6.866 projects

Proposals to us by today. We will ok them by Oct. 31.

3 possible project types:

- Original implementation of an existing algorithm.
- Rigorous evaluation of existing algorithm.
- Synthesis or comparison of several research papers.

12	10/17	Bayesian Analysis			Freeman Slides
13	10/22	Optic Flow and Direct SFM	Req: H 12, 17	Exam #1 Due	Darrell Slides
14	10/24	Affine Reconstruction	Req: FP 12	Pset #3 Assigned	Darrell Slides
15	10/29	TBD			
16	10/31	Statistical Classifiers I			
17	11/5	Statistical Classifiers II		Pset #3 Due	

12	10/17	Bayesian Analysis			Freeman Slides
13	10/22	Optic Flow and Direct SFM	Req: H 12, 17	Exam #1 Due	Darrell Slides
14	10/24	Affine Reconstruction	Req: FP 12	Pset #3 Assigned	Darrell Slides
15	10/29	Cameras looking at people: cheap tricks using the tools we've seen so far.			
16	10/31	Statistical Classifiers I			
17	11/5	Statistical Classifiers II		Pset #3 Due	

12	10/17	Bayesian Analysis			Freeman Slides
13	10/22	Optic Flow and Direct SFM	Req: H 12, 17	Exam #1 Due	Darrell Slides
14	10/24	Affine Reconstruction	Req: FP 12	Pset #3 Assigned	Darrell Slides
15	10/29	Cameras looking at people: cheap tricks using the tools we've seen so far.			
16	10/31	Statistical Classifiers I (for faces)			
17	11/5	Statistical Classifiers II		Pset #3 Due	

Does anyone mind…

If I use your photographed face for a simple face-detection demo program that we'll run in class next time?

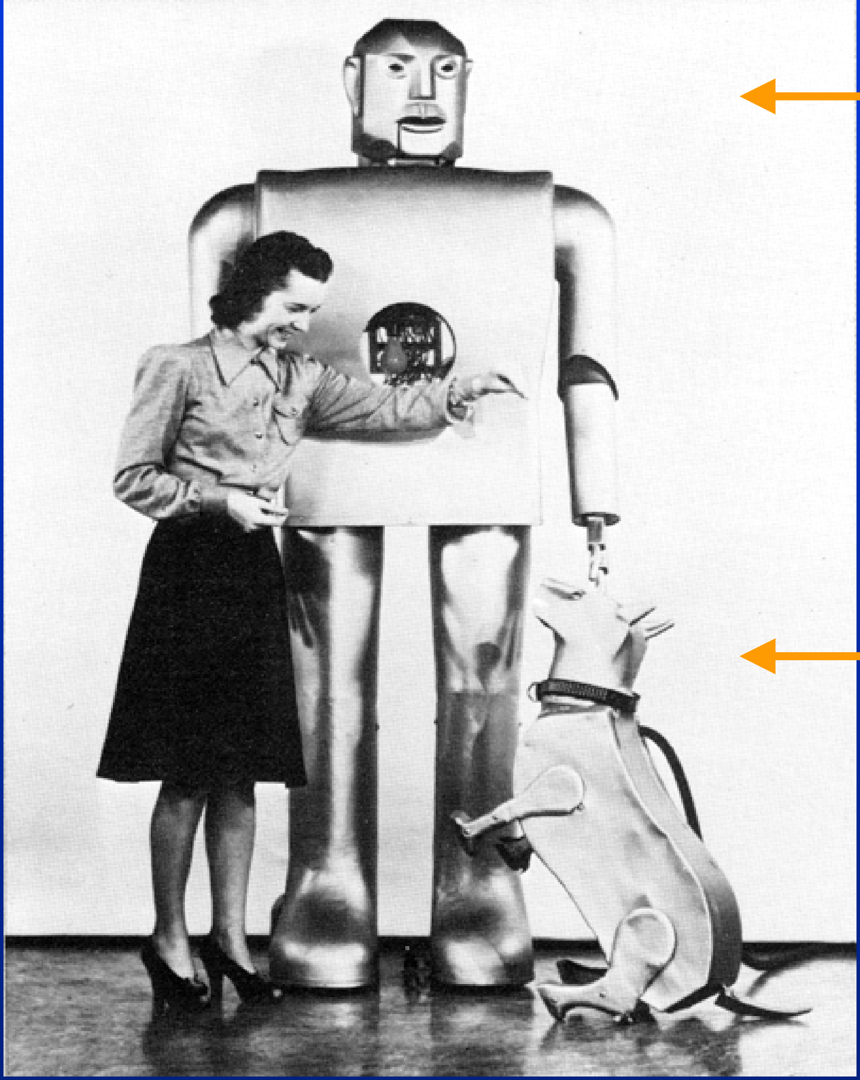
If you do mind, please let me know (before Thursday).

Today: Cameras looking at people

A mini-application lecture: under controlled conditions (not general conditions), what human interaction applications can you build with the tools we've developed so far?
To be compared with: more sophisticated detection, classification, and tracking tools that we'll study over the rest of the course.

MIT 6.801/6.866
Oct. 29, 2002

Yesterday's tomorrow



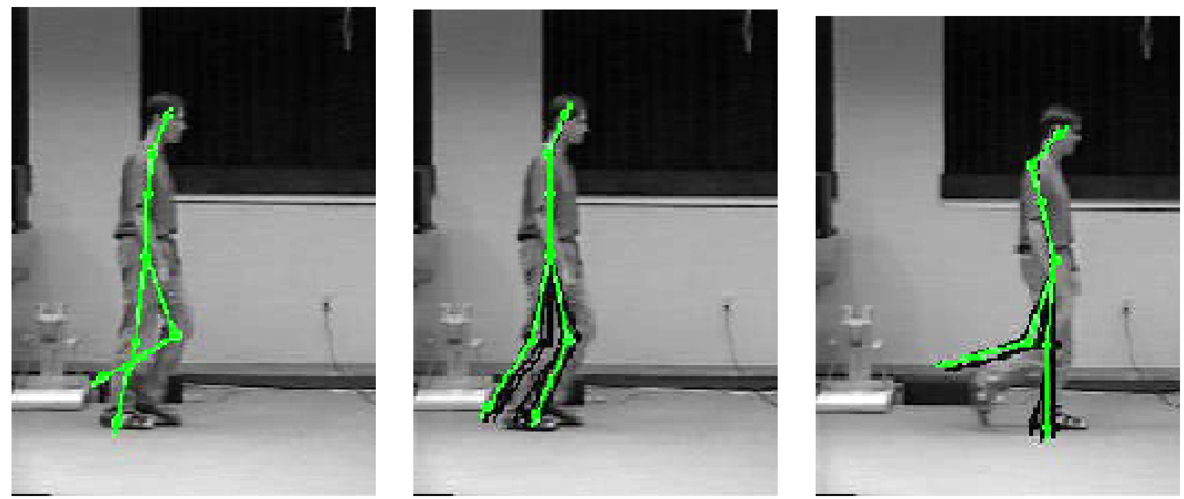
Elektro

Sparko

New York Worlds Fair, 1939

(Westinghouse Historical Collection)

Computer vision still needs to become more robust



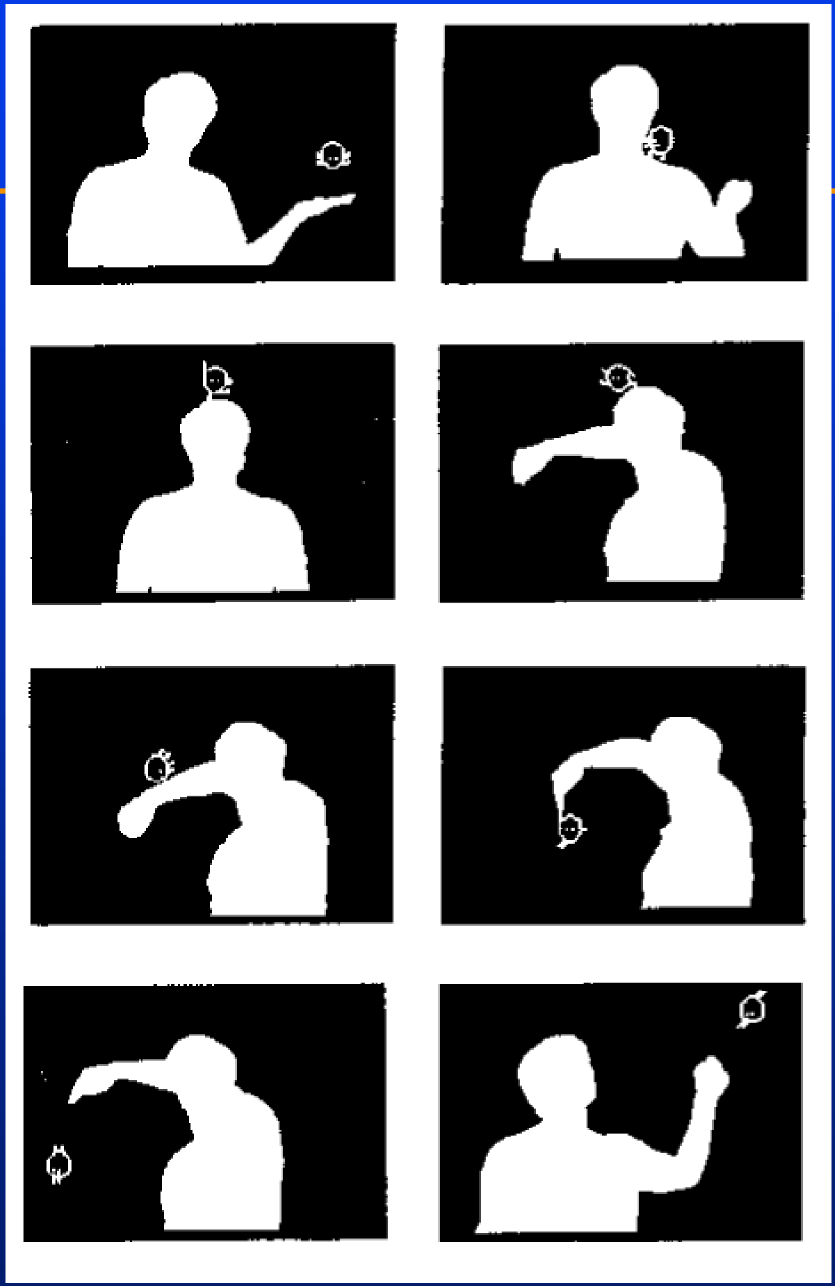
(a) (b) (c)

Figure 4: (a) Tracker (in white) using constant velocity predictor drifts off track by frame 7. (b) SLDS-based tracker is on track at frame 7. Model (switching state) 3 has the highest likelihood. Black lines show prior mean and observation. (c) SLDS tracker at frame 20.

Pavlovic, Rehg, Cham, and Murphy, Intl. Conf. Computer Vision, 1999

But we can fake it with clever system design

M. Krueger,
"Artificial Reality",
Addison-Wesley, 1983.



Research at MERL on fast, low-cost vision systems

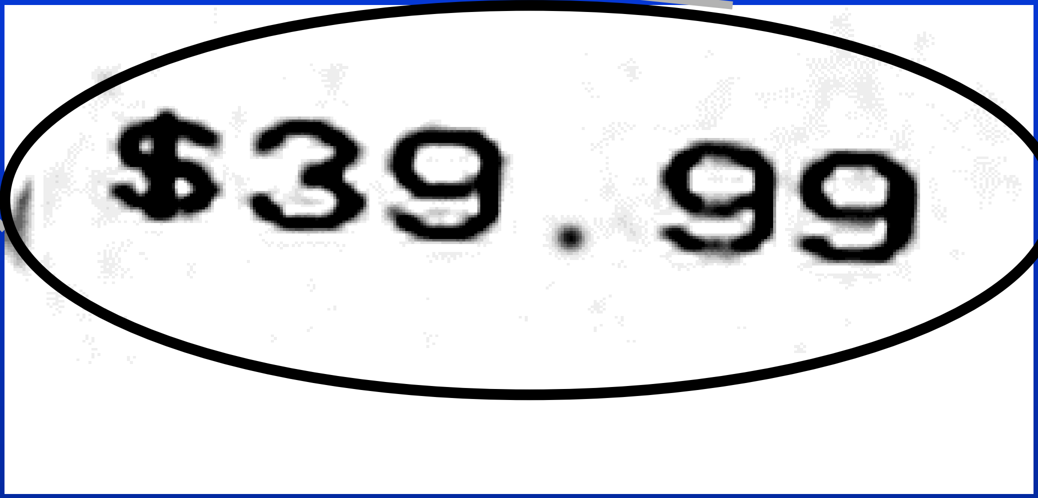
From MERL and Mitsubishi Electric:

David Anderson, Paul Beardsley, Chris Dodge, William Freeman, Hiroshi Kage, Kazuo Kyuma, Darren Leigh, Neal McKenzie, Yasunari Miyake, Michal Roth, Ken-ichi Tanaka, Craig Weissman, William Yerazunis

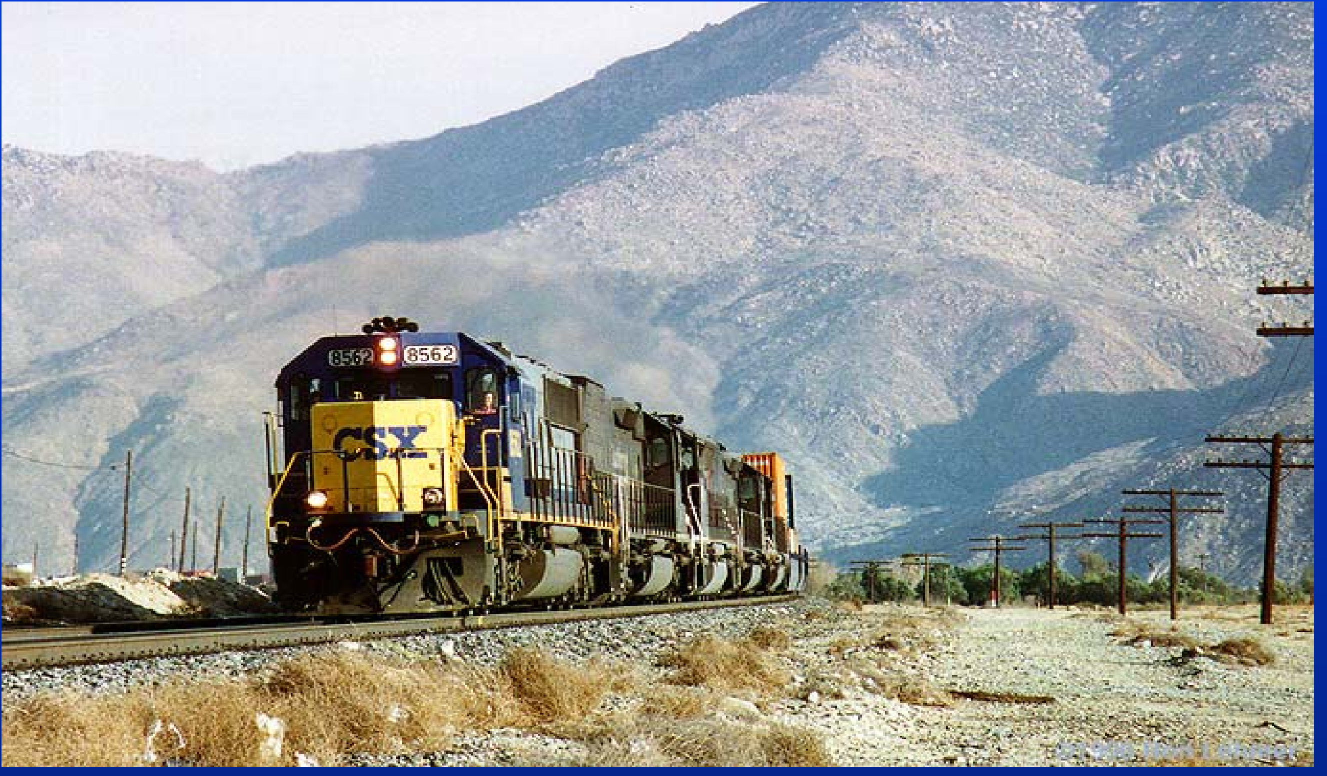
Computer vision based interface

The hope: video input will give a more expressive, natural or engaging interface.

Existing interfaces devices are fast & low-cost.



Applications make the vision easier.



Constraints simplify recognition--
if you know where the tracks are, it’s easy to guess where the train is.

There is a human in the loop.

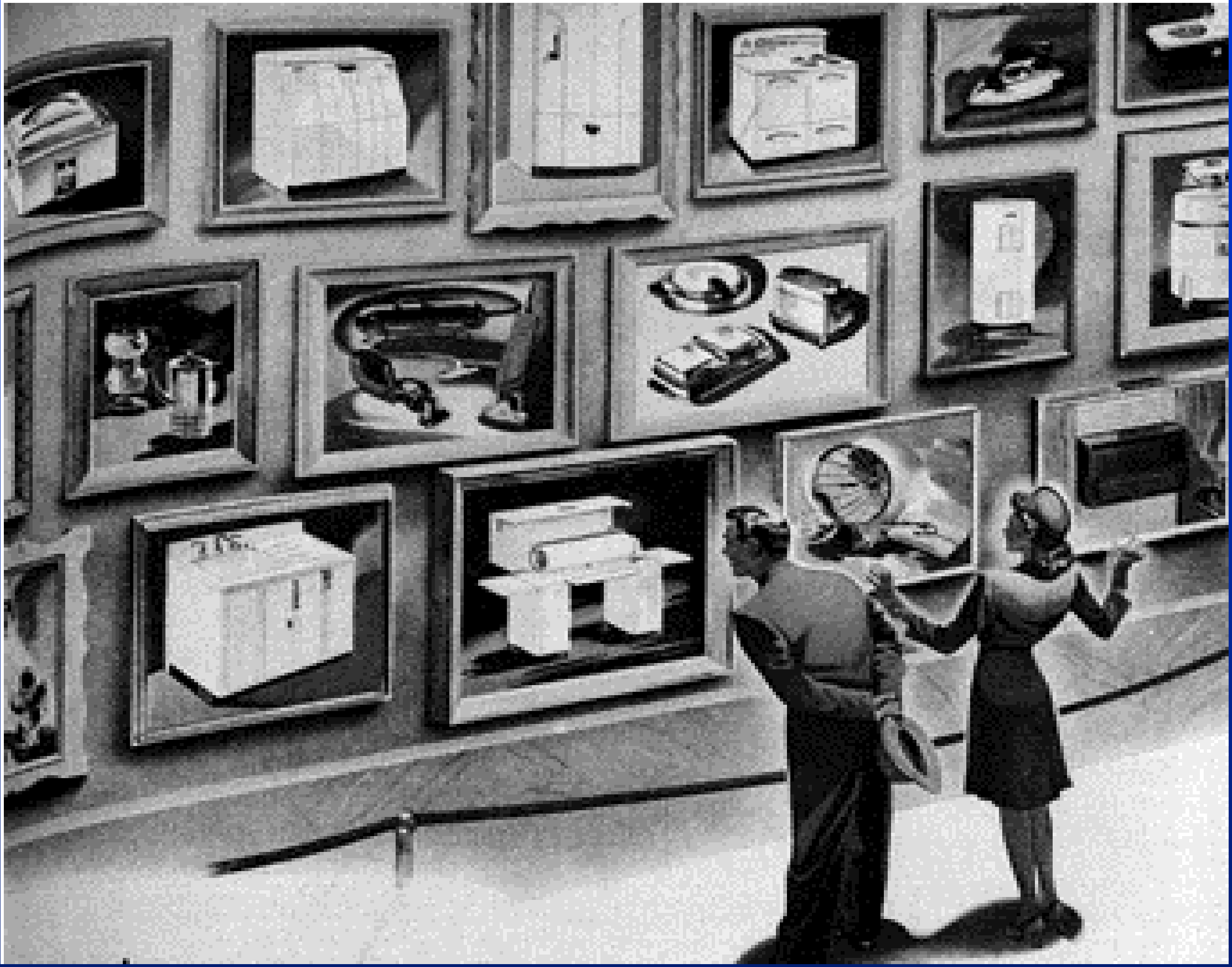
- Rich, immediate visual, audio feedback.
- The player can correct for algorithm imperfections.

