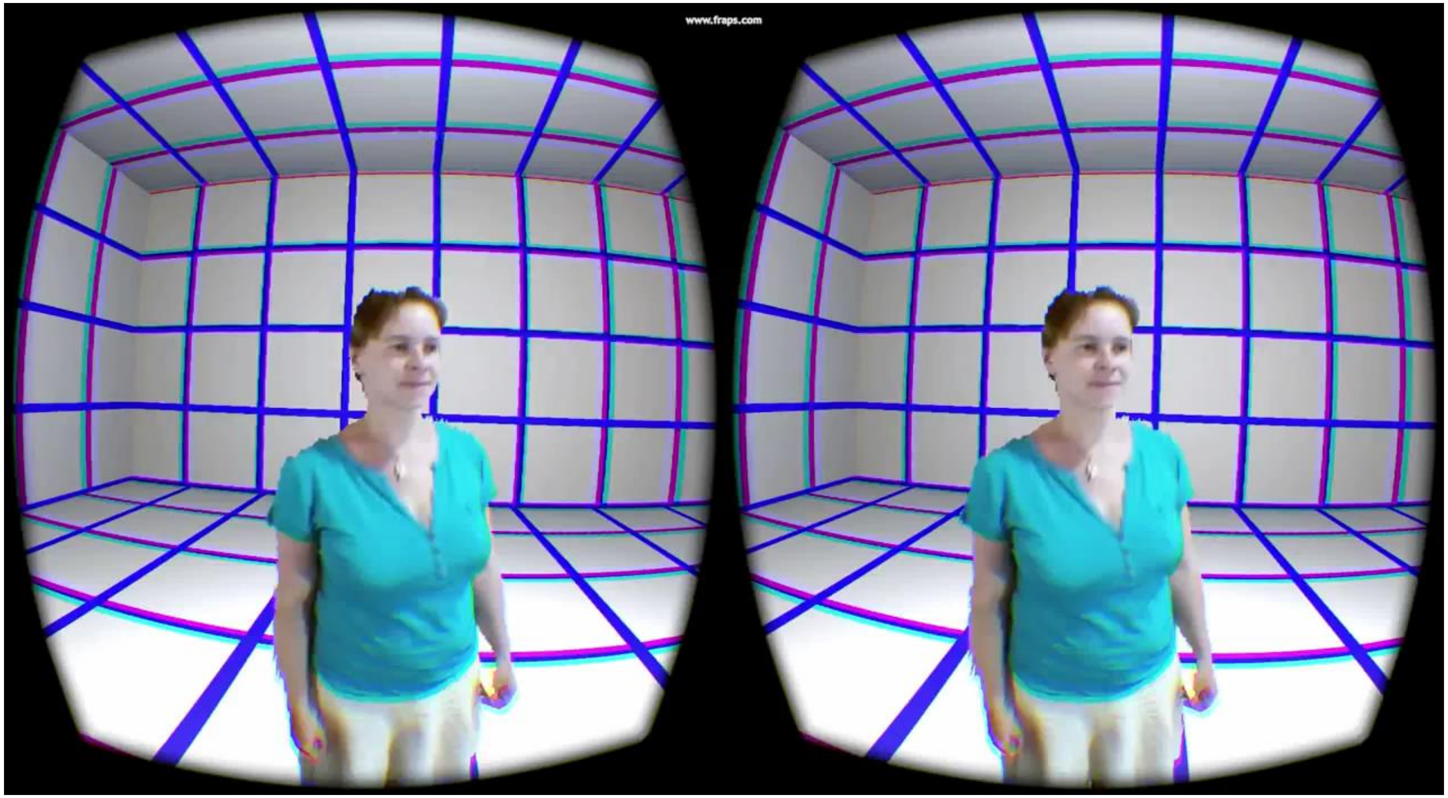
- Tracking the torso
- To properly render a full-body avatar requires either torso tracking, or some odd assumptions about the human body.

