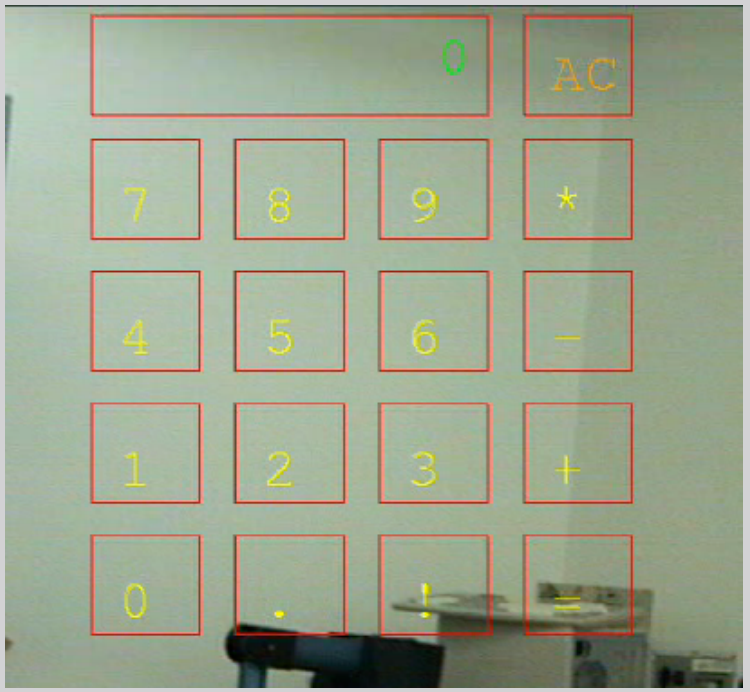
VICs: A Modular Vision-Based HCI Framework

The Visual Interaction Cues Project

Guangqi Ye, Jason Corso
Darius Burschka, & Greg Hager
CIRL, The Johns Hopkins University

Visual Interaction Cues (VICs)



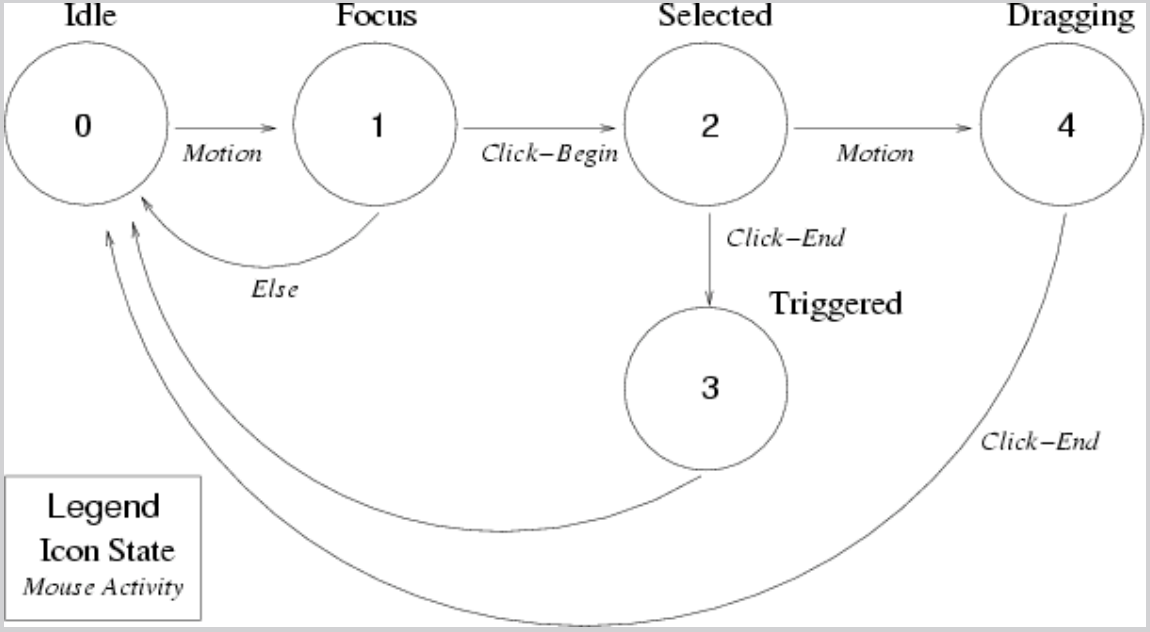
Talk Structure

- Modeling Interaction
- The VICs Paradigm
- The VICon:the core component
- Examples VICons
- Modes of Interaction
- Video and Conclusion