Computer vision algorithms as ocean-going vessels

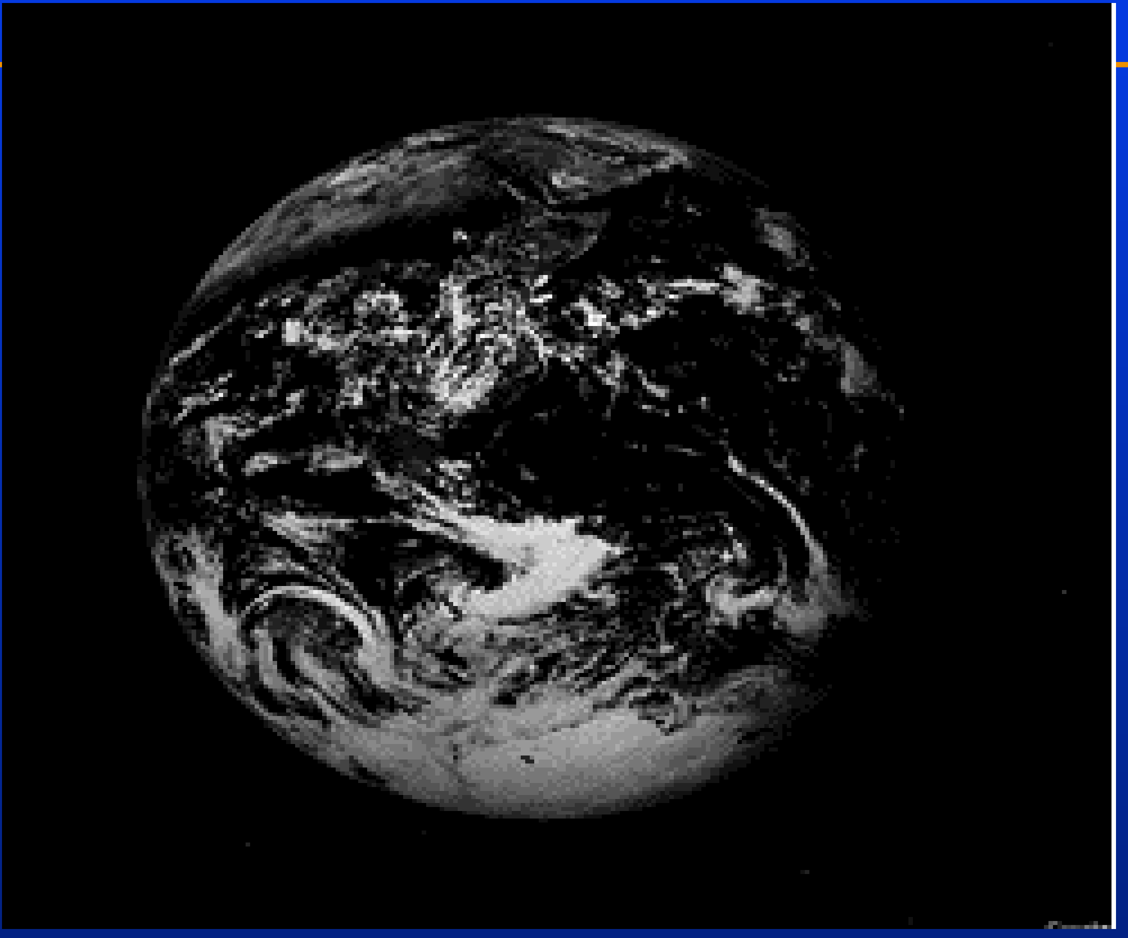
Computer vision algorithms as ocean-going vessels

this work ←

1. Selected appliance: television



television market



~1 billion television sets

Survey

"What high technology gadget has improved the quality of your life the most?"

What two things were mentioned most?

Survey results

"What high technology gadget has improved the quality of your life the most?"

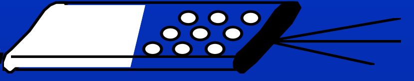
Microwave ovens and TV remote controls
--Porter/Novelli survey, 1995

message:
People value the ability to control a television from a distance.

Control of television set from a distance

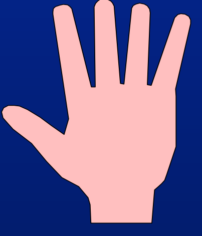
Wired remote control.

Infra-red remote control.



Voice control.

Gesture control.



Design constraints

- ***From the user's point of view***
- ***From the computer's point of view***

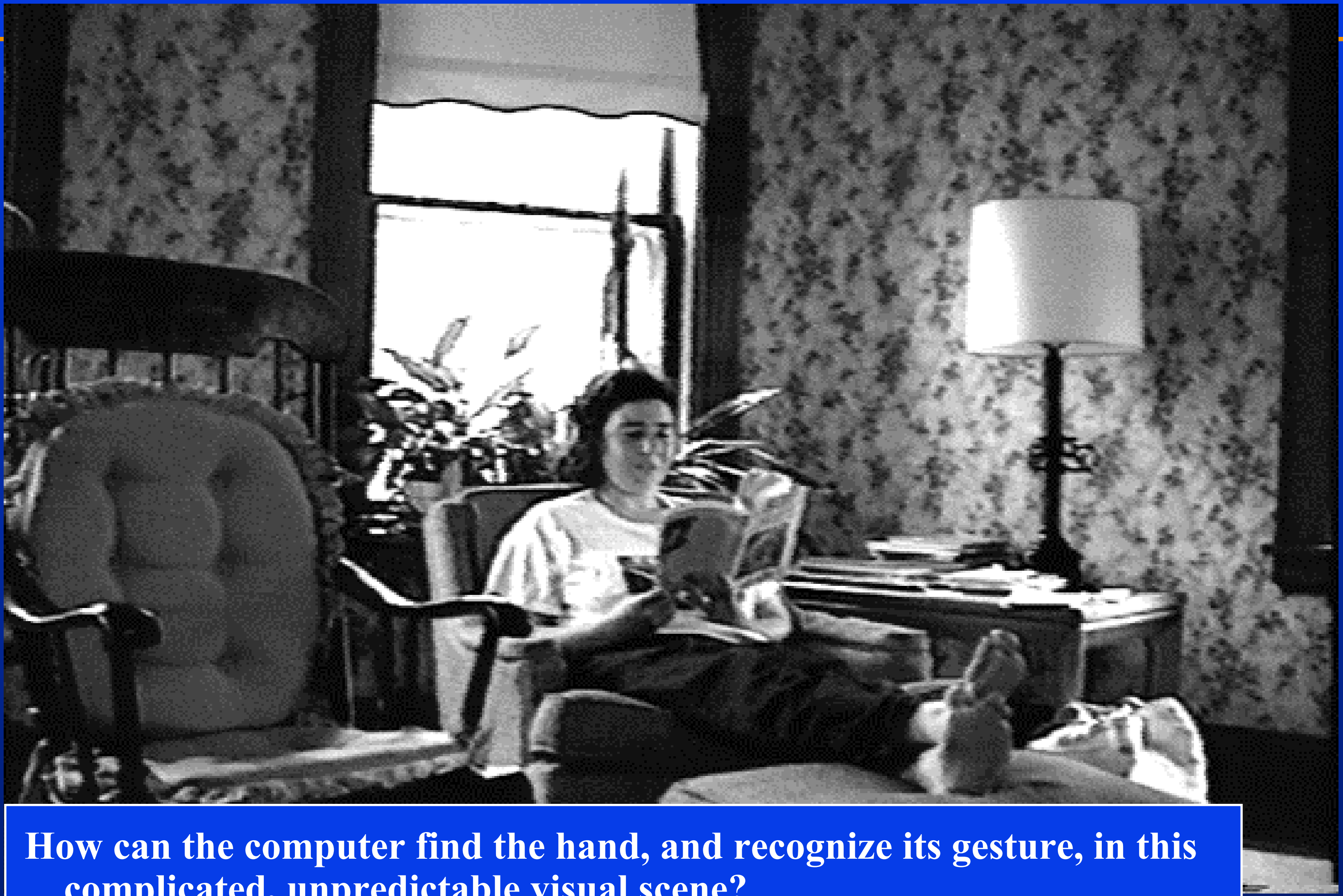
From the user's point of view:

Complex commands require complicated gestures?

"mute"

From the computer's point of view:

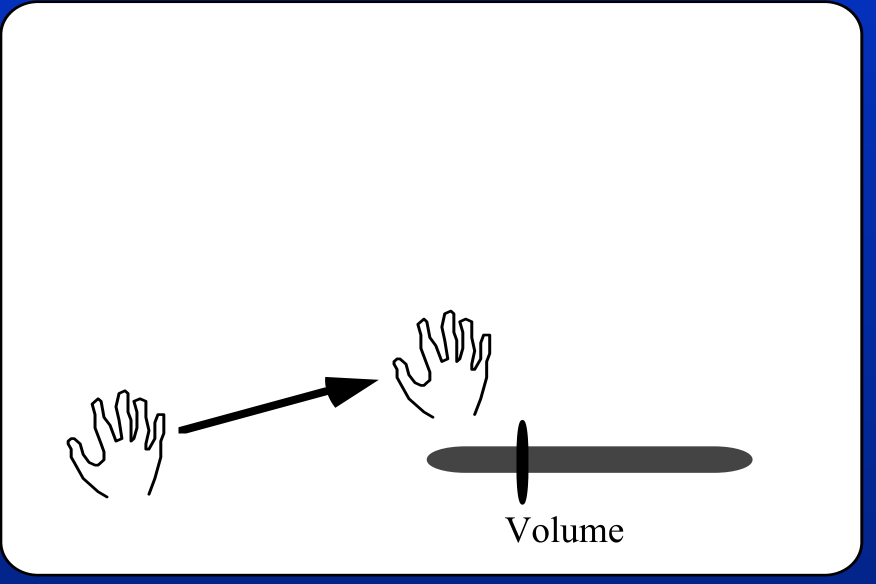
Living room scene is difficult



How can the computer find the hand, and recognize its gesture, in this complicated, unpredictable visual scene?

Our solution: exploit the visual feedback from the television

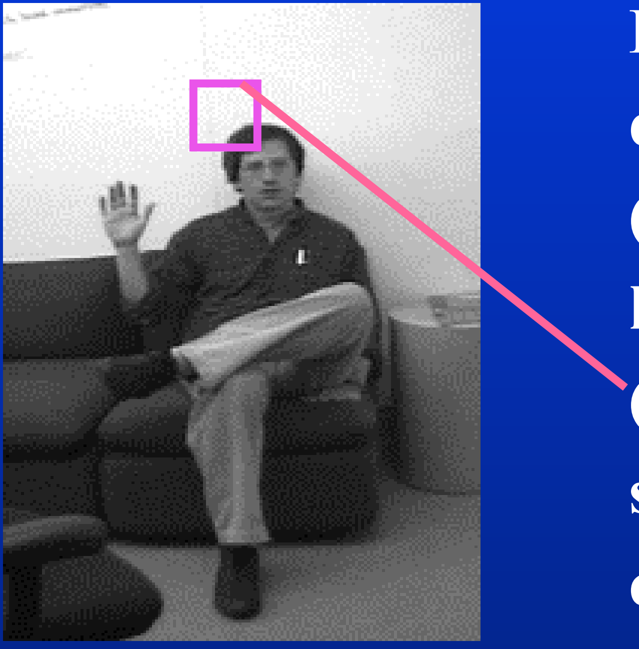
user



television

hand recognition method: template matching

template



image

Examine the squared difference between
(a) pixel values in the hand template, and
(b) pixel values in a square centered at each possible position in the image.

hand recognition method: normalized correlation

template

image

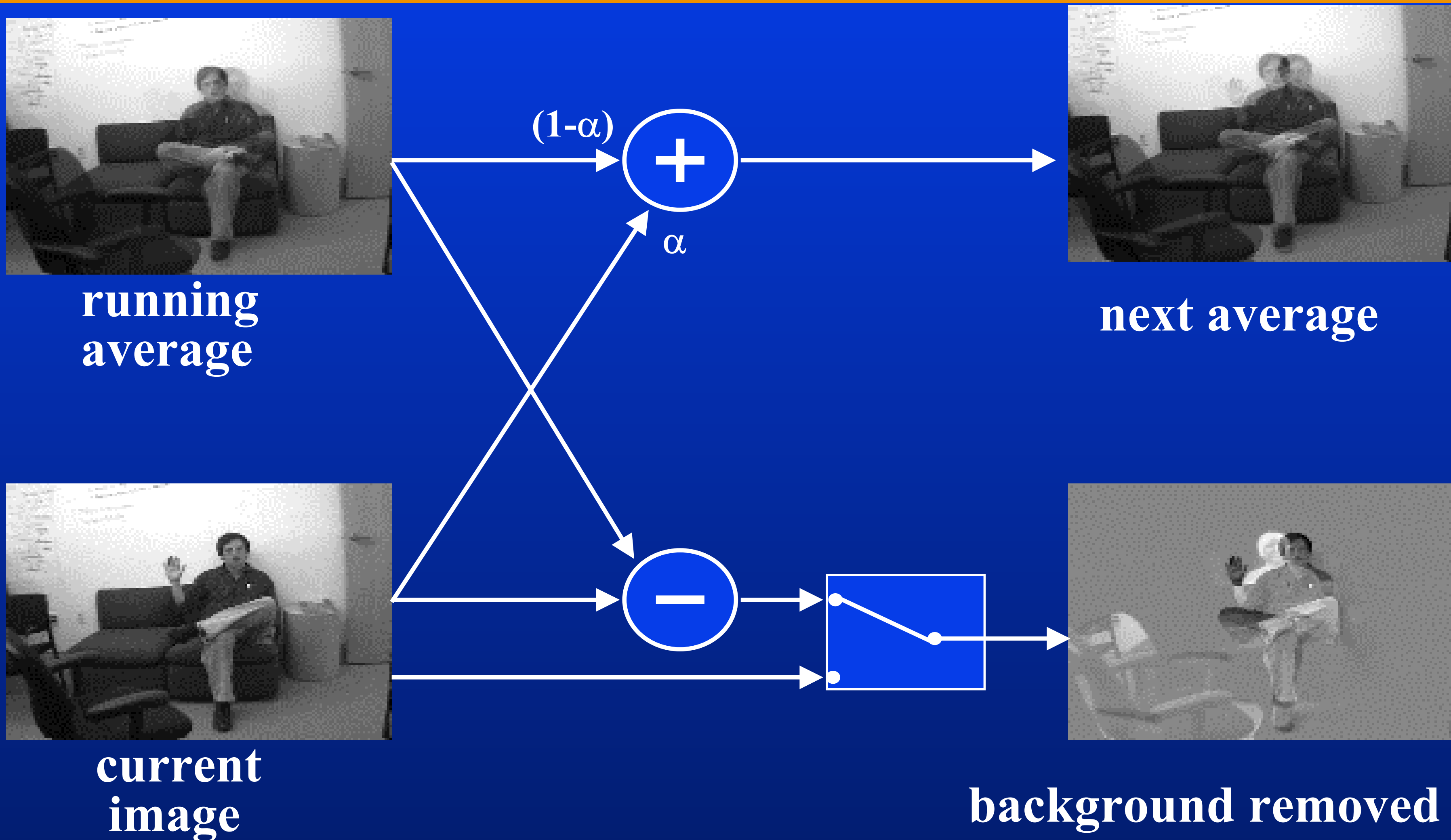
normalized correlation

Normalized correlation

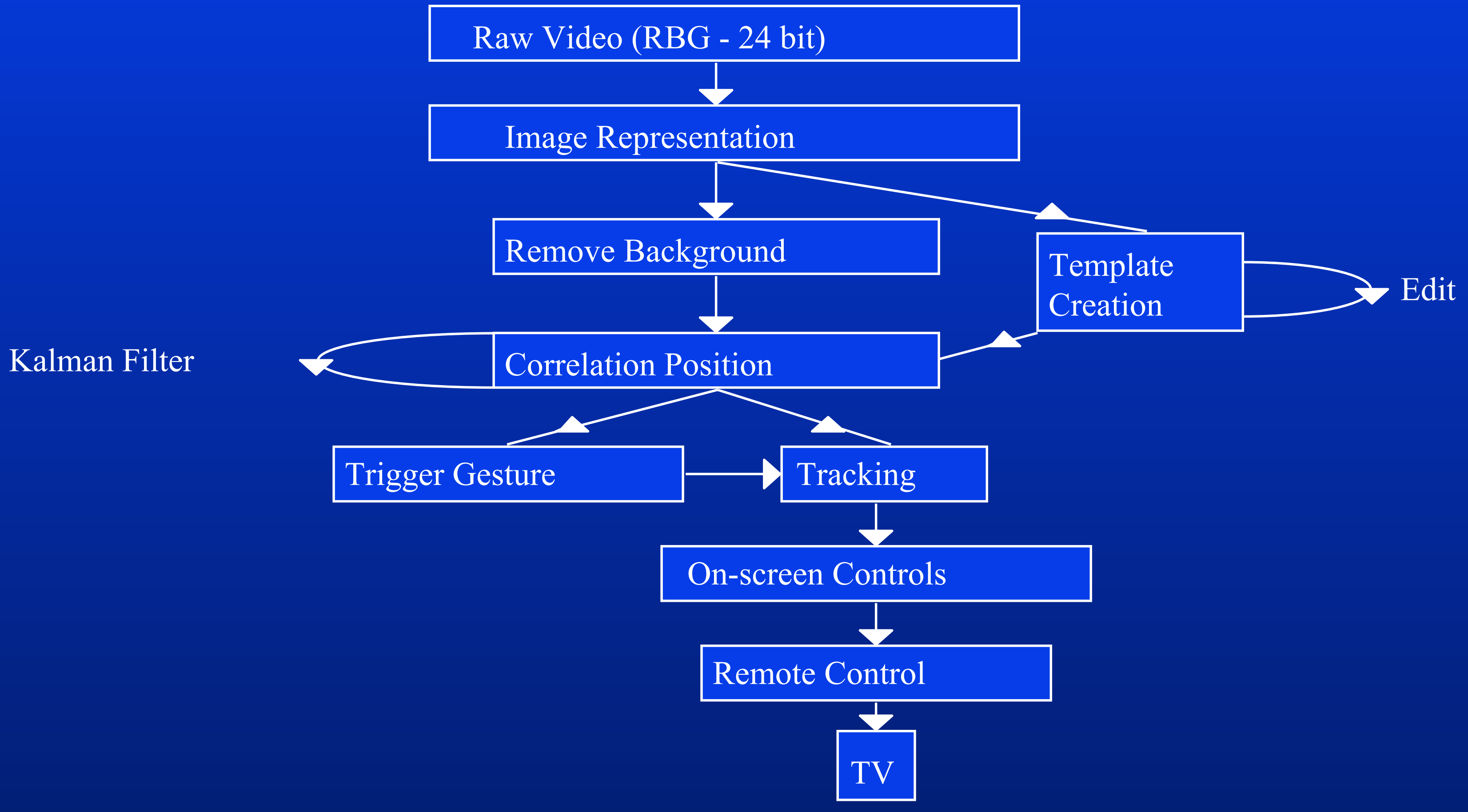
$$\frac{\vec{a} \cdot \vec{b}}{\sqrt{(\vec{a} \cdot \vec{a})(\vec{b} \cdot \vec{b})}}$$