Body tracking - Kinect

Kinect ToolKit

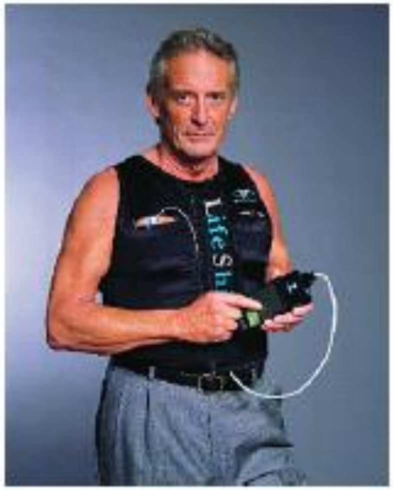



Oculus


www.fraps.com

User tracking

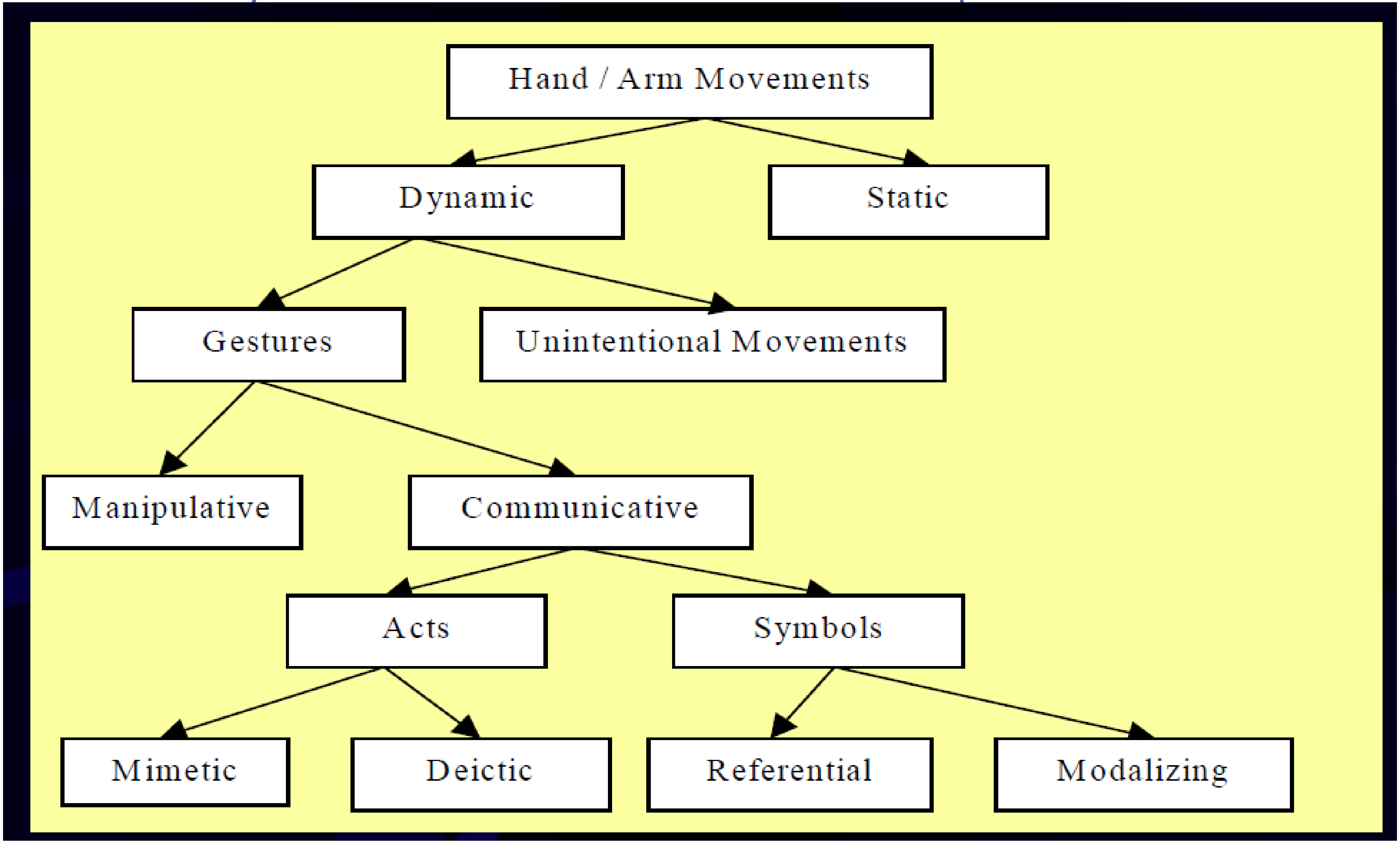
- Tracking other body parts
- Other aspects of a participant's "situation“ can be measured -- temperature, heart rate, perspiration, brain waves.
- Perhaps useful for medical/athletic evaluation, or done for artistic reasons, or training.



Gesture Recognition

- What is a gesture?
  - Body motion used for communication.
  - Body motion used for control.
- There are different types of gestures.
  - Hand gestures (e.g., waving goodbye).
  - Head gestures (e.g., nodding).
  - Body gestures (e.g., kicking).

