# 6.869 projects

Projects due Thursday, May 12 (3 weeks from today).

On that day, you'll give us a 5 minute, informal presentation about your project. This is to have fun, to see what other people did, and to do something different on the last day of class (we'll have refreshments). It will also help me and Xiaoxu see on overview of your project before we read your write-up.

The write-up of the project is the main thing. It should be about the length and style of a conference paper submission: about 6 to 8 double-column, single-spaced pages.

# 6.869 projects, continued

The write-up should have an introduction, where you explain why the reader should be interested in the problem, and frame the problem in context.

For a presentation and papers on writing conference papers, see the Weds, April 10, 2002 lecture and readings on this course web page:
http://www.ai.mit.edu/courses/6.899/doneClasses.html

# Next week: a field trip to a guest lecture

*Prof. Dan Huttenlocher, from Cornell*

*Graphical Models for Object Recognition*

*Kiva 32-G449, Tuesday, April 26, 2005, 3-4pm, refreshments at 2:45. I'll come down here at 2:30 to remind anyone who forgets the one-time shift in class location.*

# Today: Cameras looking at, and tracking, people

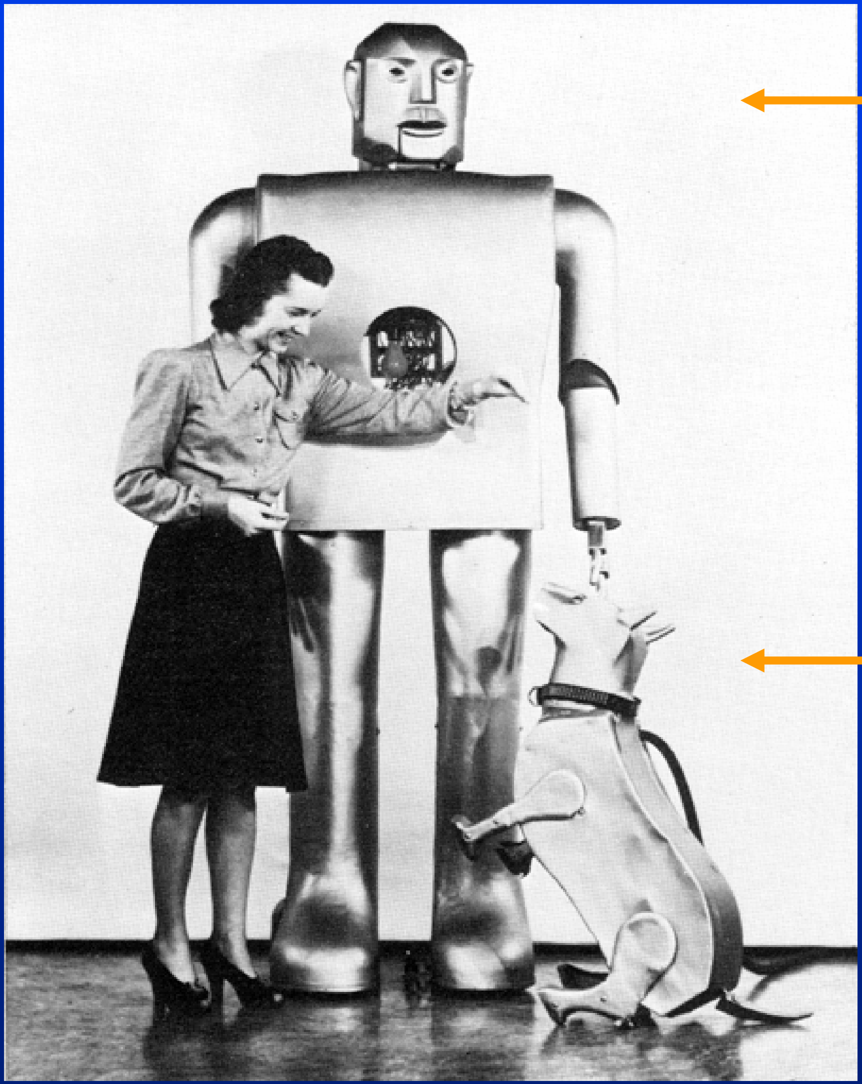
A mini-application lecture: under controlled conditions (not general conditions), what human interaction applications can you build with the tools we've developed so far?

To be compared with: more sophisticated detection, classification methods that we've studied, and the tracking tools that we'll study next.

***MIT 6.869***

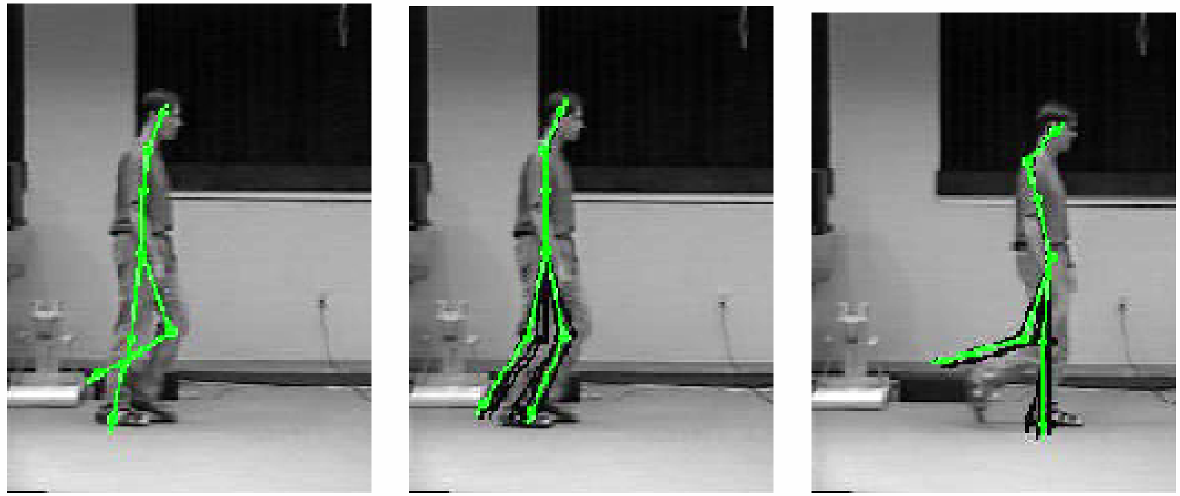
***April 21, 2005***

# Yesterday's tomorrow




New York Worlds Fair, 1939
(Westinghouse Historical Collection)

# Computer vision still needs to become more robust



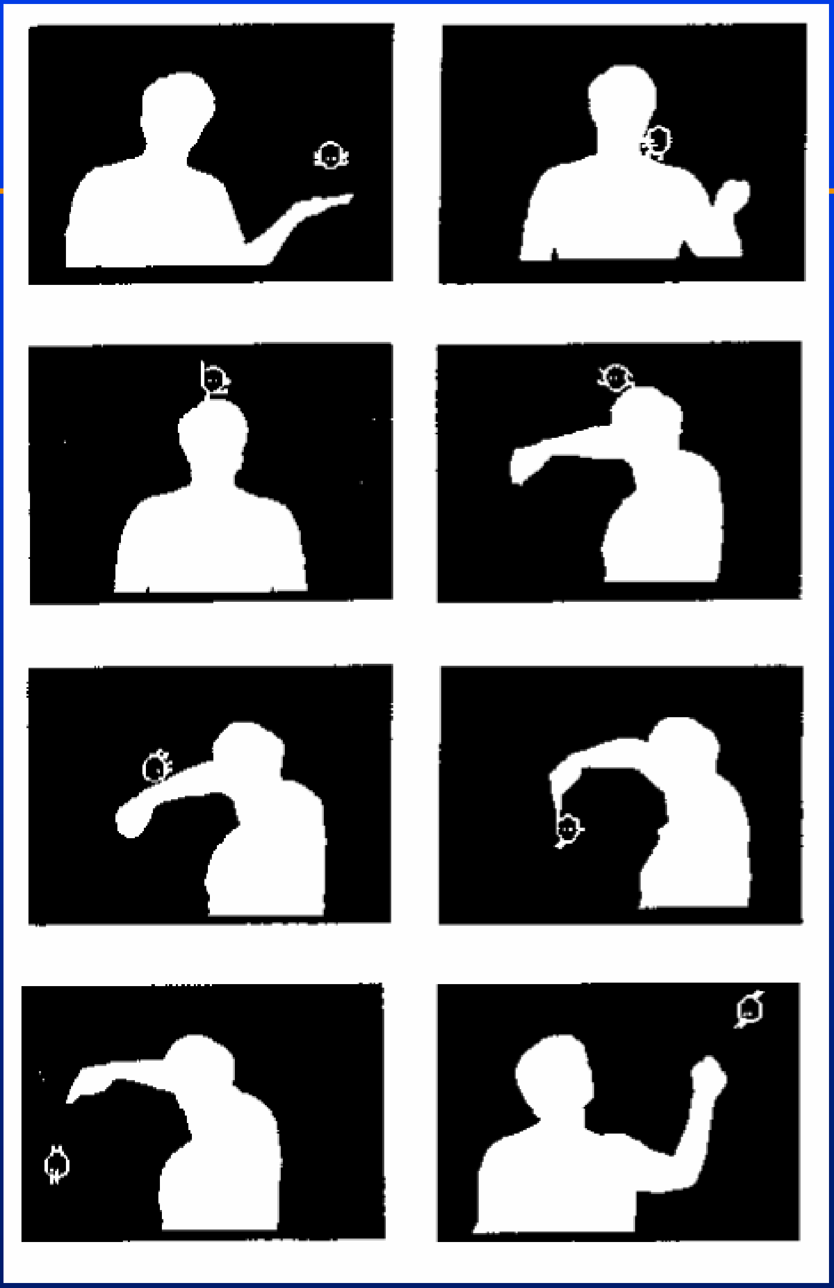
(a) (b) (c)

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:*

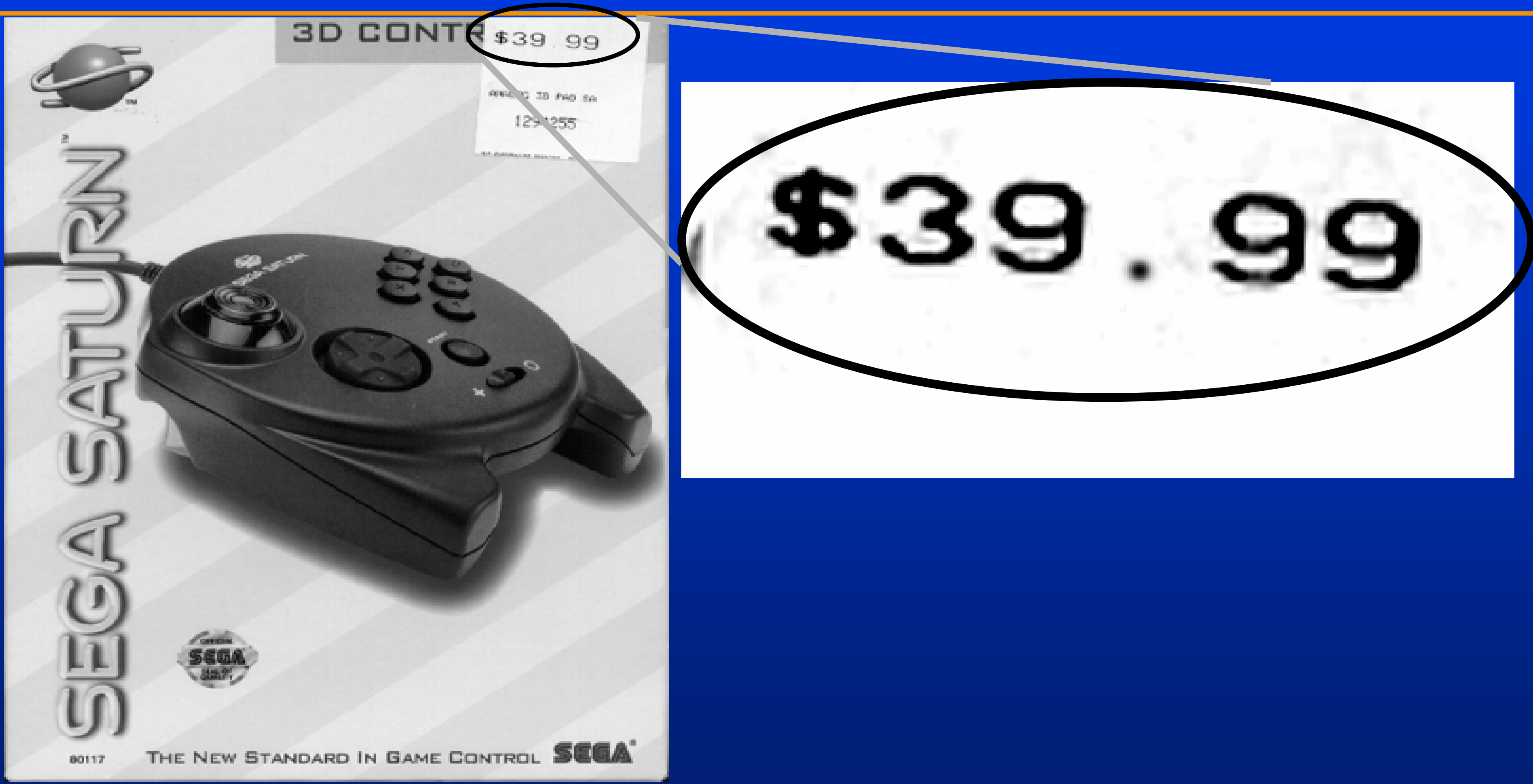
*David Anderson, Paul Beardsley, Chris Dodge, William Freeman, Hiroshi Kage, Kazuo Kyuma, Darren Leigh, Neal McKenzie, Yasunari Miyake, Michal Roth, Ken-ichi Tanaka, Craig Weissman, William Yerazunis*