Where a and b are vectors from rasterized patches of the image and template

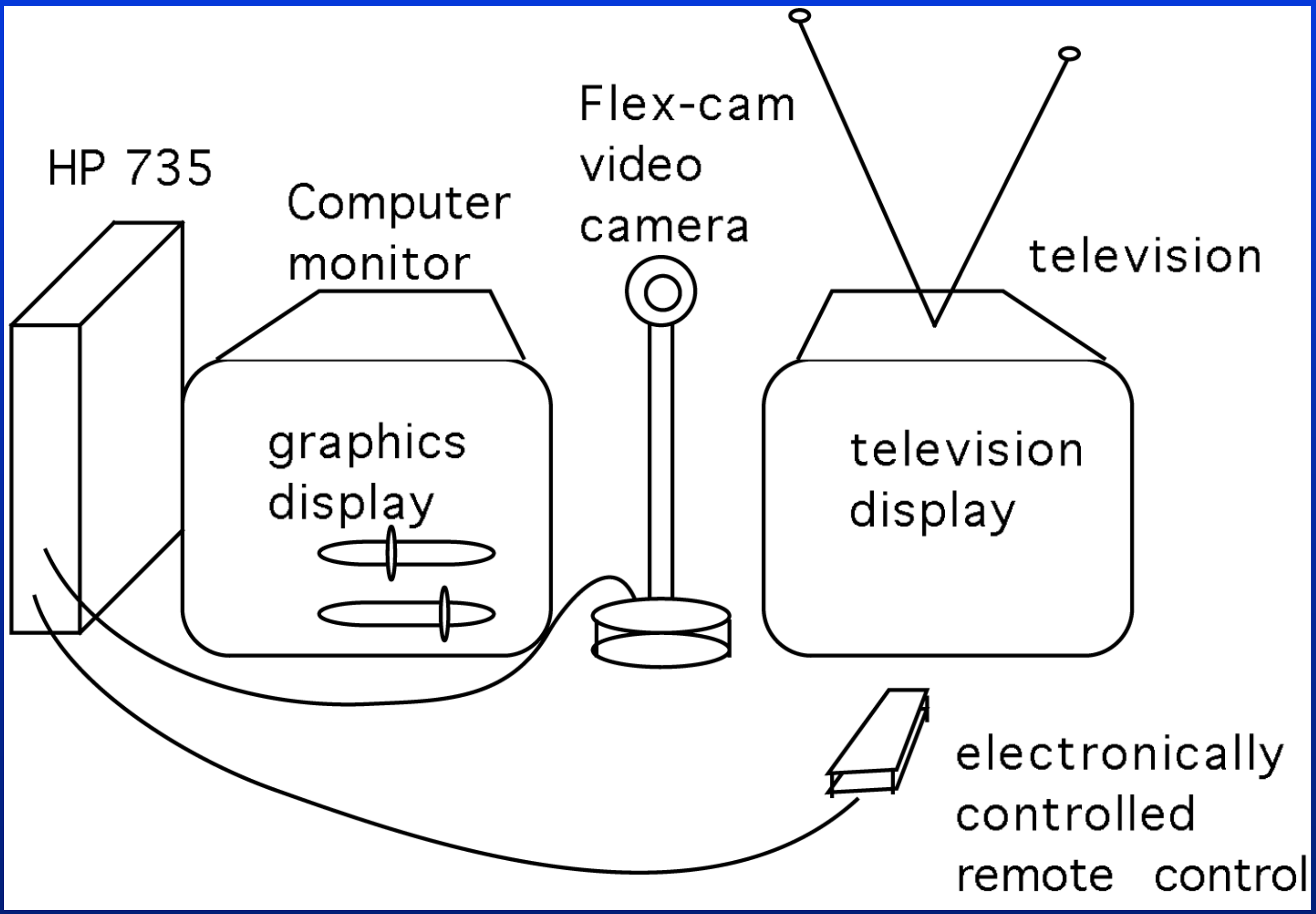
Background removal



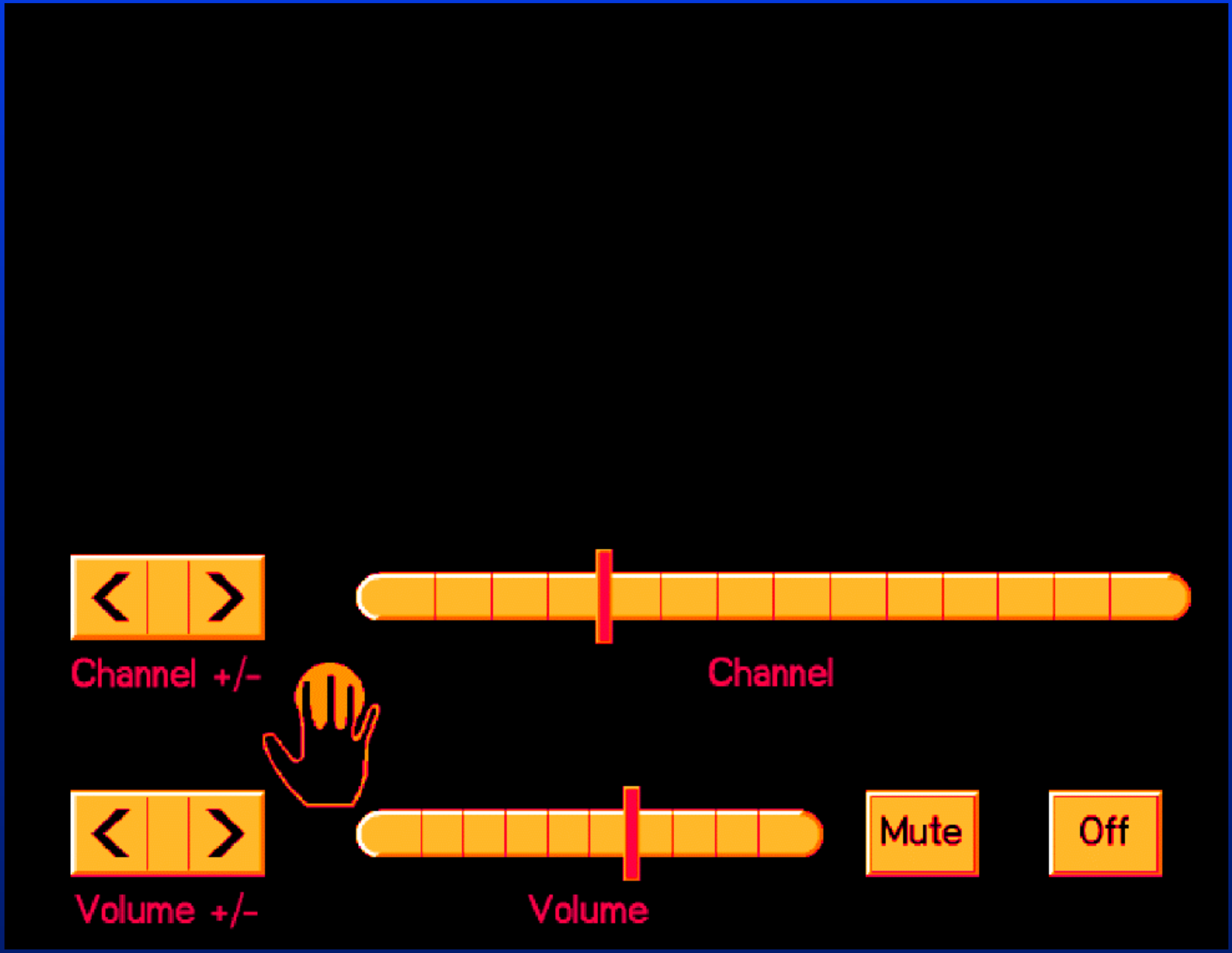
Processing block diagram



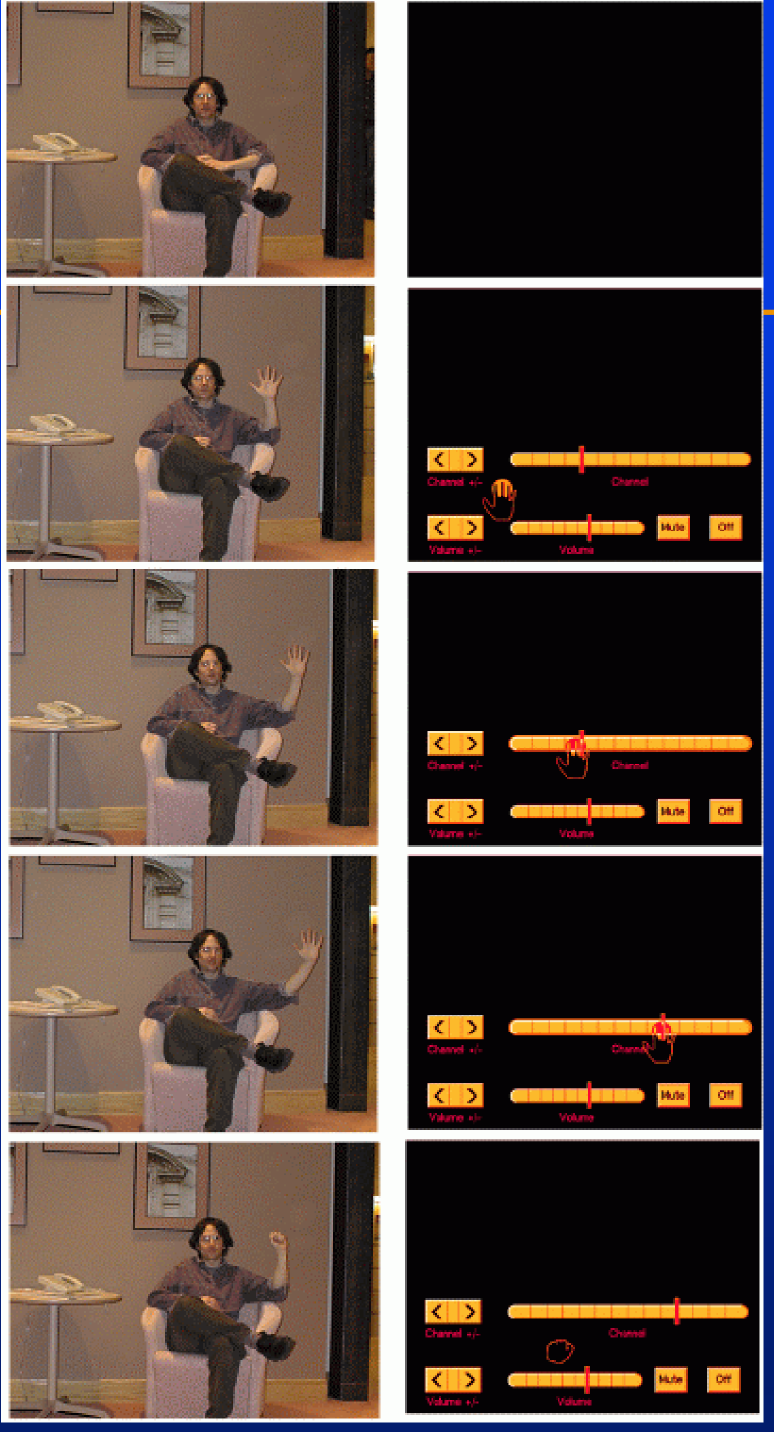
Prototype of television controlled by hand signals.



TV screen overlay



TV control



Video

Prototype limitations

- ***Distance from camera:***

  6 - 10 feet.

- ***Field of view:***

  trigger gesture: 15 ^{o} tracking: 25 ^{o}

- ***Coupling to television is loose.***
- ***Two screens instead of one.***
- ***Robustness during operation:***

  no template adaptation to different users.

  background removal may need variable contrast control.

Product hardware requirements

Short term

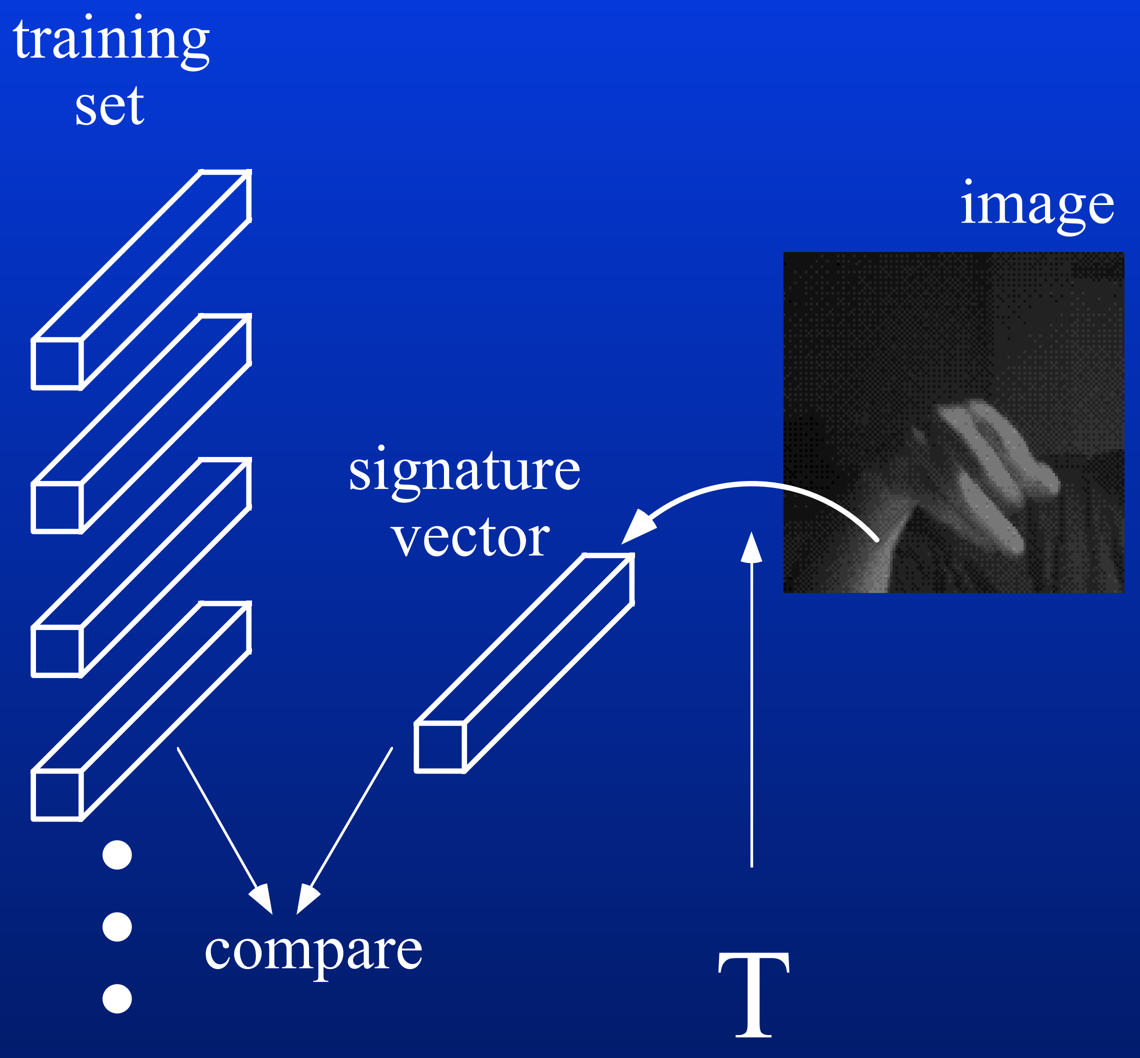
- camera
- video digitizer
- computer

Long term

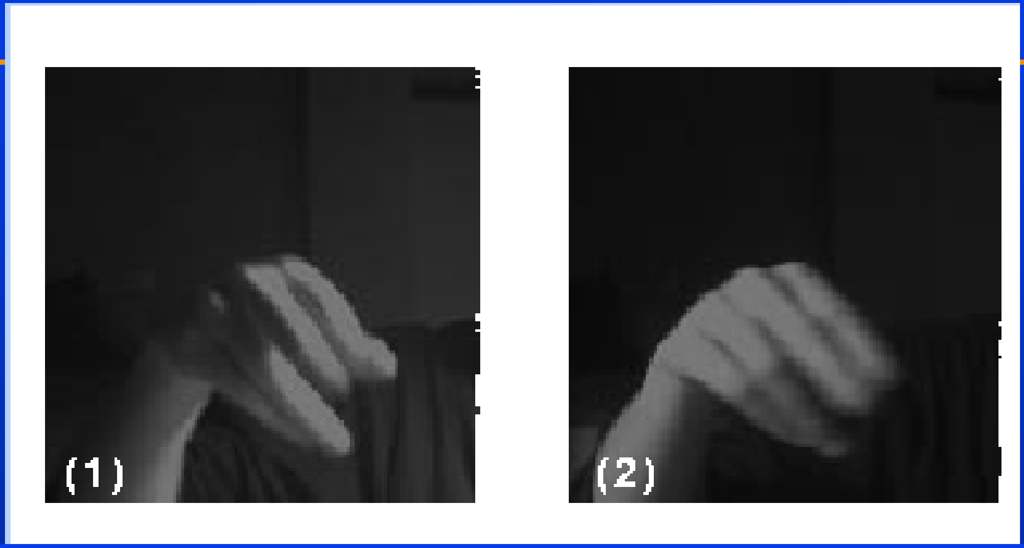
- TV's / computers / browsers will have cameras and powerful computers.
- <u>a software product.</u>

2. Simple gesture recognition method

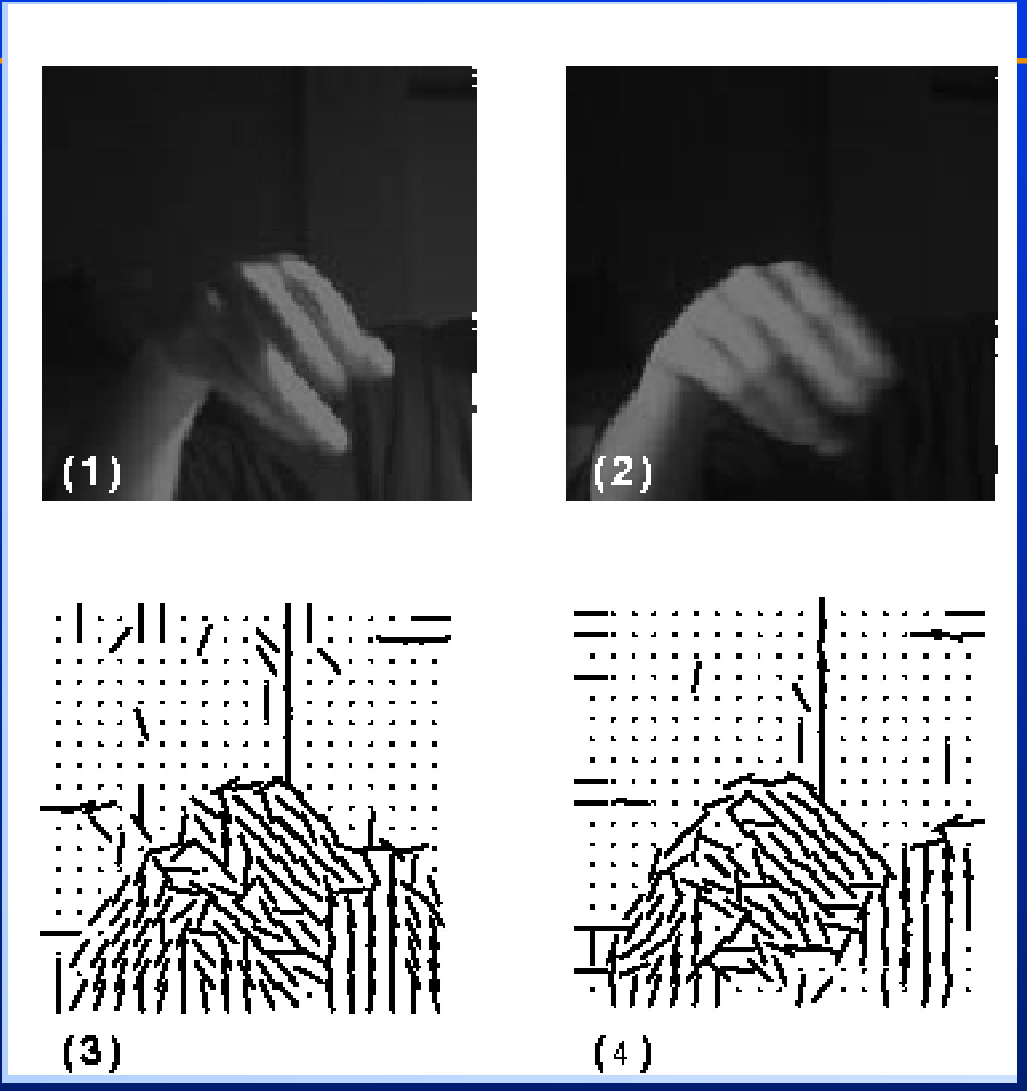
Real-time hand gesture recognition by orientation histograms



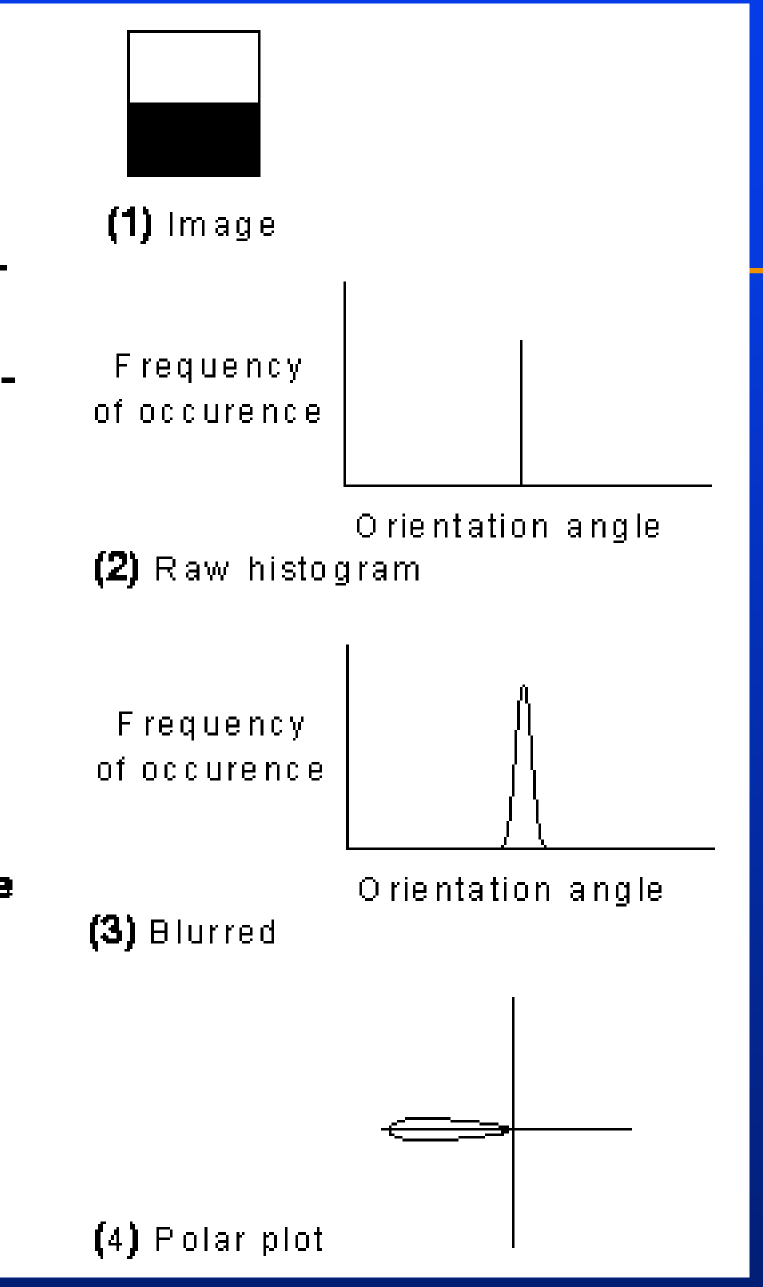
Orientation measurements (bottom) are more robust to lighting changes than are pixel intensities (top)