Taxonomy of Gesture for Human computer Interaction



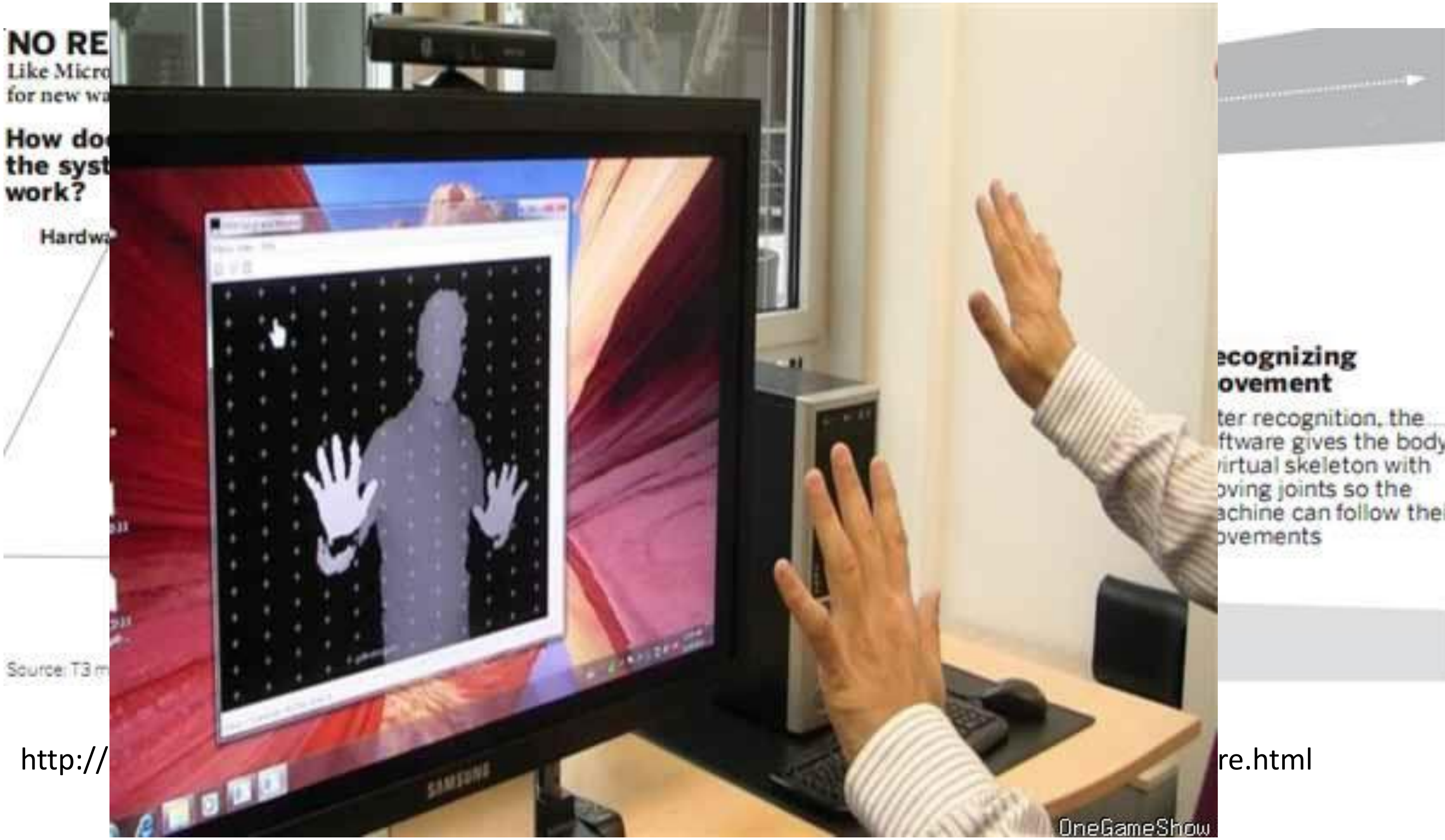
Input devices (1)

- The ability to track the hand movements and determine what gestures they may be performing can be achieved through various tools:
  - **Wired gloves**.

http://www.idrt.com/GestureRecognition.php

Input devices (2)

- Depth-aware cameras, stereo cameras or single camera.



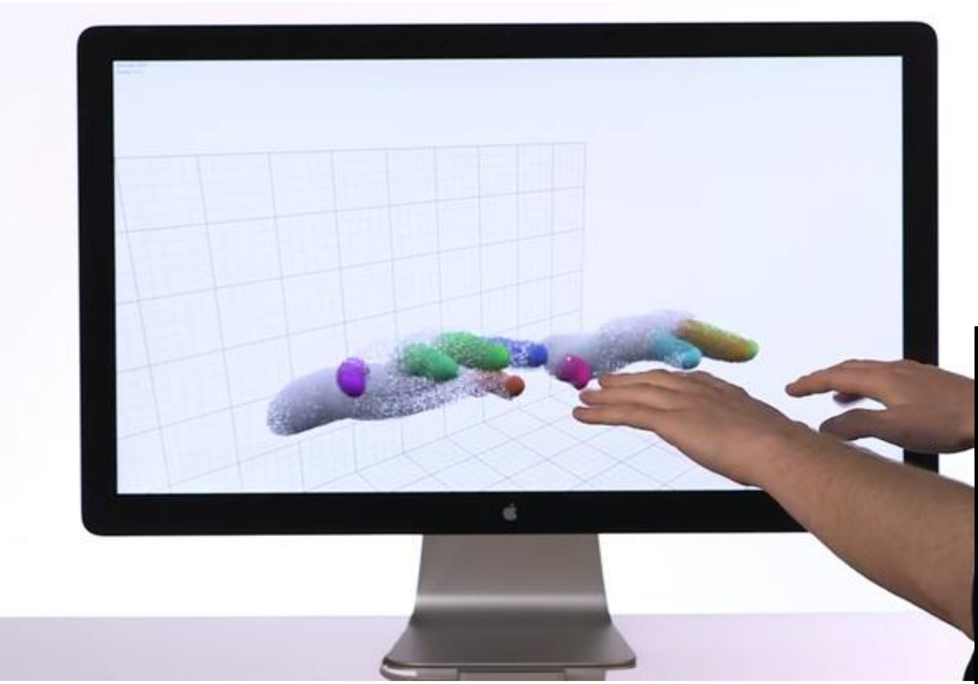
http:// re.html

http://www.wreporter.com/usa/microsoft-opens-up-its-system-kinect-pc-to-developers/