# Computer vision based interface

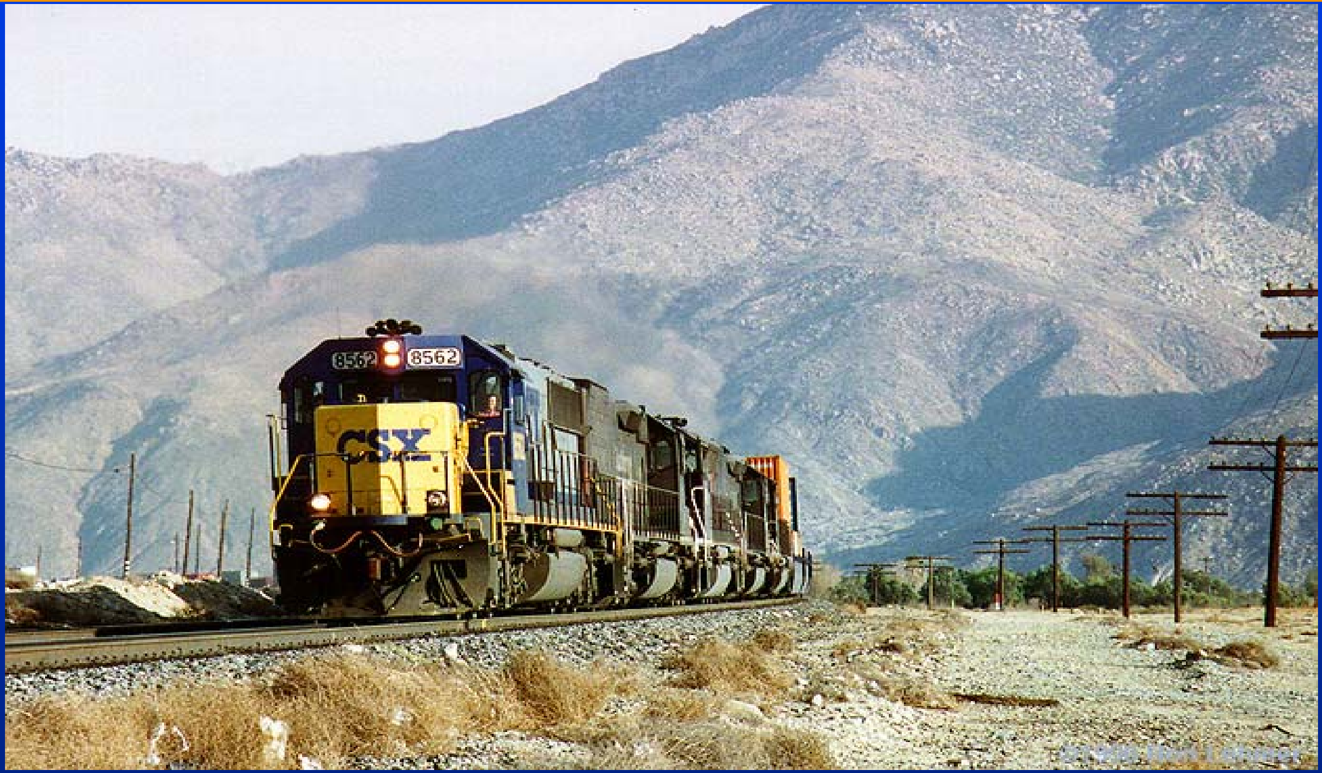



***The hope: video input will give a more expressive, natural or engaging interface.***

# 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.

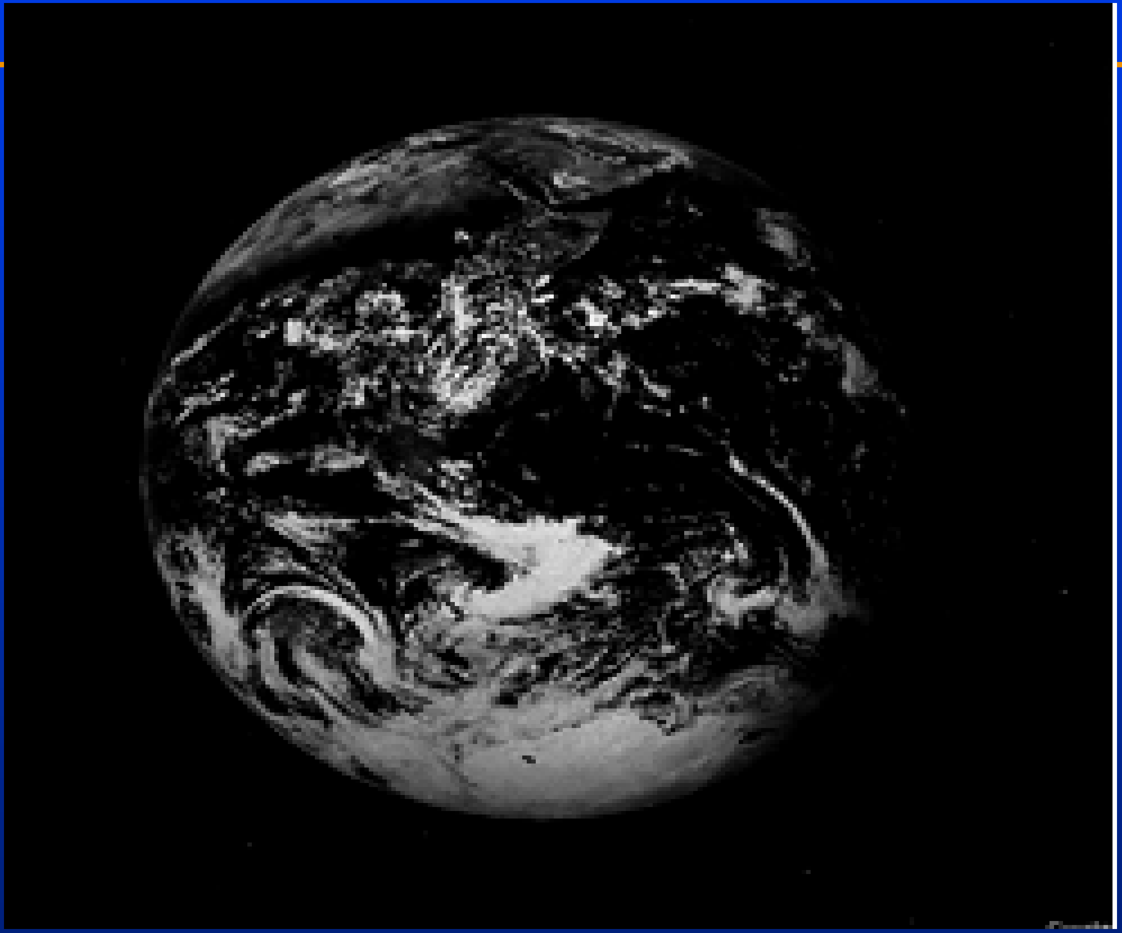
# Computer vision algorithms as ocean-going vessels

# 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*

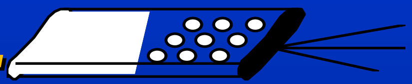
*message:*

*People value the ability to control a television from a distance.*

# Control of television set from a distance

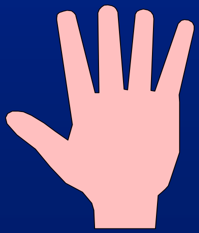
***Wired remote control.***

***Infra-red remote control.***



***Voice control.***

***Gesture control.***



# Design constraints

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

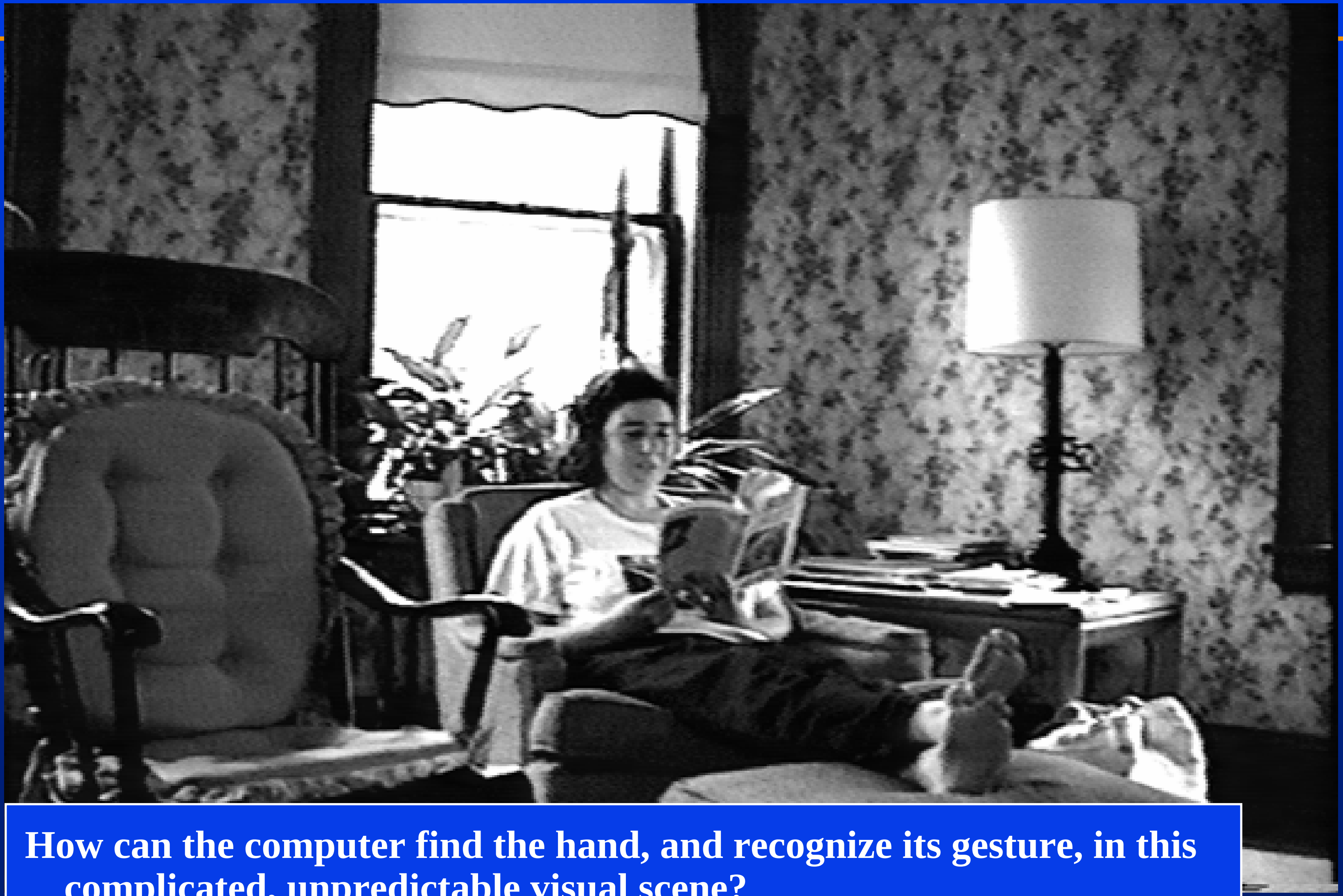
From the user's point of view:

# Complex commands require complicated gestures?

*"mute"*

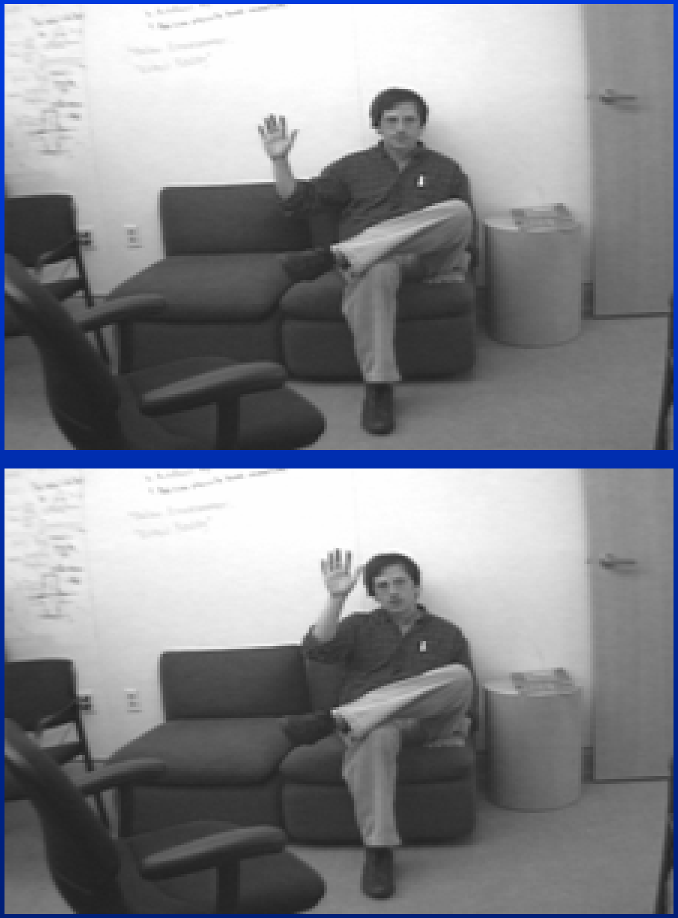
From the computer's point of view:

# Living room scene is difficult

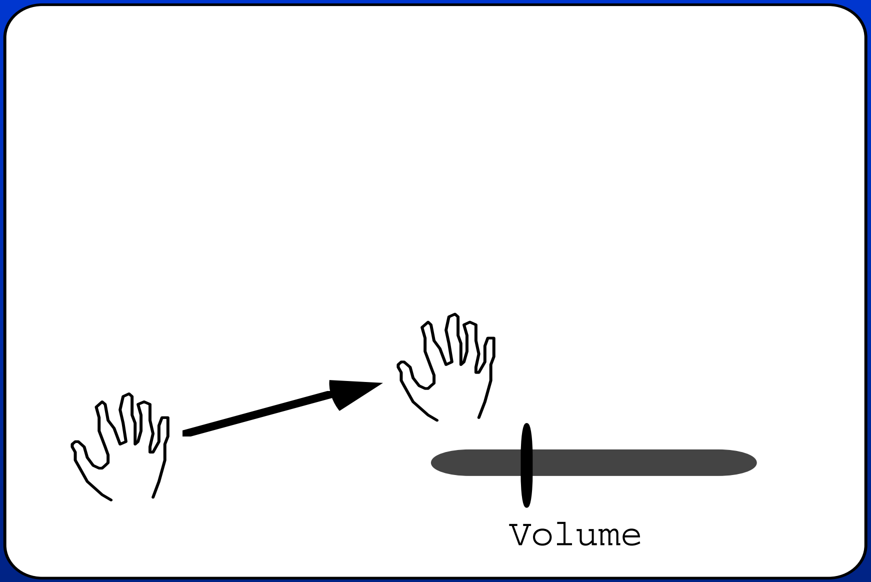


How can the computer find the hand, and recognize its gesture, in this complicated, unpredictable visual scene?