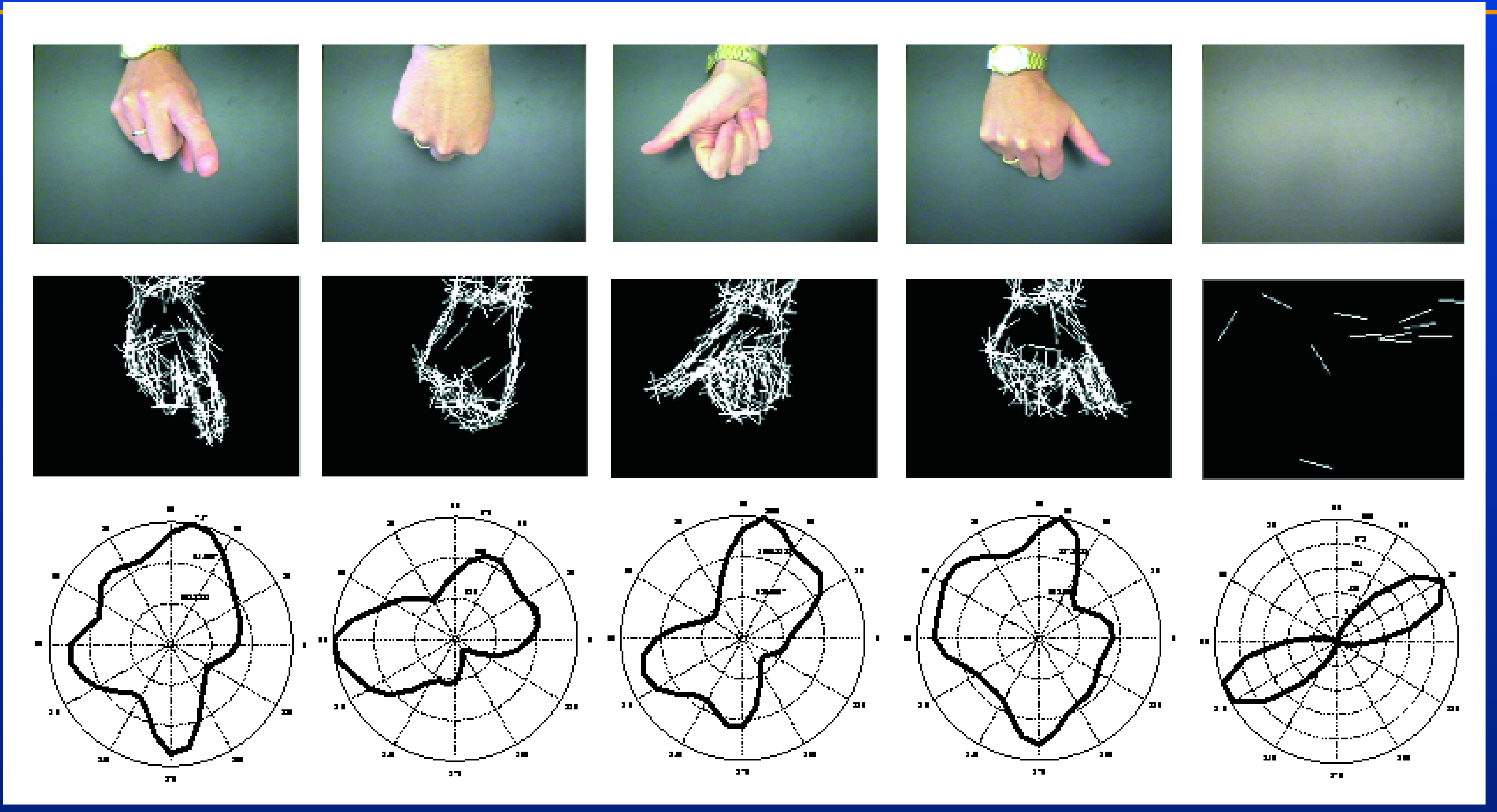
Orientation measurements (bottom) are more robust to lighting changes than are pixel intensities (top)



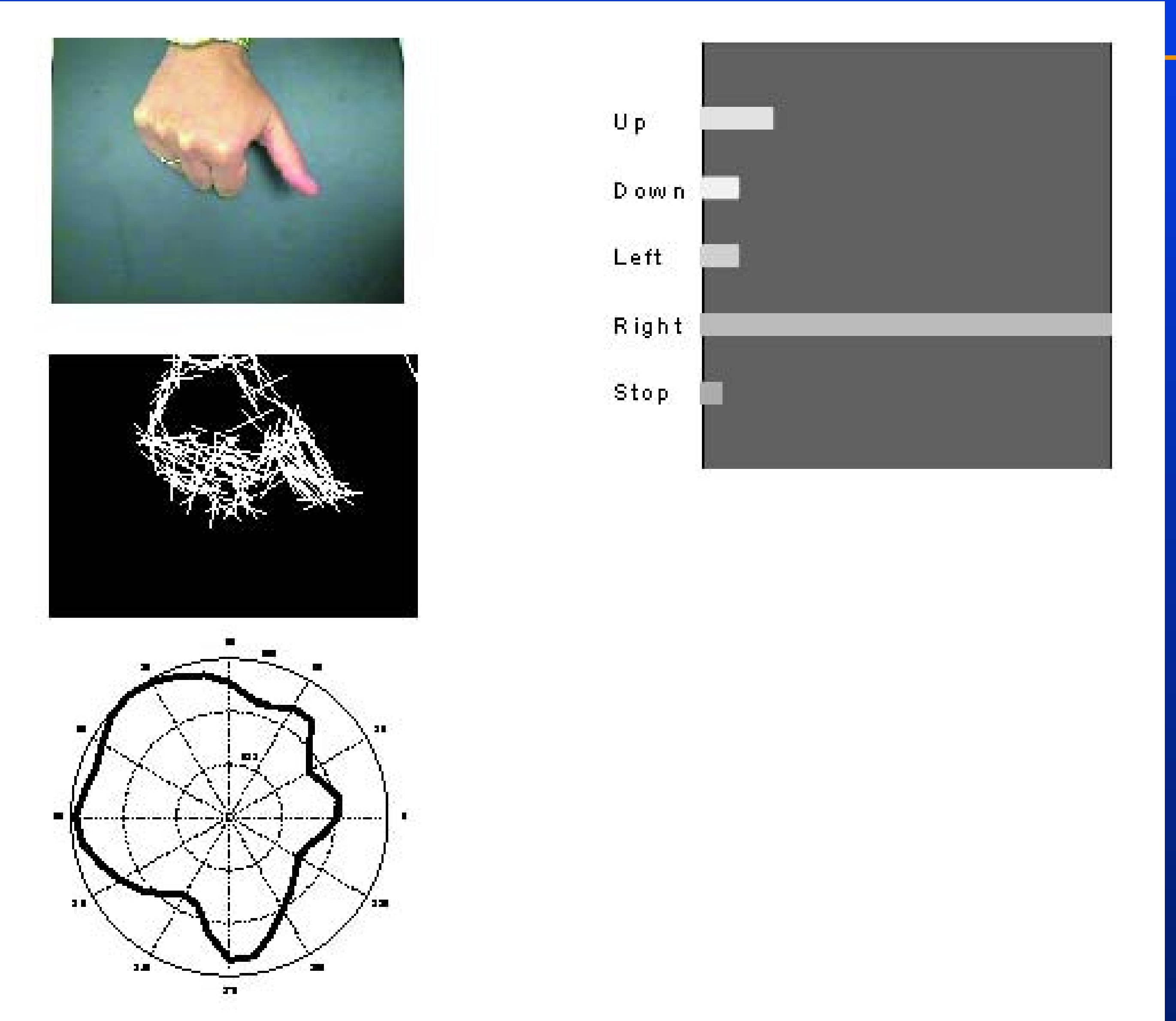
C Simple illustration of an orientation histogram. (1) An image of a horizontal edge has only one orientation at a sufficiently high contrast. (2) Thus the raw orientation histogram has counts at only one orientation value. (3) To allow neighboring orientations to sense each other, we blurred the raw histogram. (4) The same information, plotted in polar coordinates. We define the orientation to be the direction of the intensity gradient, plus 90 degrees.

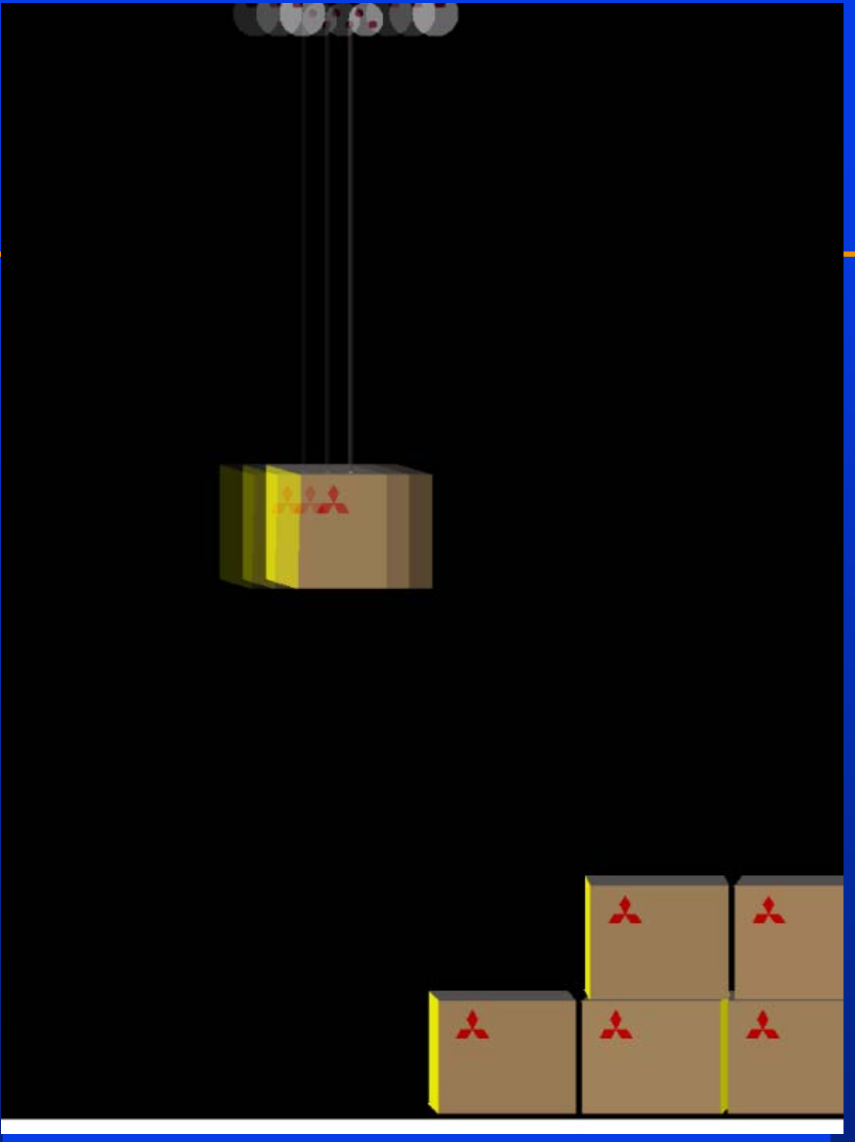


Images, orientation images, and orientation histograms for training set

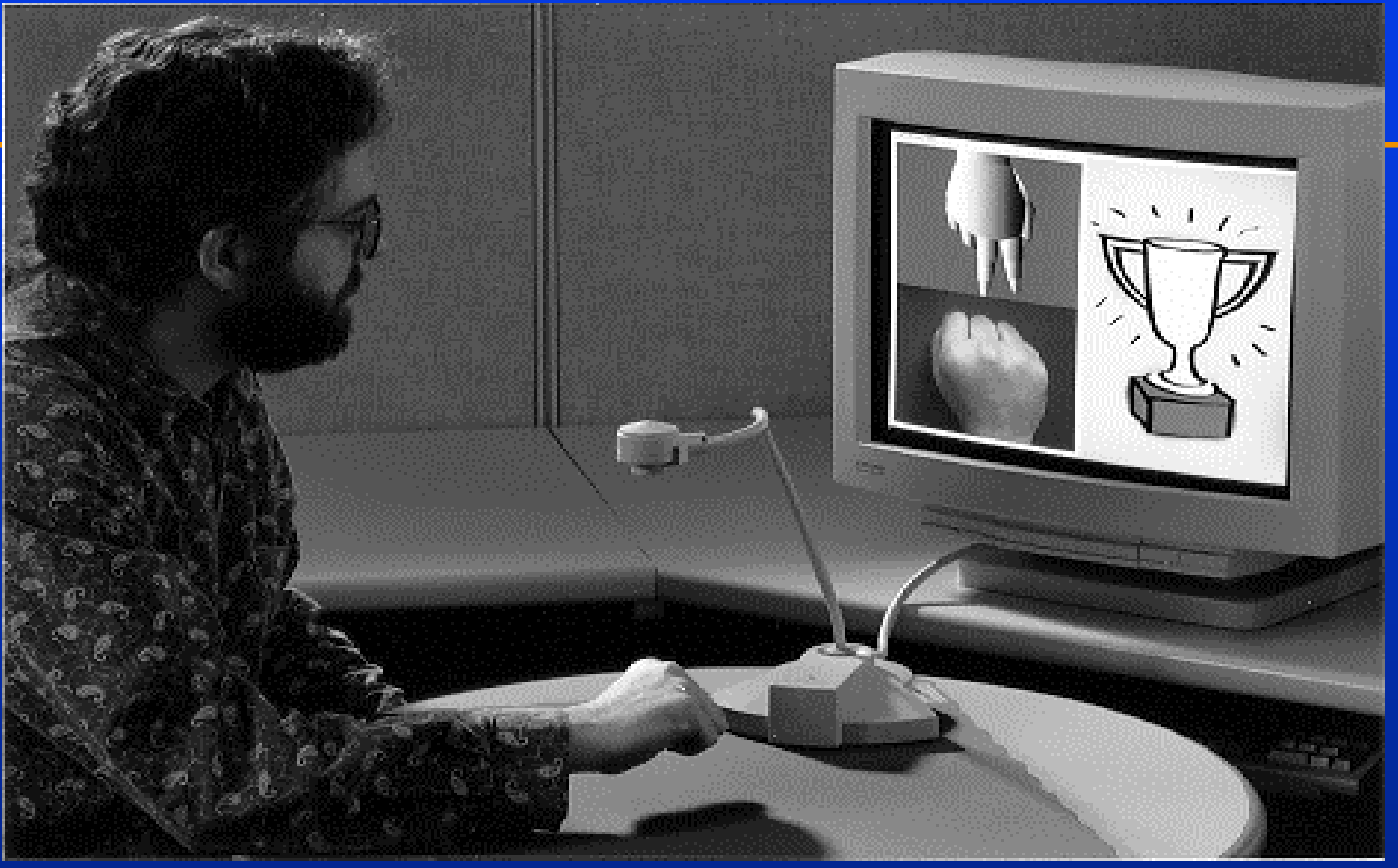


Test image, and distances from each of the training set orientation histograms (categorized correctly).





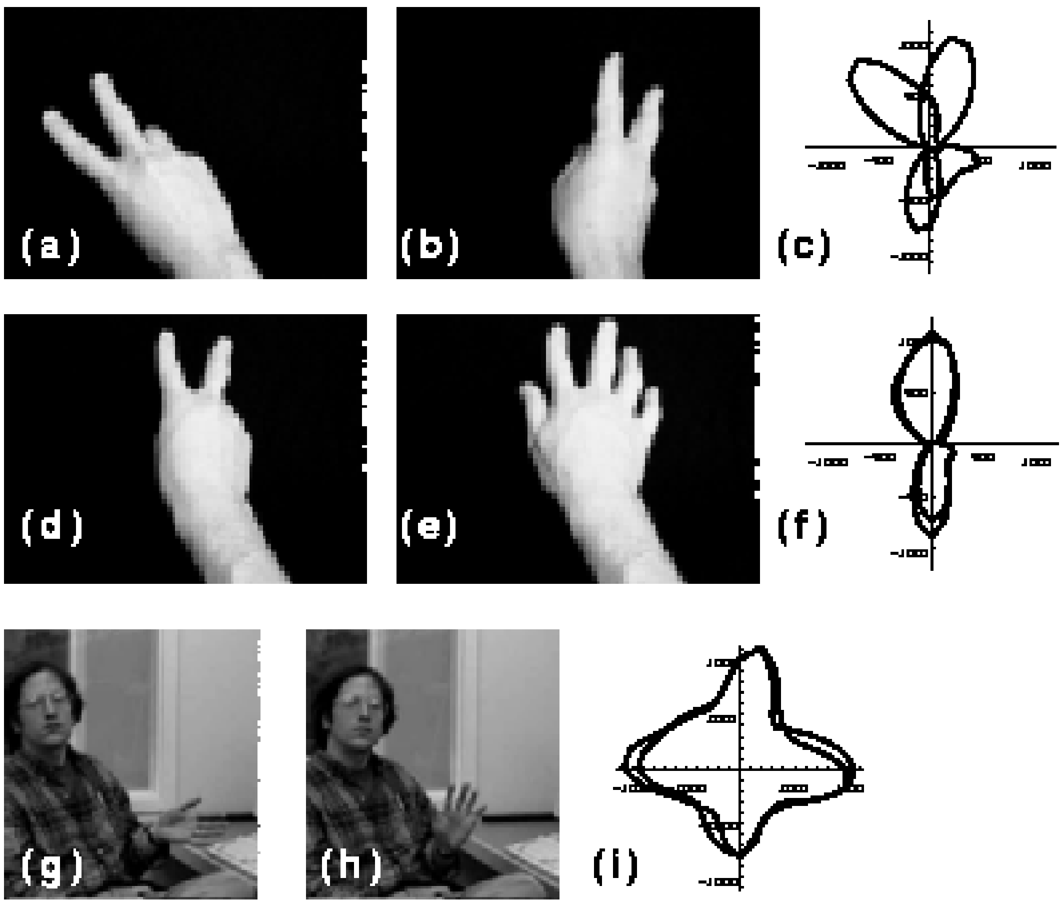
Crane movements controlled by hand gestures



Janken game

video

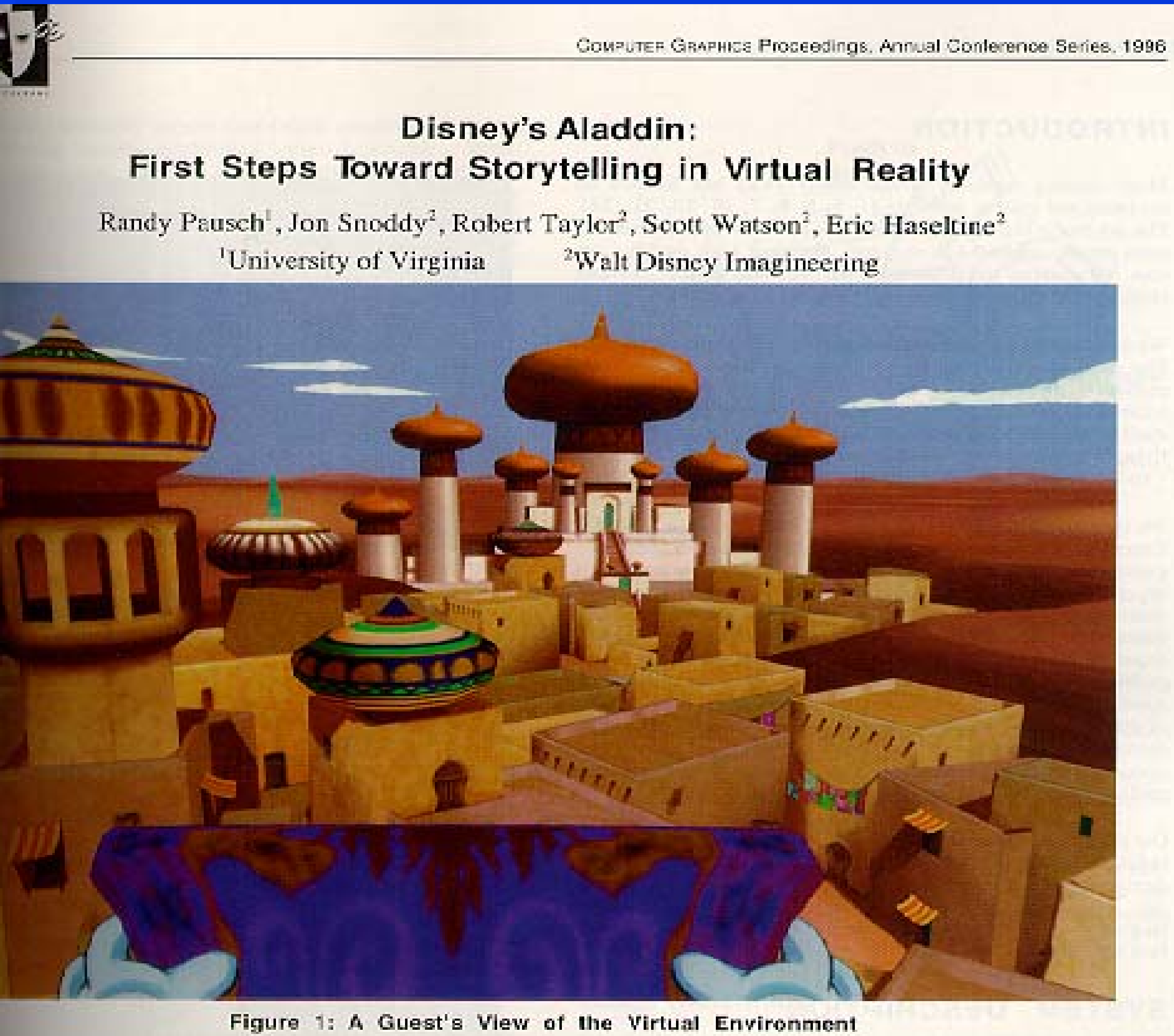
7 Problem images for the orientation histogram-based gesture classifier.