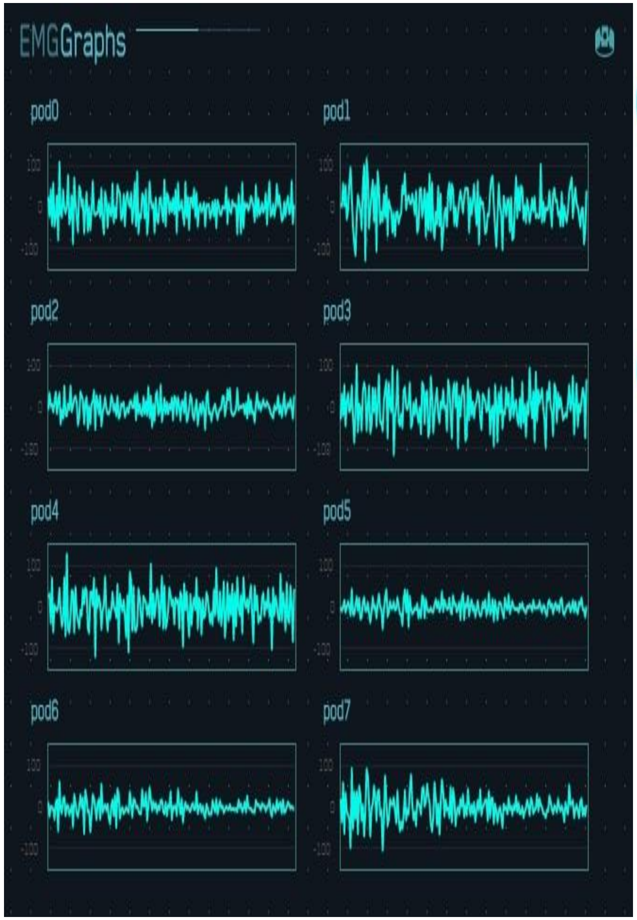
Input devices (3)