Modeling Interaction

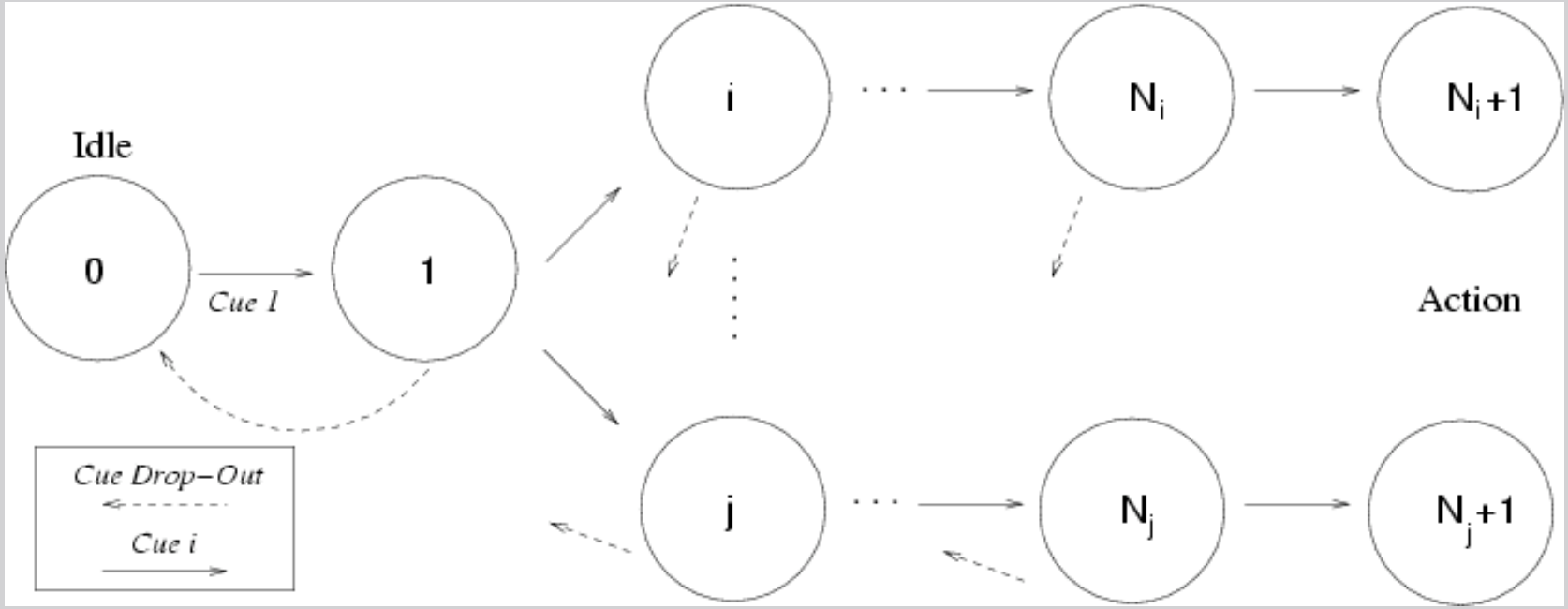
- Mainstream Interface Technology:

  WIMP - Windows, Icons, Menus, and Pointers. [van Dam 97]



Modeling Interaction

- A more general model:



Harnessing the Power of Vision

- Difficult
- Tracking-based approaches
  - Gaze, Head, Full-body tracking
- We differ by
  - Placing the focus on the interface.
  - Kjeldsen et al. (Session 5)

The VICs Paradigm

- Two governing principles:
  - Site-centric interaction.
  - Simple-to-Complex processing.
- Modular structure
  - Visual Interaction Cue Components - VICons.

Site-centric Interaction

- Reverse the interaction problem:

  Center processing about the components not the user.

- Each VICon observes its local neighborhood for *important* cues.