3. Computer vision for computer games.

Games add fun and purpose: "Get the sprite through the golden rings."

Field test results from Disney's VR Aladdin.



"Guests cared about the experience, not the technology."

Games selected for vision interface

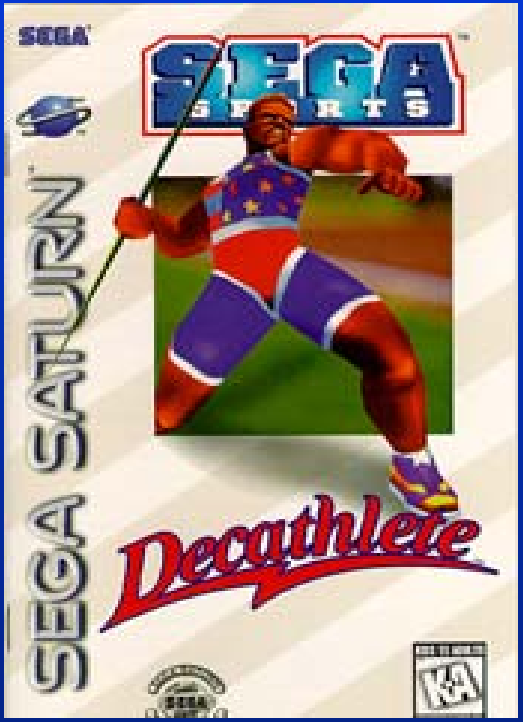


Image moments give a very coarse image summary.

$$M_{00} = \sum_x \sum_y I(x,y)$$

$$M_{10} = \sum_x \sum_y x\, I(x,y)$$

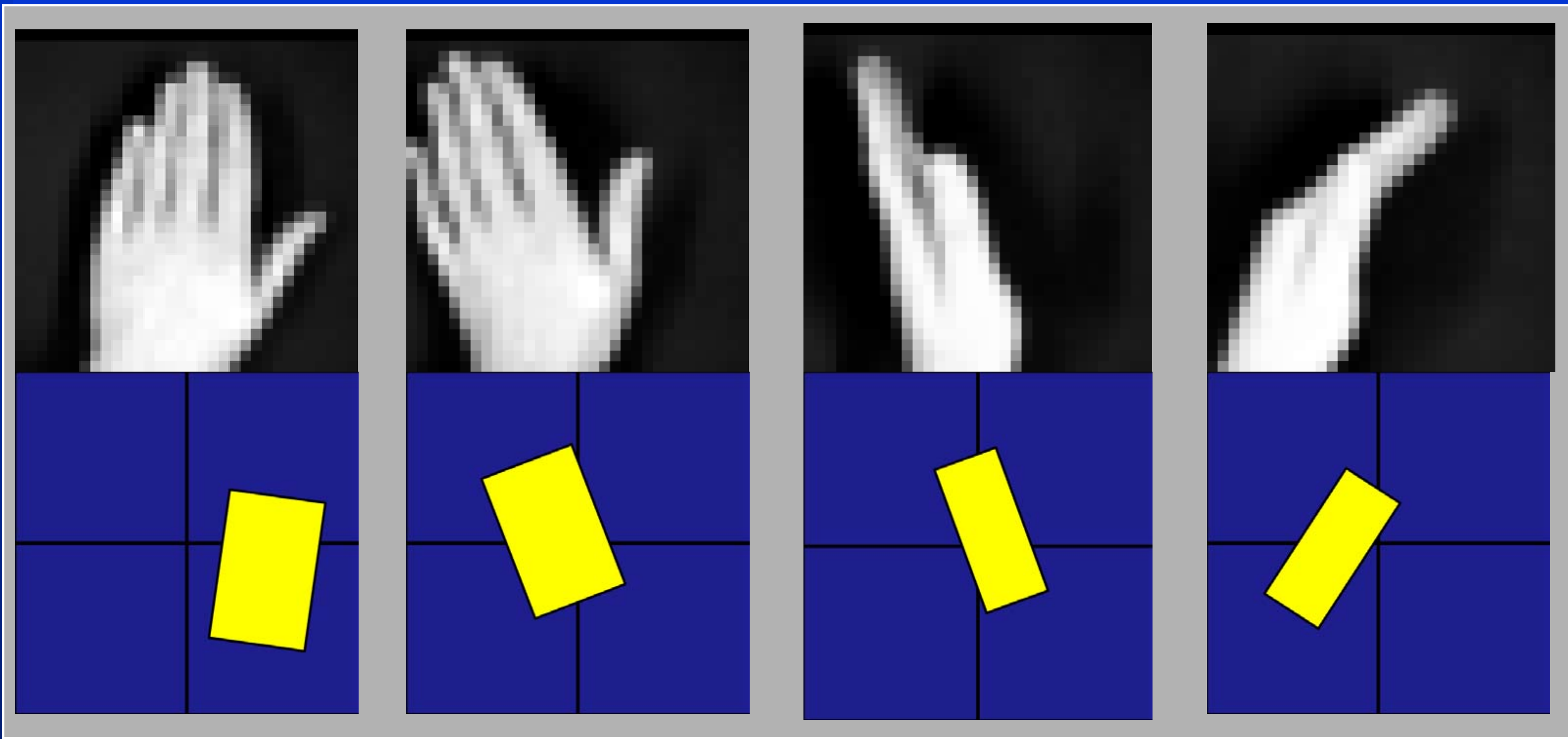
$$M_{01} = \sum_x \sum_y y\, I(x,y)$$

$$M_{20} = \sum_x \sum_y x^2\, I(x,y)$$

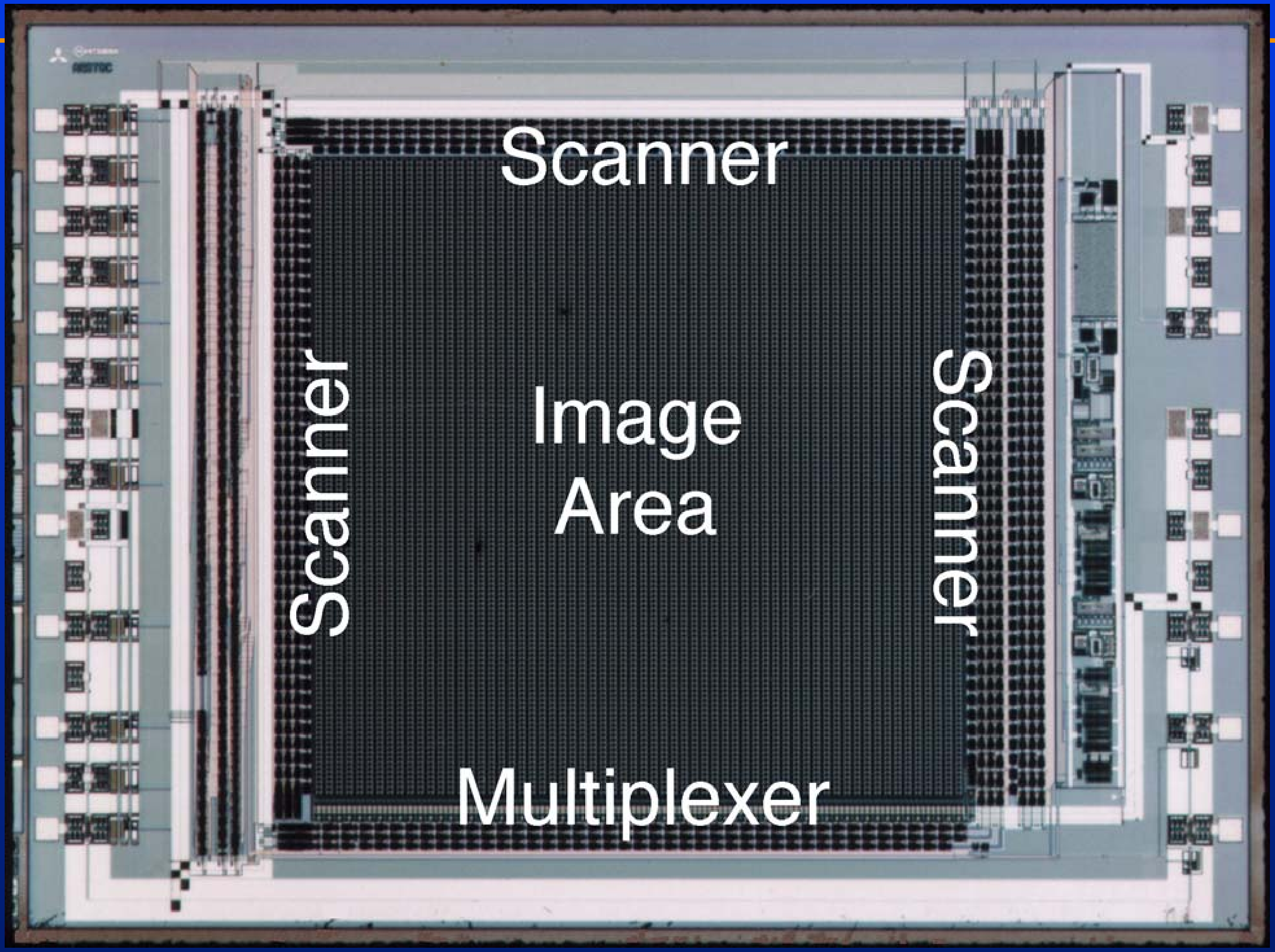
$$M_{11} = \sum_x \sum_y xy\, I(x,y)$$

$$M_{02} = \sum_x \sum_y y^2\, I(x,y)$$

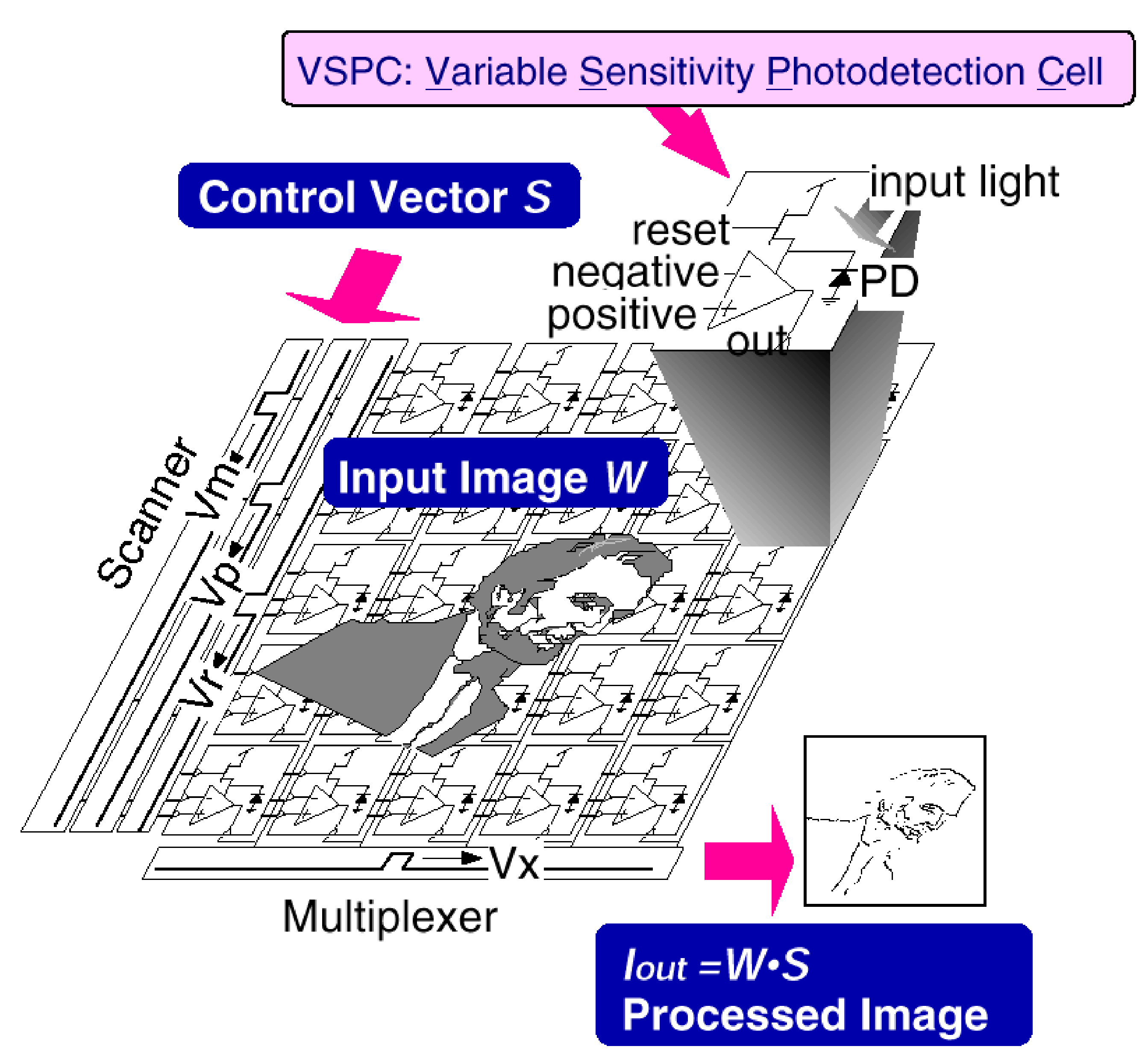
Hand images and equivalent rectangles having the same image moments



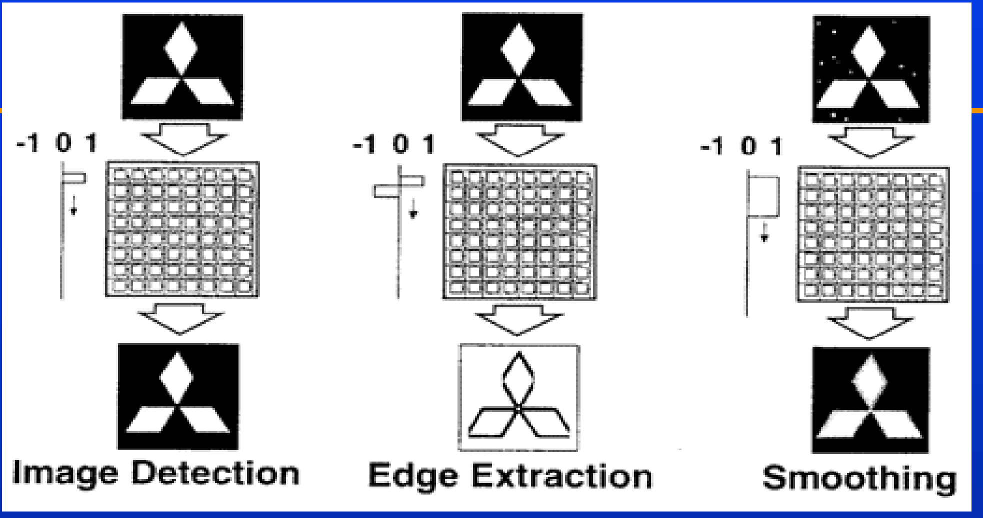
Artificial Retina chip for detection and low-level image processing.