# Our solution: exploit the visual feedback from the television



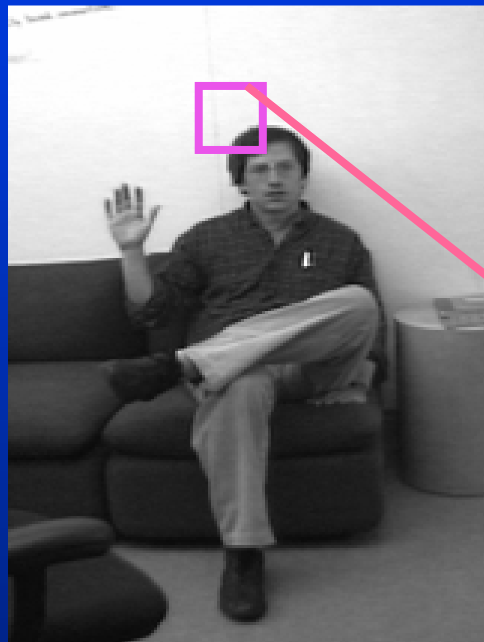
user




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

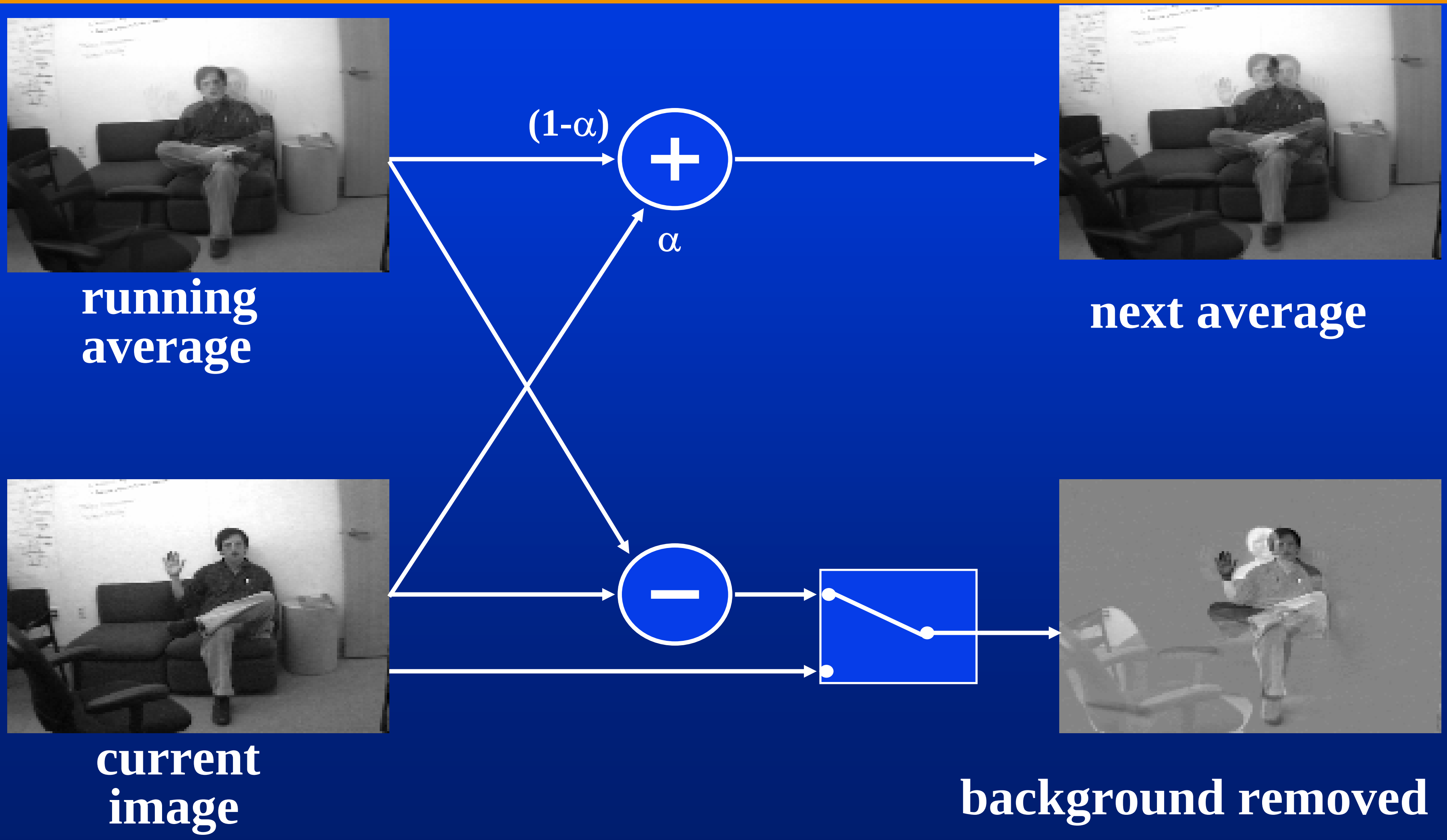
normalized correlation

# Normalized correlation

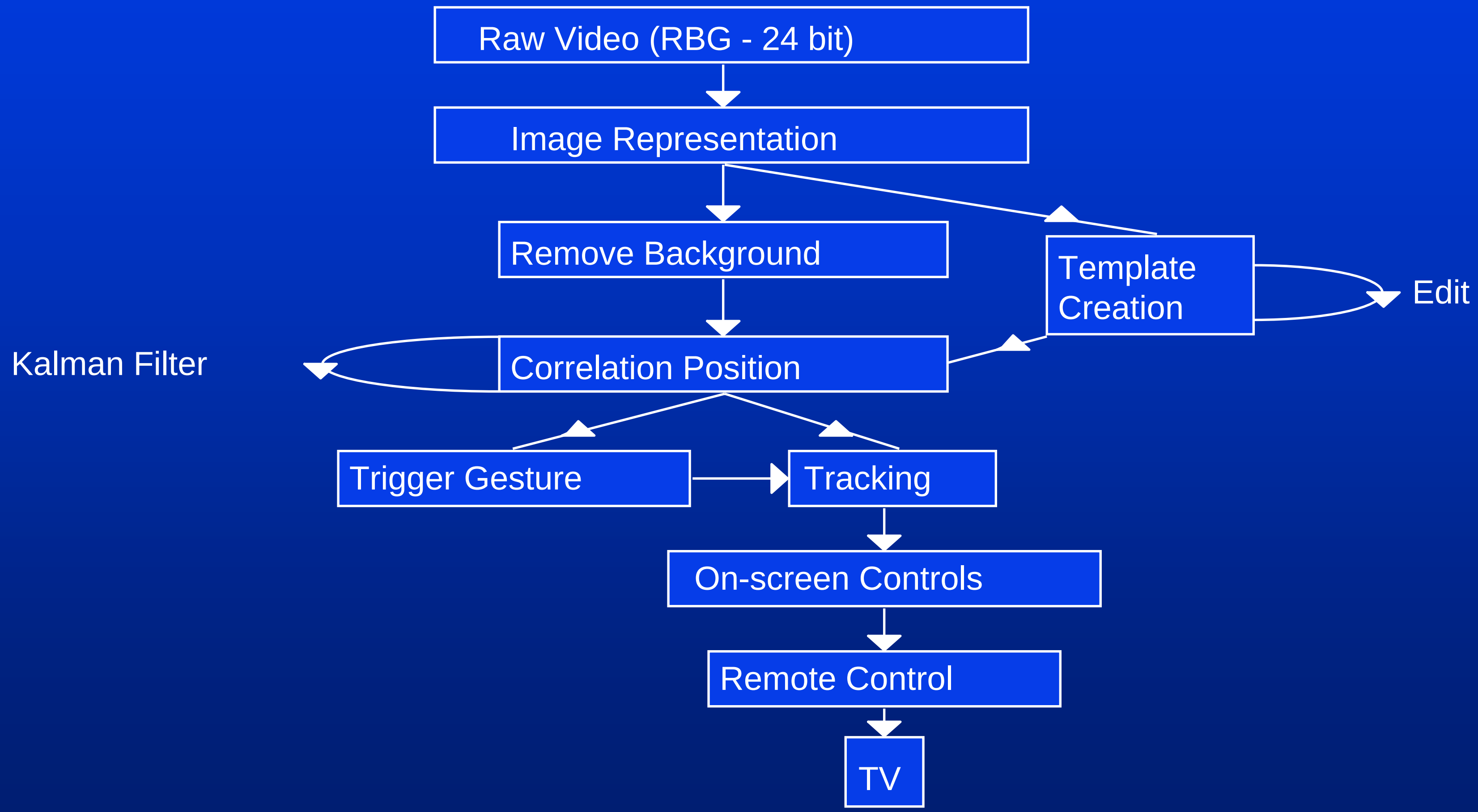
$$\frac{\vec{a} \cdot \vec{b}}{\sqrt{(\vec{a} \cdot \vec{a})(\vec{b} \cdot \vec{b})}}$$