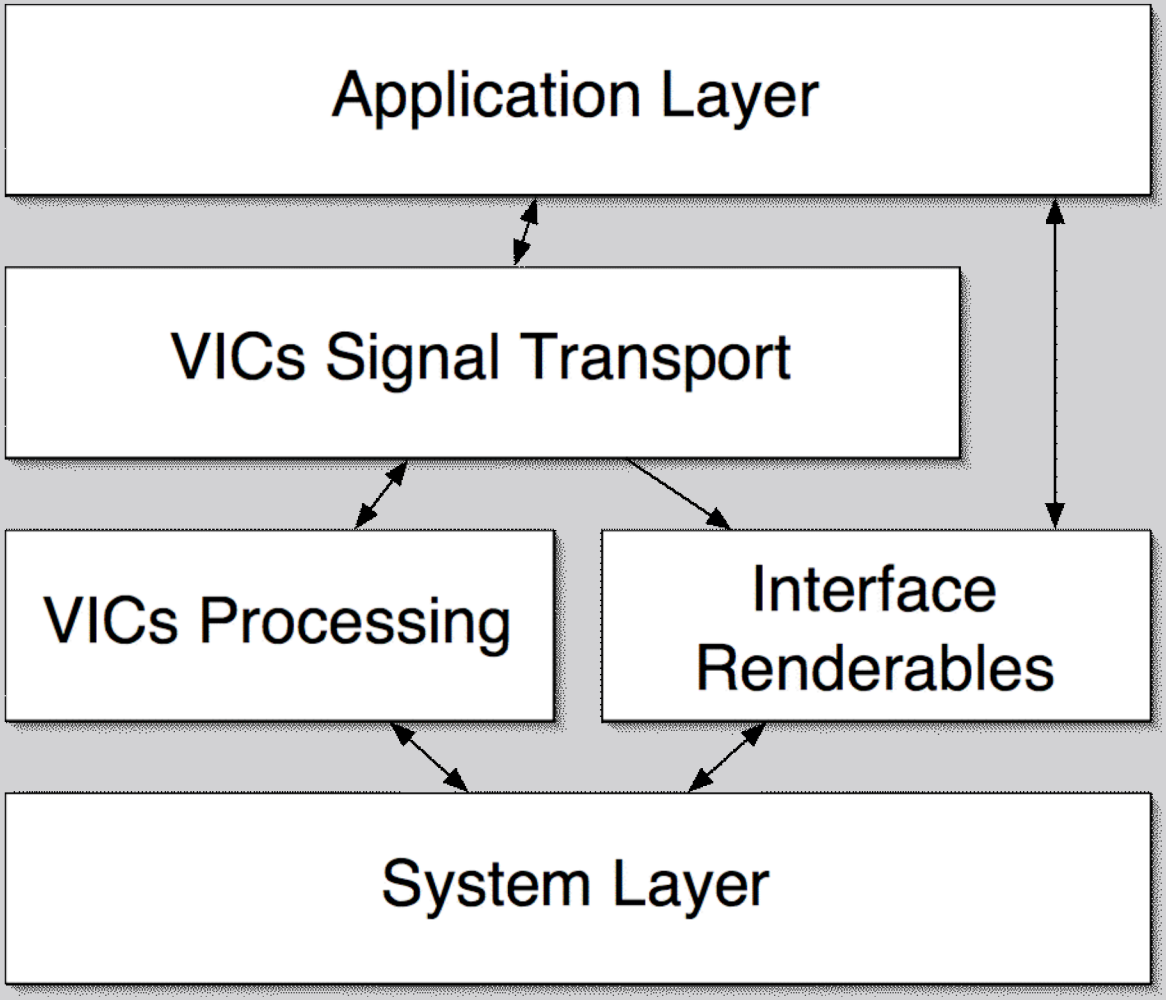
Simple-to-Complex Processing

- Maximize detection vs. minimize computation
- Typical approach - template matching
  - Prone to false-positives
  - Potentially wasteful
- Watch for a stream of cues structured from simple-to-complex
  - E.g.. Motion detection : Hue Blob : Shape Test : …

The VICon's Structure

1. A tangible representation: graphical, audible, haptic.
2. A set of signals to provide application specific functionality.
3. A visual processing engine.
   - The core of the VICon - parses the cue stream

VICs Architecture at a Glance



An Example VICon - A Button

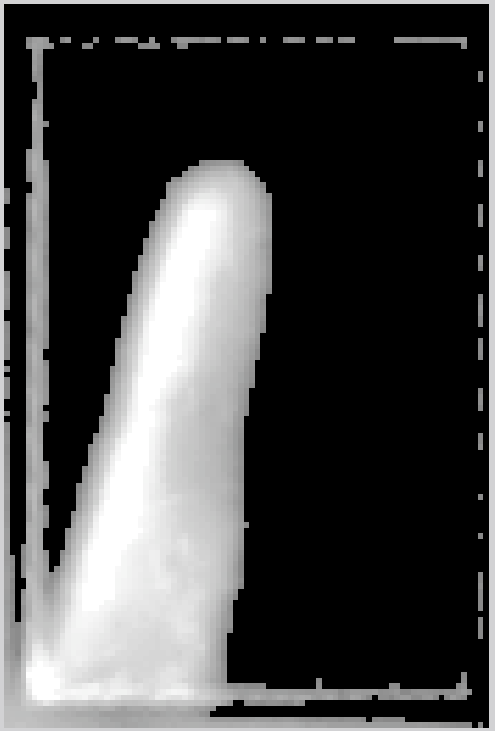
- The order of the cue-parser
  - Motion
  - Hue Blob
  - Shape

An Example VICon - A Button

- The order of the cue-parser
  - Motion
  - Hue Blob
  - Shape

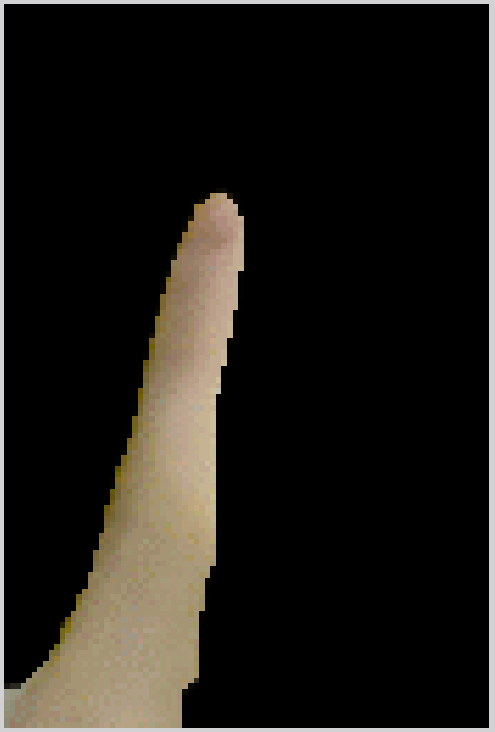
An Example VICon - A Button

- The order of the cue-parser
  - Motion
  - Hue Blob
  - Shape
- Background Subtraction