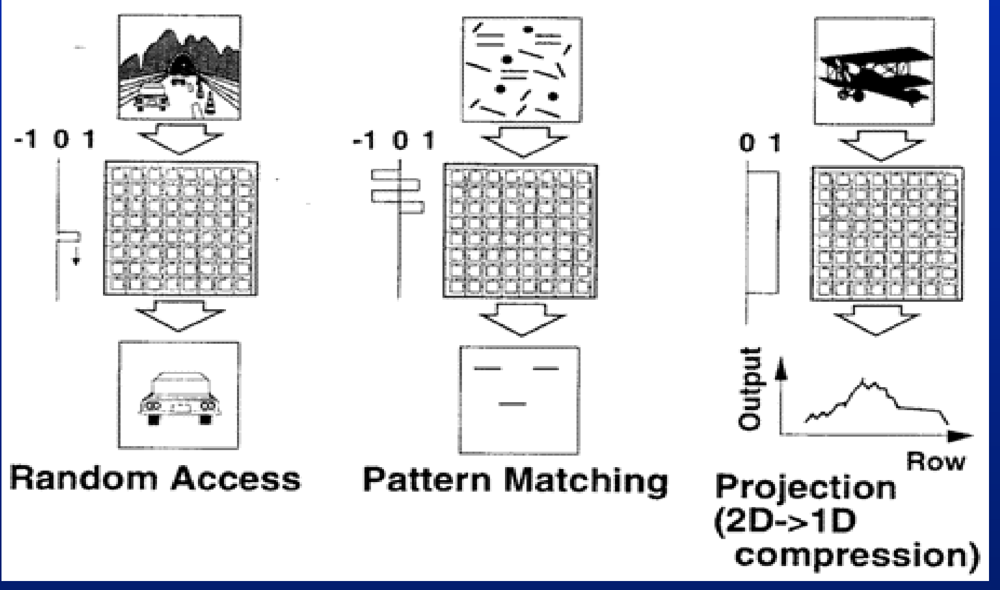


Artificial Retina chip



Artificial Retina functions



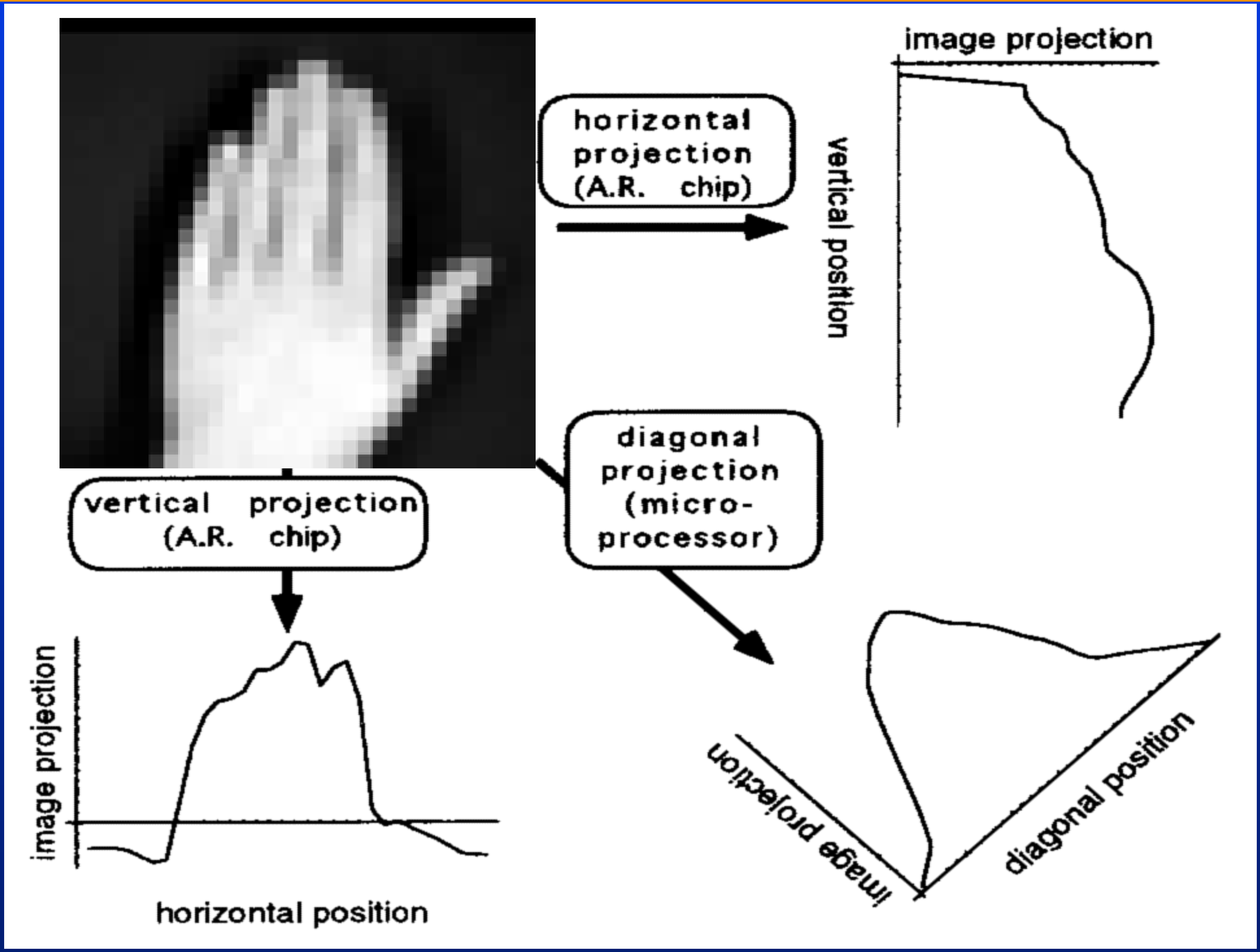


Fast image moment calculation with artificial retina chip

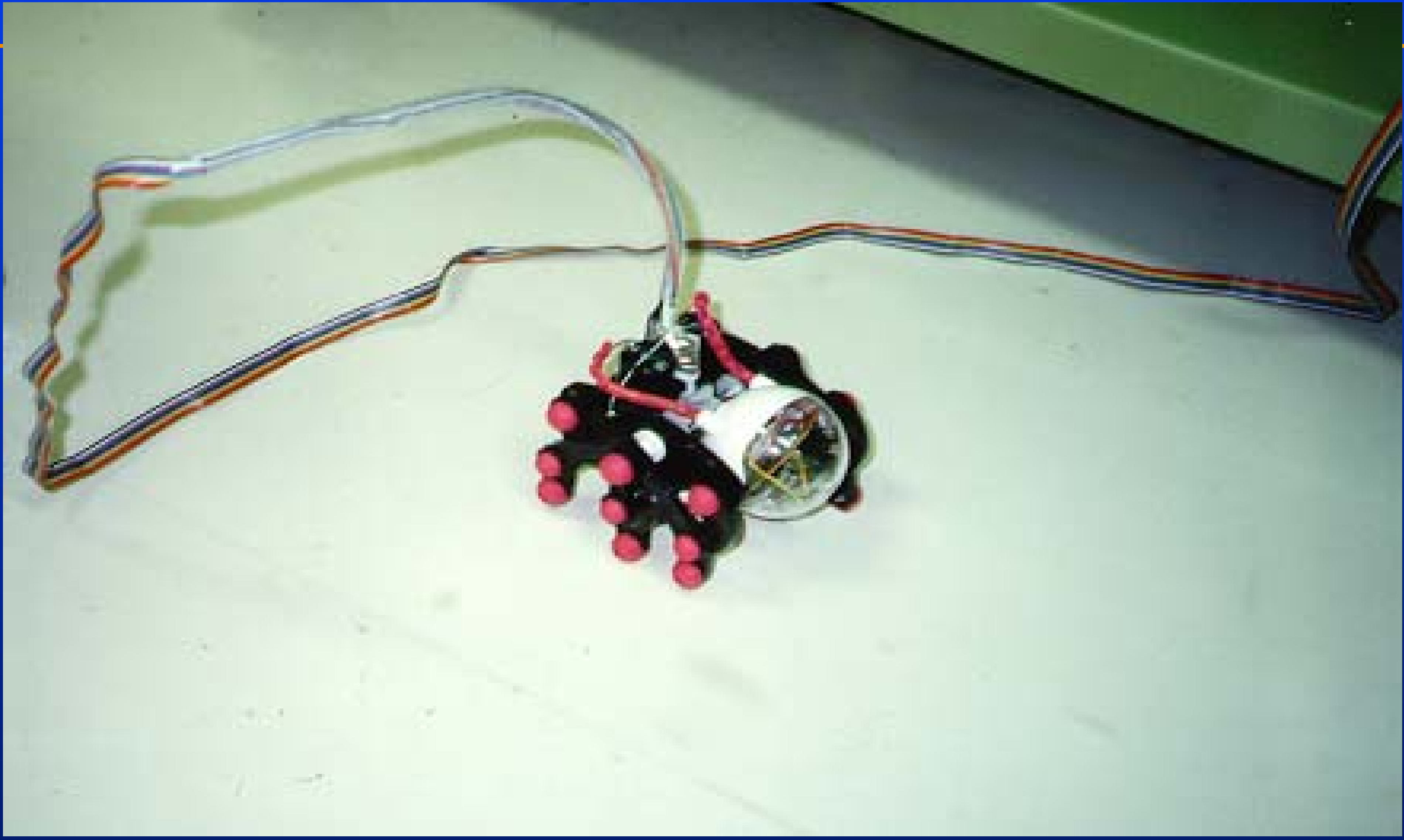
Processing time for image projections:

w/o AR chip: 10 msec

with AR chip: 0.3 msec



Hand gesture-controlled robot

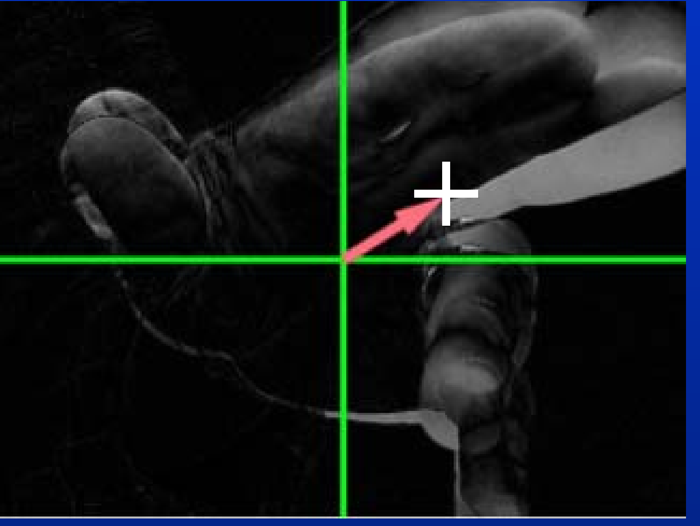


Game: Nights

Moment-based pointing control

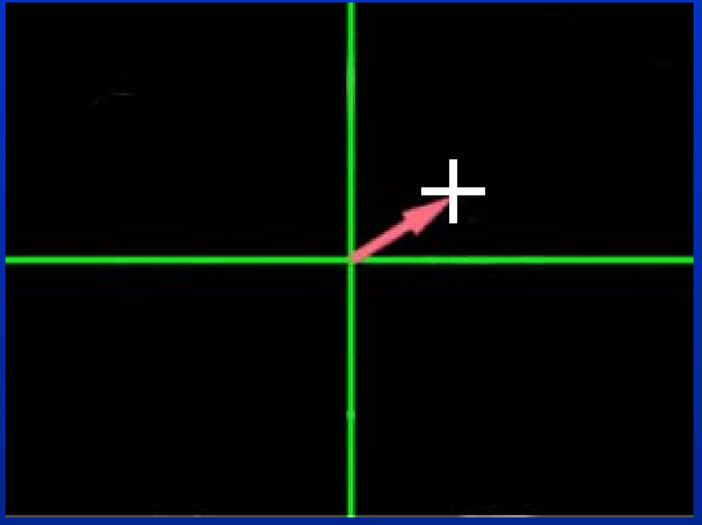
time 1

time 2



Center-of-mass of absolute value of difference-image

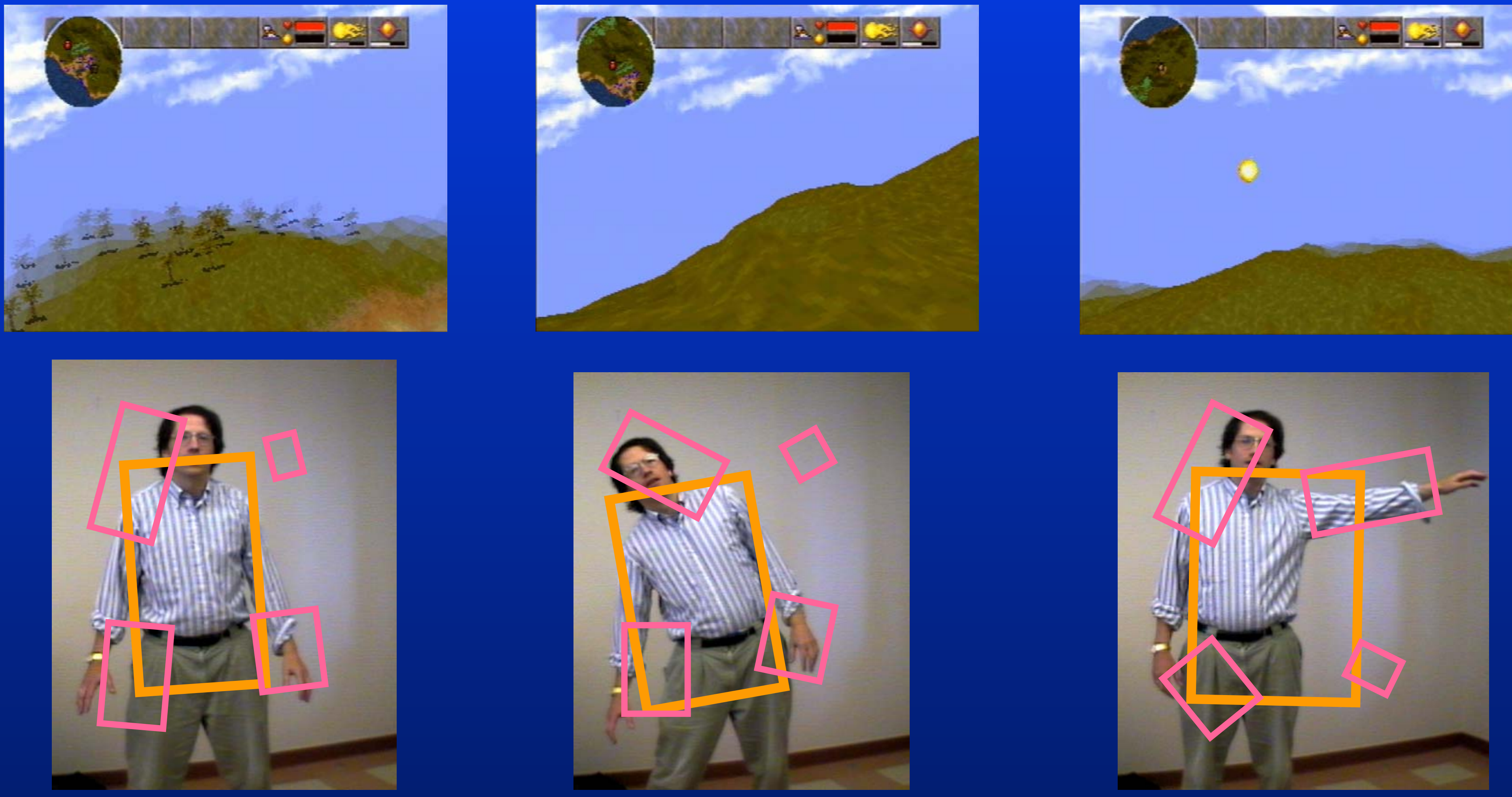
Moment-based pointing control



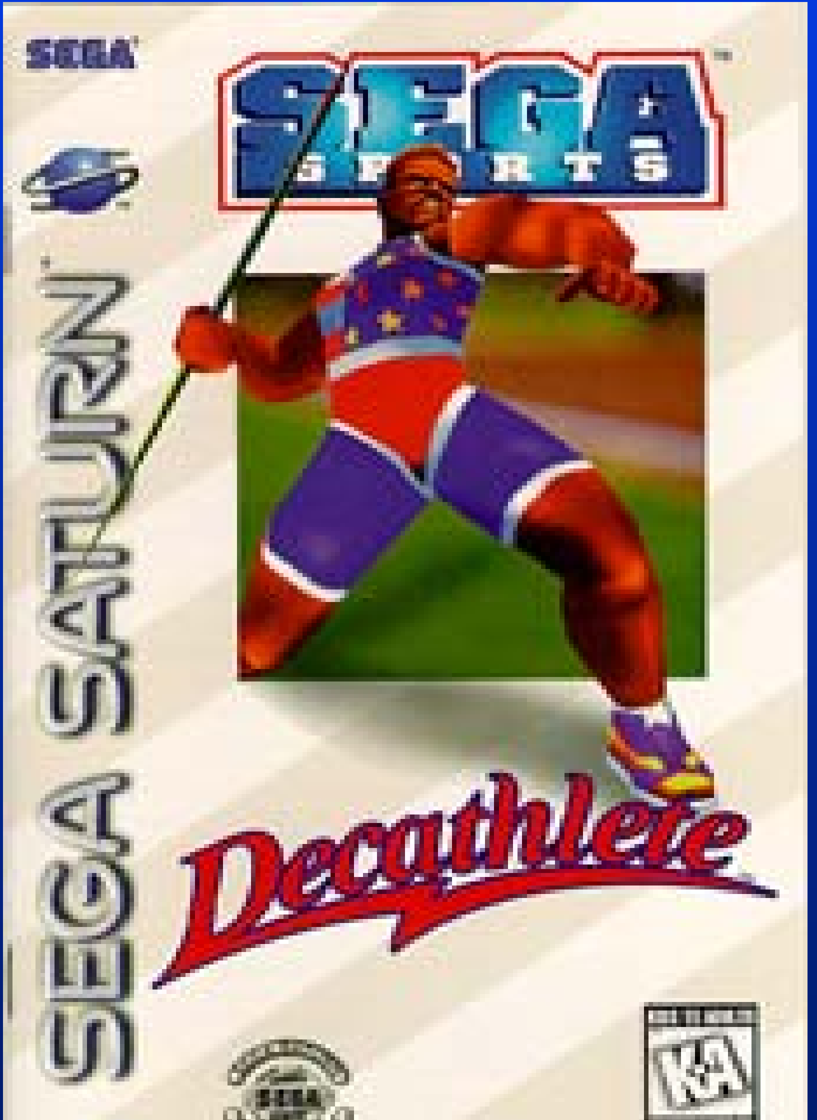
Line to difference-image center-of-mass
determines flight direction.

Game: Magic Carpet

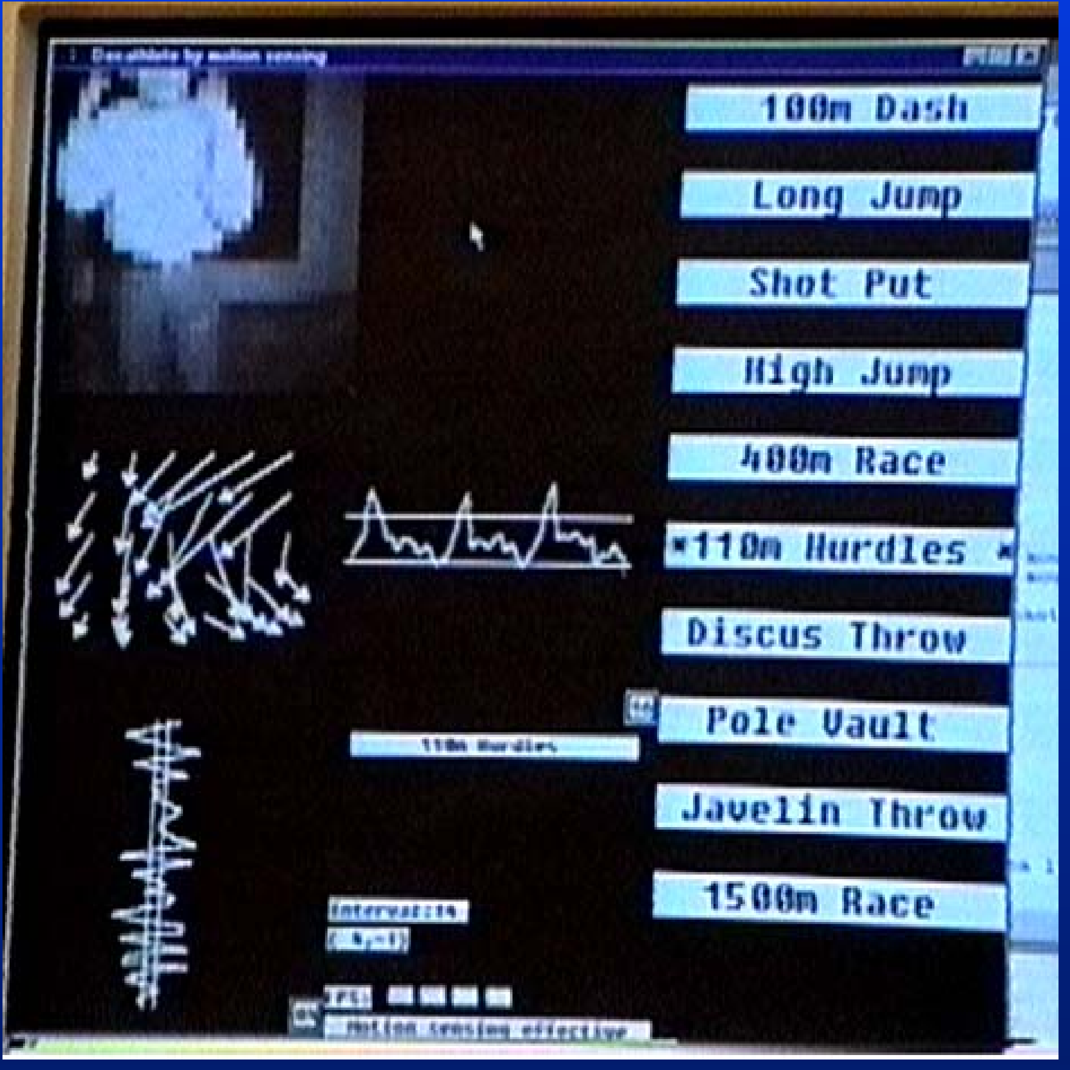
Magic carpet game--figure analysis by hierarchical image moments



Game: Decathlete

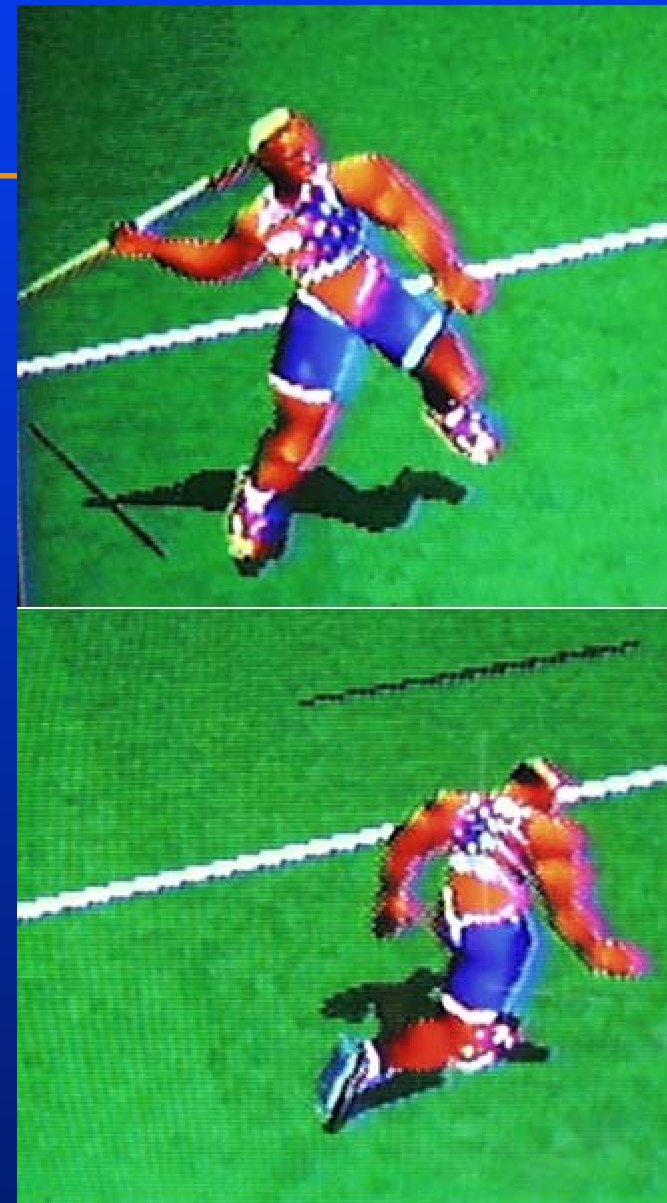


Optical-flow-based Decathlete figure motion analysis



Decathlete 100m hurdles

Decathlete javelin throw



Decathlete javelin throw

video

Nintendo Game Boy Camera

Several million sold (most of any digital camera). Imaging chip is Mitsubishi Electric's "Artificial Retina" CMOS detector.

video

Summary

- ***Fast, simple algorithms and low-cost hardware are well-suited to interactive graphics applications.***
- ***We followed this approach to make a television controlled by hand gestures, simple hand gesture recognition, and vision-based computer game interfaces.***