InnoSoc

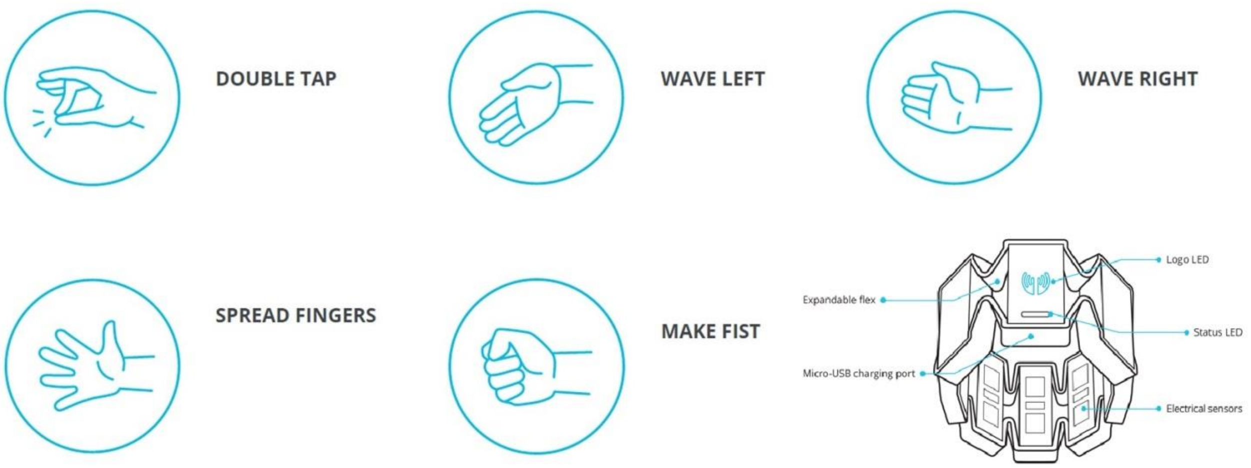
- Controller-based gestures

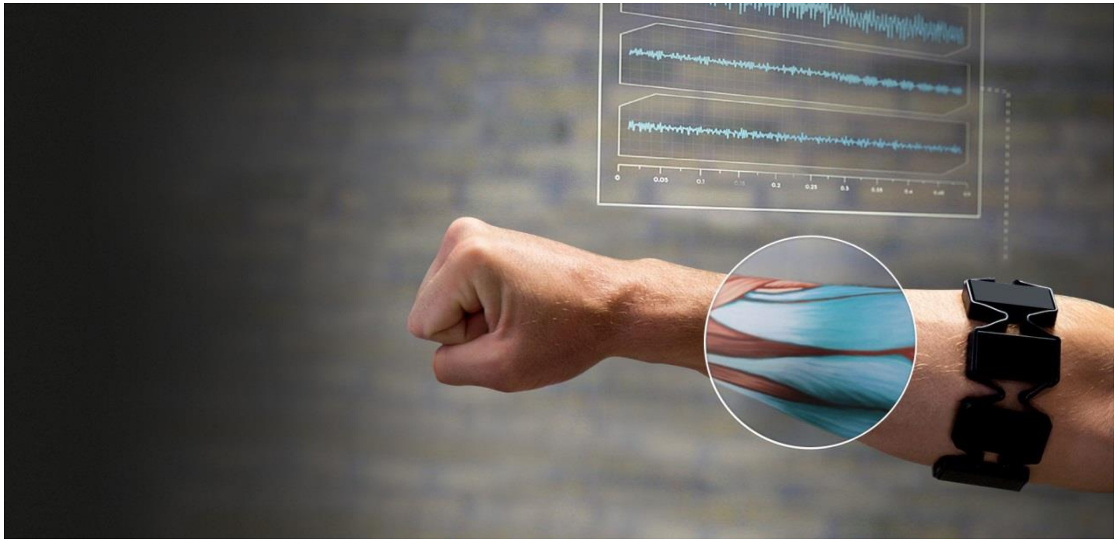


Input devices (4)