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

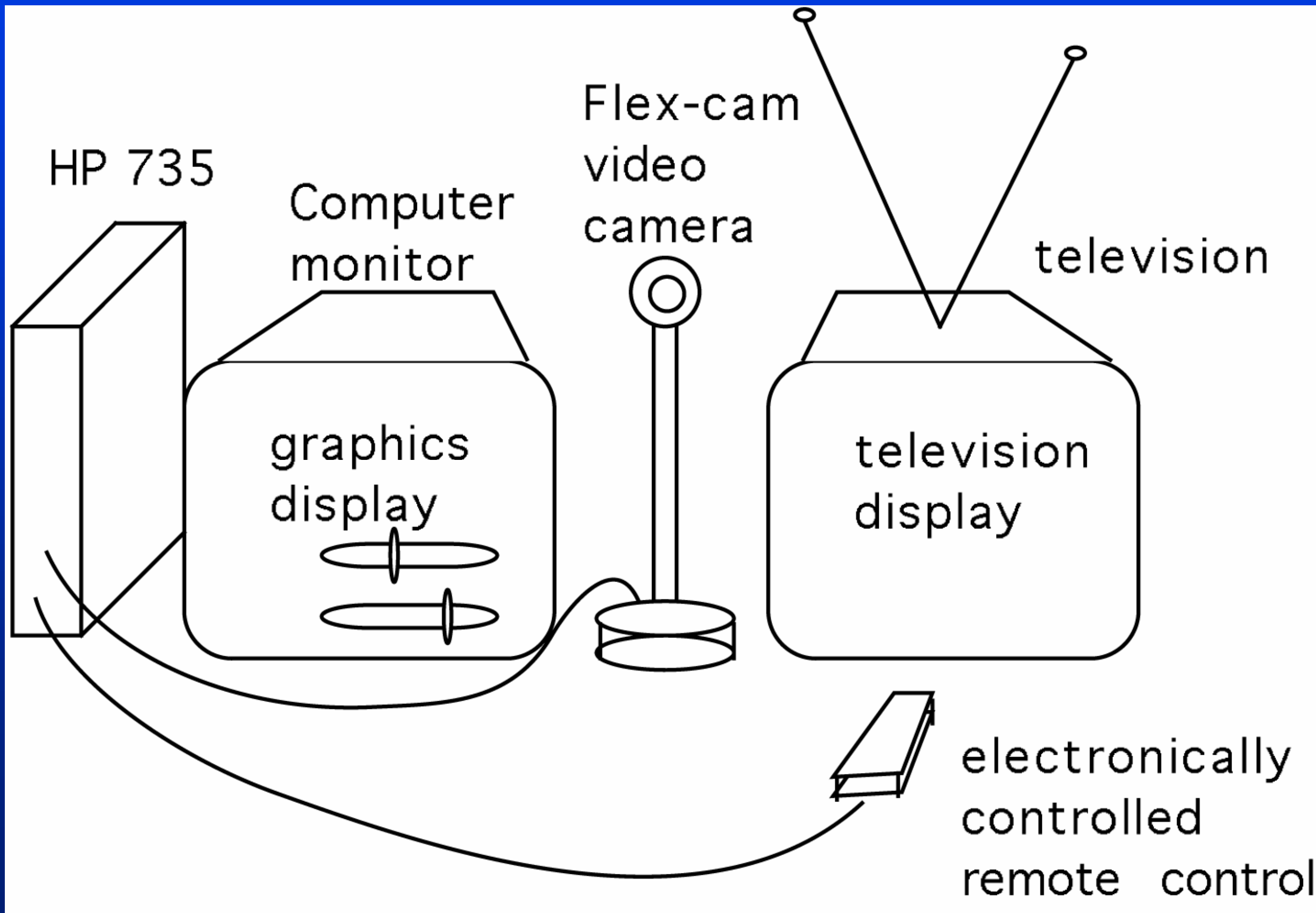
# Background removal



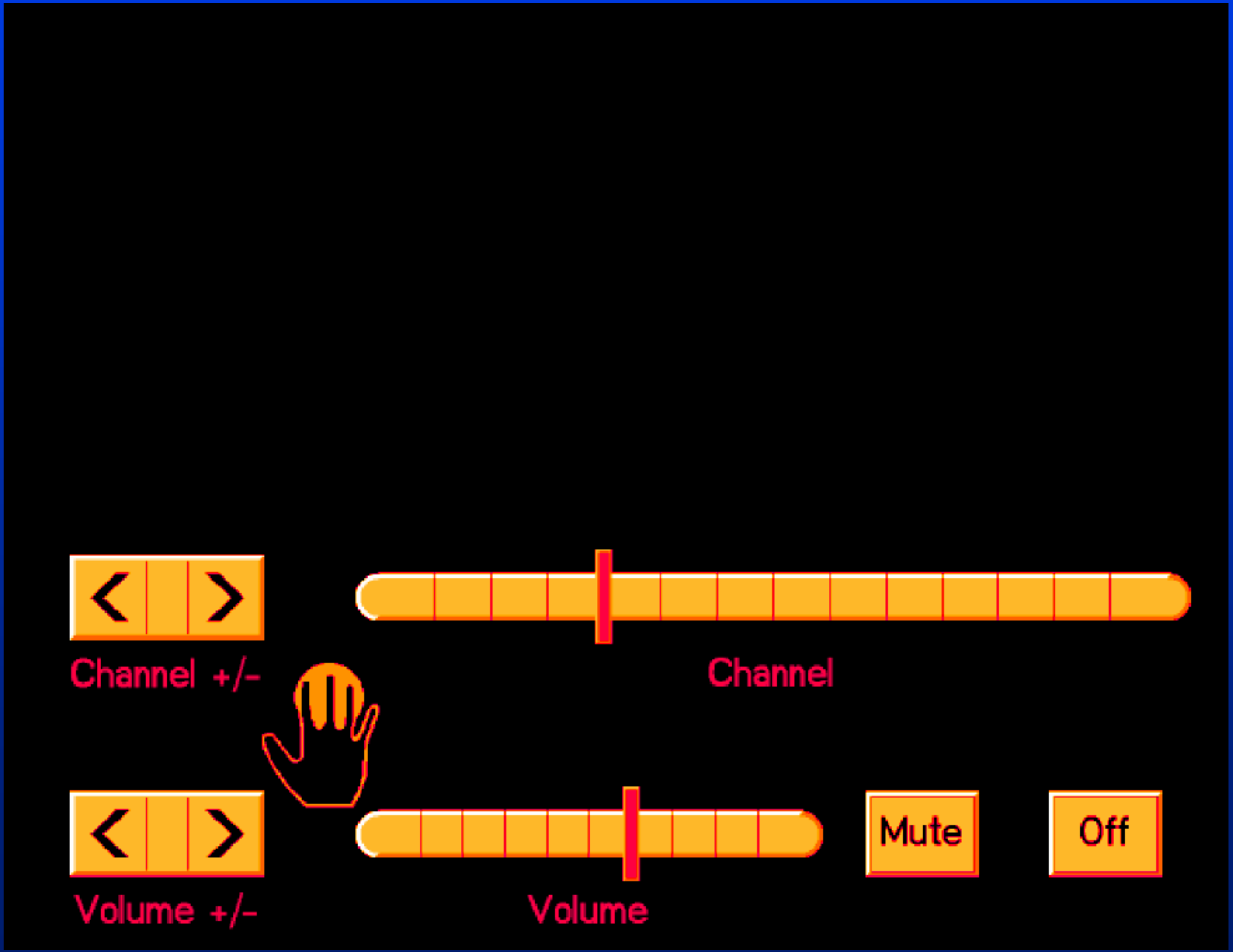
# Processing block diagram



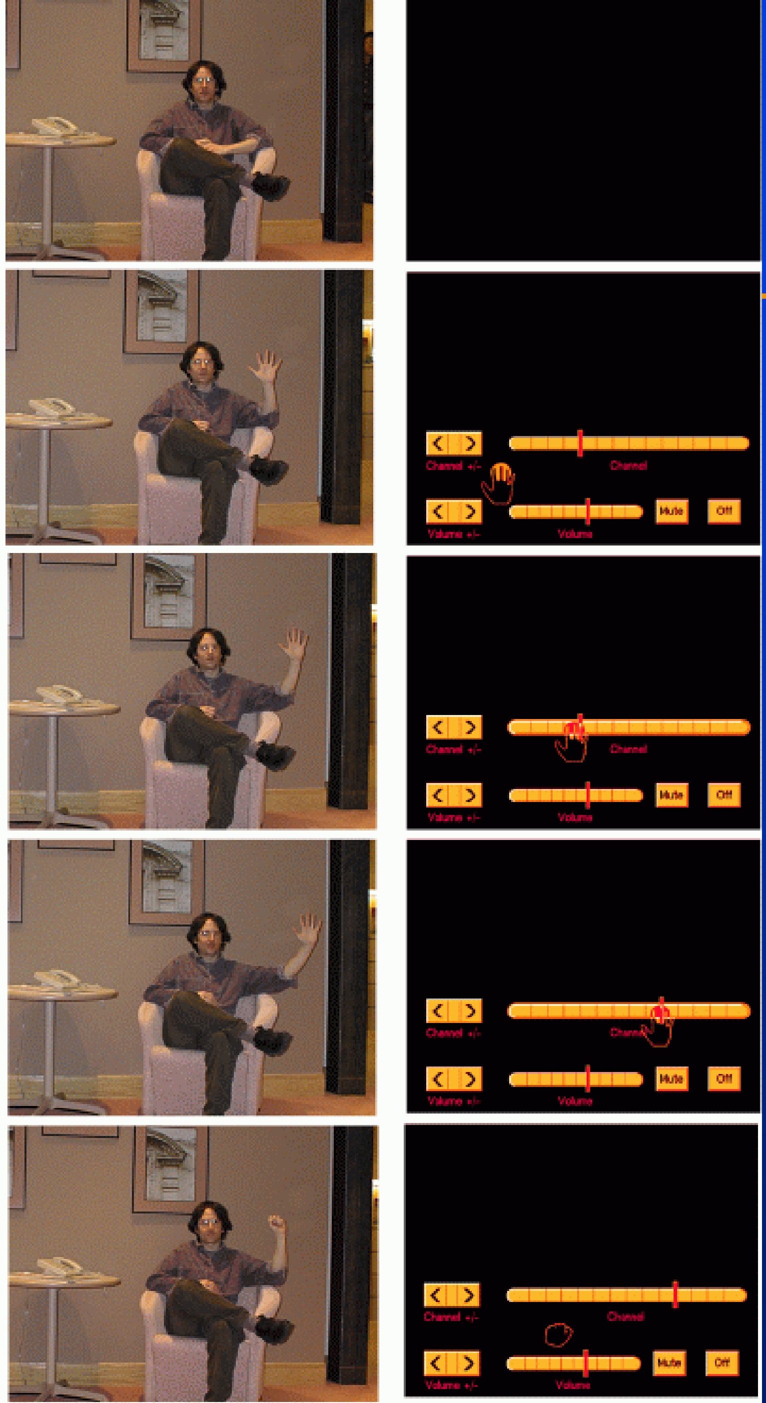
# Prototype of television controlled by hand signals.



# 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***

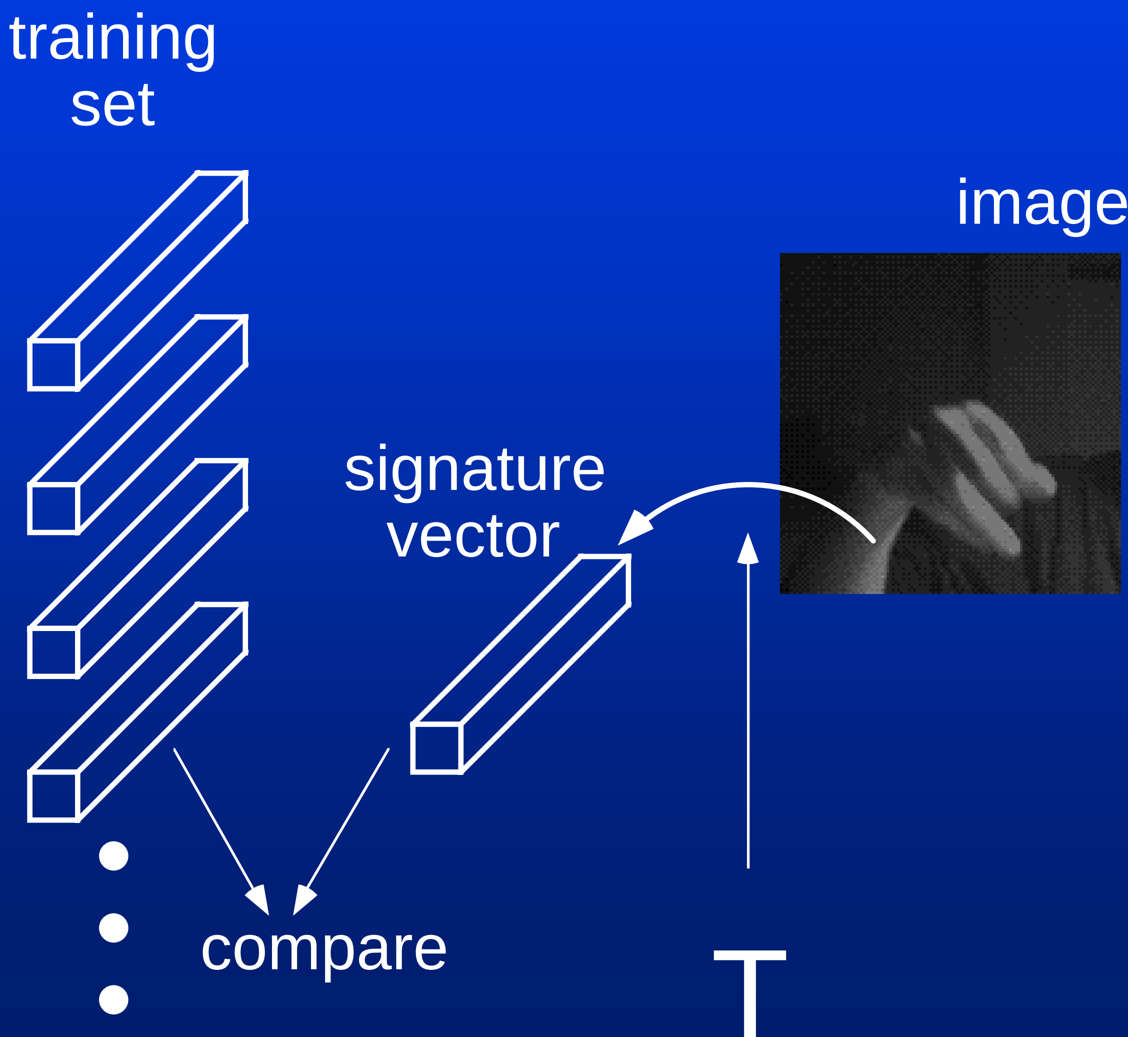
- camera
- video digitizer
- computer

***Long term***

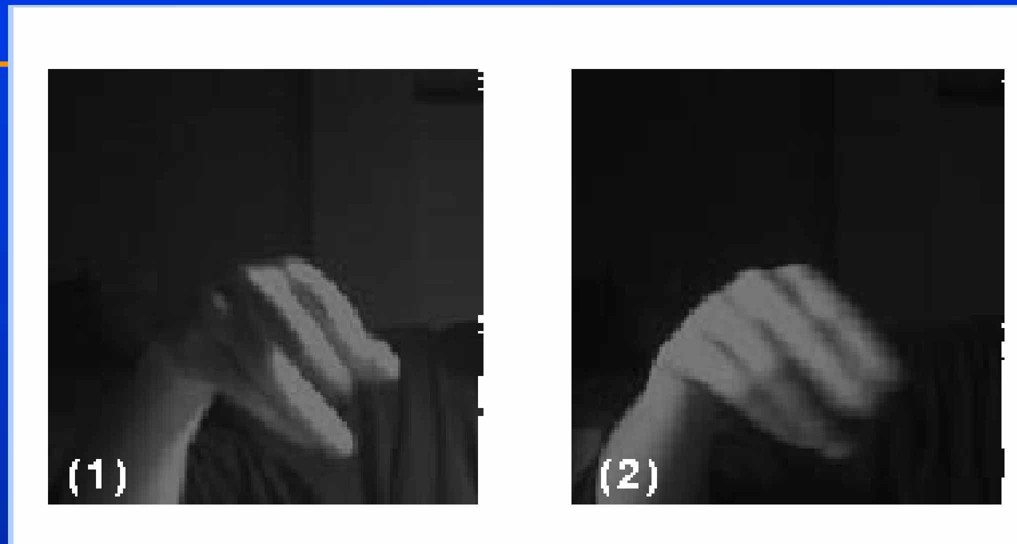
- TV's / computers / browsers will have cameras and powerful computers.
- a software product.

# 2. Simple gesture recognition method

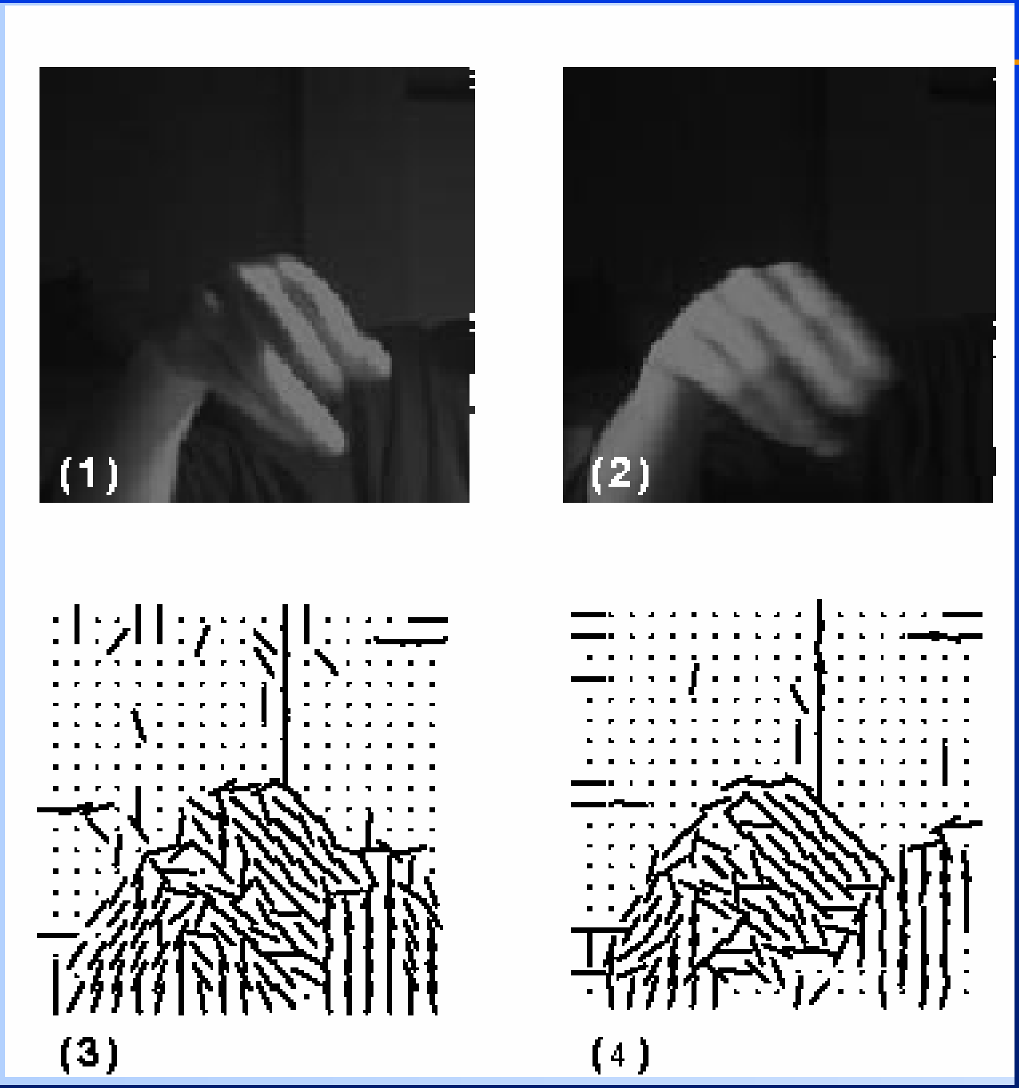
# Real-time hand gesture recognition by orientation histograms



