Alternative Input Modalities

Dr Pradipta Biswas, PhD (Cantab)
Assistant Professor
Indian Institute of Science
https://cambum.net/

Content

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

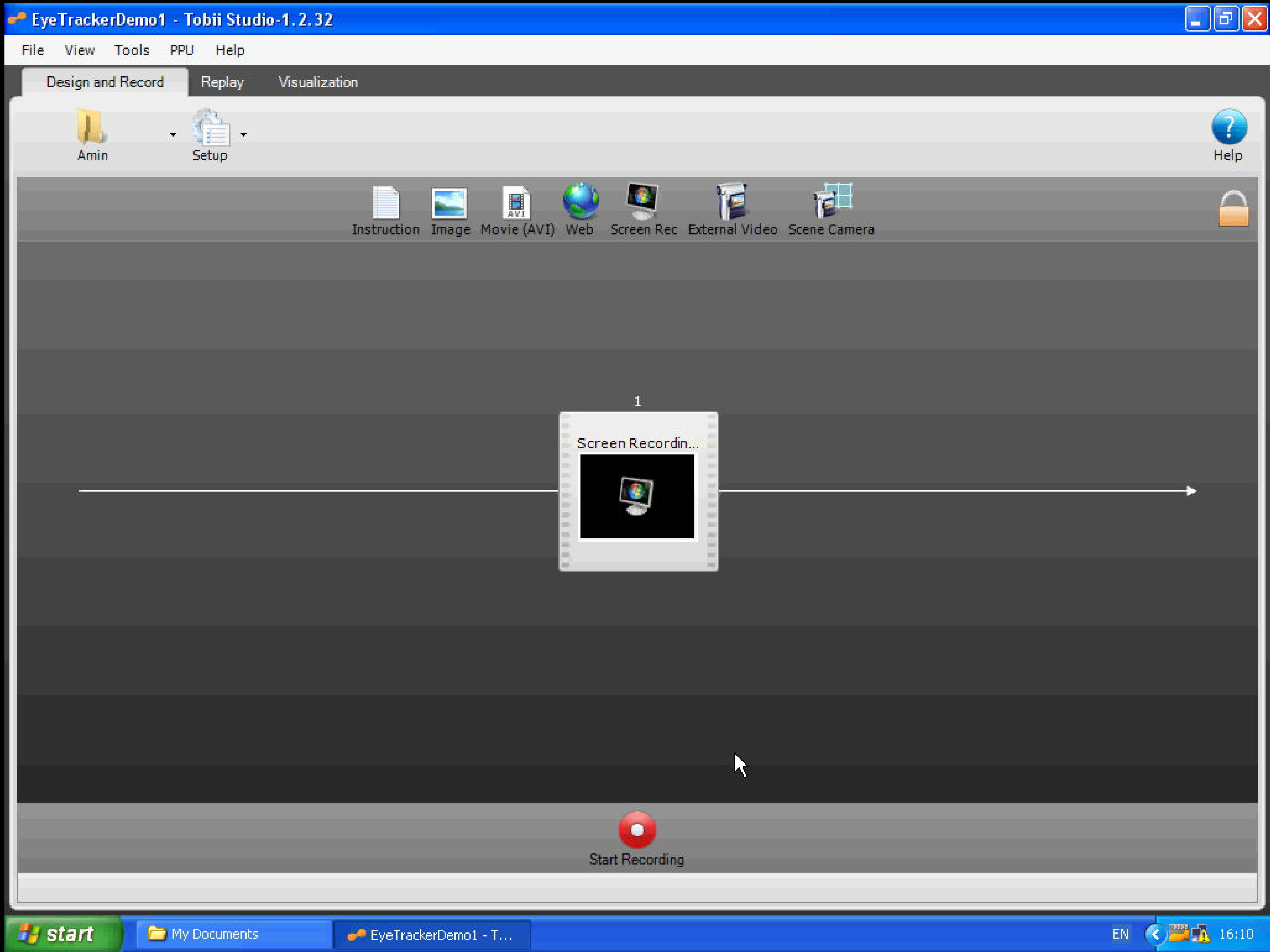
What is Eye Tracking & Gaze Control

- **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.

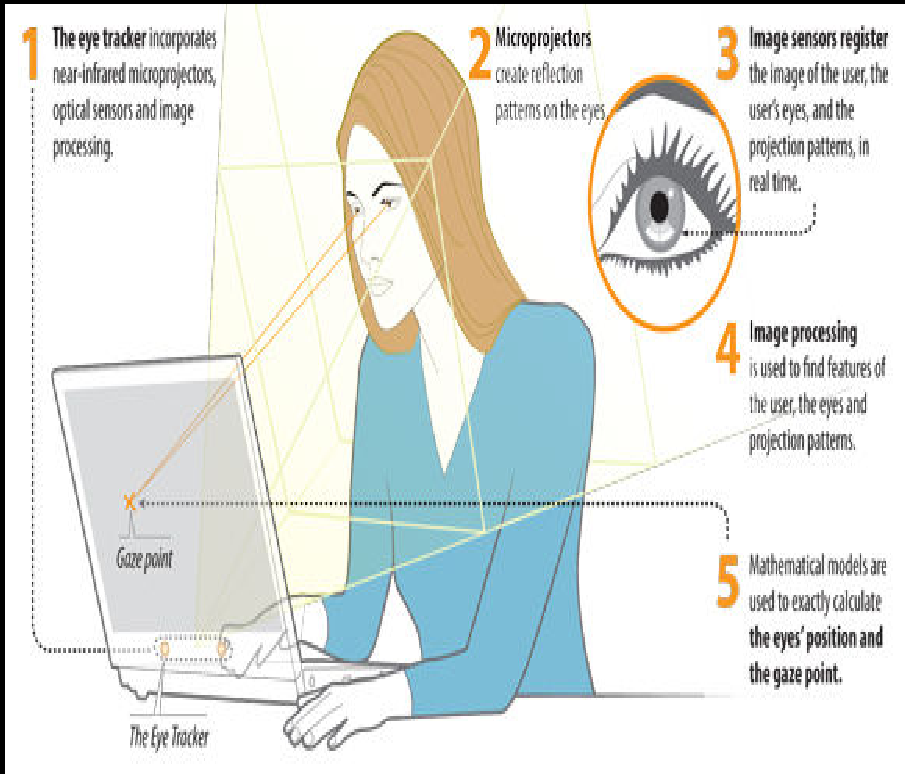
3

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



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)