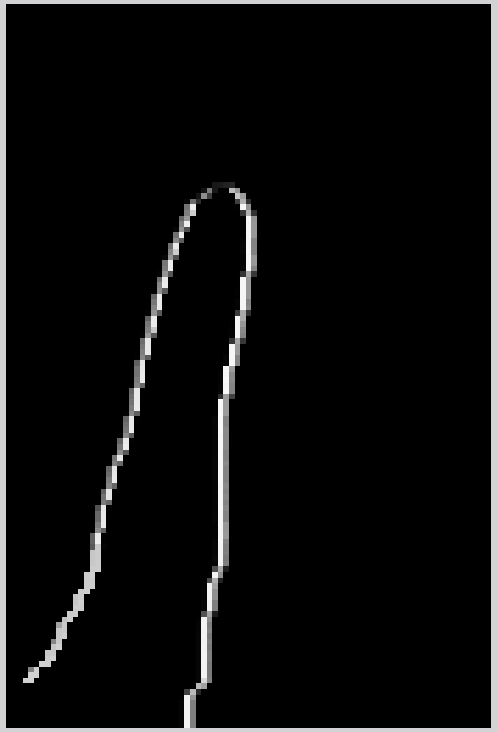
An Example VICon - A Button

- The order of the cue-parser
  - Motion
  - Hue Blob
  - Shape



An Example VICon - A Button

- The order of the cue-parser
  - Motion
  - Hue Blob
  - Shape



Computation Minimized

- Constant computations per-pixel.
  - In this case, a difference and a threshold.
- With action, increased computation only occurs near the action.
- Unnecessary computation removed.

Example using Motion Dynamics

- A Stochastic VICon via Hidden Markov Model.
  - Commonly used in Speech Recognition.

- Emulates a simple button
  - 2 state VICon model

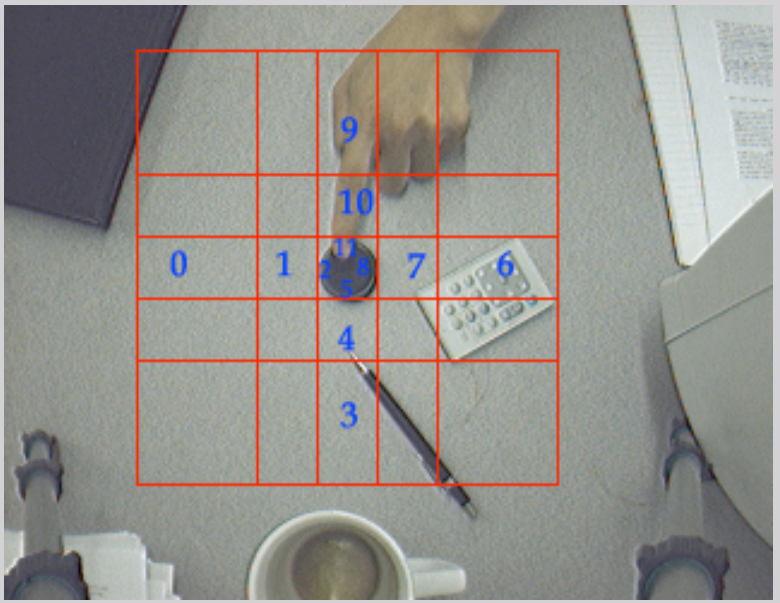
The HMM State-Space

- Convert input image stream into a series of symbols that describes the system state.
- Discrete feature describing current position and orientation of the finger tip.