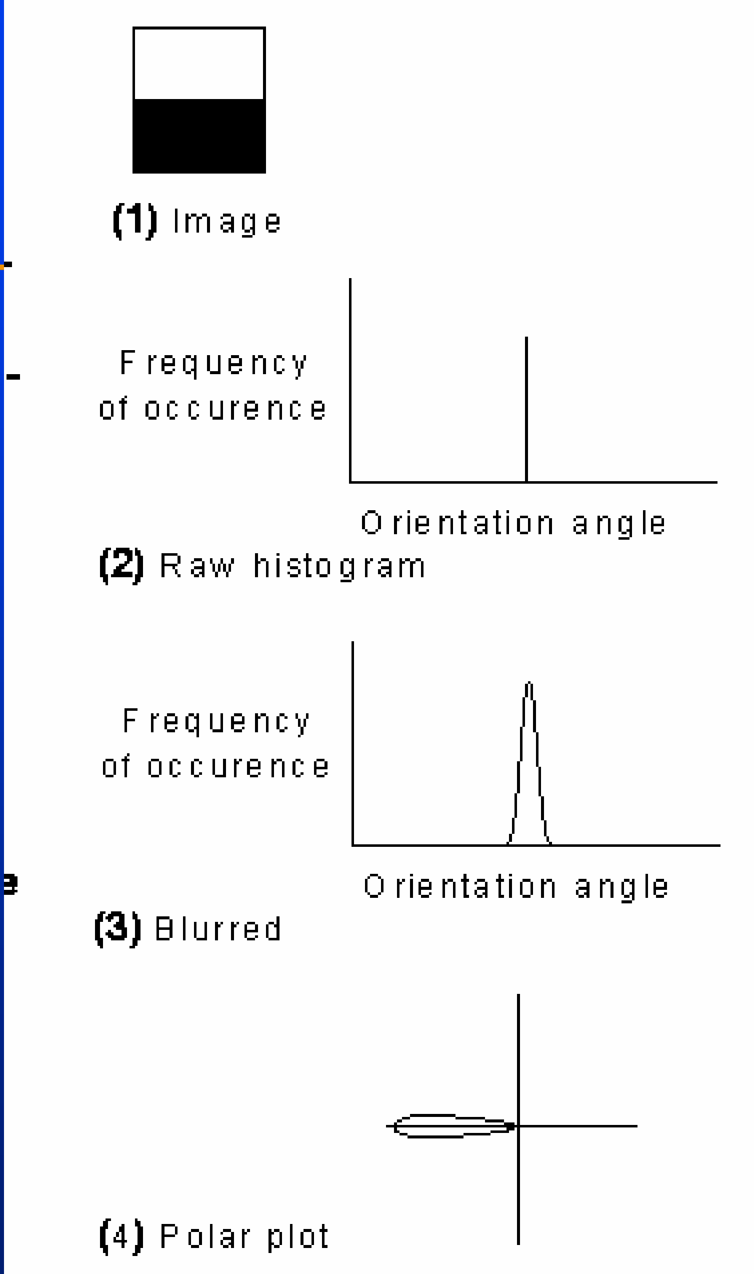
## Orientation measurements (bottom) are more robust to lighting changes than are pixel intensities (top)



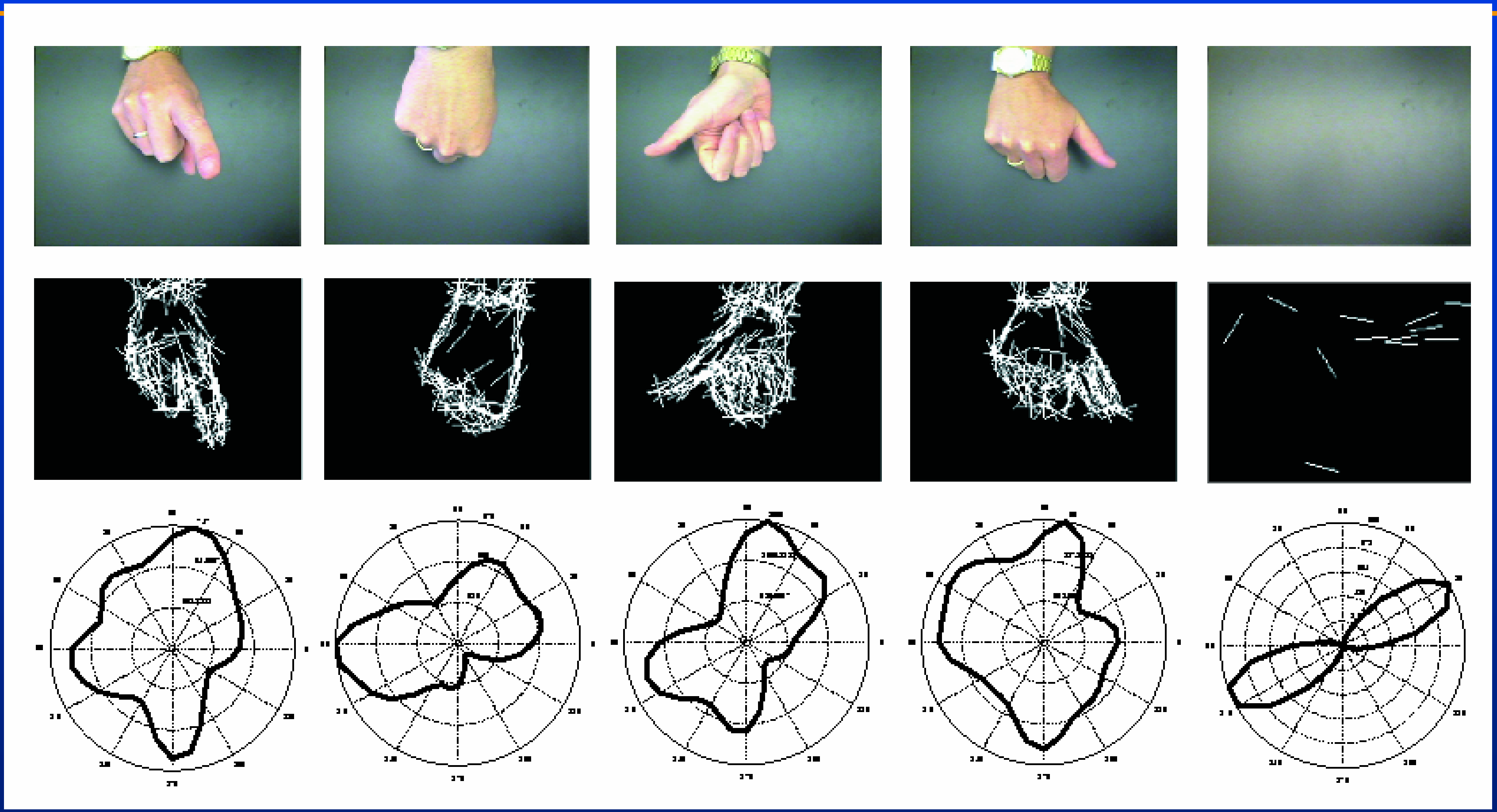
# Orientation measurements (bottom) are more robust to lighting changes than are pixel intensities (top)



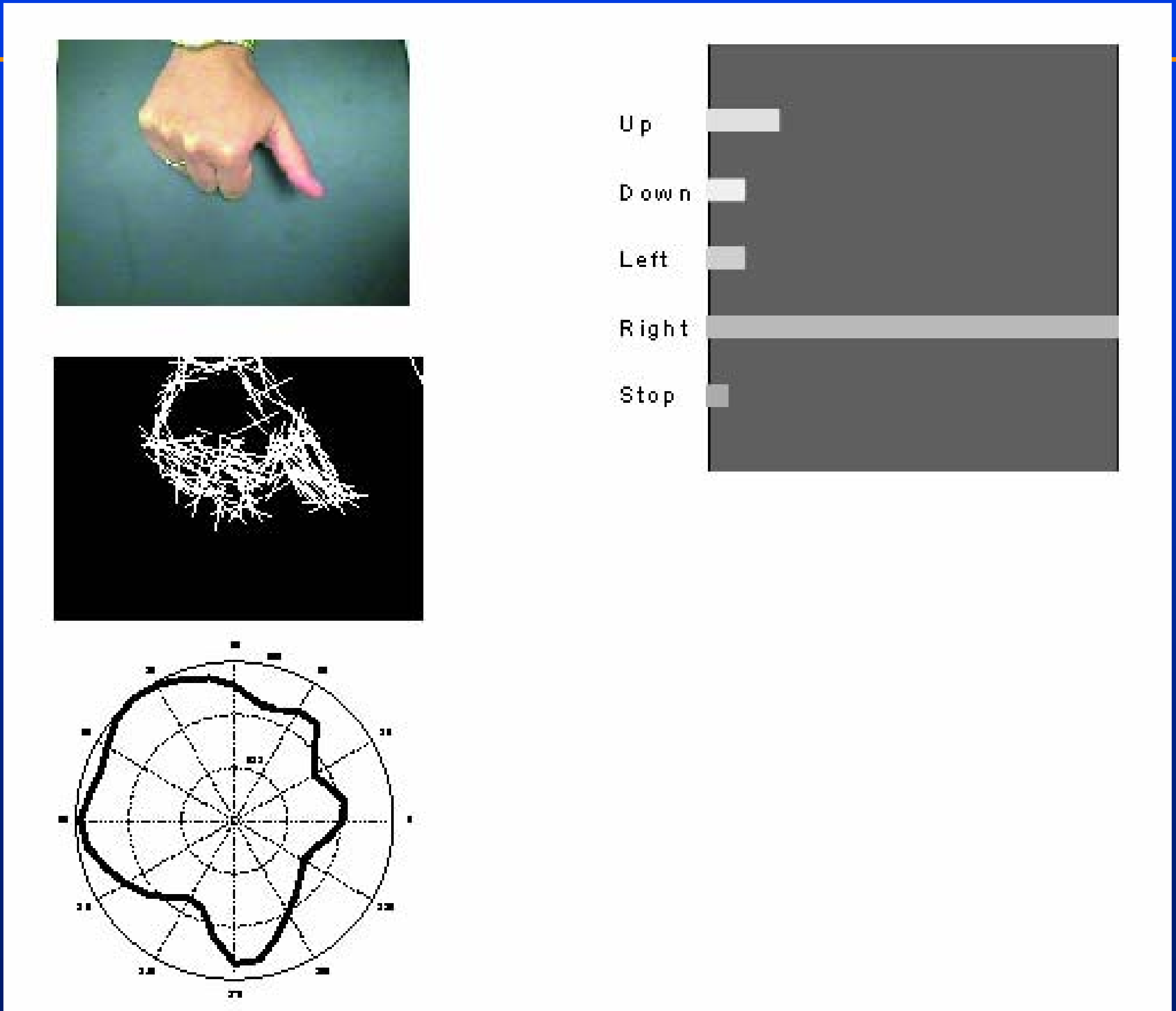
**C** Simple illustration of an orientation histogram. (1) An image of a horizontal edge has only one orientation at a sufficiently high contrast. (2) Thus the raw orientation histogram has counts at only one orientation value. (3) To allow neighboring orientations to sense each other, we blurred the 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.