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

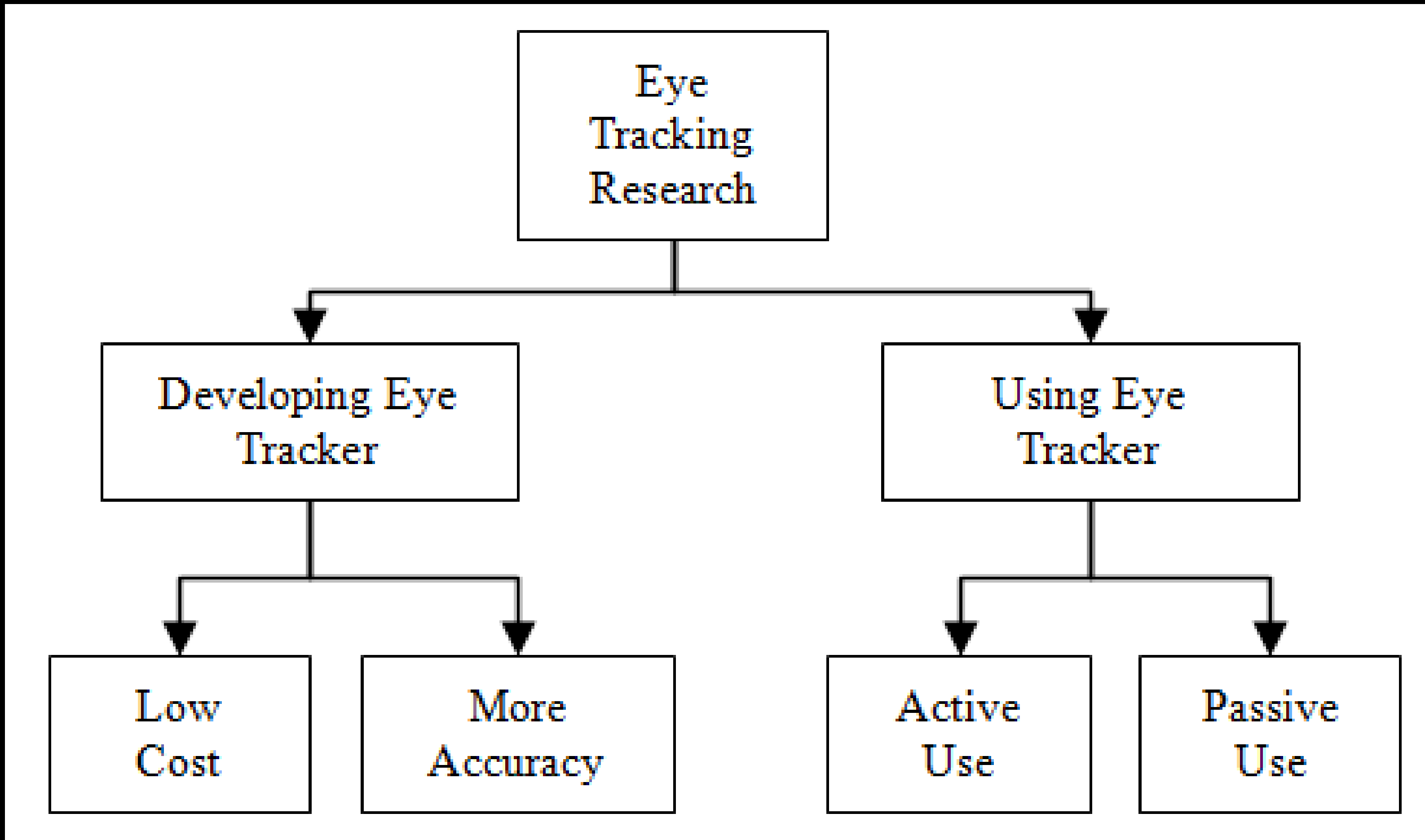
Types of Technology

- Infra red based
- Visible Spectrum based
- Electrode / Lens based (early models)

Comparison

- Infra red based
  - Accurate
  - Needs to install infra red trackers
  - Costly
- Visible Spectrum based
  - Less accurate
  - Works with existing webcam
- Recent work also investigating use of low cost infrared tracker (e.g.: EyeTribe Technology, $99 infrared ET)