To Trevor's slides…

Perceptive Context for Pervasive Computing

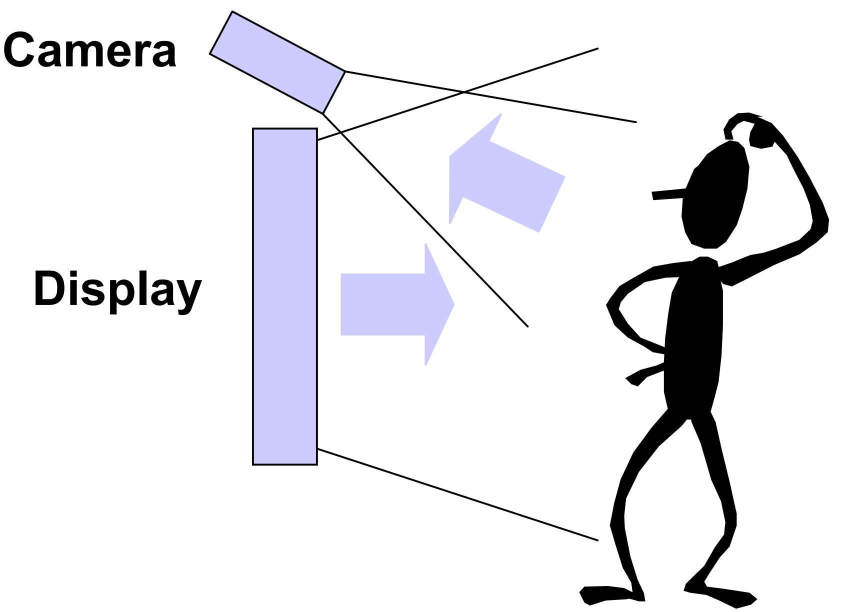
Trevor Darrell
Vision Interface Group
MIT AI Lab

Perceptually Aware Displays

Camera associated with display

Display should respond to user

- font size
- attentional load
- passive acknowledgement



e.g., "Magic Mirror", Interval

Compaq's Smart Kiosk

ALIVE, MIT Media Lab

Example: A Face Responsive Display

- Faces are natural interfaces!
  - Ubiquitous, fast, expressive, general.
  - Want machines to generate and perceive faces.
- A Face Responsive Display...
  - Knows when it's being observed
  - Recognizes returning observers
  - Tracks head pose
  - Robust to changing lighting, moving backgrounds…

A Face Responsive Display

Tasks

- Detection
- Identification
- Tracking

How? Exploit multiple visual modalities:

- Shape
- Color
- Pattern

Tasks and Visual Modalities

	shape	***color***	***pattern***
detection	silhouette classifier	skin classifier	face detection
identification	biometrics	flesh hue	face recognition
tracking	coarse motion estimation	clothing histogram	fine motion estimation / pose tracking

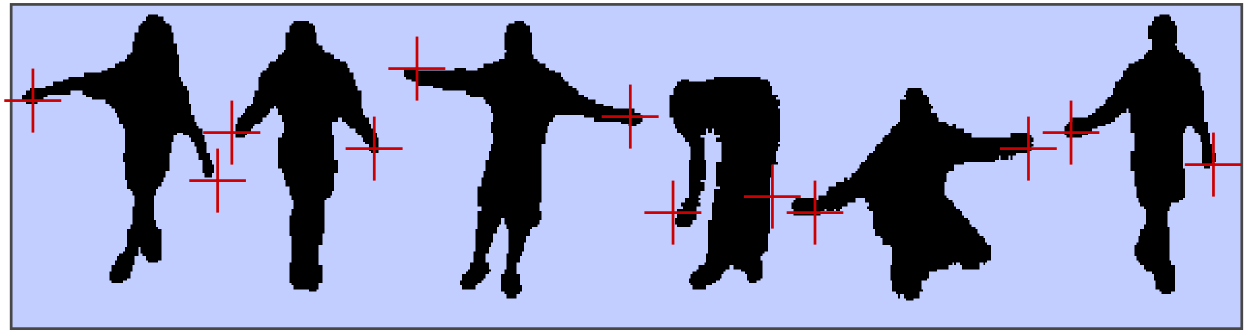
Mode and Task Matrix

	shape	***color***	***pattern***
detection	silhouette classifier	skin classifier	face detection
identification	biometrics	flesh hue	face recognition
tracking	Shape change	clothing histogram	Appearance change

Finding Features

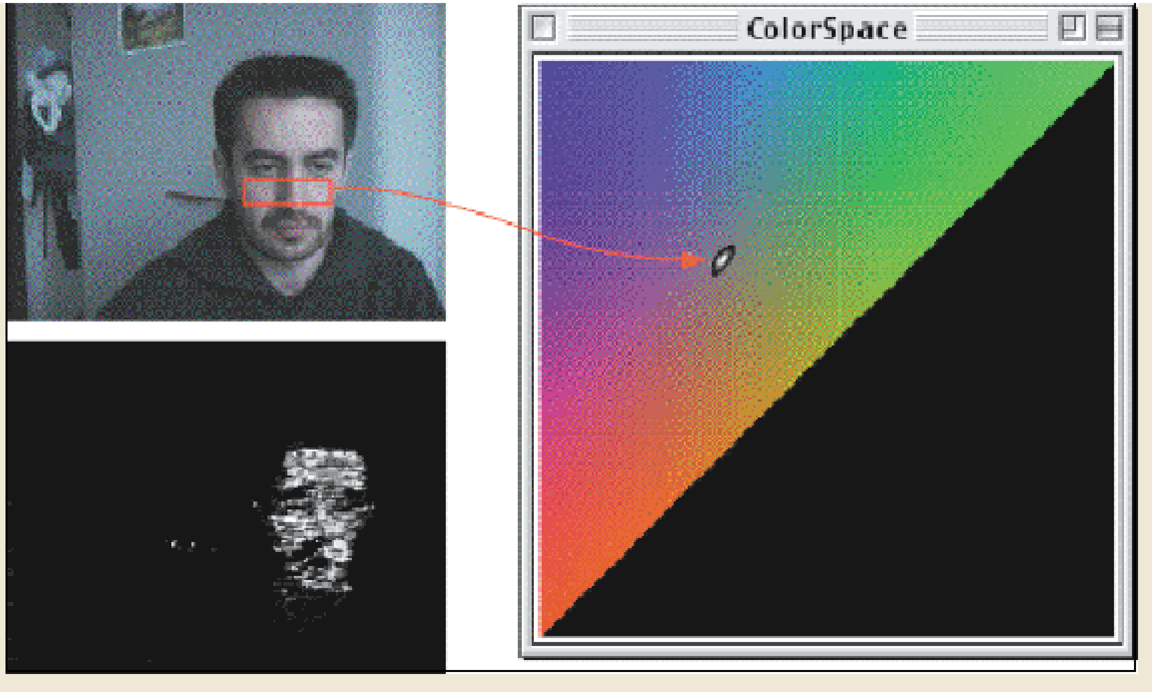
2D Head / hands localization

- contour analysis: mark extremal points (highest curvature or distance from center of body) as hand features
- use skin color model when region of hand or face is found (color model is independent of flesh tone intensity)



Flesh color tracking

- Often the simplest, fastest face detector!
- Initialize region of hue space



[Crowley, Coutaz, Berard, INRIA]

Color Processing

- Train two-class classifier with examples of skin and not skin
- Typical approaches: Gaussian, Neural Net, Nearest Neighbor
- Use features invariant to intensity

  Log color-opponent [Fleck et al.]

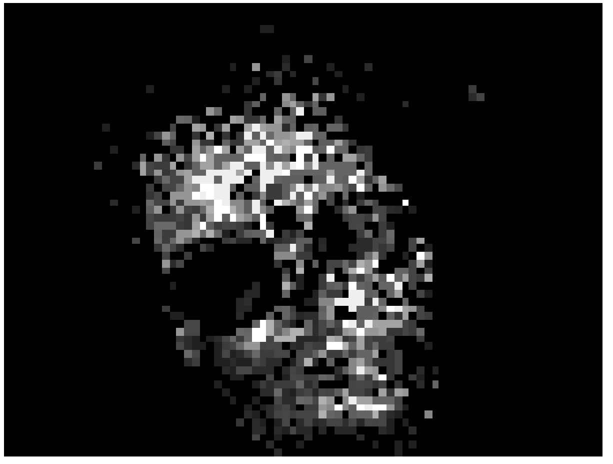
  (log(r) - log(g), log(b) - log((r+g)/2))

  Hue & Saturation

Flesh color tracking

Can use Intel OpenCV lib's CAMSHIFT algorithm for robust real-time tracking.

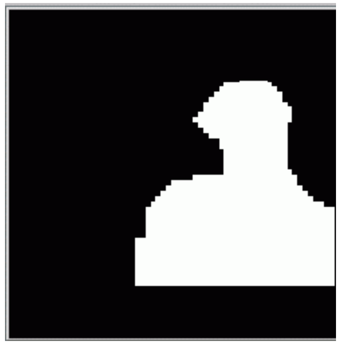
(open source impl. avail.!)



[Bradsky, Intel]

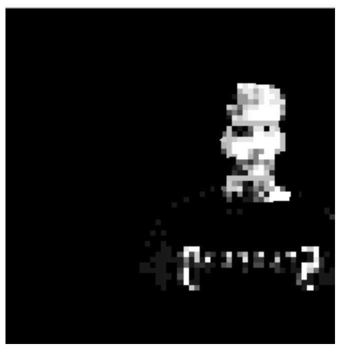
Detection with multiple visual modes

Shape



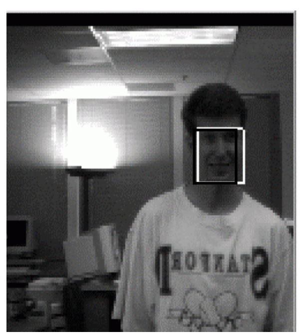
Find head sized peaks in 2-D or 3-D.

Flesh Color Detection



Detect skin pigment in hue-based color space

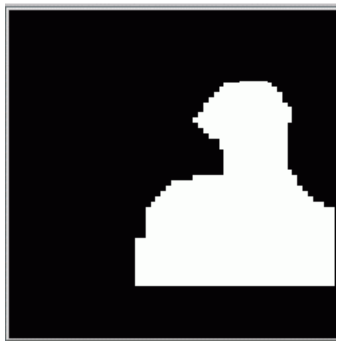
Face Pattern Detection



Classify intensity vector corresponding to face class

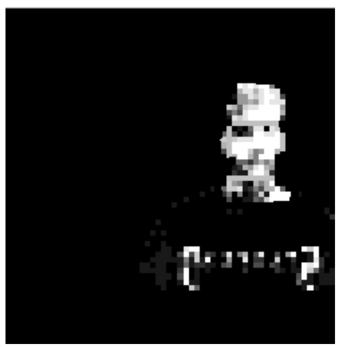
Common Detection Failure Modes

Shape



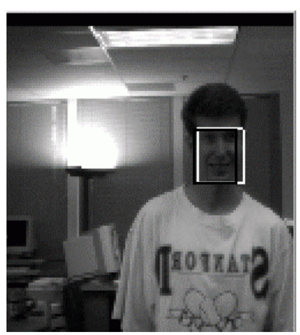
Fooled by head shaped peaks

Flesh Color Detection



Fooled by flesh colored objects

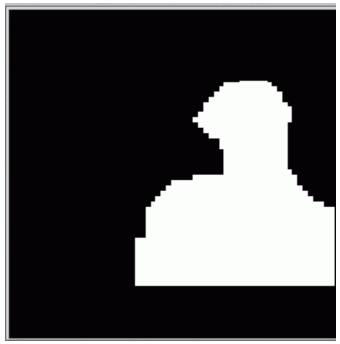
Face Pattern Detection



Misses out of plane rotation or expression

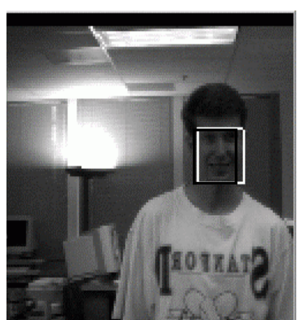
Robust real-time performance

Shape



Flesh Color Detection

Face Pattern Detection



Integrated Face Detection Algorithm
(temporally asynch. voting scheme)

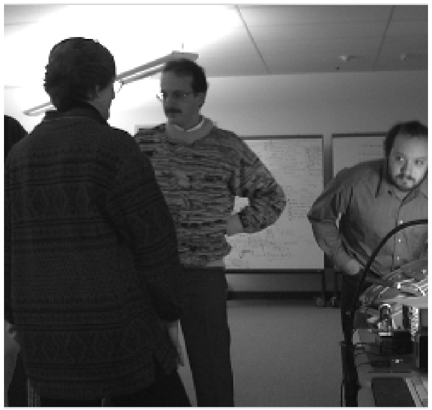
Mode and Task Matrix

	shape	***color***	***pattern***
detection	silhouette classifier	skin classifier	face detection
identification	biometrics	flesh hue	face recognition
tracking	Shape change	clothing histogram	Appearance change

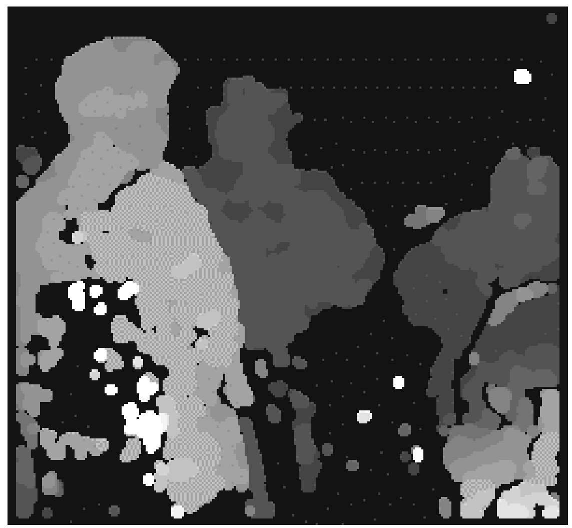
A Key Technology: Video-Rate Stereo

- Two cameras –> stereo range estimation; disparity proportional to depth
- Depth makes tracking people easy
  - segmentation
  - shape characterization
  - pose tracking
- Real-time implementations becoming commercially available.

Video-rate stereo



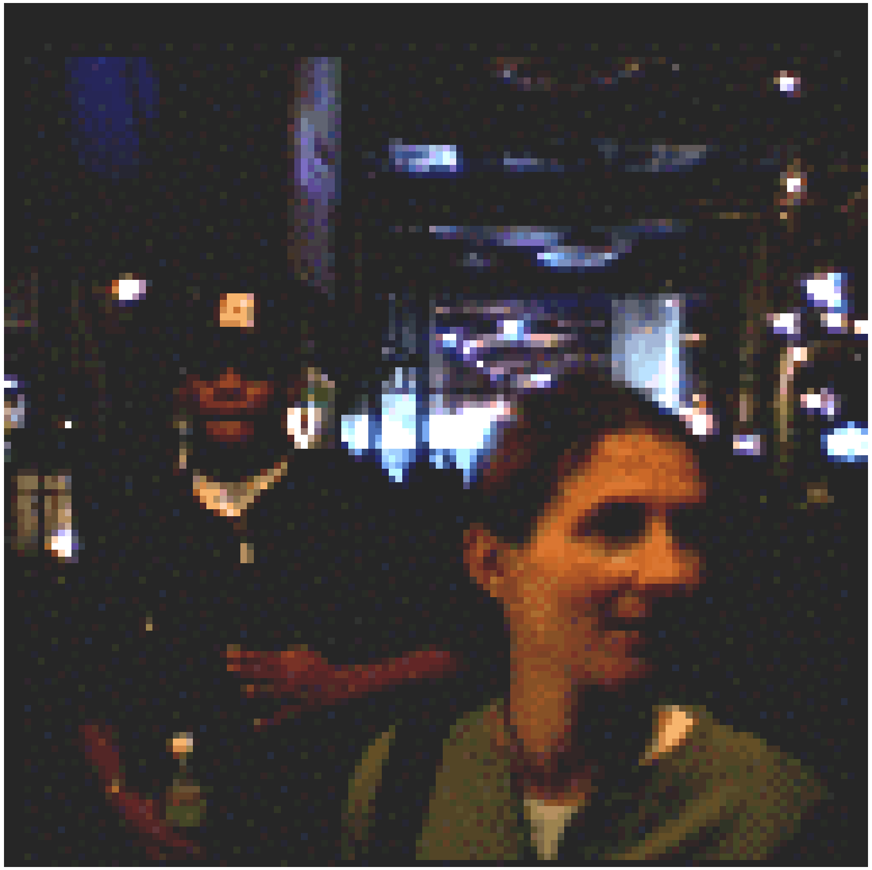
Left and right images

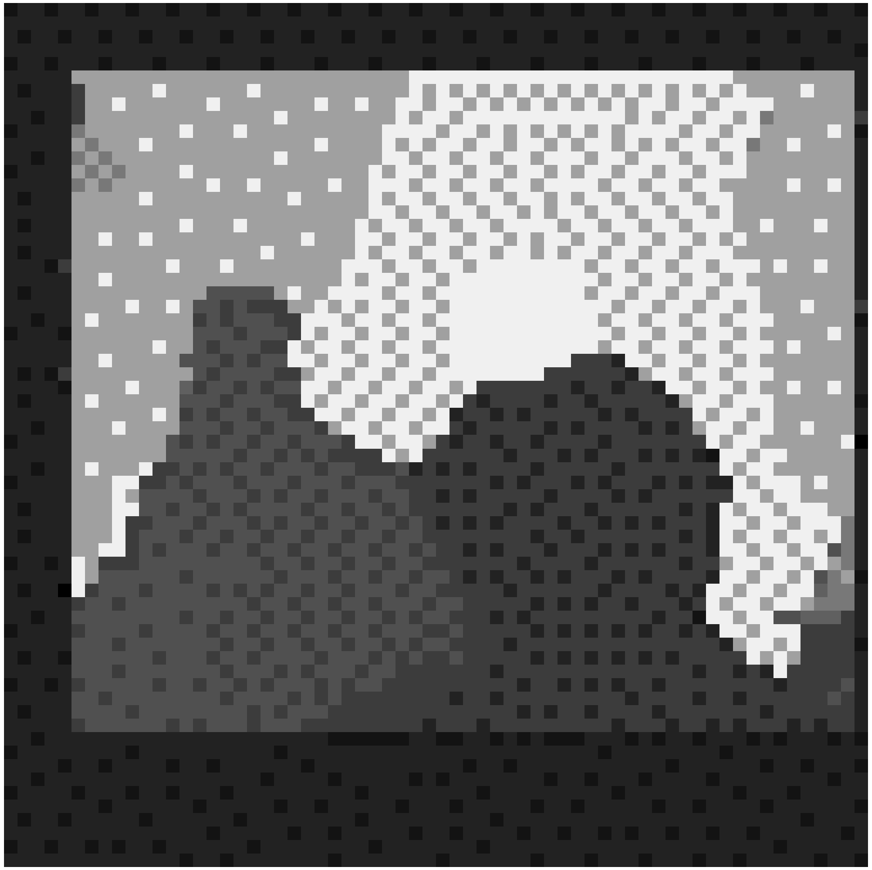


Computed disparity

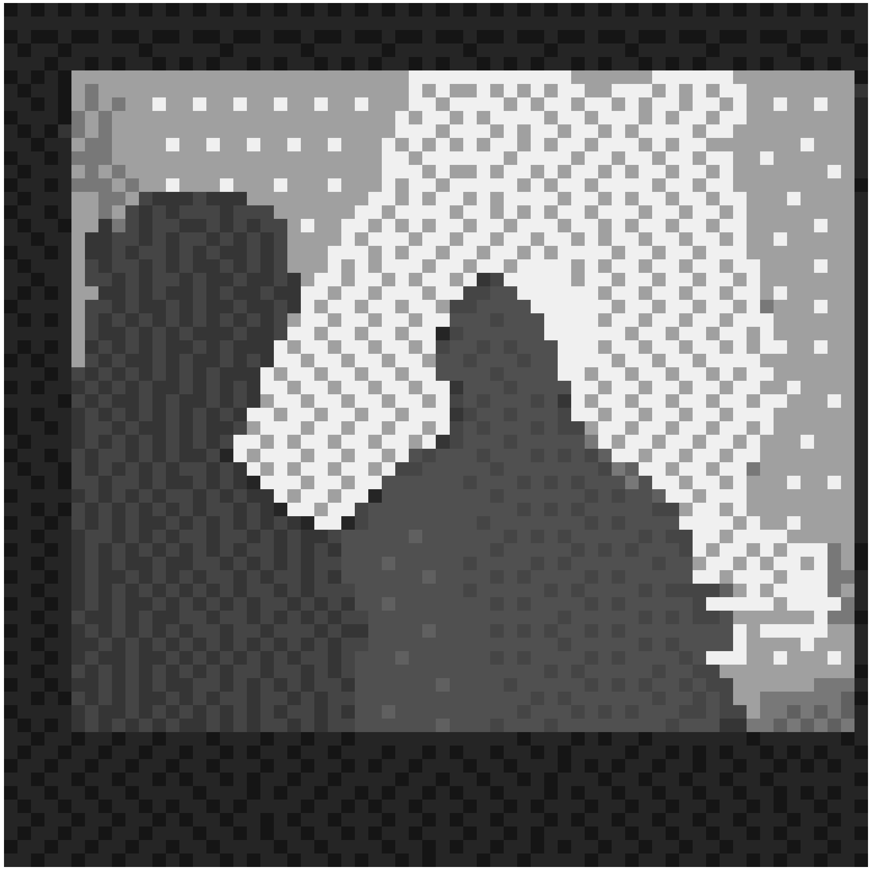
Foreground pixels; grouped by local connectivity