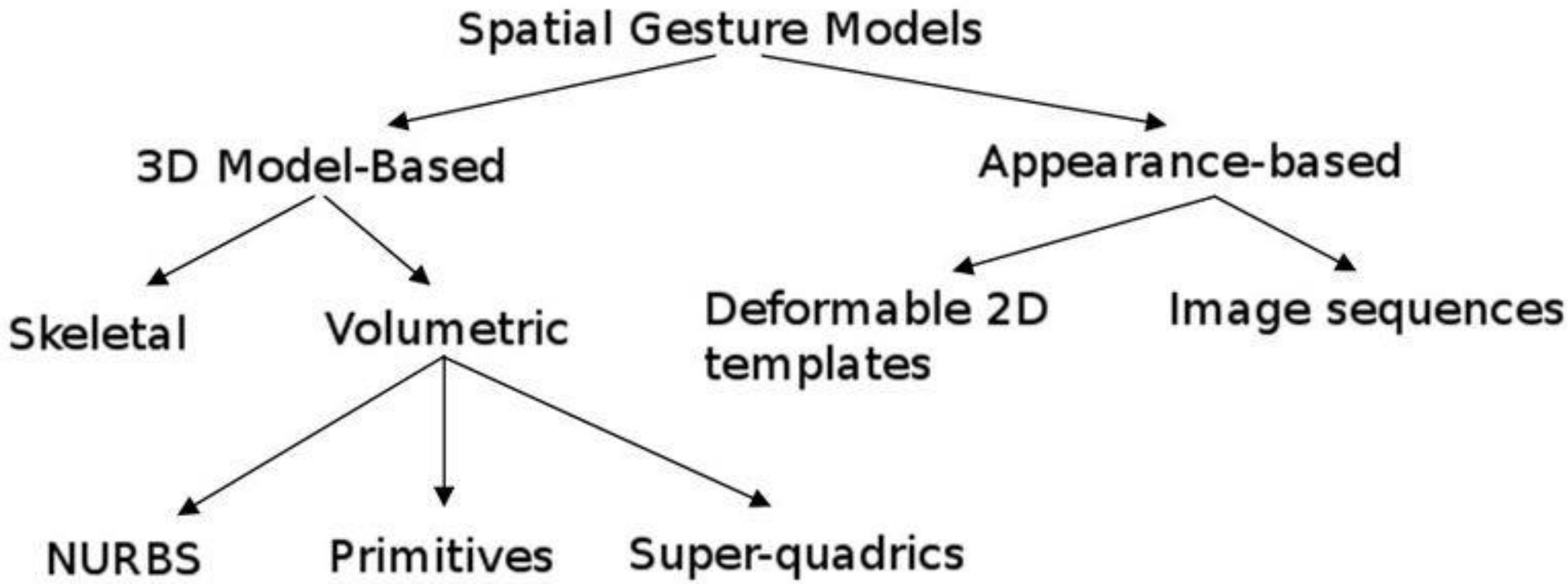


Output for each EMG pod for gesture “wave left” for one user





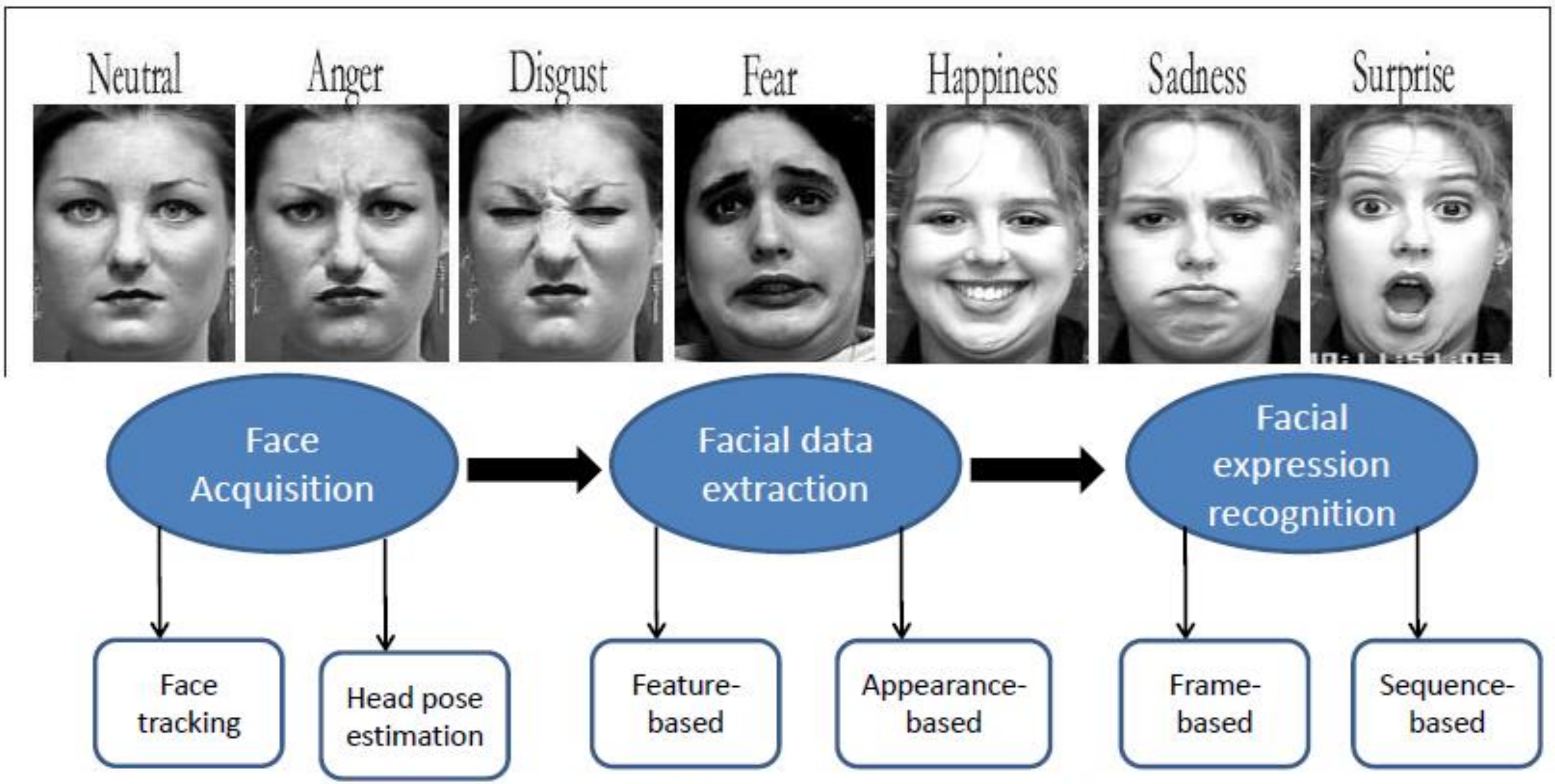
Algorithms for gesture recognition



http://en.wikipedia.org

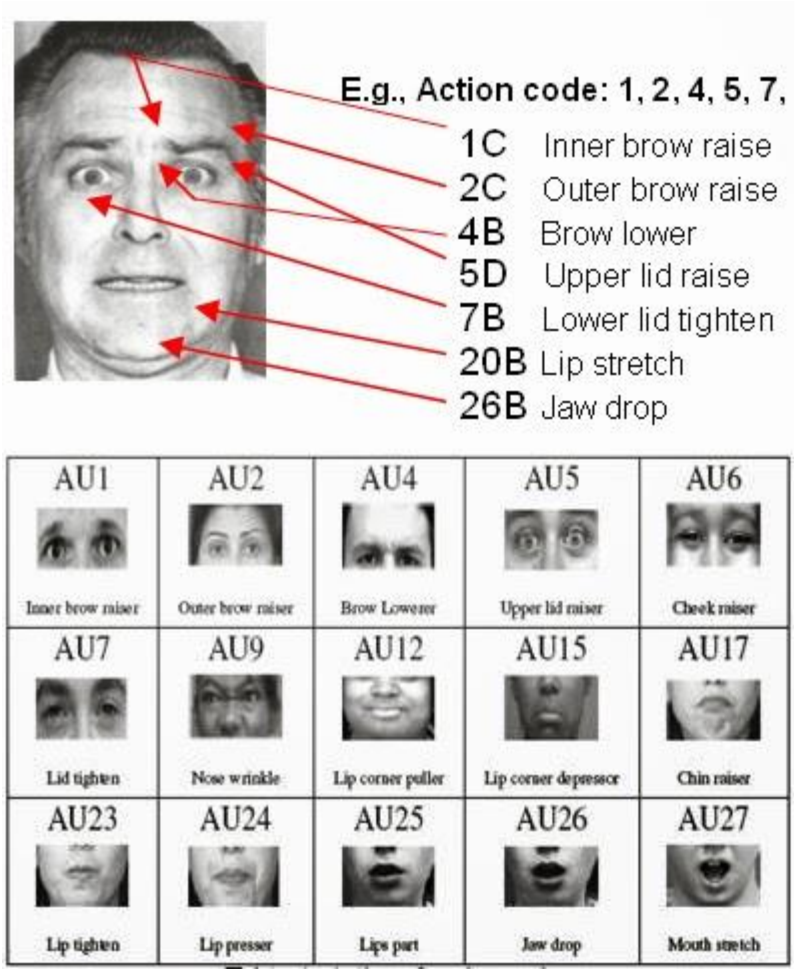
Gesture manipulation of 3D objects

Face and emotion recognition



Basic structure of facial expression analysis systems.

What is emotion made from?



Emotion recognition in real time

Physical Input Devices

- Other Physical Input Devices
  - In addition to user tracking, physical devices are another part of the interface between the user and the virtual world.
  - Physical Controls
  - Props
  - Platforms
  - Speech control

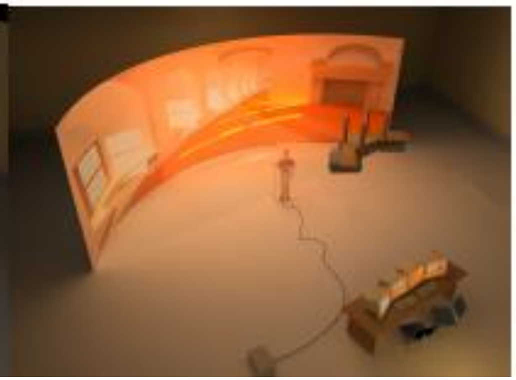
Physical Controls

- Props: A physical object used as an interface to a virtual world (generic or specific).
- The physical properties (shape, weight, texture, center of gravity, solidity) give a limited amount of haptic feedback, and often suggest how the device is used.
- Real nature of props allows user to easily manipulate the object. (e.g., hand it to another user).
- Realness of the prop may make the entire world seem more real.




Stationary VR Display platform

- Stationary VR Display platform
- CAVEs, ImmersaDesks
- Function as a type of generic platform.

World monitoring

World monitoring

- Information can be gathered and brought into the experience from sources not directly related to the participant.
- Applies to both:
  - Persistent virtual world applications.
  - Real world.

World monitoring

Bringing the Real world into the Virtual World:

- Real-world data can be gathered by video cameras and other measuring devices.
- There are several applications where the inclusion of real-world data may be advantageous (e.g., can help make a virtual world more realistic).
- Useful for scientists/engineers to monitor some phenomenon of the real world.
- Beneficial for steering a user away from real-world hazards.
- For creating a virtual replica of a real space (because it's too dangerous or difficult to get to the real space).

