

Multimodal Interaction

Gesture and Affect State Recognition

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

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.

5

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

6

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.

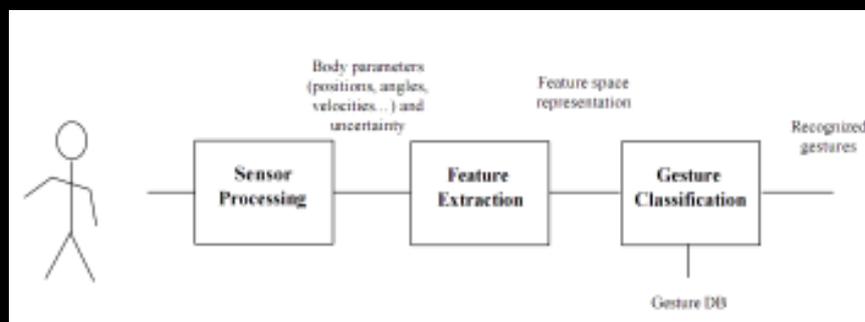
Examples

- 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?
 - Occulsion: can only see from a certain point of view. Multiple cameras create integration and correspondence issues.

Gesture Recognition System



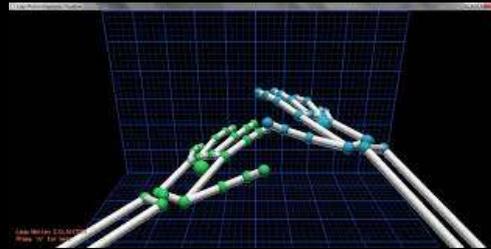
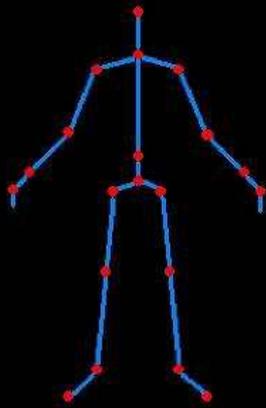
Issues

- **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



13

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

$$\text{ScreenX} = \frac{\text{ScreenWidth}}{w} \times (\text{finger.TipPosition.x} + a)$$
$$\text{ScreenY} = \frac{\text{ScreenHeight}}{h} \times (b + c \times \text{finger.TipPosition.y} - d \times \text{finger.TipPosition.z})$$

Jitter Removal

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

15

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.

Affective Computing

What does Affective Computing mean...

- Building systems and devices that –
Recognize emotion, Have emotions & Process emotions
- By studying manifestation/expression in human behaviour -
Visual, Vocal & Tactile Modalities
- Designing 'truly' intelligent technologies –
Represent, Detect & Analyse affective phenomenon

Non - Verbal Behaviour ...



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



Body Gestures

... It's not so simple



Bio-signal Analysis



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

Practical Issues

- Appropriate sources: acted, induced or naturalistic
- Description of emotional content
- Which modalities & in what combinations
- Description of emotion related features
- Coding schemes
- Technical settings / Recording setups
- Selection of appropriate coders

Bigger Ethical Issues

- Level of control & reliability
- User acceptance
- Replacing human communication

Applications

- Education
- Interactive gaming and entertainment
- Security and surveillance
- Healthcare
- Sales & adverts - Retail kiosks and market research
- Remote collaboration
- Smart homes
- Research – automated behaviour analysis

Take away points

- Basics on different types of gesture
- Gesture recognition from multiple body parts
- Basic structure of a gesture recognizer
- Introduction to Affective Computing
- Using MS Kinect and LeapMotion to track hand and finger movement

26