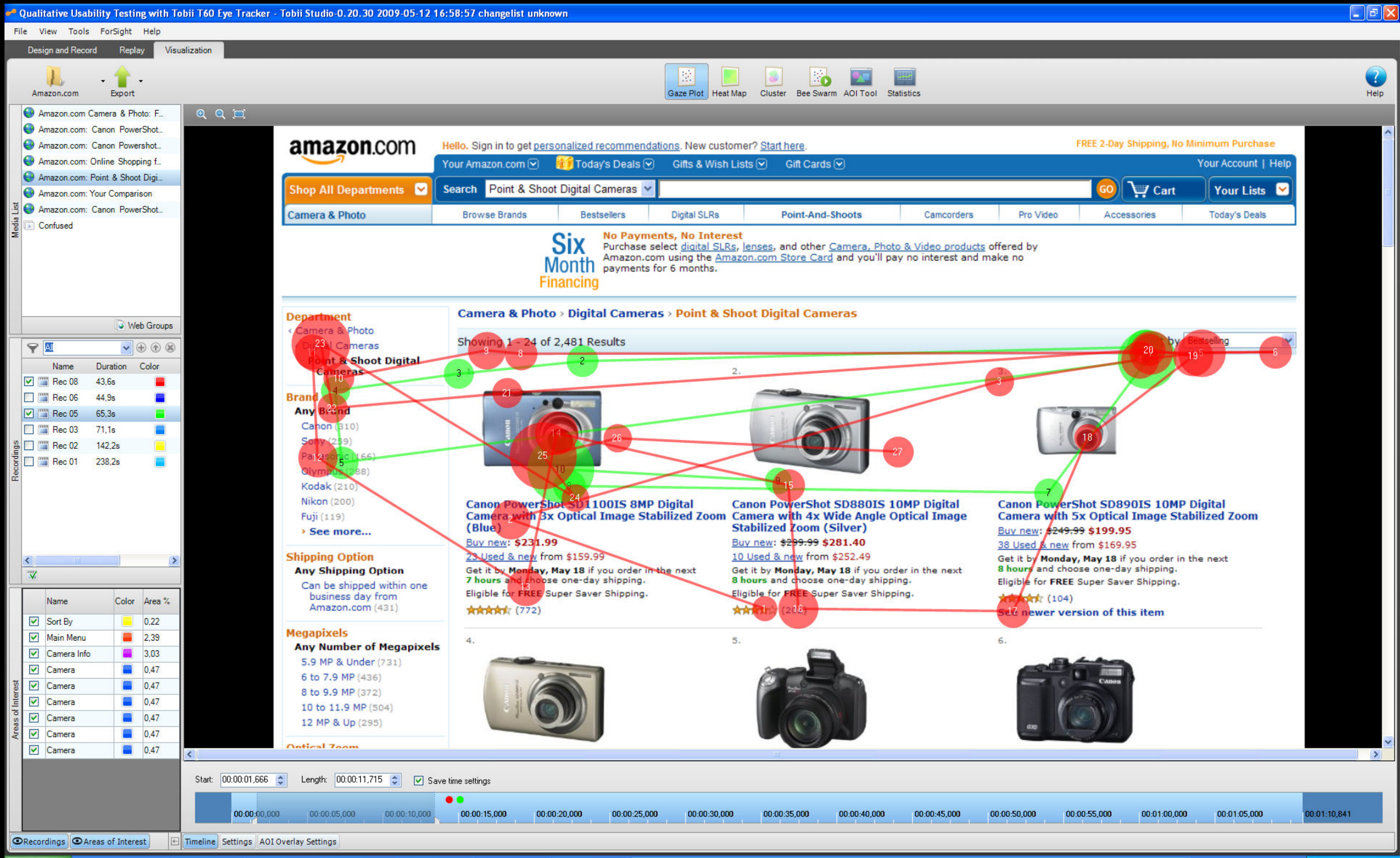
Types of Applications



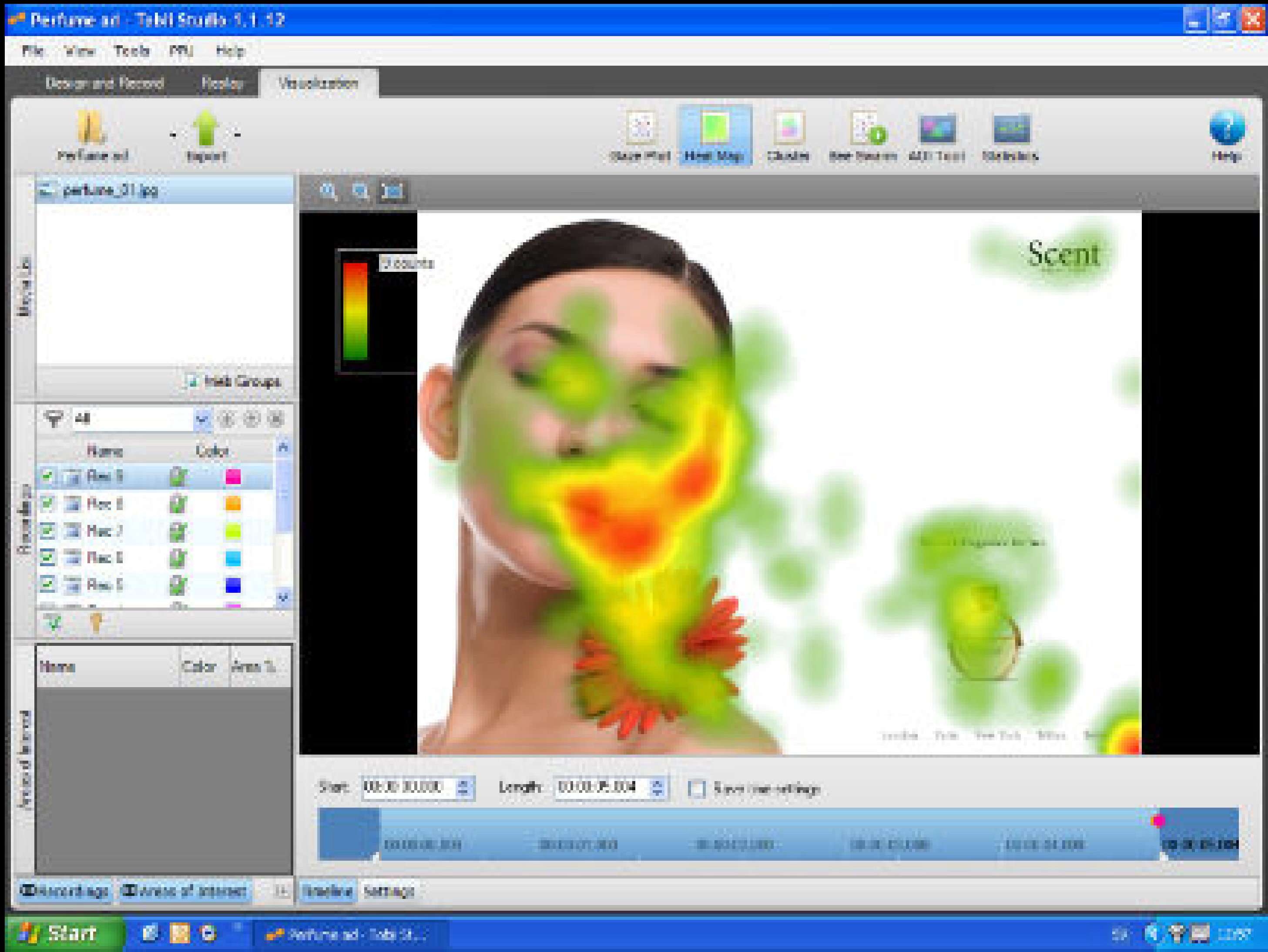
Passive Eye Tracking

Theory of Visual Perception