- 3 Distances
- 4 Directions

  Up,left,etc

- Yields 12 states

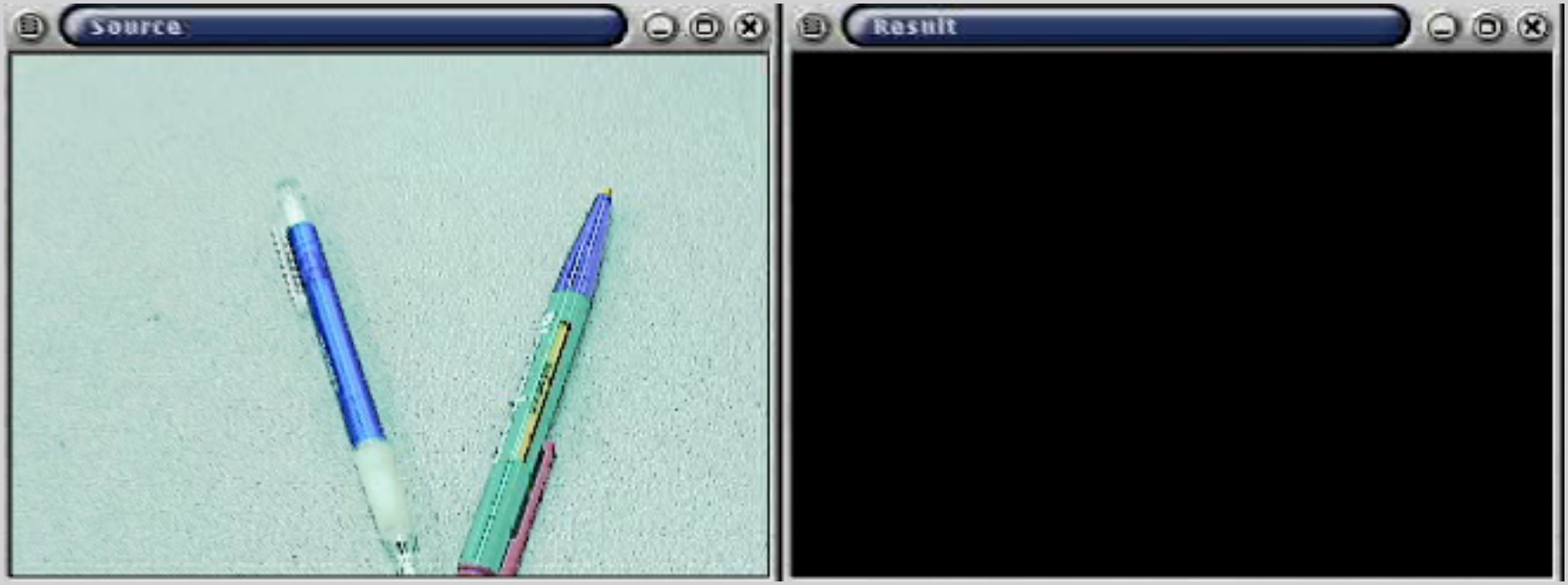


BG/FG Modeling & Segmentation

- Assume static camera.
- Hue Histogram to model appearance on-line.
- Segmentation based on histogram intersection.

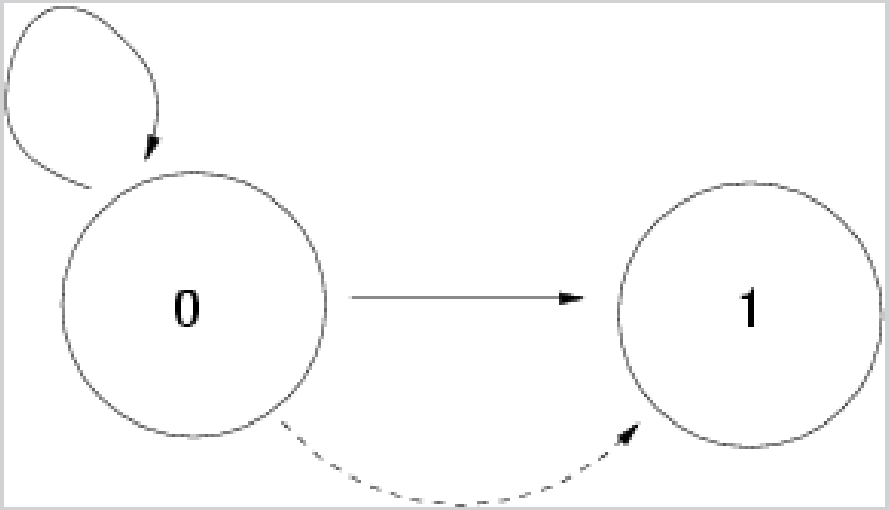
$$HI(Measure, Model) = \frac{\sum_{i=1}^{n} \min(Measure_i, Model_i)}{\sum_{i=1}^{n} Model_i}$$

Foreground Segmentation : Example



The HMM Structure

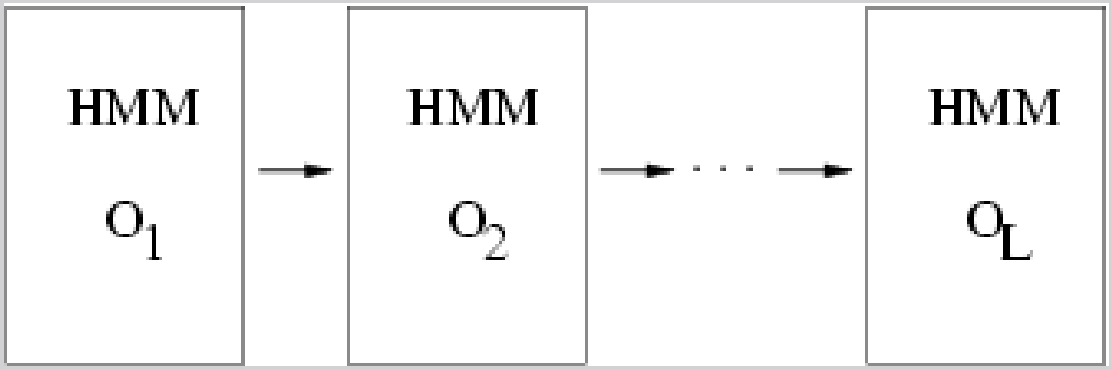
- Building block: singleton HMM.



- For each of the 12 states, define basic HMM to represent it.

The HMM Structure

- Build an HMM for each action category (up,down,etc).
- Concatenate singletons based on a representative sequence and fix a length L.