Interface to the Virtual World - OUTPUT

- A key component of a virtual reality experience is how the user perceives the environment.
- Their physical perception of the virtual world is based entirely on what the computer displays.
- There are three basic arrangements for all sensory displays:
  - Stationary
  - Head-based
  - Hand-based

Display

- Stationary displays
  - Are fixed in place.
  - The output is rendered to reflect the changing position of the user’s input sensory organs.
- Head-based displays
  - Are in some way attached to the user’s head and move in conjunction with the head.
  - Thus, visual screens remain in front of the user’s eyes and headphones on their ears.

Portable Displays

- Oculus Rift
- HTC Vive
- Samsung Gear VR
- Sony VR
- Fove

FOVE Eye Tracking Virtual Reality Headset

- FOVE uses a combination of hardware and software in order to identify where the iris of the user is pointed at;
- The system knows where the user is looking allows it to focus its graphical processing power on that specific area;
- It has a higher resolution than both Oculus Rift and the HTC Vive - 2560 X 1440 DISPLAY, FRAME RATE: 70FPS;

FOVE

- http://eyeplaythepiano.com/en/

Properties of Visual Display Paradigms

- Color
- Resolution
- Contrast
- Brightness
- Number of display channels
- Focal distance
- Opacity
- Masking
- Field of view
- Field of regard
- Head position information
- Cruciality of graphics latency
- Frame rate

Properties of Visual Display Paradigms

- Number of channels
- Two for stereoscopic displays
- Sometimes two display channels, but the same image on both (binocular-monoscopic)
- Can combine techniques to introduce more channels (perhaps for two viewers)
- Many ways to transmit multiple channels

Properties of Visual Display Paradigms

- Focal distance: Distance at which images seem to appear
  - Typically the screen in stationary displays
  - Can be infinite via optics in an head-based display
  - Incongruity between focal distance and disparity can lead to eye strain, or worse.