RGBZ input

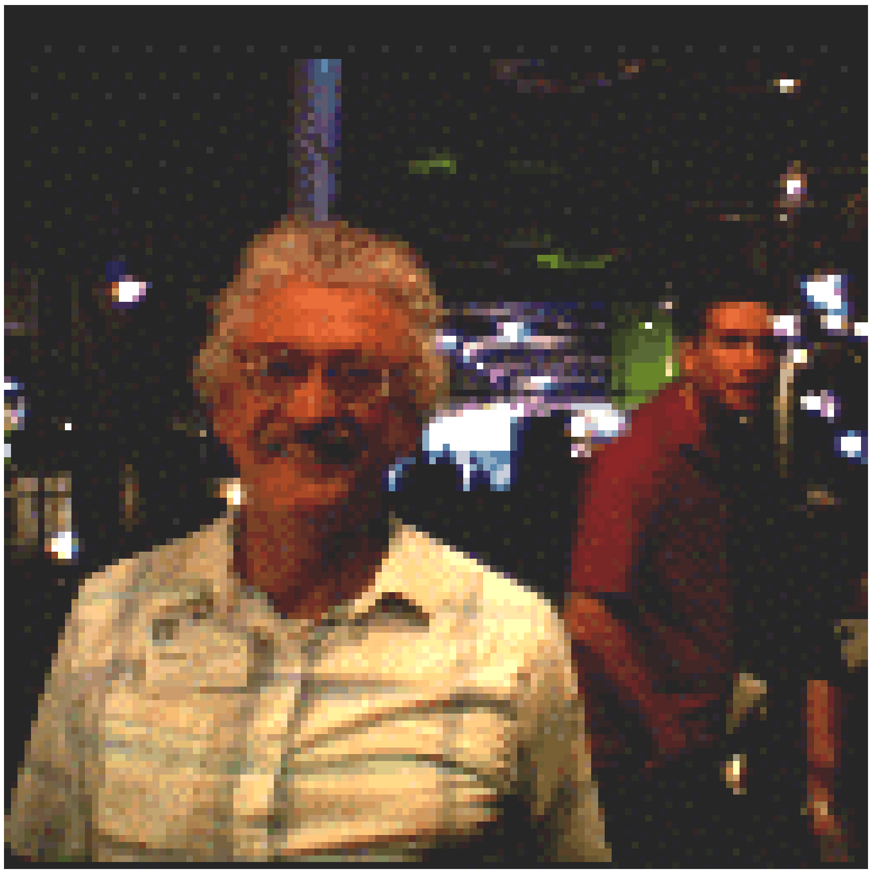




RGBZ input



RGBZ input

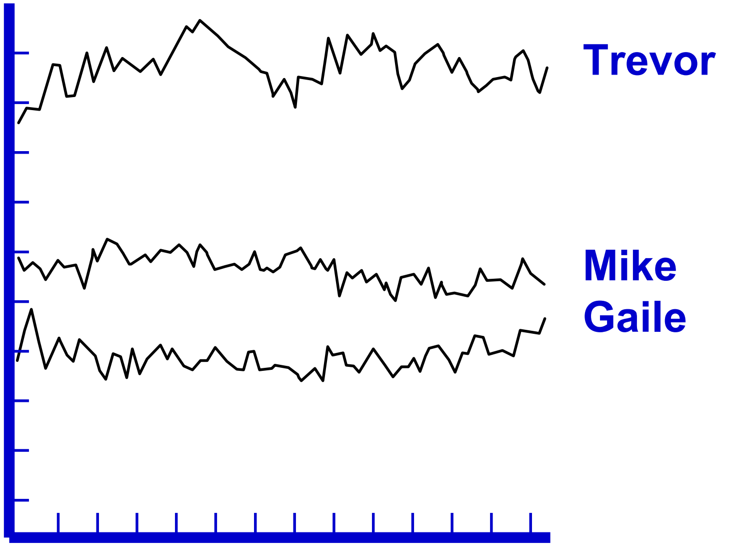


Video-Rate Stereo

- Multiple cameras –> stereo range estimation; disparity proportional to depth
- Real-time implementations becoming commercially available.
- Depth makes tracking people easier
  - segmentation
  - shape characterization
  - pose tracking

Range feature for ID!

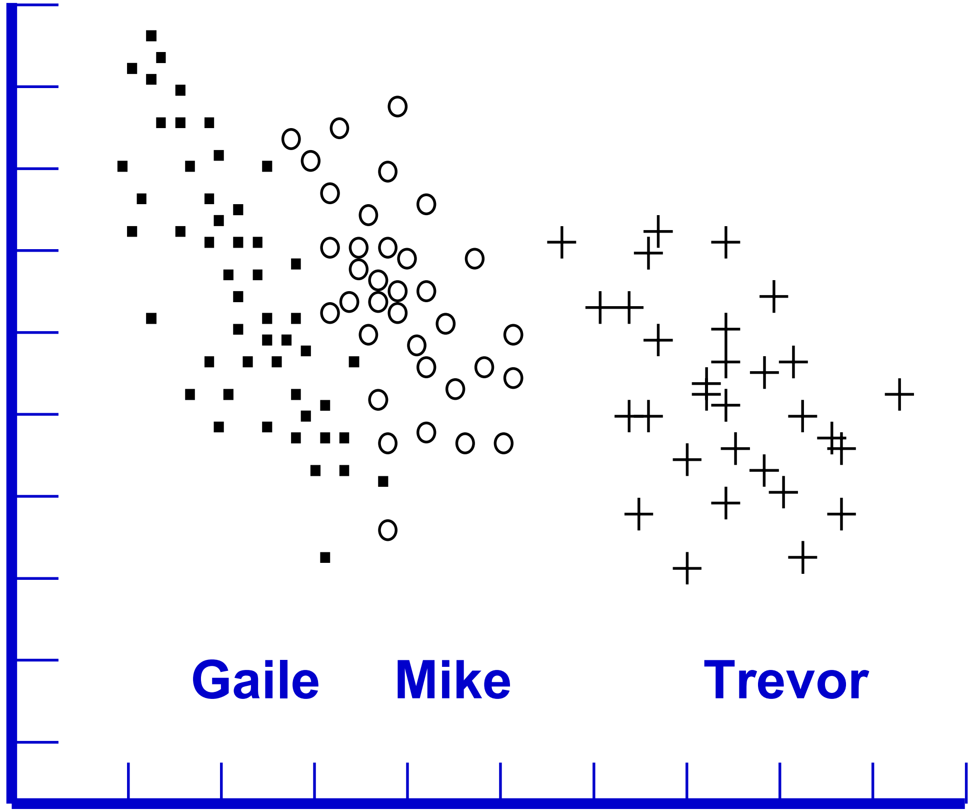
- Body shape characteristics -- e.g., height measure.
- Normalize for motion/pose: median filter over time



- Near future: full vision-based kinematic estimation and tracking-- active research topic in many labs.

Color feature for ID!

For long-term tracking / identification, measure color hue and saturation values of hair and skin….



For same-day ID, use histogram of entire body / clothing

Mode and Task Matrix

	shape	*color*	*pattern*
detection	silhouette classifier	skin classifier	face detection
identification	biometrics	flesh hue	face recognition
tracking	Shape change	clothing histogram	Appearance change

See lectures by Trevor later in the course

Robust, Multi-modal Algorithm

Combine modules for detection:

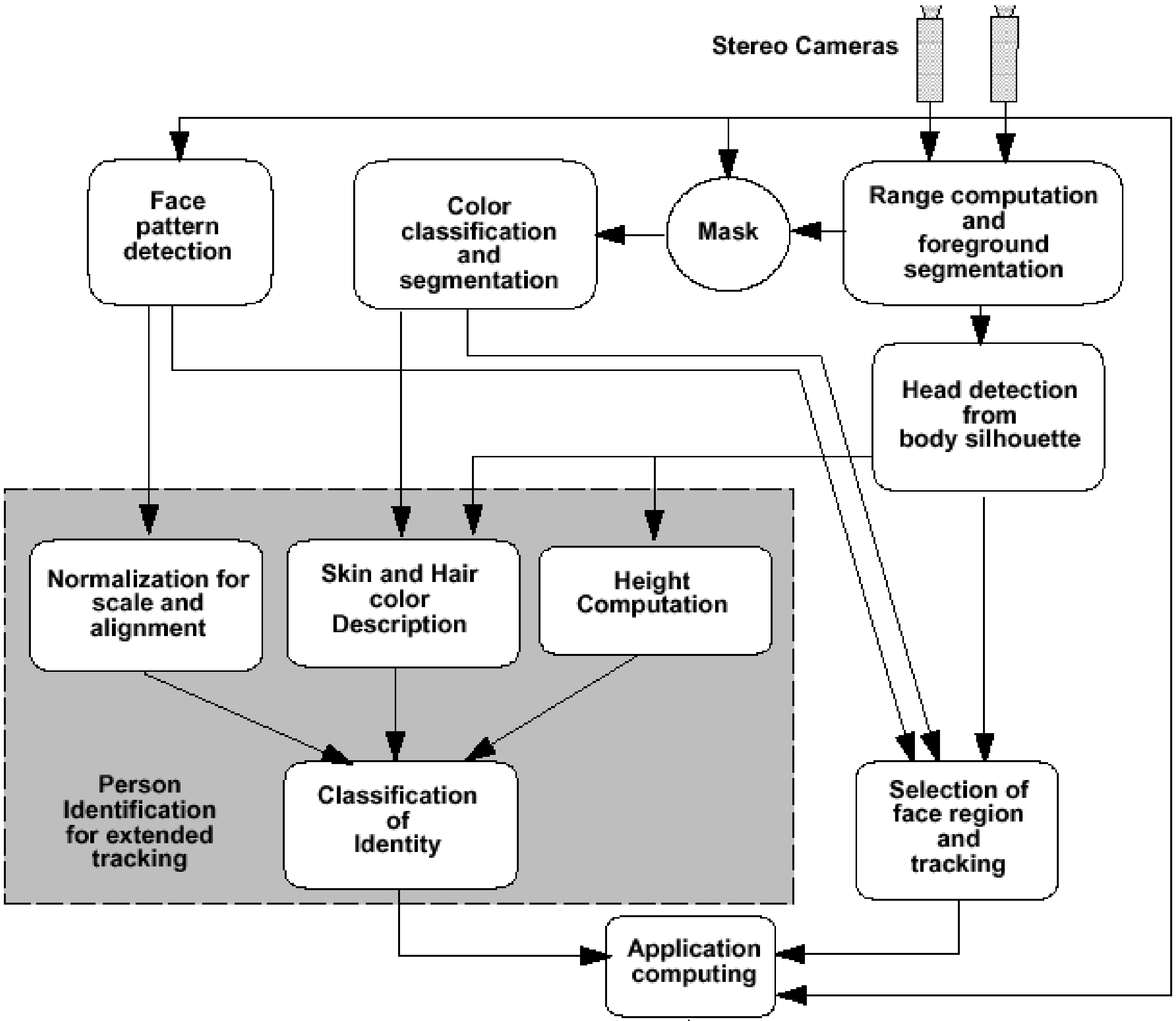
- Silhouette finds body
- Color tracks extremities
- Pattern discriminates head from hands.

Use each also to recognize returning people:

- Face recognition
- Biometrics (skeletal structure)
- Hair and Skin hue
- Clothing (intra-day.)

[CVPR ‘98; T. Darrell, G. Gordon, M. Harville, J. Woodfill]

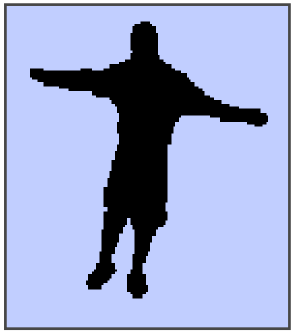
System Overview



Classic Background Subtraction model

- Background is assumed to be mostly static
- Each pixel is modeled as by a gaussian distribution in YUV space
- Model mean is usually updated using a recursive low-pass filter

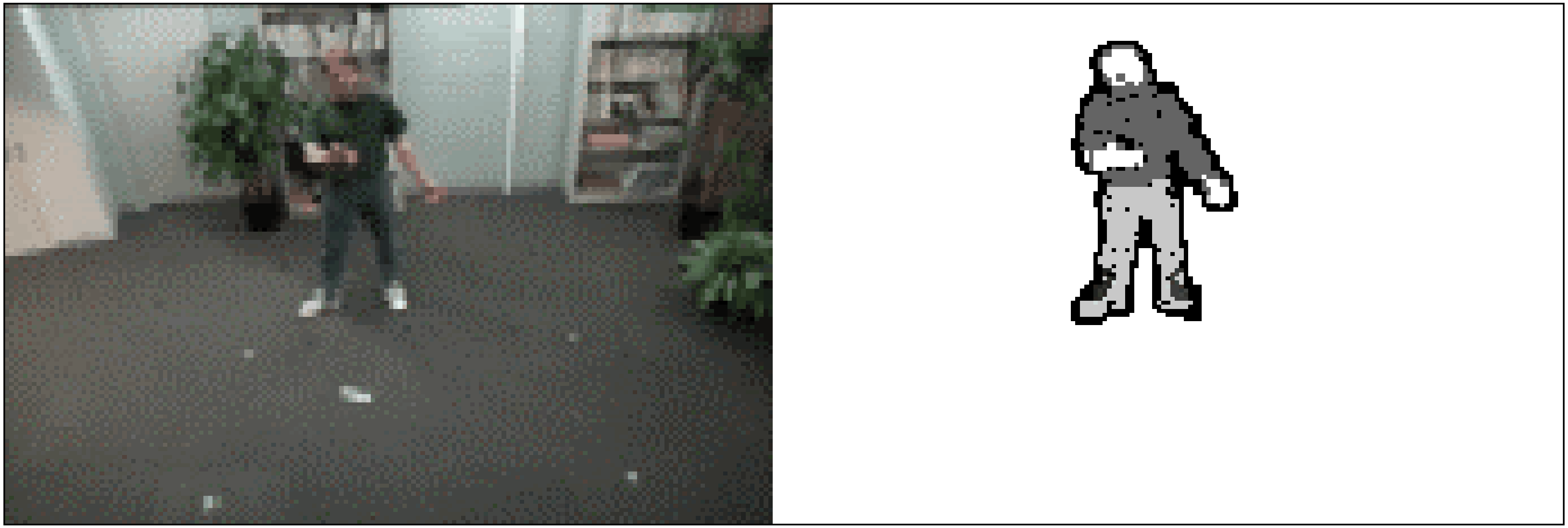
Given new image, generate silhouette
by marking those pixels that are significantly
different from the “background” value.



Static Background Modeling Examples

[MIT Media Lab Pfinder / ALIVE System]

Static Background Modeling Examples



[MIT Media Lab Pfinder / ALIVE System]

Static Background Modeling Examples

[MIT Media Lab Pfinder / ALIVE System]

The ALIVE System

Camera

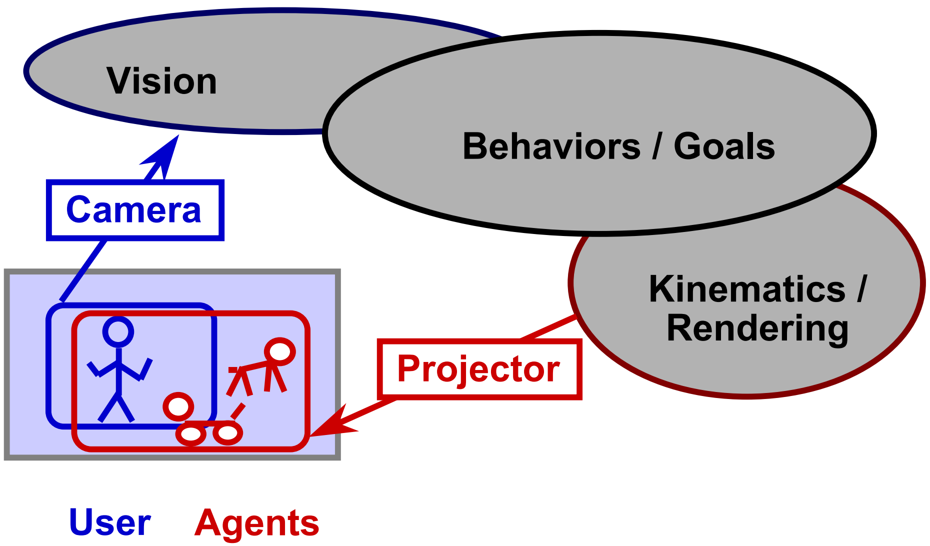
User

Video
Screen

Autonomous Agents

ALIVE

- Real sensing for virtual world
- Tightly coupled sensing-behavior-action
- Vision routines: body/head/hand tracking



[Blumberg, Darrell, Maes, Pentland, Wren, … 1995]

ALIVE system, MIT

M.I.T. Media Laboratory Perceptual Computing Technical Report No. 257
(To appear, ACM Multimedia Systems)

The ALIVE System:

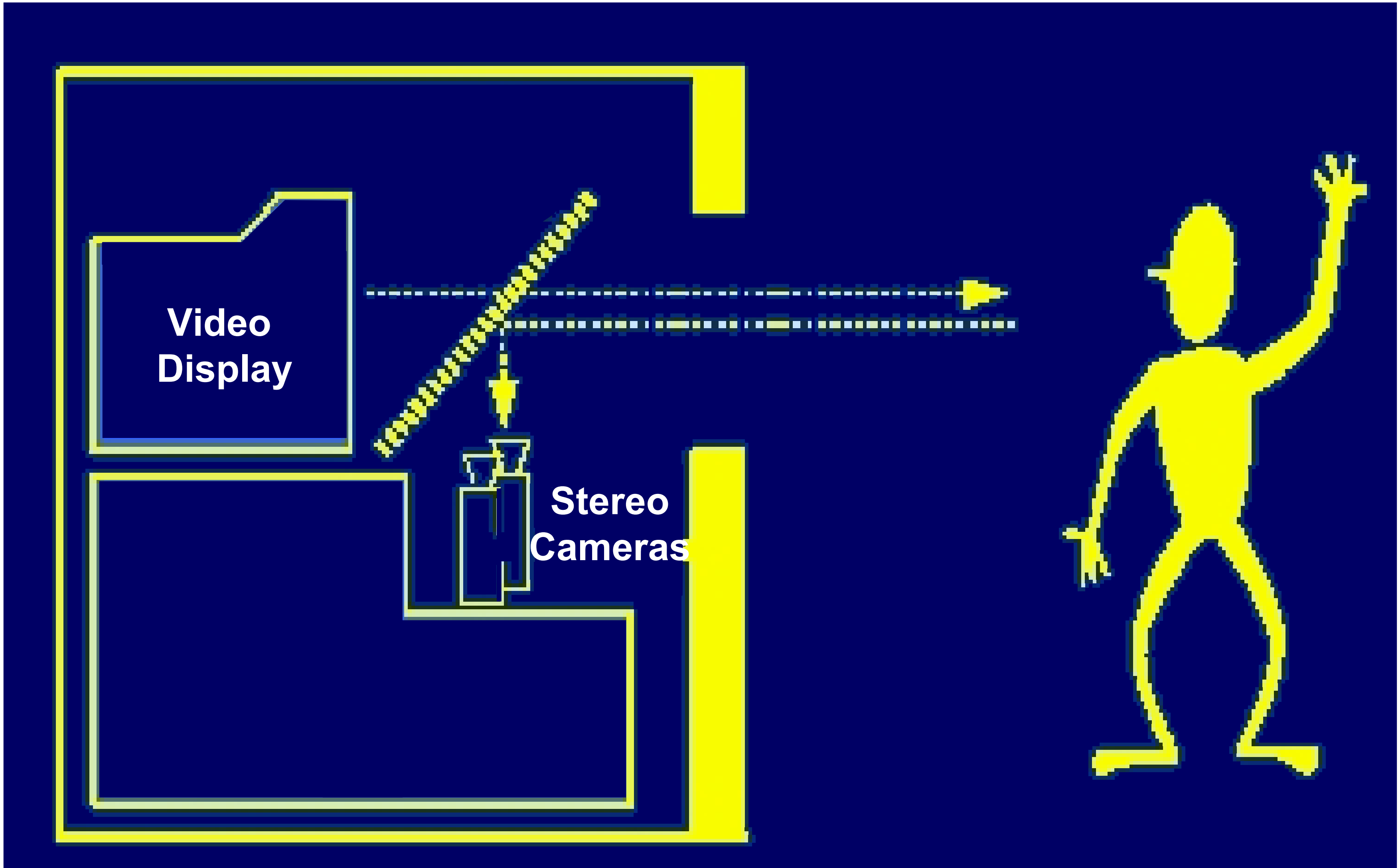
Wireless, Full-body Interaction with Autonomous Agents

Pattie Maes, Trevor Darrell, Bruce Blumberg, Alex Pentland
MIT Media Laboratory

http://vismod.www.media.mit.edu/cgi-bin/tr_pagemaker (TR 257)

http://vismod.www.media.mit.edu/cgi-bin/tr_pagemaker (TR 257)

A Face Responsive Display



Vision-only Application: Interactive Video Effects

end