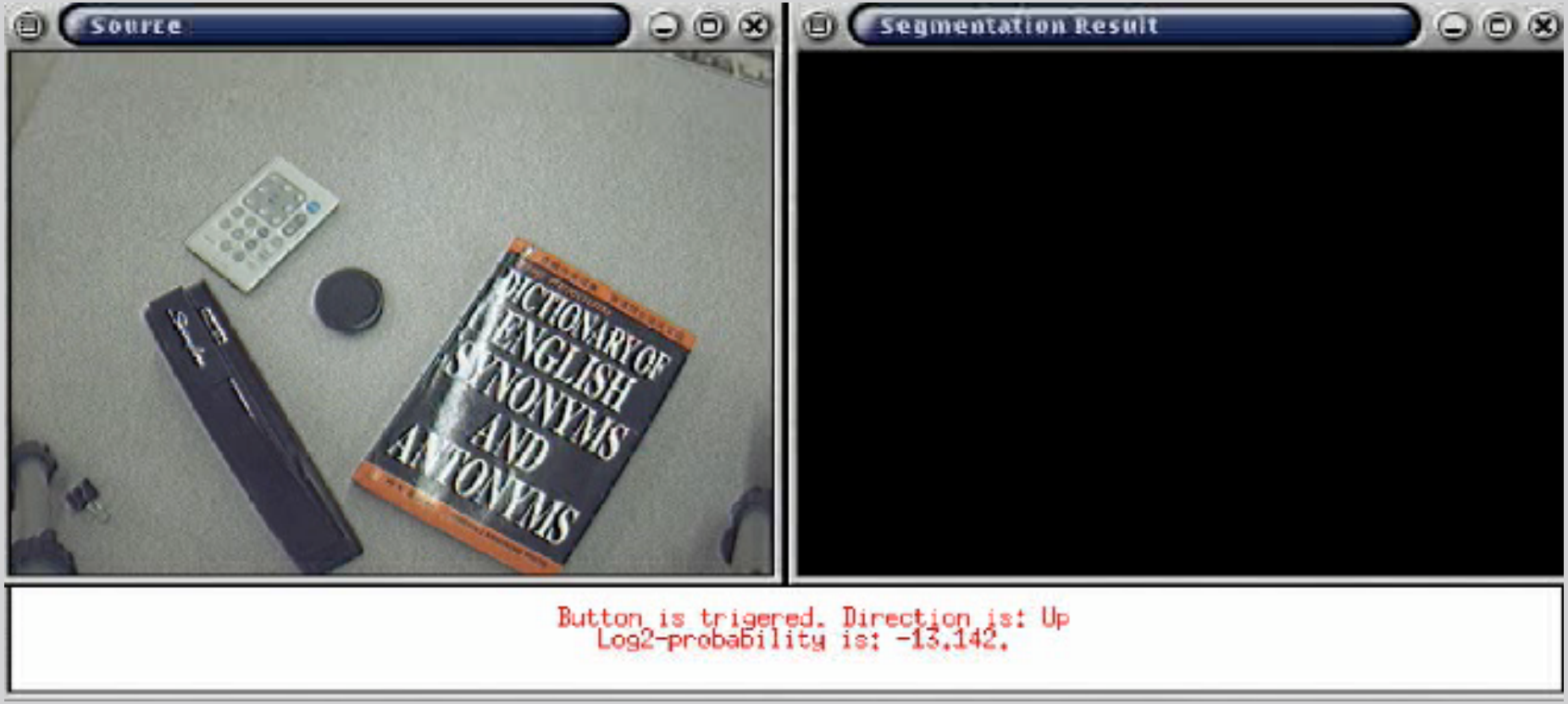
- If likelihood for a sequence is too low, consider it an illegal sequence.

HMM Training and Recognition

- Training set of valid actions.
- Select a characteristic sequence for each of the 4 directions.
- Run the Baum-Welch Algorithm.
- At run-time, for each length L sequence, attempt recognition.
  - If valid, trigger correct signal.