Points of Fixation



Area of interest



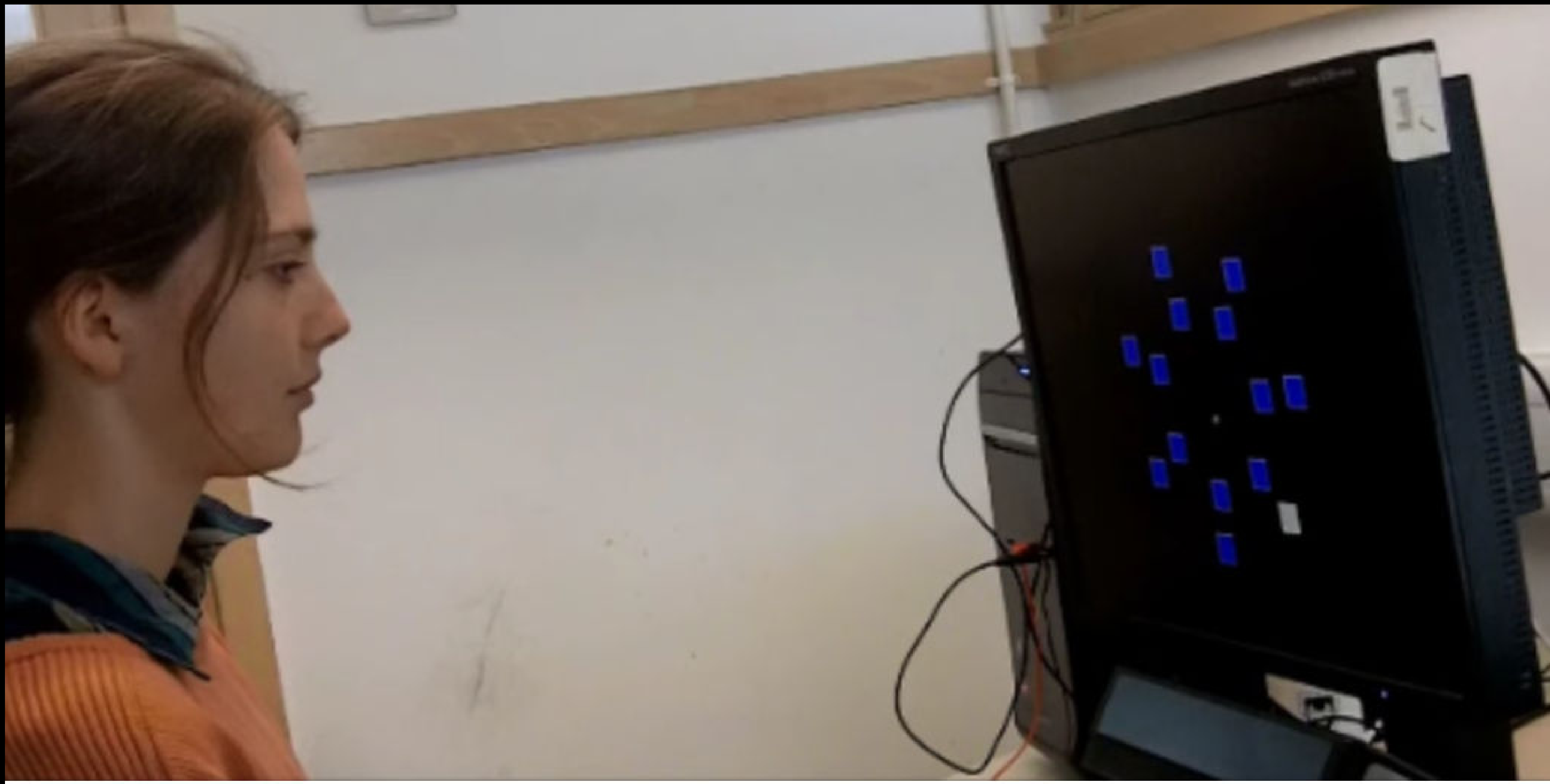
Applications

- 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.

