Human Computer Interaction

Alternative Input Modalities

Dr Pradipta Biswas, PhD (Cantab)
Assistant Professor
Indian Institute of Science
http://cpdm.iisc.ernet.in/PBiswas.htm

Content

- Eye Gaze Tracker
- Head Tracker
- Gesture Recognition
- Hand/ Finger Movement Tracker
- Speech Recognition

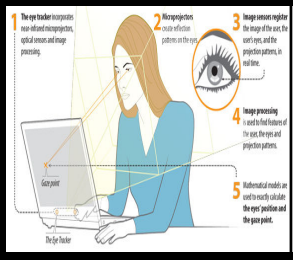
What is Eye Tracking & Gaze Control

- **Eye tracking** is the process of measuring either the point of gaze (where one is looking) or the motion of an eye relative to the head. An eye tracker is a device for measuring eye positions and eye movement.
- **Gaze control** is about effecting computer action by changing the direction of one's gaze (eye movement), blinking or dwelling on an object.

Eye movement

How Eye Tracking Works

- Most Commonly used technique is Pupil Centre and Corneal Reflection Technique.
- Simple calibration procedure (usually following a shape around screen required for each user.
- Infrared-sensitive video takes rapid pictures of eye.
- Infrared LED illuminates the eye.
- LED reflects small amount of light off the cornea & through the pupil onto the retina.
- Bright pupil allows image processor to locate centre of pupil.
- Tracker can then locate where the person is looking on the screen based on the relative positions of the pupil centre and corneal reflection within the video image of the eye



5

Types of Eye Tracker

- Non-intrusive
  - Attached to device (e.g.: Facelab)
  - Mobile (e.g.: Tobii X series)
- Intrusive
  - Glass based (e.g.: SMI Eye Glass)
  - Head attached
  - Lens based (very early models)
  - Electrodes (early models)

6

Comparison

- Non-Intrusive
  - Records natural interaction
  - Have issues with ambient illumination, screen size and head movement
- Intrusive
  - Needs to wear glasses or head mounted device
  - Supports head movement
  - Works for small and big screen devices
    - Mobile phone, big display etc

7

Types of Technology

- Infra red based
- Video based
- Electrode / Lens based (early models)

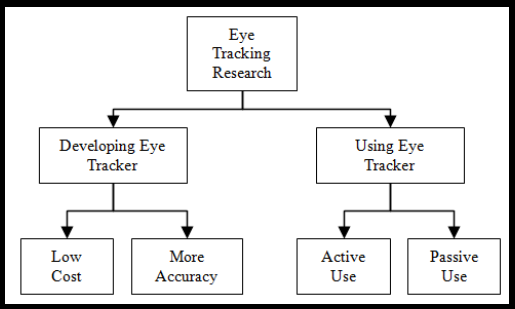
8

Comparison

- Infra red based
  - Accurate
  - Needs to install infra red trackers
  - Costly
- Video based
  - Less accurate
  - Works with existing webcam
- Some video based eye trackers need special camera though it is still less costly than infra red ones
- Recent work also investigating use of low cost infrared tracker (e.g.: EyeTribe Technology, $99 infrared ET)

9

Types of Applications



10

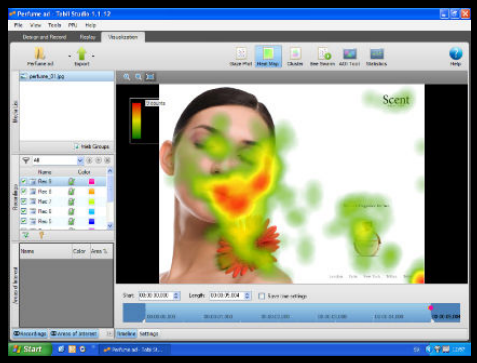
Passive Eye Tracking

Theory of Visual Perception

Points of Fixation

12

Area of interest



13

Applications

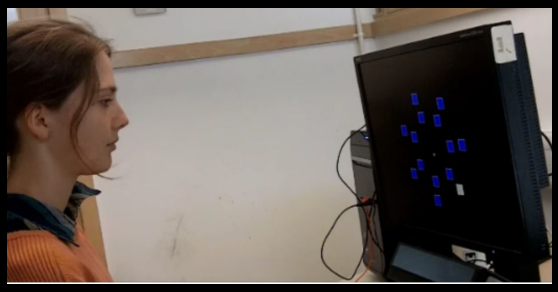
- Analyzing points of fixation and eye movement to investigate
  - Areas of interest in a display
  - Reading behaviour
  - Affect state of user
  - Visual impairment
    - Nystagmus – irregular eye movement
  - Design of billboard, traffic sign etc.