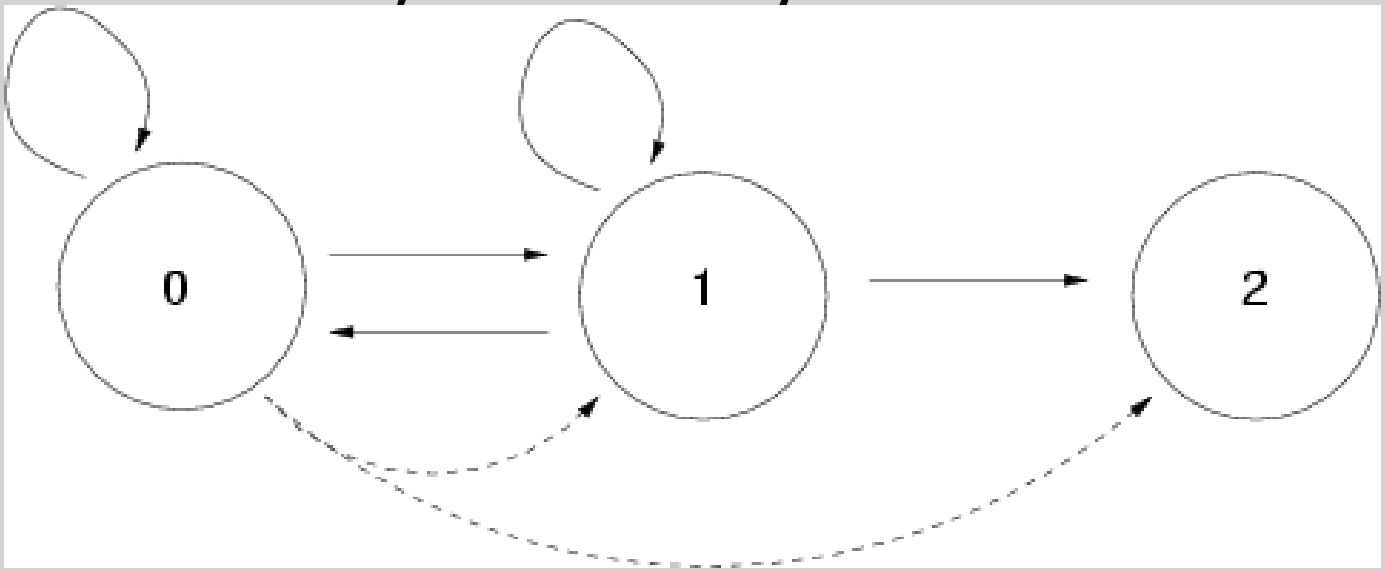
Experiment Results

- 76 sequences for training, over 300 for testing.
- 100% on training set; 96.8% on test set.



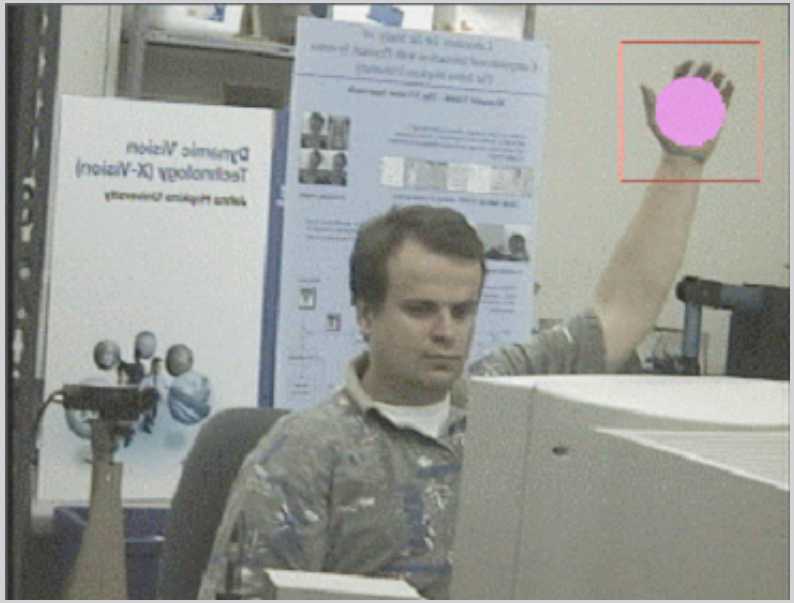
Improving the HMM Structure

- Singleton-based HMM is rudimentary
- Incorporate time dynamics into 1 multi-state, forward/backward HMM.



Interaction Modes 1

- 2D-2D Mirror
  - One camera observes user
  - Video stream displayed in interface.
  - VICons composited into video stream.