Active Eye Tracking

Gaze Control Interface

Demonstration – Gaze Control Interface



Types of Eye Gaze Movement

- Saccades
- Smooth Pursuits
- Vergance

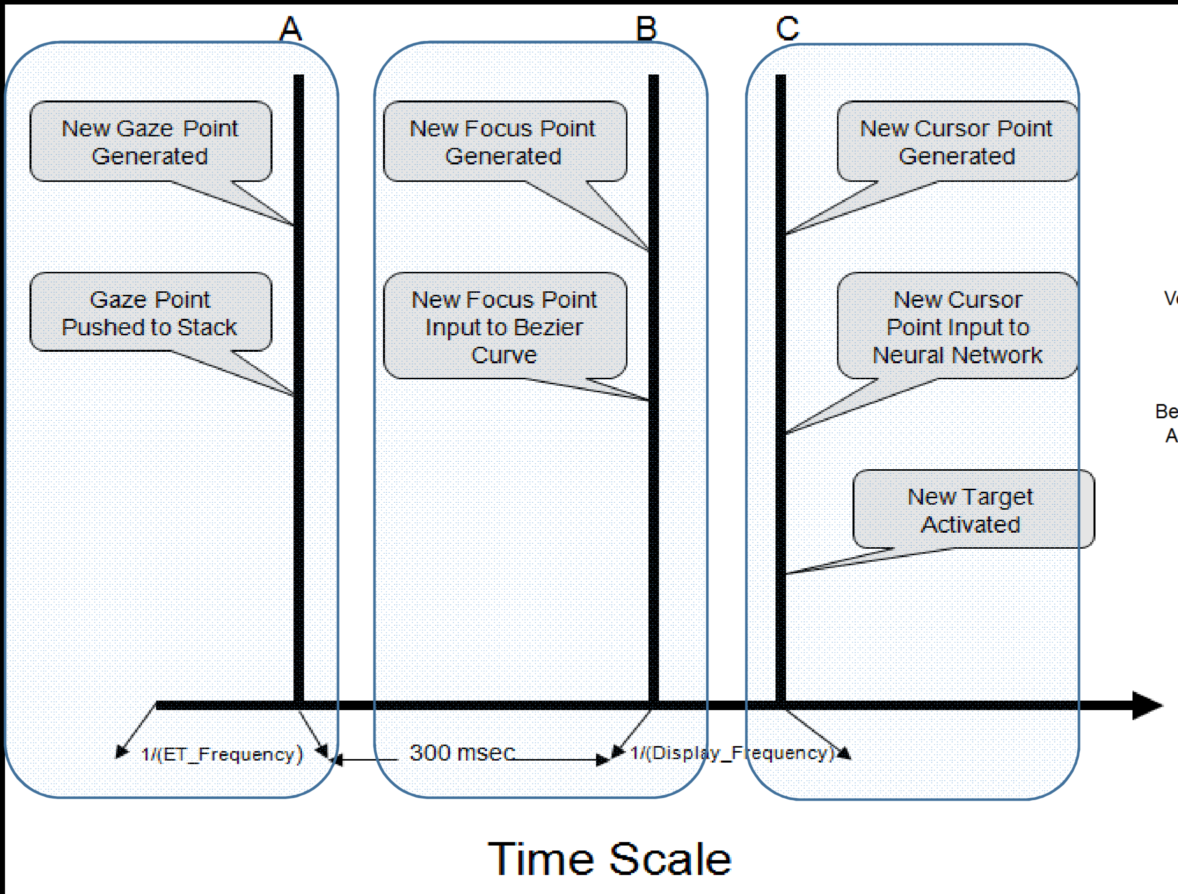
Issues with gaze control

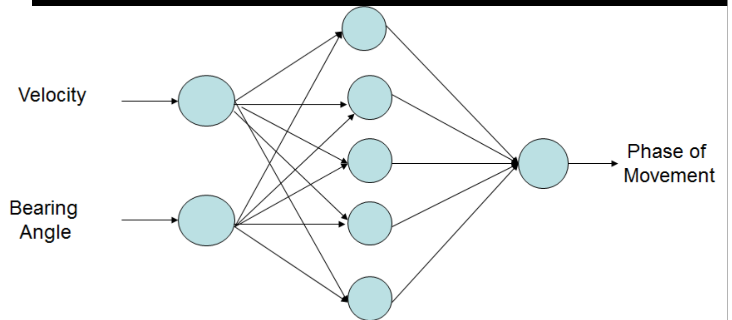
- Strain
- Accuracy
- Selection
  - Midas Touch Problem

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

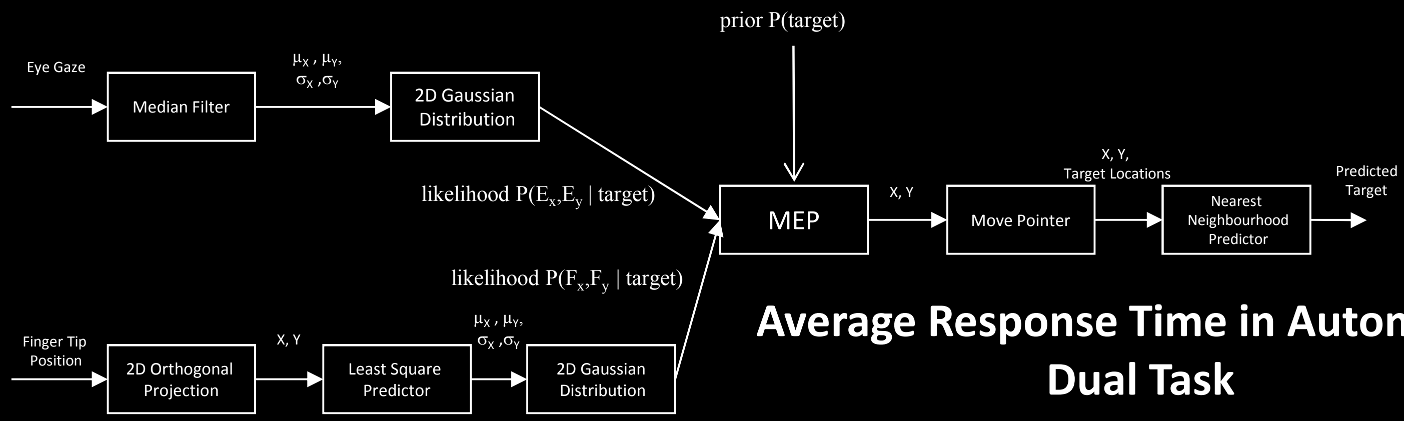
Target Prediction for Eye Gaze Tracking System



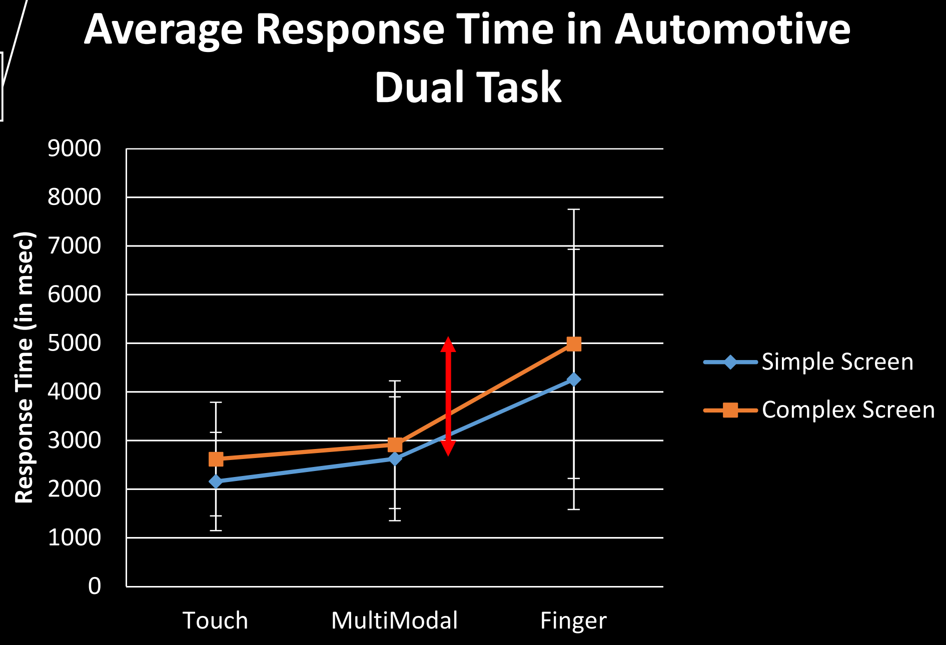


A Neural Network is used to detect the phase of movement (ballistic vs homing)

Multimodal Fusion



The multimodal system significantly reduced response times compared to unimodal finger tracking system




Finished
1st
2
77 kmh

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

Demonstration – Head Tracker

Gesture Recognition

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

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.

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

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.