Properties of Visual Display Paradigms

- Opacity: Occlude the real world, or not
- CAVE does not occlude the real world
- Most head-based displays occlude the real world
- AR applications require displays that do allow the real world to be seen

Properties of Visual Display Paradigms

- Masking: Hiding things behind an object
- Body parts and objects held by user mask the real world in stationary displays
- Okay when objects are out of reach, but a problem when they are (virtually) between the participant's eyes and the masking part. In this case, the virtual object should occlude the mask (e.g. hand) but it doesn't.
- Not a concern for occlusive head-based displays
- See-through head-based displays can accommodate this if the masking objects are tracked

Properties of Visual Display Paradigms

- Field of regard: Amount surrounding space where virtual world is displayed
- Measured as a percentage
- Head-based and hand-based displays are typically 100%
- Stationary displays are often much less (except for 6-sided CAVEs)

Properties of Visual Display Paradigms

- Head position information
- Typically, position trackers monitor six "degrees of freedom/ movement" (DOF) of the participant's head.
- Some position trackers are only capable of monitoring a subset
- 3-DOF orientation is necessary for head-based displays
- 3-DOF location is necessary for stationary displays. But for proper stereoscopic viewing of stationary displays, full 6-DOF tracking information is required (3-DOF of each eye)

Properties of Visual Display Paradigms

- Cruciality of graphics latency
- How detrimental is lag to quality of the display?
- Lag is very noticeable when rotating head in head-based display
- Lag is less noticeable when rotating head in stationary display
- The latter is true because a view of world from the new direction was already there.

Properties of Visual Display Paradigms

- Frame rate
- Image updates per second (measured as FPS or Hz)
- Motion pictures capture 24 FPS
- 15 Hz is considered marginally acceptable
- 10 Hz and below causes brain to notice that it is seeing a series of still images

Logistic qualities of visual displays

- User mobility
- Interface with tracker methods
- Environment requirements
- Associability with other sense displays
- Portability
- Throughput
- Encumbrance
- Safety
- Cost

Logistic qualities of visual displays

- User mobility
  - Effect on (mental) immersiveness and usefulness
  - E.g. cables that tether the user or screens that prevent further physical movement
- Environment requirements
  - Conditions of the surrounding space necessary to provide a good VR experience
  - Projected images require low light
  - Large stationary displays require big rooms
  - Head-based displays work best in tight quarters (e.g. a training task used on a submarine)

Logistic qualities of visual displays

- Portability
- Large stationary displays are not portable
- Head-based displays can practically be run off a laptop
- Throughput
- Many people can see the screen in a CAVE
- Easier for people to enter/exit a CAVE, and to pass stereo glasses from one person to another
- Head-based displays often require a minute or two to change viewers

Logistic qualities of visual displays

- Encumbrance
  - Generally more wires associated with head-based displays
  - Stereo shutter glasses operate wirelessly
  - Wireless tracking can be done
- Safety
  - Can't see what you're doing (in the real world) in an occlusive headbased display
  - Eye fatigue and nausea can result from poor optics in an head-based display

Stationary displays

- Monitor based VR
- Fewer components / lower cost
- Easy to "suit-up"
- Standard interface devices available (keyboard, mouse, joystick)
- Limited FOR
- Limited FOV
- Generally less mentally immersive

Stationary displays

- Projection-based VR
- Larger, more costly displays
- Longer range tracking systems
- Greater FOV and FOR
- Not isolated from the real world
- Usually, less stressful on the eyes = more time immersed
- Not very encumbering
- Multi-user friendly
- More graphics power required (for multi-screened displays)
- Reflectivity problems (with multi-screened displays)
- More floor space required

Non Stationary displays

- Head-Based VR
- Small, lightweight screens
- Mental immersion relies mostly on head orientation
- 100% FOR
- Wide range of resolution (correlated with cost & weight)
- More portable than stationary displays
- Lag in tracking is more detrimental
- Generally lacking in FOV
- Generally more encumbering than stationary displays
- Lengthy use difficult due to increased eye strain

360 degree videos

- Shooting video in 360° is not like shooting regular video. It has its own set of rules and considerations.
- Now, the first thing to remember is that 360° means EVERY THING is in the view of the camera and will be in the shot. This includes you, your crew, any lights, microphones, equipment, every crack on the ceiling, and dirty footprints on the floor. This type of shot limits and gives freedom to the director at the same time.

360 degree videos

- Placement of the camera in the scene is key. To capture the action of the scene, you need to consider what role the person wearing the VR headset will take. Are they participants or observers? Meaning, should the camera be placed at an odd position, like hung from the ceiling or low to the ground? Or mounted six feet from the floor to equal to an average person's height? In either case, the camera should be placed in or around the center of the activity. That way the viewer can explore the scene as they wish, and will have something of interest in all directions.

Parrot Dron with Kinect and Oculus Rift

Thank you!

Questions?

InnoSoc
Innovative ICT Solutions for the Societal Challenges

Co-funded by the Erasmus+ Programme of the European Union

Unska 3, HR-10000 Zagreb, Croatia
innosoc@fer.hr

sociallab.education/innosoc
facebook.com/innosoc
twitter.com/innosoc