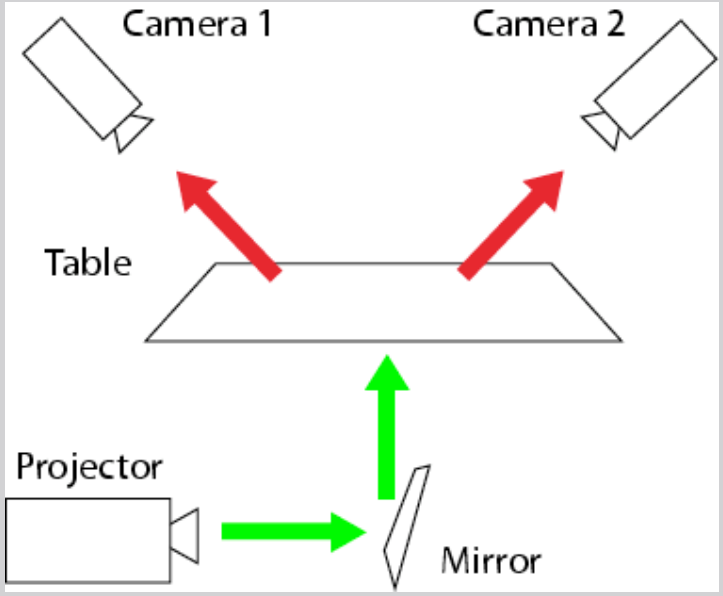


Interaction Modes 2 & 3

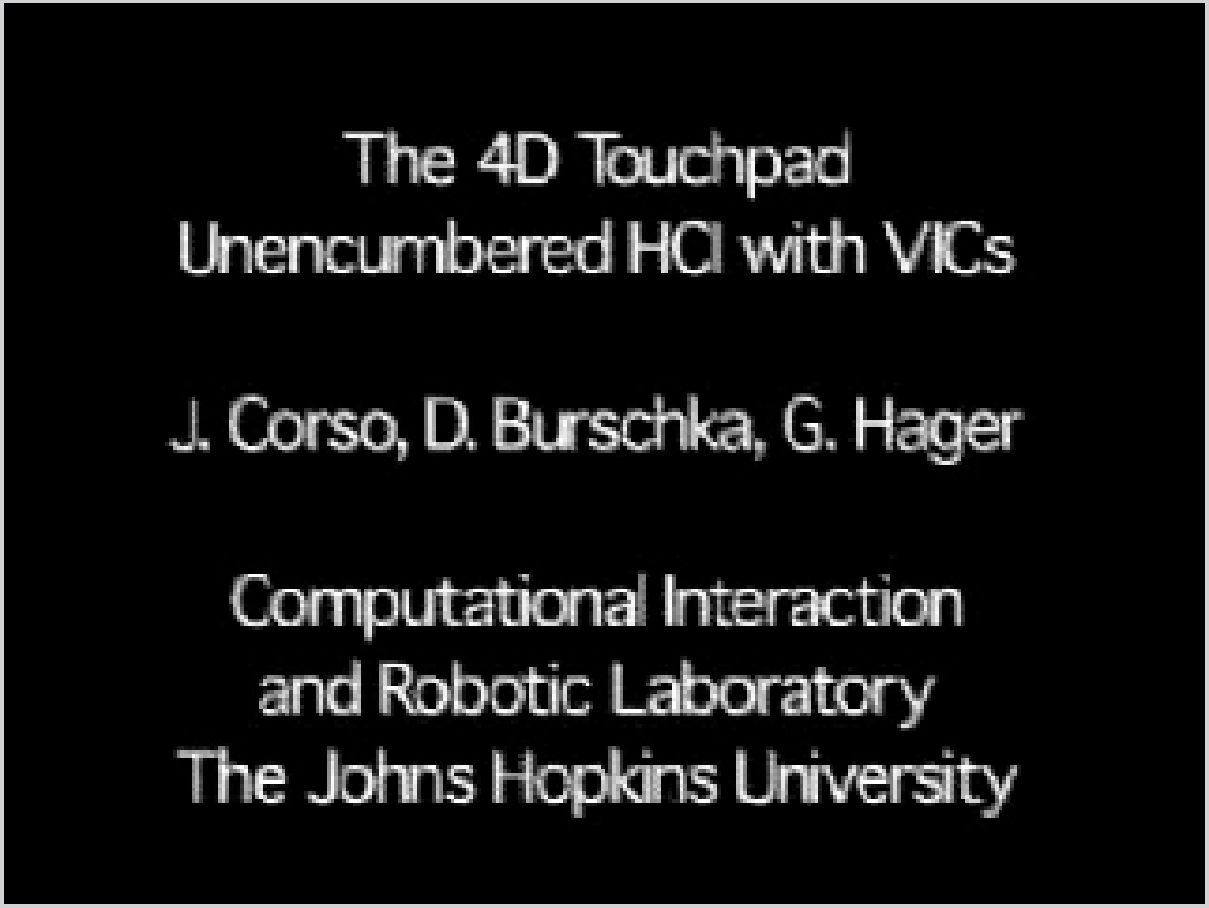
- 2.5D Augmented Reality
  - Video see-through
  - Constrain interface to a surface
- 3D Augmented Reality
  - Allow VICons to be fully 3D
- Examples
  - Surgery for 3D Microscopes; e.g. retinal
  - Motor-function training for young children.

Interaction Modes 4 & 5

- 2D-2D & 3D-2D Projection
  - 1, 2 or more cameras
  - The 4D-Touchpad [CVPRHCI 2003]
    - Provisional Patent Filed.



Video Example
3D-2D Projection - 4DT



Conclusions

- A new framework for transparently incorporating vision-based components into interface design.
- Our first system to incorporate motion dynamics in a formal manner.
- Can we fuse the higher spatial dimension and temporal nature of interaction in a structured way?
- A language of interaction?

- Thank You.
- Questions/Comments?
- Acknowledgements:
  - This material is based upon work supported by the National Science Foundation under Grant No. 0112882. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

VICon Click Recognition

- A basic interface task.
- Problems to resolve:
  1. Visual cues to detect and describe finger's movement
  2. Reliable model to recognize button-push action; or generally, more gestures to interact with VICon
  3. Computation efficiency to fulfill real-time interaction

HMM-based Gesture Recognition

- Hidden Markov Model's
  - Dynamic stochastic procedures.
  - Speech, gestures, etc.
- Discrete HMMs
  - Model the pattern of a button-push
  - The finger moves toward the button and stays on the button for a short period of time.

HMM Models

- Singleton HMM for each of the 12 features: a simple 2-state forward HMM structure.
- Big picture: Select a characteristic sequence to model the action. Concatenate corresponding singletons to form the model HMM.
- Training and recognition: standard Baum-Welch algorithm (Forward/Backward) . Normalized probability as the recognition criterion.
- Rejection threshold selected as the lowest probability among training sequences.