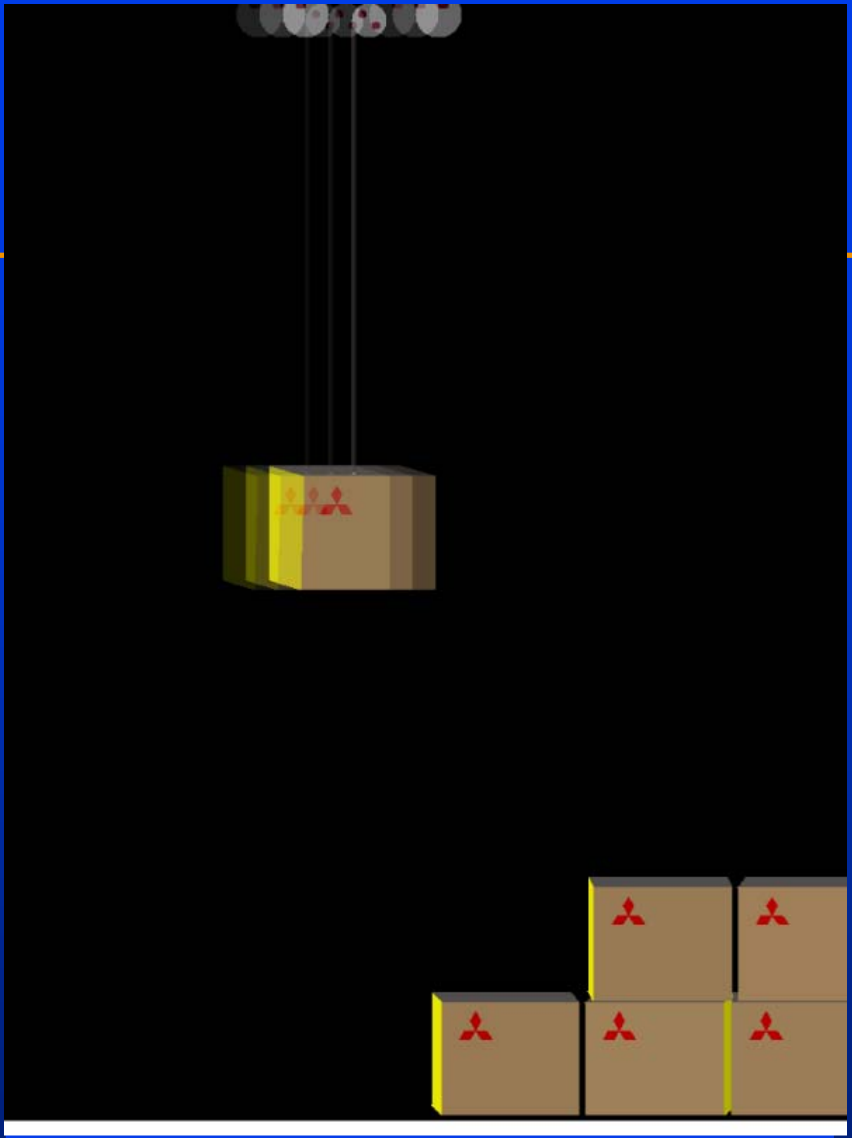


# Images, orientation images, and orientation histograms for training set

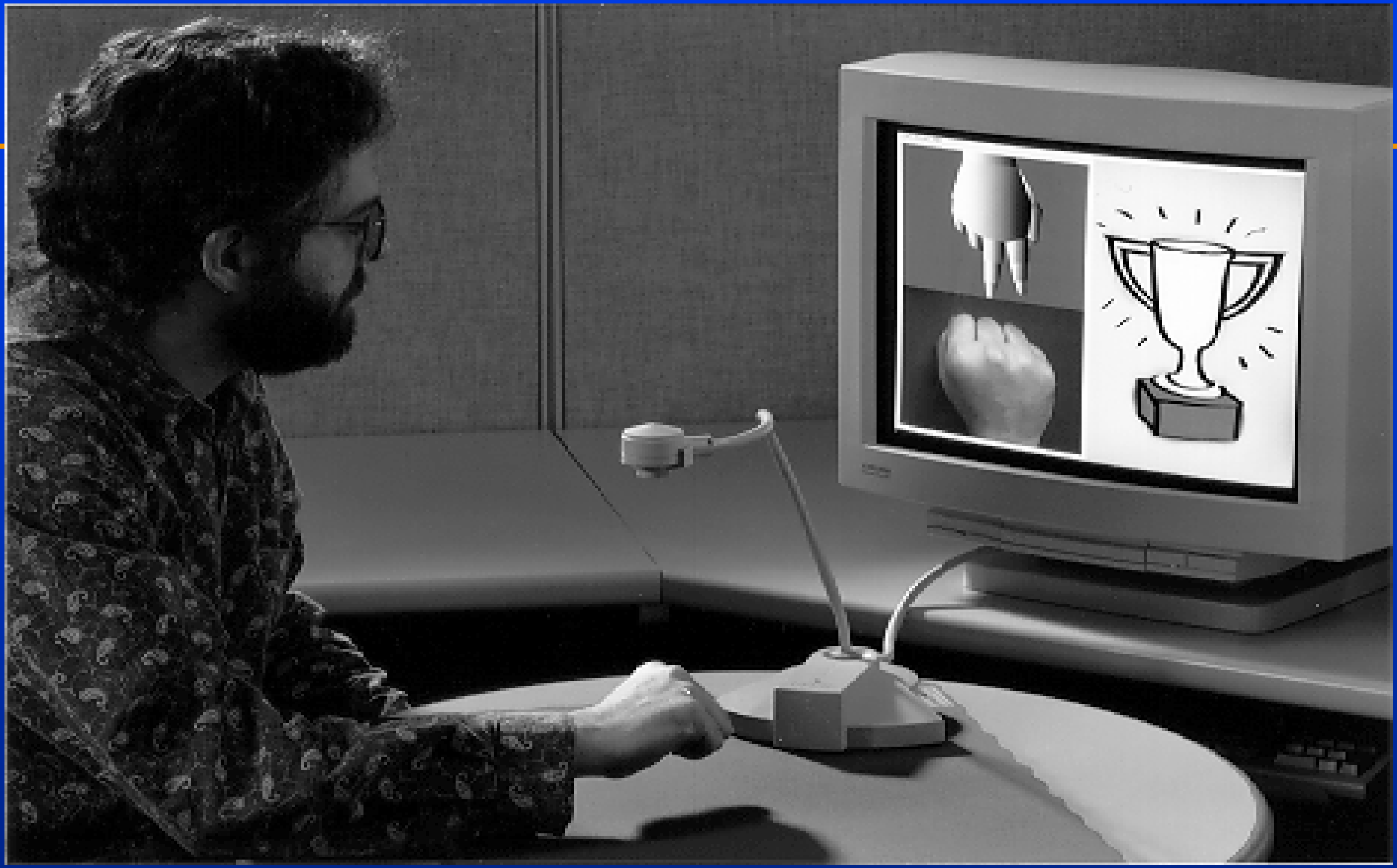


# Test image, and distances from each of the training set orientation histograms (categorized correctly).





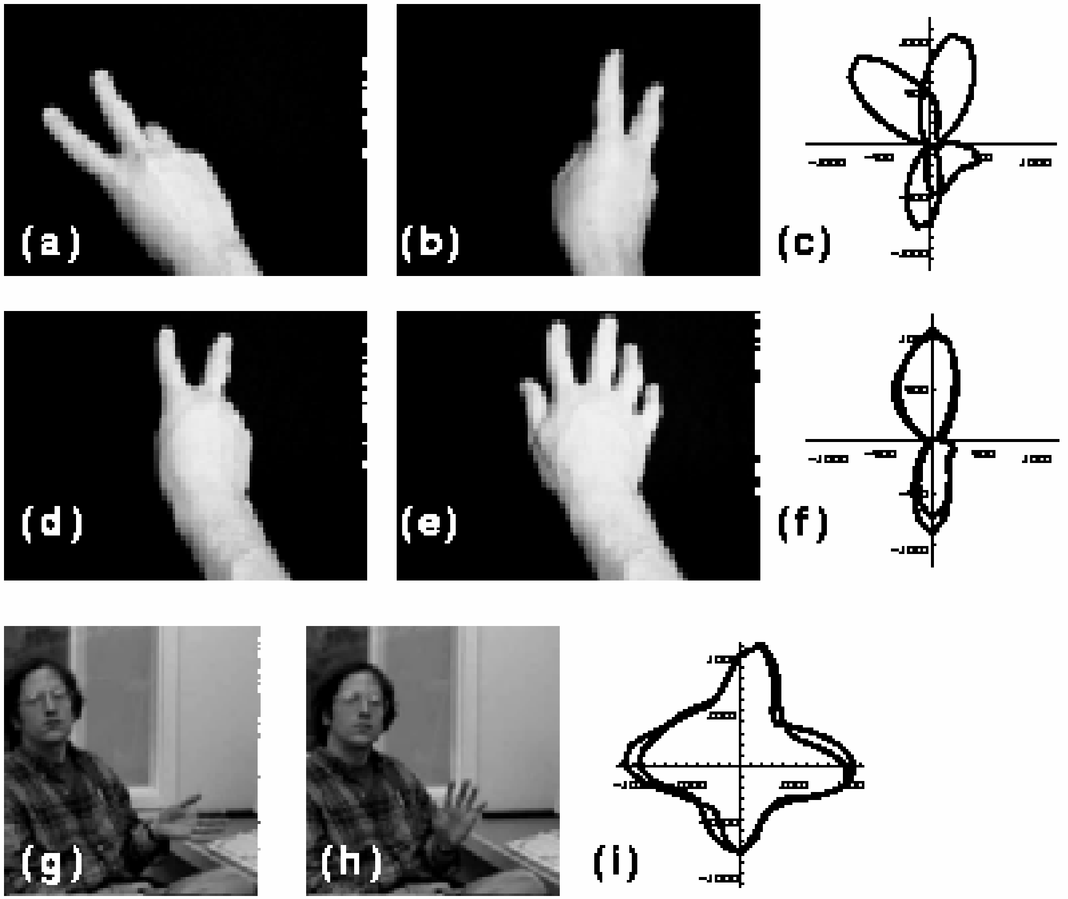
Crane movements controlled by hand gestures



**Janken game**

# video

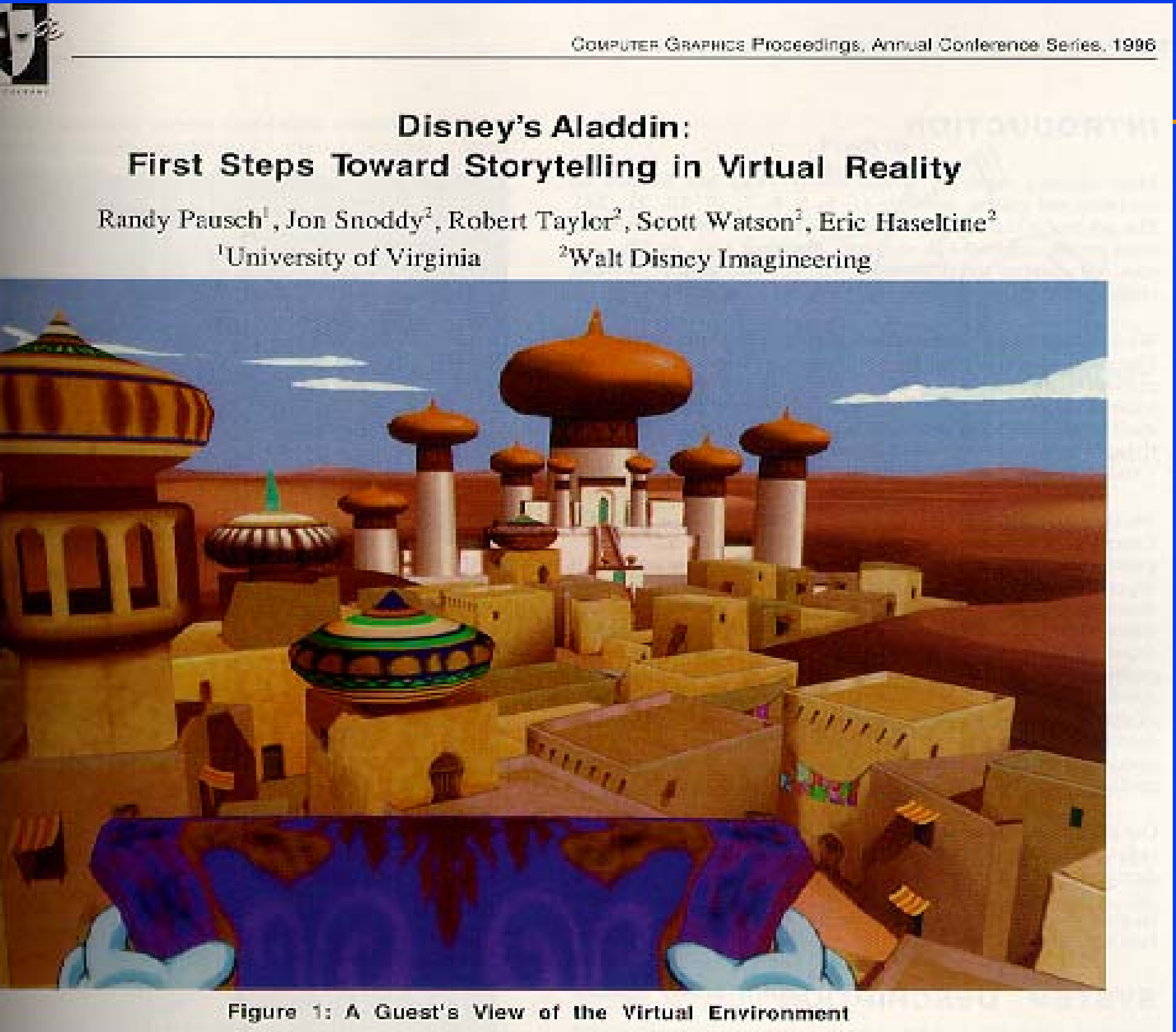
**7** Problem images for the orientation histogram-based gesture classifier.



# 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.


COMPUTER GRAPHICS Proceedings, Annual Conference Series, 1996

Disney's Aladdin:
First Steps Toward Storytelling in Virtual Reality

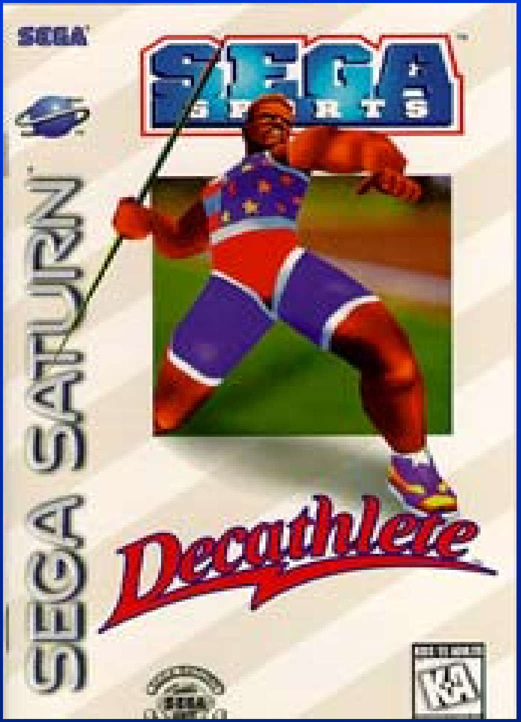
Randy Pausch[1], Jon Snoddy[2], Robert Tayler[2], Scott Watson[2], Eric Haseltine[2]
[1]University of Virginia [2]Walt Disney Imagineering

Figure 1: A Guest's View of the Virtual Environment

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

# Games selected for vision interface





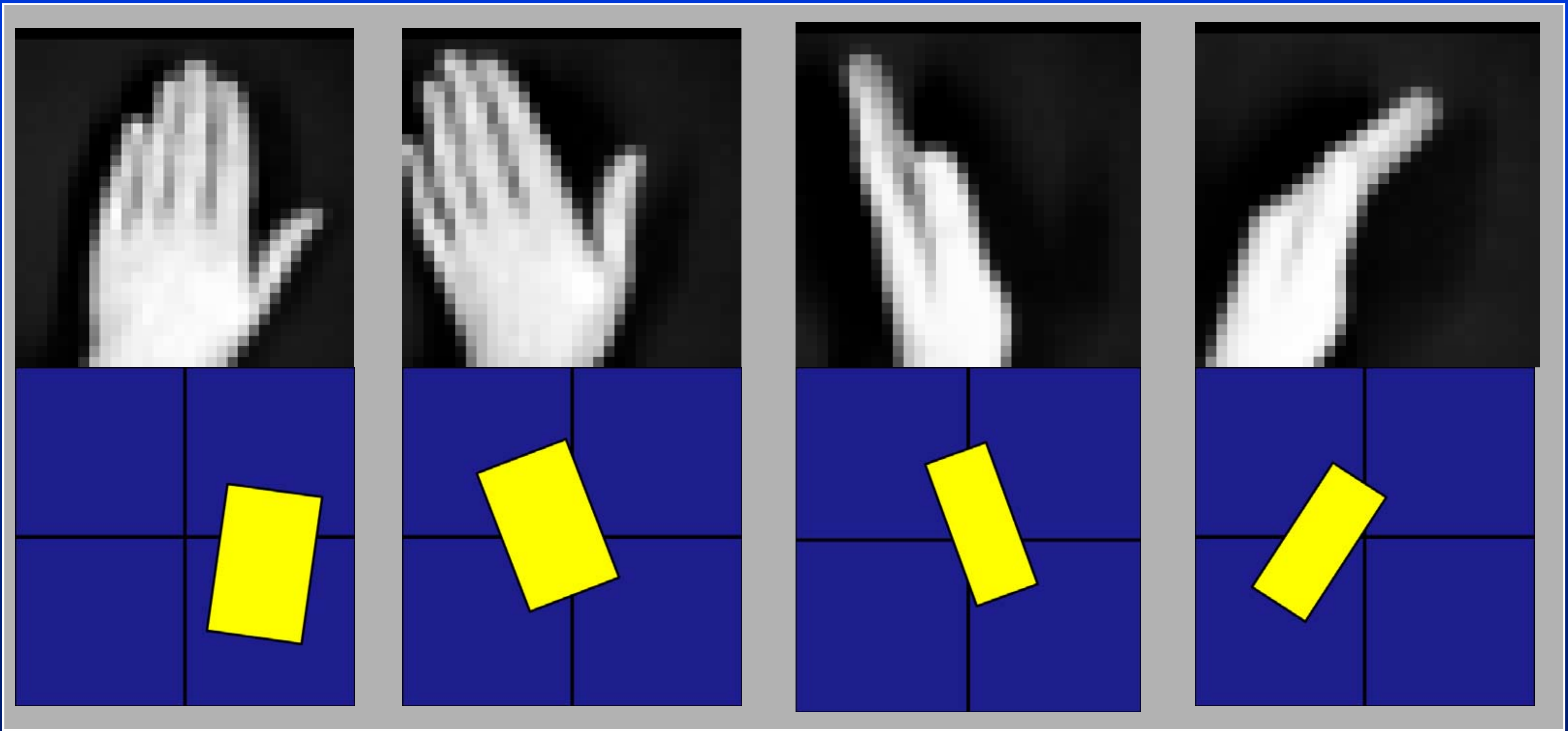
# Image moments give a very coarse image summary.