14

Active Eye Tracking

Gaze Control Interface

Demonstration – Gaze Control Interface



16

Types of Eye Gaze Movement

- Saccades
- Smooth Pursuits
- Vergance

17

Issues with gaze control

- Strain
- Accuracy
- Selection
  - Midas Touch Problem

18

Multimodal Eye Tracking

- MAGIC System
  - Selection using mouse
- Eye Tracking and BCI
  - Selection through imagined action detected through EEG
- Eye Tracking and Assistive Technology
  - Selection through single switch scanning

19

Head Tracker

Related technology to gaze control

Types of Head tracker

- Helmet Based
- Video based
  - http://www.cameramouse.com
- Attaching Gyroscopic Tracker
- Similar issues with intrusive and non-intrusive head trackers as with gaze control

21

Demonstration – Head Tracker

22

Gesture Rcognition

23

The Nature of Gesture

- Gestures are expressive, meaningful body motions, i.e., physical movements of the fingers, hands, arms, head, face, or body with the intent to convey information or interact with the environment.

Functional Roles of Gesture

- Semiotic: to communicate meaningful information
- Ergotic: to manipulate the environment
- Epistemic: to discover the environment through tactile experience.

Semiotic Gesture

- The semiotic function of gesture is to communicate meaningful information. The structure of a semiotic gesture is conventional and commonly results from shared cultural experience. The good-bye gesture, the American sign language, the operational gestures used to guide airplanes on the ground, and even the vulgar ``finger'', each illustrates the semiotic function of gesture.
- HCI Example: Blooming signal to MS Hololens

26

Ergotic Gesture

- The ergotic function of gesture is associated with the notion of work. It corresponds to the capacity of humans to manipulate the real world, to create artefacts, or to change the state of the environment by ``direct manipulation''. Shaping pottery from clay, wiping dust, etc. result from ergotic gestures.
- HCI examples: typing on a keyboard, moving a mouse, and clicking buttons.

27

Epistemic Gesture

The epistemic function of gesture allows humans to learn from the environment through tactile experience. By moving your hand over an object, you appreciate its structure, you may discover the material it is made of, as well as other properties.

HCI Example: Haptic Interface

28

Gesture vs. Posture

- Posture refers to static position, configuration, or pose.
- Gesture involves movement. Dynamic gesture recognition requires consideration of temporal events. This is typically accomplished through the use of techniques such as time-compressing templates, dynamic time warping, hidden Markov models (HMMs), and Bayesian networks.

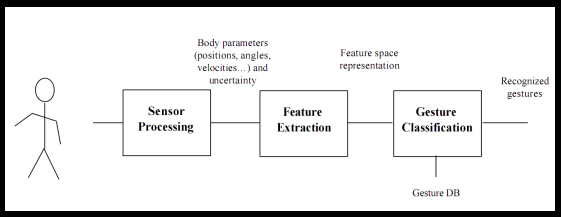
Examples

- Pen-based gesture recognition
- Tracker-based gesture recognition
  - Instrumented gloves
  - Body suits
- Passive vision-based gesture recognition
  - Head and face gestures
  - Hand and arm gestures
  - Body gestures

Vision-based Gesture Recognition

- Advantages:
  - Passive and non-obtrusive
  - Low-cost
- Challenges:
  - **Efficiency:** Can we process 30 frames of image per second?
  - **Accuracy:** Can we maintain robustness with changing environment?
  - **Occlusion:** can only see from a certain point of view. Multiple cameras create integration and correspondence issues.

Gesture Recognition System



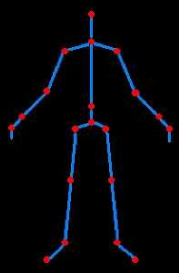
Issues

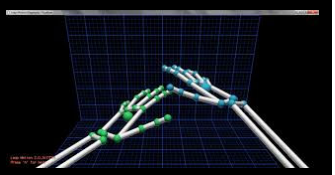
- **Number of cameras**. How many cameras are used? If more than one, are they combined early (stereo) or late (multi-view)?
- **Speed and latency.** Is the system real-time (i.e., fast enough, with low enough latency interaction)?
- **Structured environment.** Are there restrictions on the background, the lighting, the speed of movement, etc.?
- **User requirements**. Must the user wear anything special (e.g., markers, gloves, long sleeves)? Anything disallowed (e.g., glasses, beard, rings)?
- **Primary features.** What low-level features are computed (edges, regions, silhouettes, moments, histograms, etc.)?
- **Two- or three-dimensional representation**.
- **Representation of time**: How is the temporal aspect of gesture represented and used in recognition?