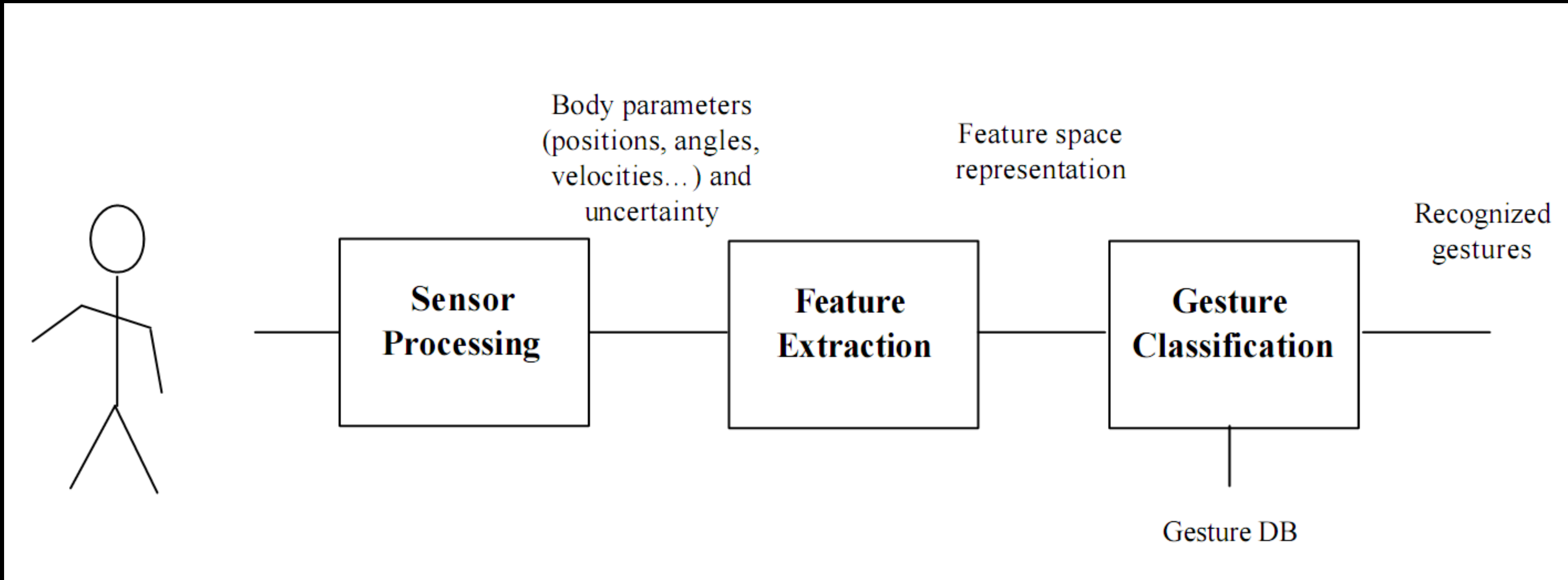
Examples

- 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



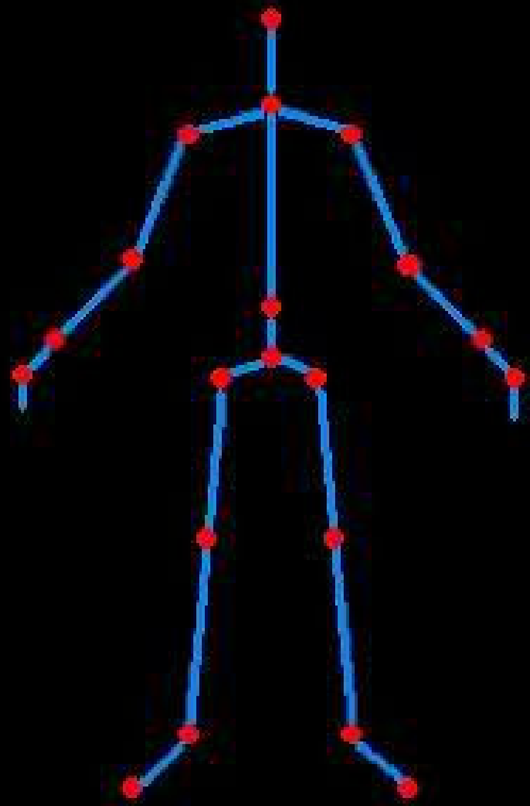
Issues

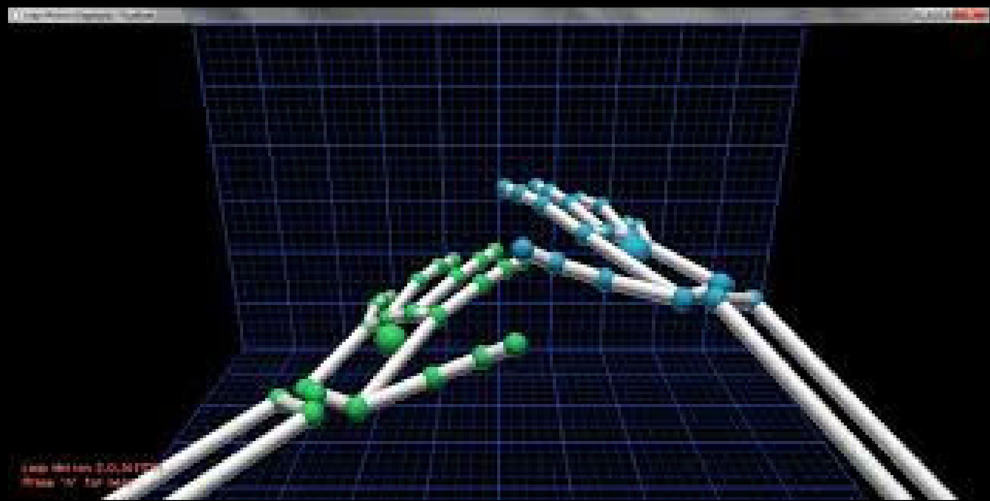
- **Number of cameras**. How many cameras are used? If more than one, are they combined early (stereo) or late (multi-view)?
- **Speed and latency.** Is the system real-time (i.e., fast enough, with low enough latency interaction)?
- **Structured environment.** Are there restrictions on the background, 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





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

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.

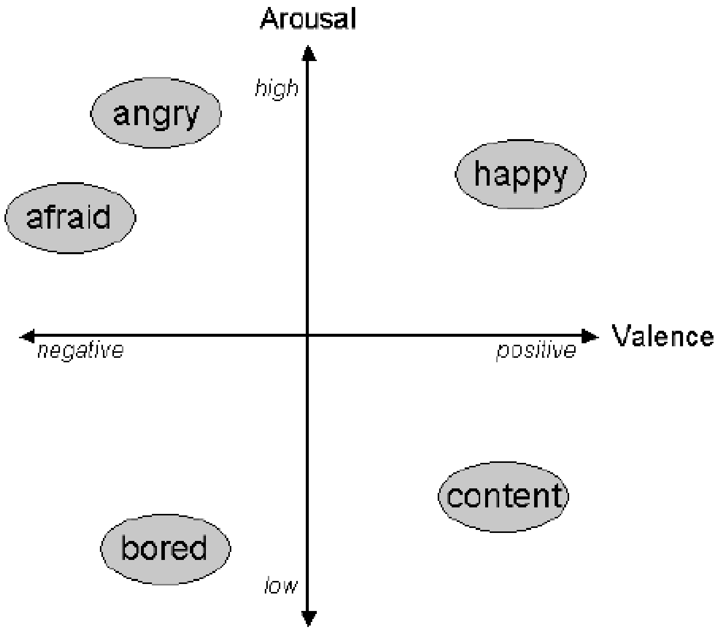
… It's not so simple

What misery!