$$M_{00} = \sum_x \sum_y I(x,y) \qquad M_{10} = \sum_x \sum_y x\, I(x,y)$$

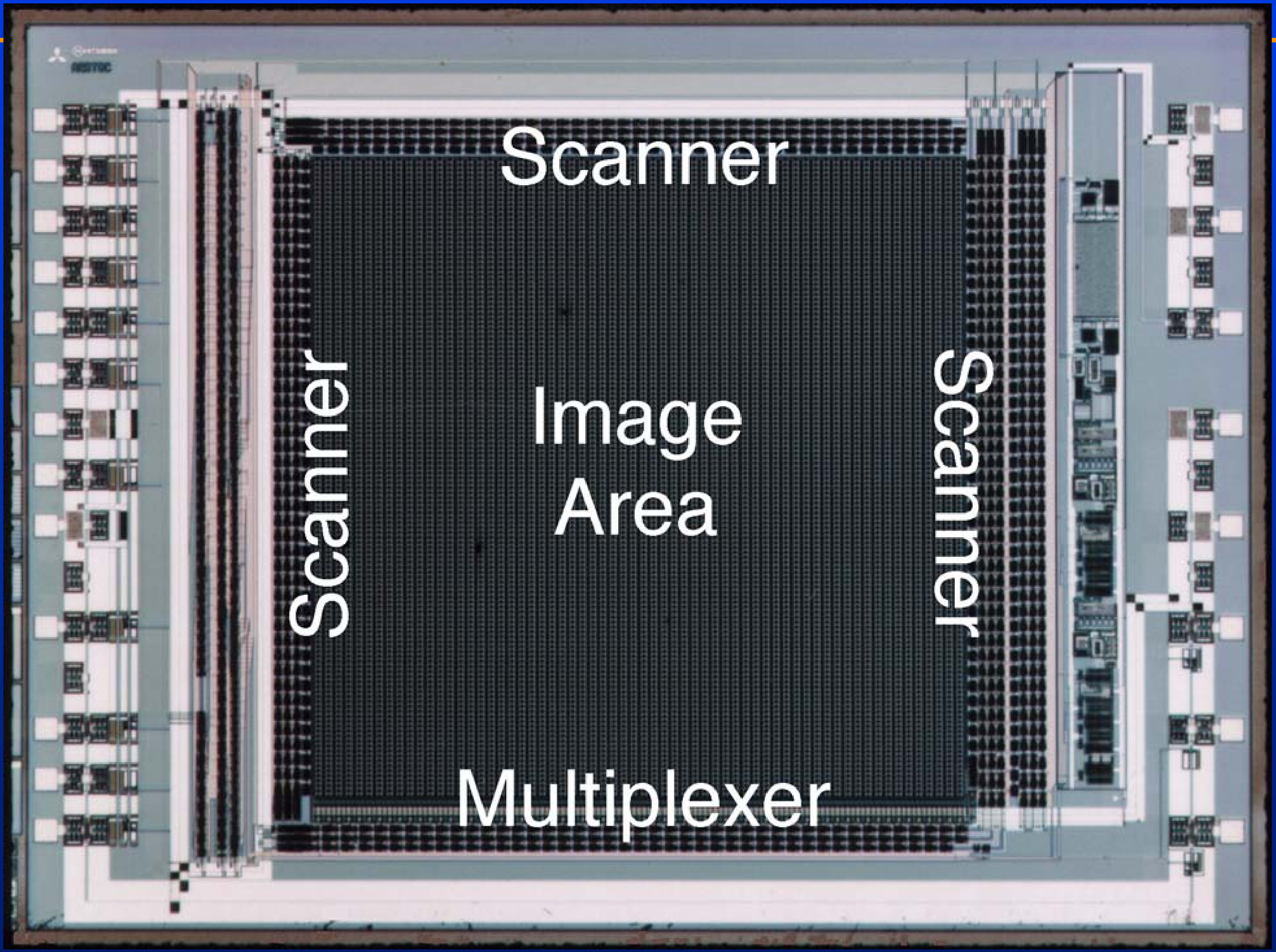
$$M_{01} = \sum_x \sum_y y\, I(x,y) \qquad M_{20} = \sum_x \sum_y x^2\, I(x,y)$$

$$M_{11} = \sum_x \sum_y xy\, I(x,y) \qquad M_{02} = \sum_x \sum_y y^2\, I(x,y)$$

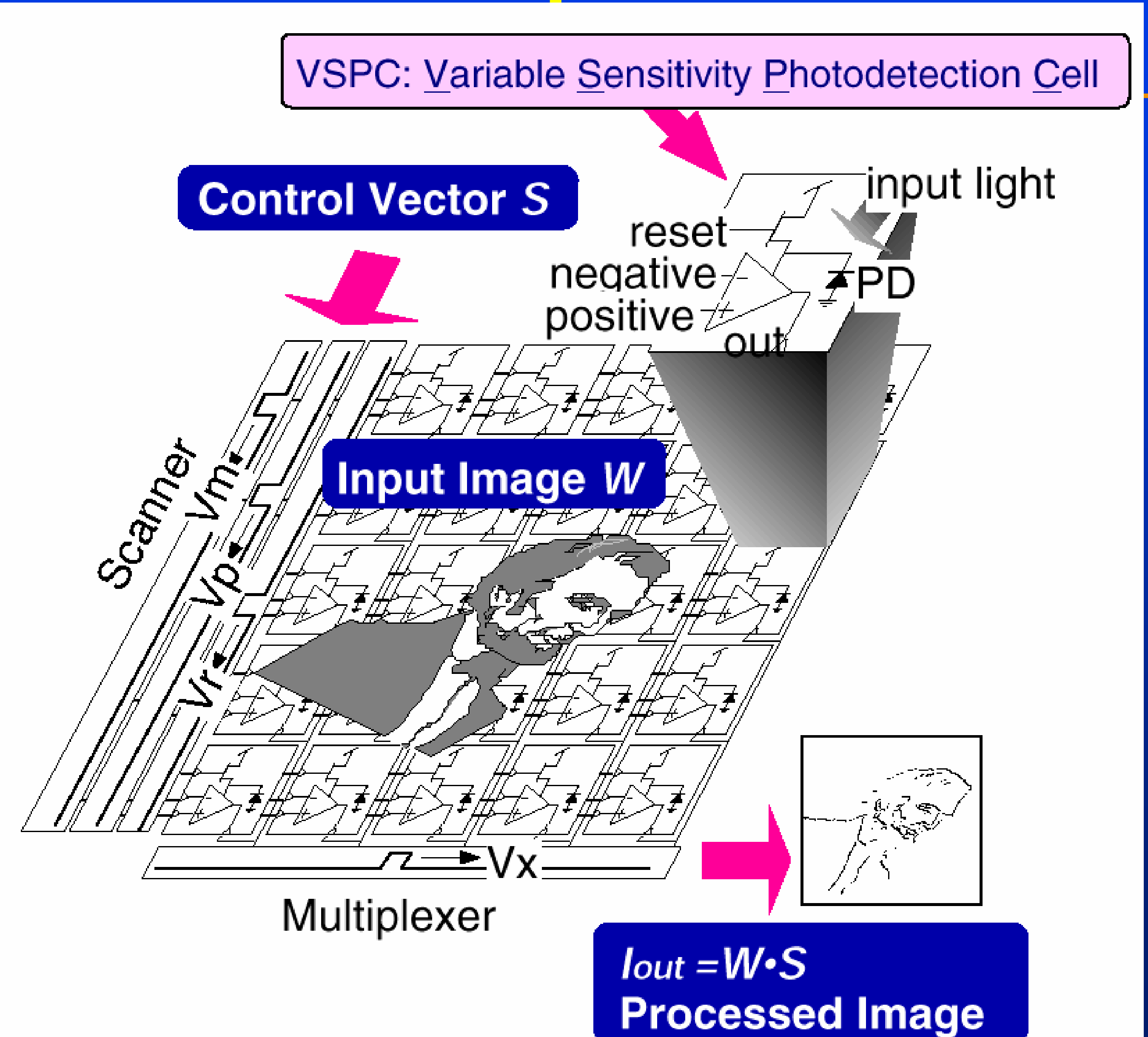
# Hand images and equivalent rectangles having the same image moments



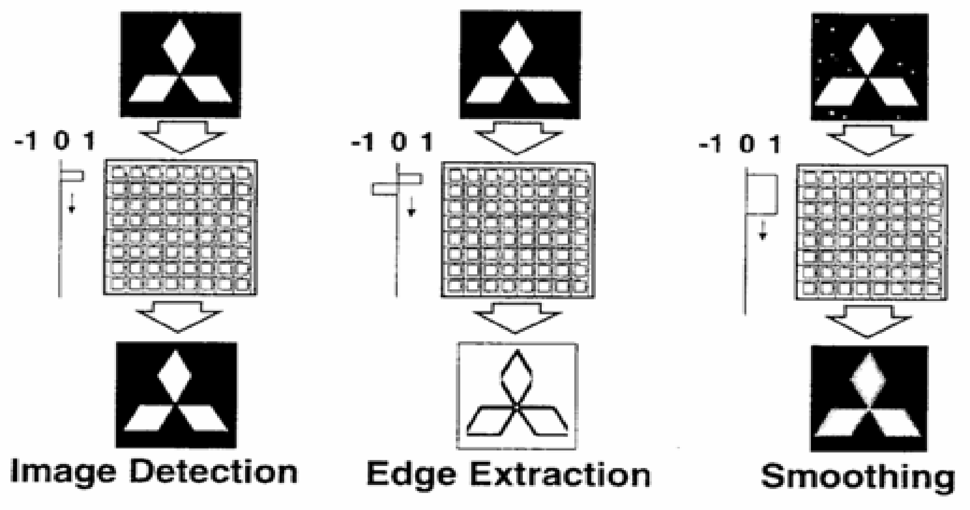
# Artificial Retina chip for detection and low-level image processing.