Tools for Gesture Recognition

- Static gesture (pose) recognition
  - Template matching
  - Neural networks
  - Pattern recognition techniques
- Dynamic gesture recognition
  - Time compressing templates
  - Dynamic time warping
  - Hidden Markov Models
  - Conditional random fields
  - Time-delay neural networks
  - Particle filtering and condensation algorithm
  - Finite state machine

Hand / Finger Tracking





35

Pointer Control

- 3-D to 2-D mapping
- Orthogonal Projection
  - Evaluate the equation of 2-D screen in tracker's coordinate system
  - Calculate projection of finger / hand position on that plane

$$ScreenX = \frac{ScreenWidth}{w} \times (finger.TipPosition.x + a)$$

$$ScreenY = \frac{ScreenHeight}{h} \times (b + c \times finger.TipPosition.y - d \times finger.TipPosition.z)$$

Jitter Removal

- Averaging filter
- Exponential averaging
- Kalman Filter
- Higher order polynomial filtering

37

Head and Face Gestures

- Nodding or shaking the head;
- Direction of eye gaze;
- Raising the eyebrows;
- Opening the mouth to speak;
- Winking;
- Flaring the nostrils;
- Facial expression: looks of surprise, happiness, disgust, anger, sadness, etc.

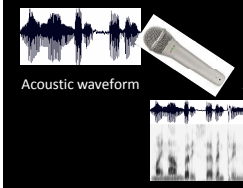
Body Gesture

- Human dynamics: tracking full body motion, recognizing body gestures, and recognizing human activity.
- Activity may be defined over a much longer period of time than what is normally considered a gesture; for example, two people meeting in an open area, stopping to talk, and then continuing on their way may be considered a recognizable activity.
- Bobick (1997) proposed a taxonomy of motion understanding in terms of:
  - Movement. The atomic elements of motion.
  - Activity. A sequence of movements or static configurations.
  - Action. High-level description of what is happening in context.

Speech Recognition

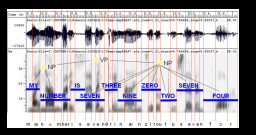
40

How might computers do it?



Acoustic waveform

Acoustic signal



Speech recognition

- Digitization
- Acoustic analysis of the speech signal
- Linguistic interpretation

41/34

Approaches to ASR

- Template matching
- Knowledge-based (or rule-based) approach
- Statistical approach:
  - Noisy channel model + machine learning

42/34

Template-based approach

- Store examples of units (words, phonemes), then find the example that most closely fits the input
- Extract features from speech signal, then it's "just" a complex similarity matching problem, using solutions developed for all sorts of applications
- OK for discrete utterances, and a single user

43/34

Rule-based approach

- Use knowledge of phonetics and linguistics to guide search process
- Templates are replaced by rules expressing everything (anything) that might help to decode:
  - Phonetics, phonology, phonotactics
  - Syntax
  - Pragmatics

44/34

Statistics-based approach

- Collect a large corpus of transcribed speech recordings
- Train the computer to learn the correspondences ("machine learning")
- At run time, apply statistical processes to search through the space of all possible solutions, and pick the statistically most likely one

45/34

Comparing ASR systems

- Factors include
  - Speaking mode: isolated words vs continuous speech
  - Speaking style: read vs spontaneous
  - "Enrollment": speaker (in)dependent
  - Vocabulary size (small <20 … large > 20,000)
  - Equipment: good quality noise-cancelling mic … telephone
  - Size of training set (if appropriate) or rule set
  - Recognition method

46/34

Remaining problems

- **Robustness** – graceful degradation, not catastrophic failure
- **Portability** – independence of computing platform
- **Adaptability** – to changing conditions (different mic, background noise, new speaker, new task domain, new language even)
- **Language Modelling** – is there a role for linguistics in improving the language models?
- **Confidence Measures** – better methods to evaluate the absolute correctness of hypotheses.
- **Out-of-Vocabulary (OOV) Words** – Systems must have some method of detecting OOV words, and dealing with them in a sensible way.
- **Spontaneous Speech** – disfluencies (filled pauses, false starts, hesitations, ungrammatical constructions etc) remain a problem.
- **Prosody** –Stress, intonation, and rhythm convey important information for word recognition and the user's intentions (e.g., sarcasm, anger)
- **Accent, dialect and mixed language** – non-native speech is a huge problem, especially where code-switching is commonplace

47/34

Take away points

- Description to new modalities of interaction
- Different types of eye trackers and their comparison
- Basics on different types of gesture
- Gesture recognition from multiple body parts
- Basic structure of a gesture recognizer
- Different approaches to Automatic speech recognizer
- Open problems

48