What joy!

(Jana Novotna, Wimbledon, 1998)

Emotion Categorization



Facial Expressions (Paul Ekman 1982, Emotion in the Human Face)

Machine Perception of Affect

Diagnostic Methods

- Based on inference from sensory channels
- Bottom-up approach
- Approximation / Estimation

Predictive Methods

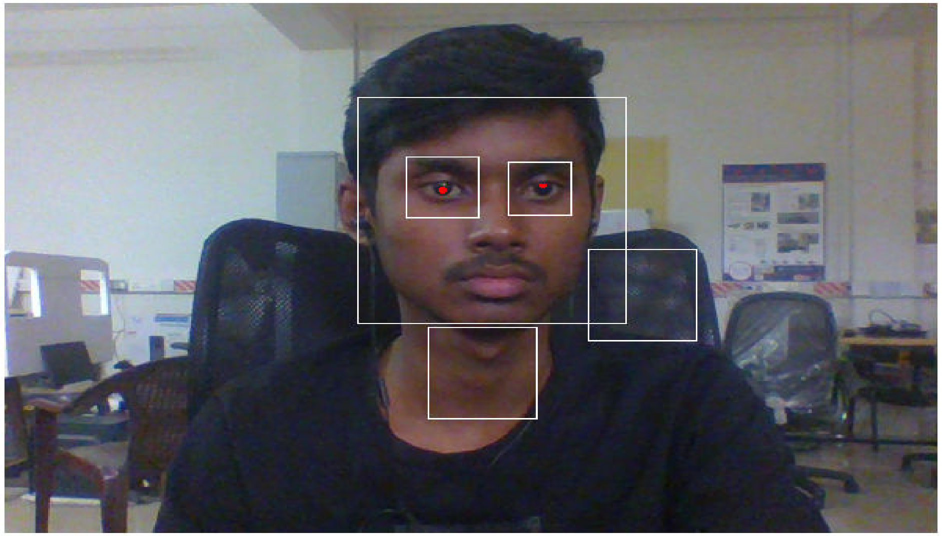
- Based on psychological theories like OCC, Appraisal
- Top-down approach
- Causal view

Hybrid

- Combination of causal & diagnostic approaches
- Leverages top-down & bottom-up evidence
- More powerful, realistic & accurate

Context-sensitive Interpretation

Face Recognition



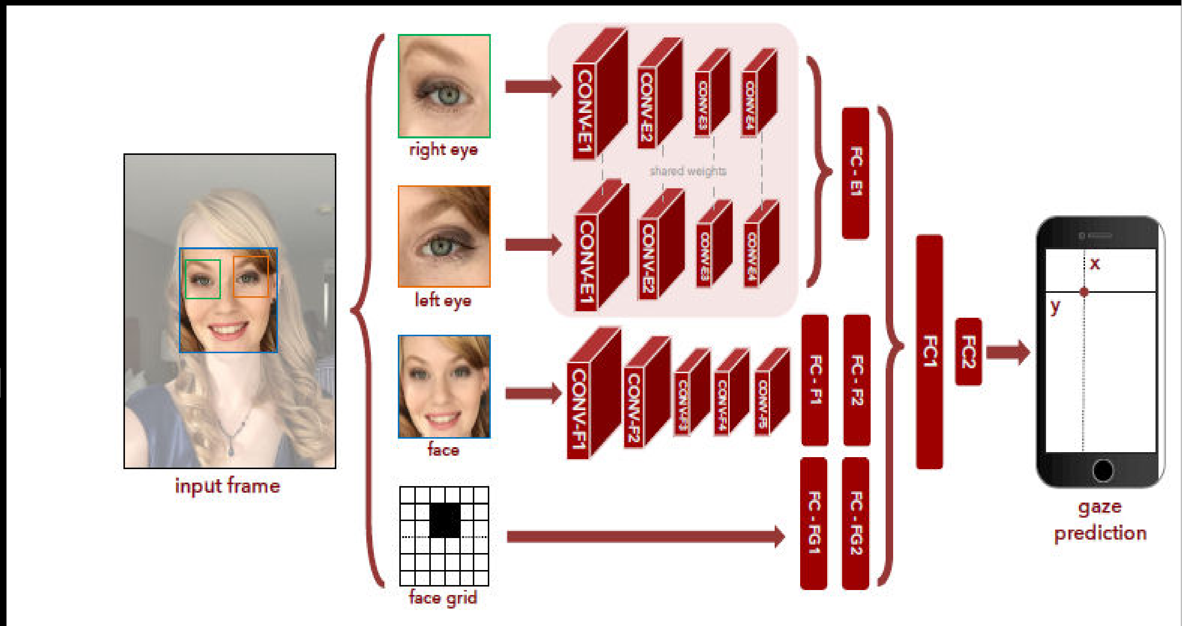
(a) Haar cascade face detection

(b) HOG based facial landmark detection

Note: Both images were taken in the same lighting condition and are only about second apart from each other . The pupil detection is added and is not part of those two detectors

Face Recognition in the Wild

- We used dlib – face detection scheme to obtain training images
- dlib – face detection uses Histogram of Oriented Gradients (HOG) feature combined with a linear classifier, an image pyramid and sliding window detection scheme.
- After training we performed ablation study using PyTorch where no face images are provided
- The accuracy of the final task in ablation study is similar to the baseline model (<5% difference)



Cognitive Load Estimation

TIME SCALE OF HUMAN ACTION		
SCALE (sec)	SYSTEM	STRATUM
10^7 10^6 10^5		SOCIAL
10^4 10^3 10^2	Task Task Task	RATIONAL
10^1 10^0 10^{-1}	Unit Task Operations Deliberate Act	COGNITIVE
10^{-2} 10^{-3} 10^{-4}	Neural Circuit Neuron Organelle	BIOLOGICAL

A Newell, Unified Theories of Cognition

- Our research estimates cognitive load from ocular parameters
- Neural processing work at a faster level than cognitive processing
- Hike in Pupil Dilation is correlated to EEG output
- SI or SWJ are clinically used to diagnose neurological problems like Alzheimer's Disease or Progressive Supranuclear Palsy

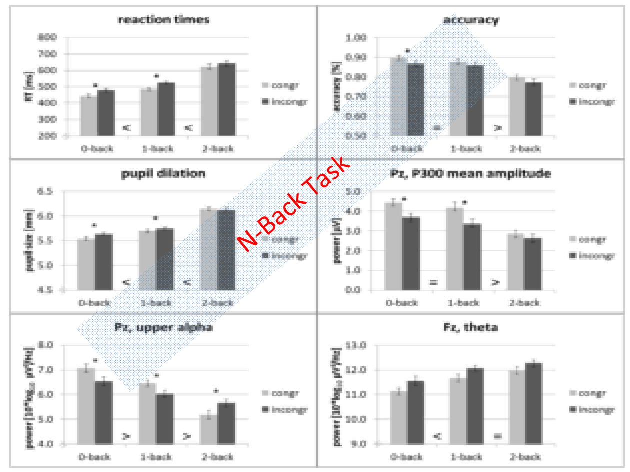


Figure 2. Mean values for reaction times, accuracy, pupil size (indicating pupil dilation), P300 amplitude at electrode Pz, upper alpha power (Pz), and theta power (Fz). Error bars: ± 1 SEM. The *, >, and < mark significant differences ($p < .05$). Light gray color symbolizes congruent trials, dark gray color incongruent trials.

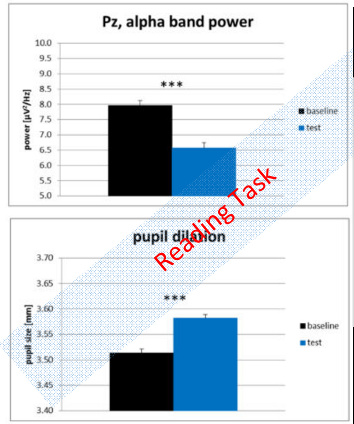
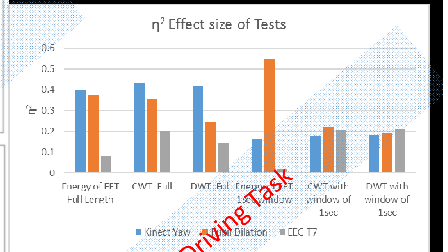


Fig 2. Mean alpha (8–13 Hz) frequency band power at electrode Pz and mean pupil dilation of Experiment 1. Note. *** indicate p <.001, black error bars indicate +1 standard error of the mean.



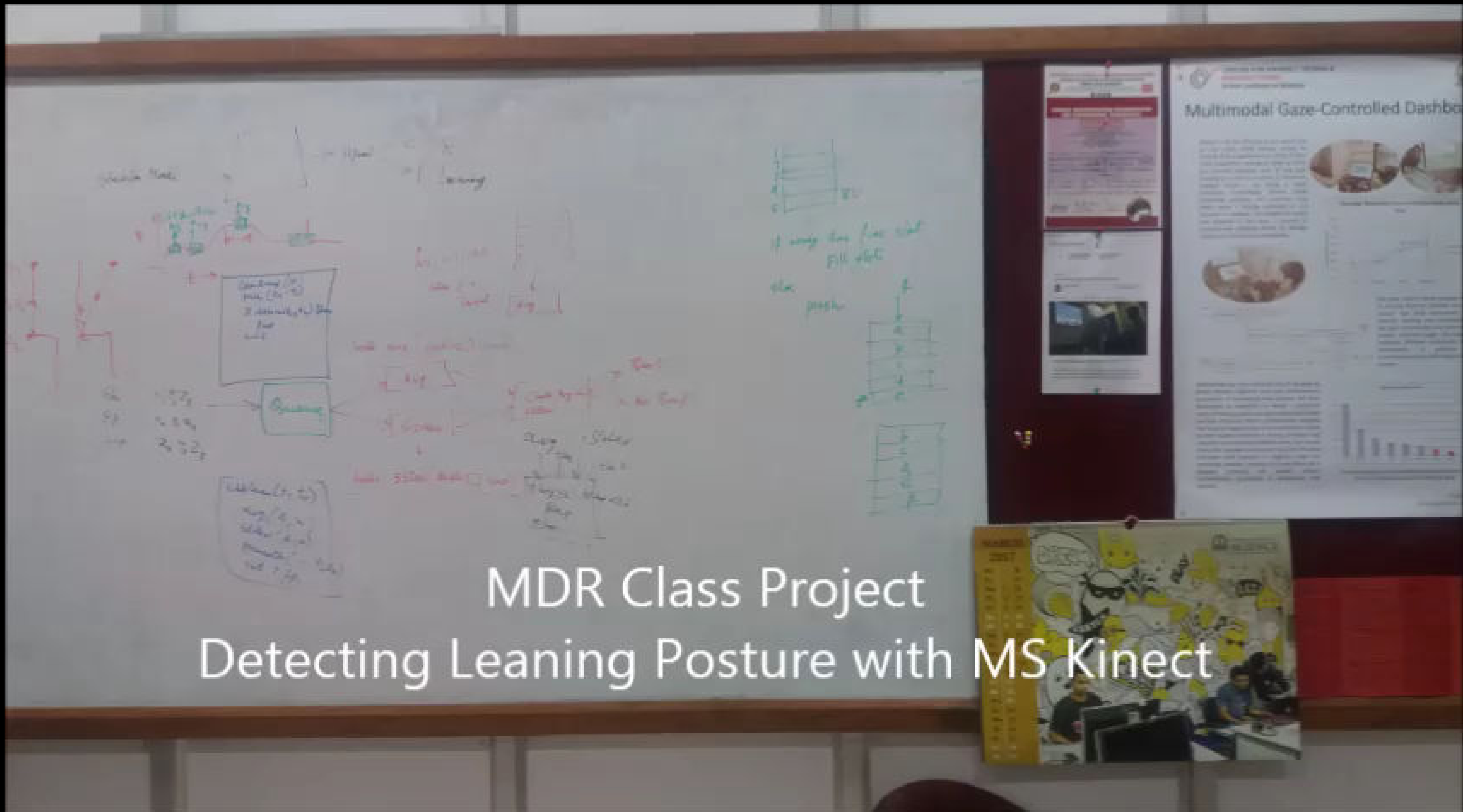
Effect size of each tests to find best performance for real-time implementation

Automotive and Aviation Use Cases

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.

Bad Posture Detection



Take away points

- Description to new modalities of interaction
- Different types of eye trackers and their comparison
- Introduction to multimodal eye tracking including head trackers
- Basics on different types of gesture
- Gesture recognition from multiple body parts
- Basic structure of a gesture recognizer
- Introduction to Affective Computing