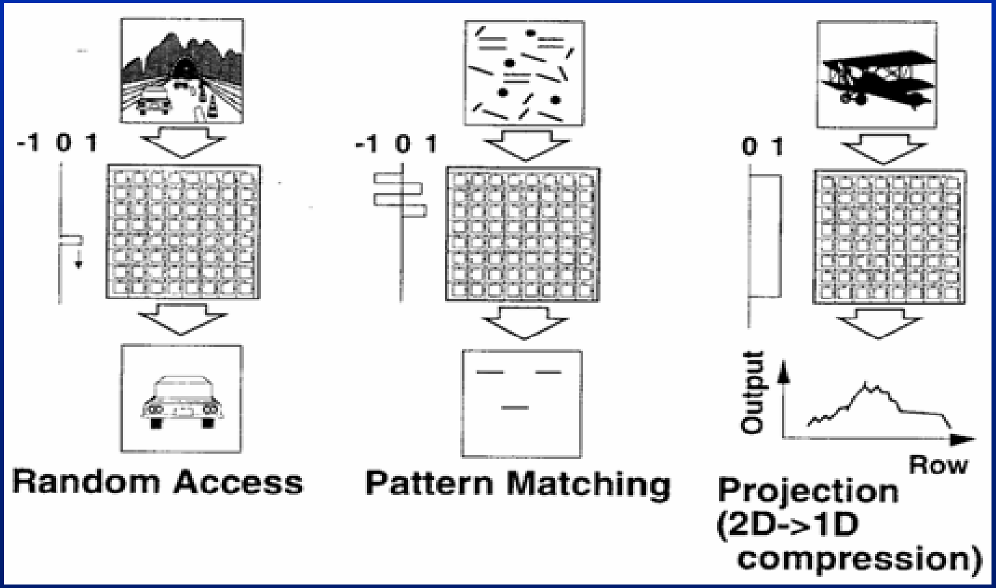


# Artificial Retina chip



# Artificial Retina functions



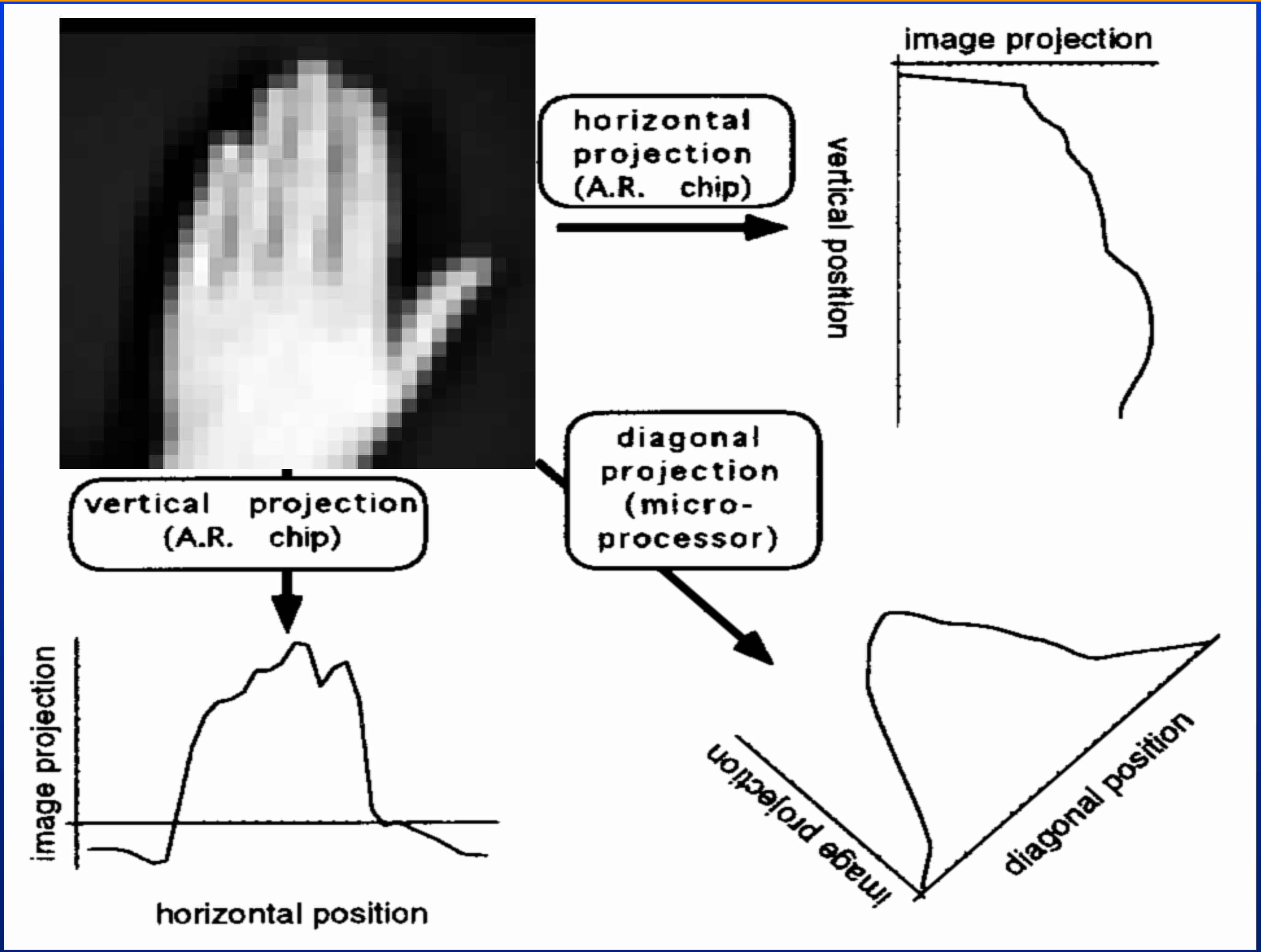



# Fast image moment calculation with artificial retina chip

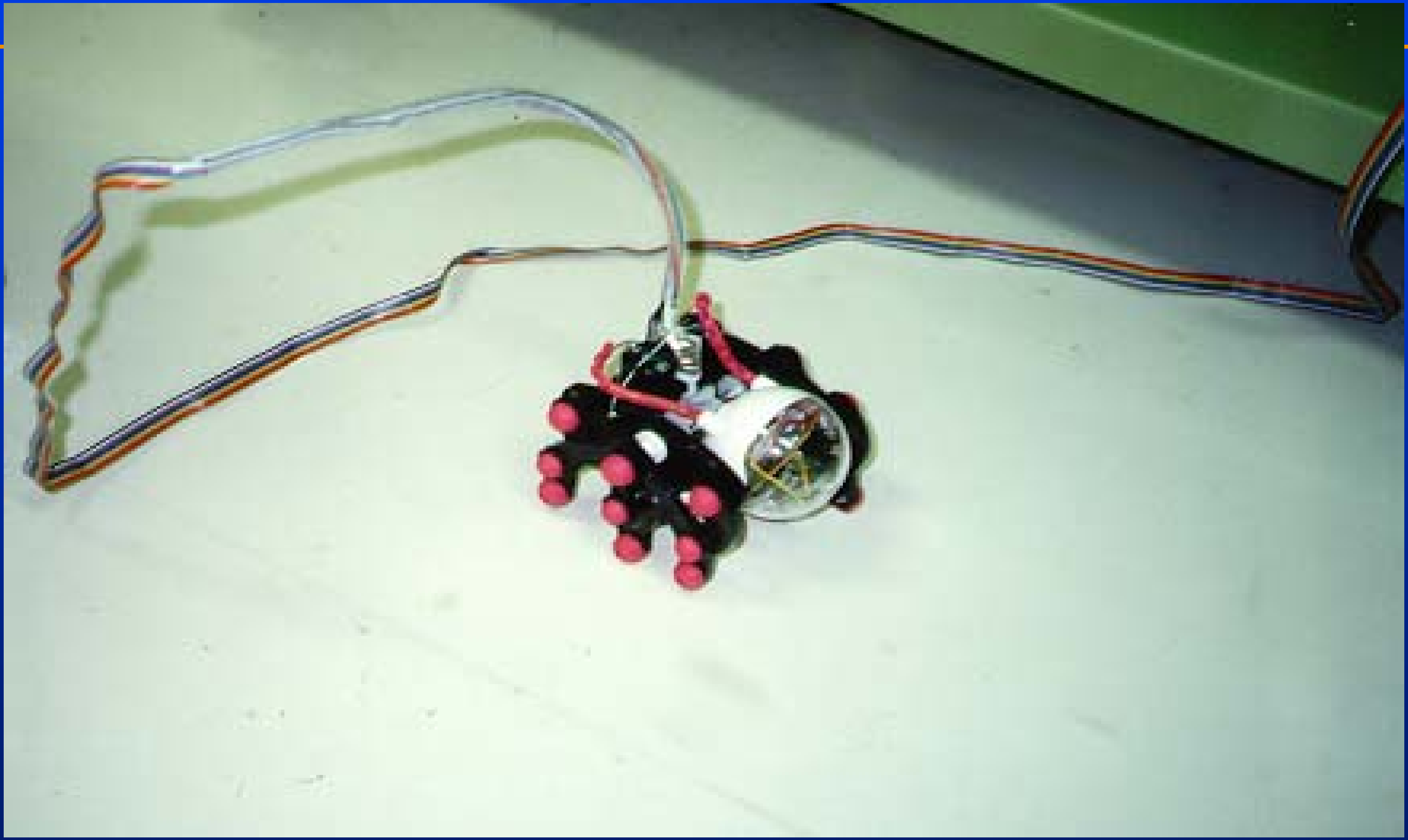
Processing time for image projections:

w/o AR chip: 10 msec

with AR chip: 0.3 msec



# Hand gesture-controlled robot

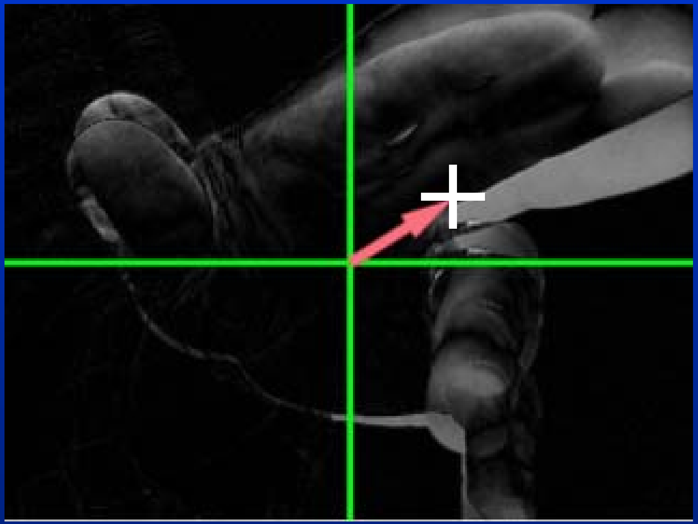


# Game: Nights

# Moment-based pointing control

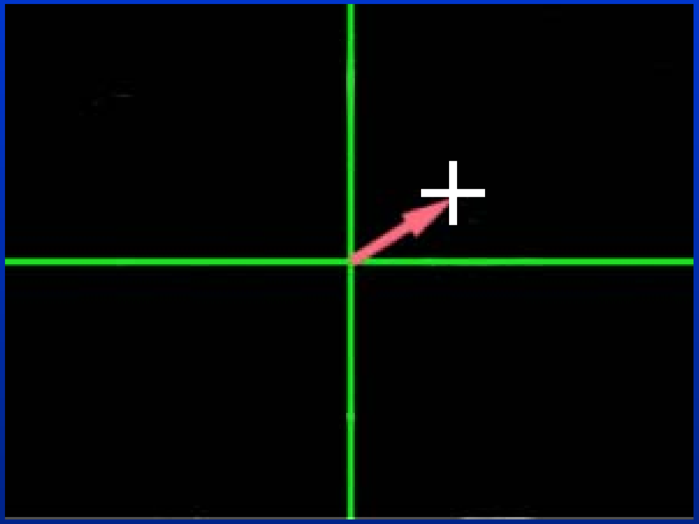
time 1

time 2



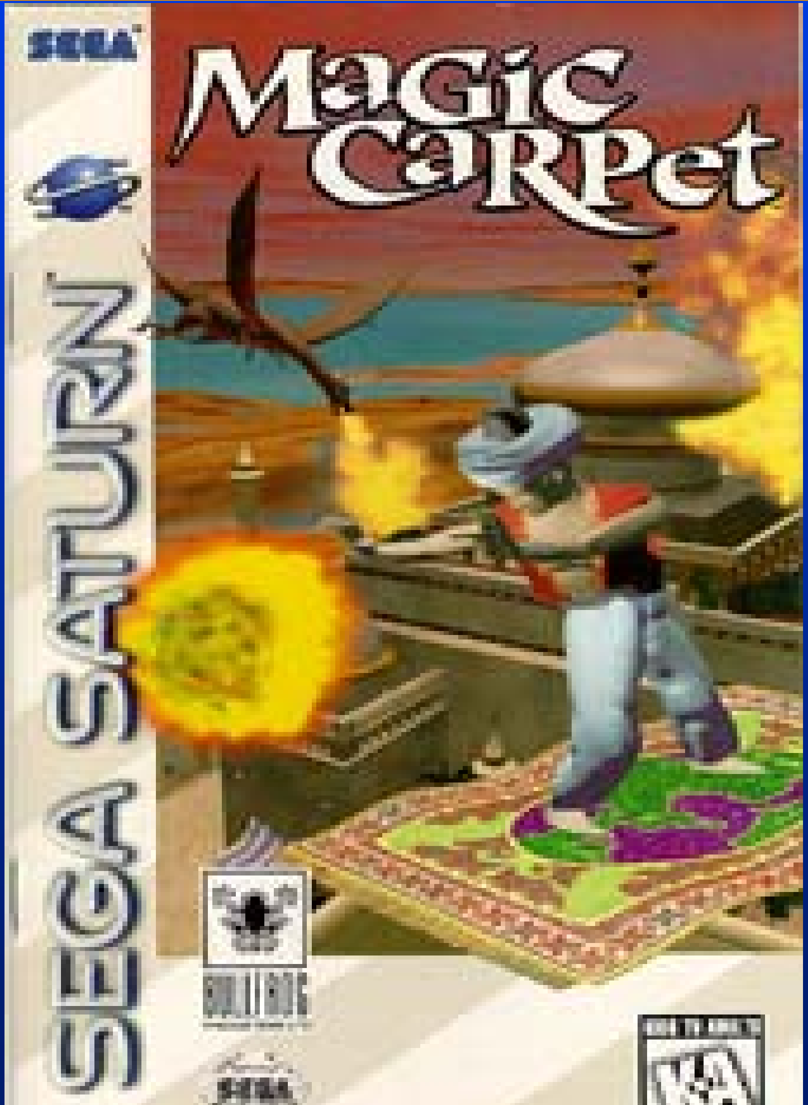
Center-of-mass of absolute value of difference-image

# Moment-based pointing control



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

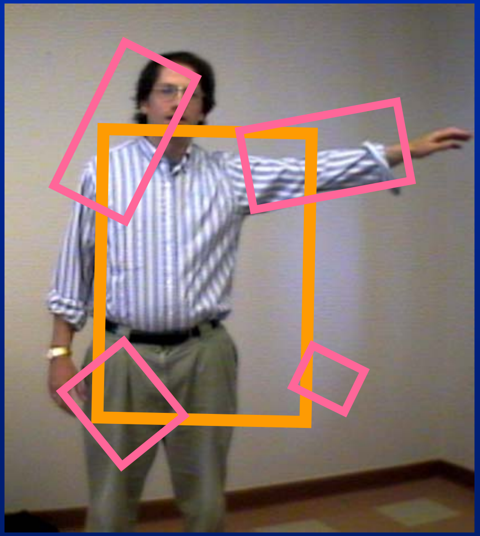
# Game:  Magic Carpet



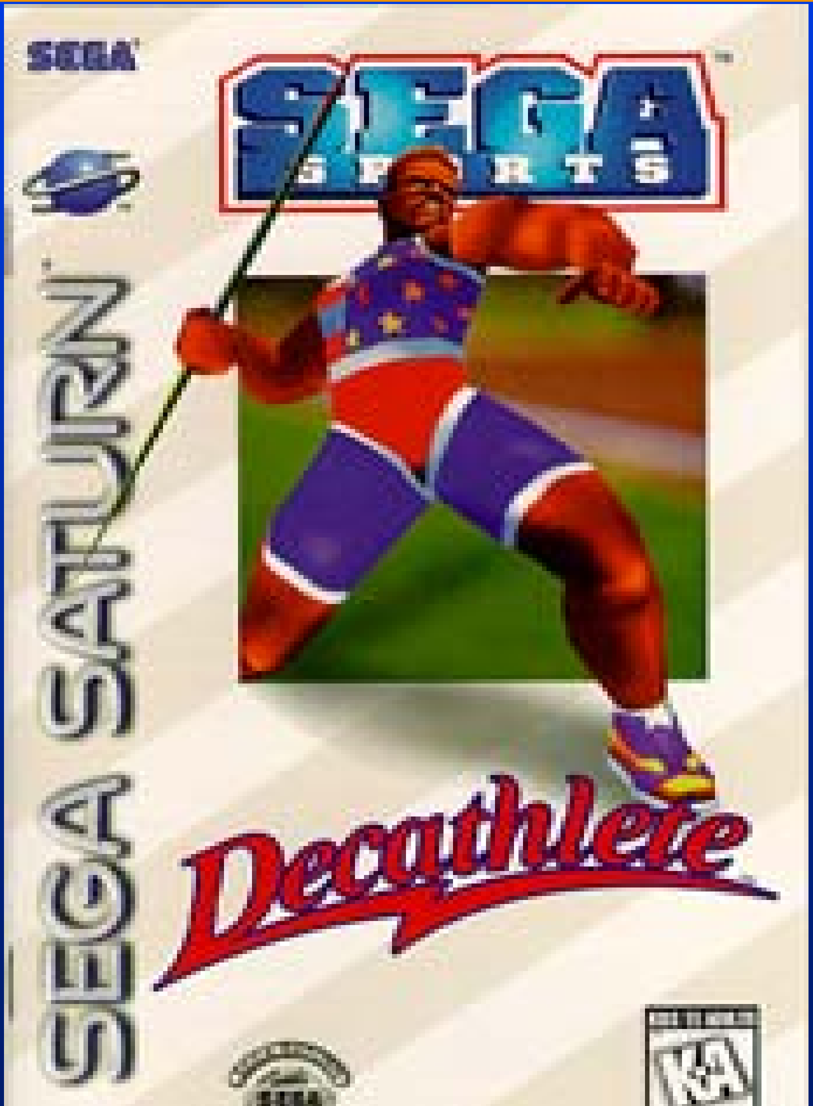
# Magic carpet game--figure analysis by hierarchical image moments



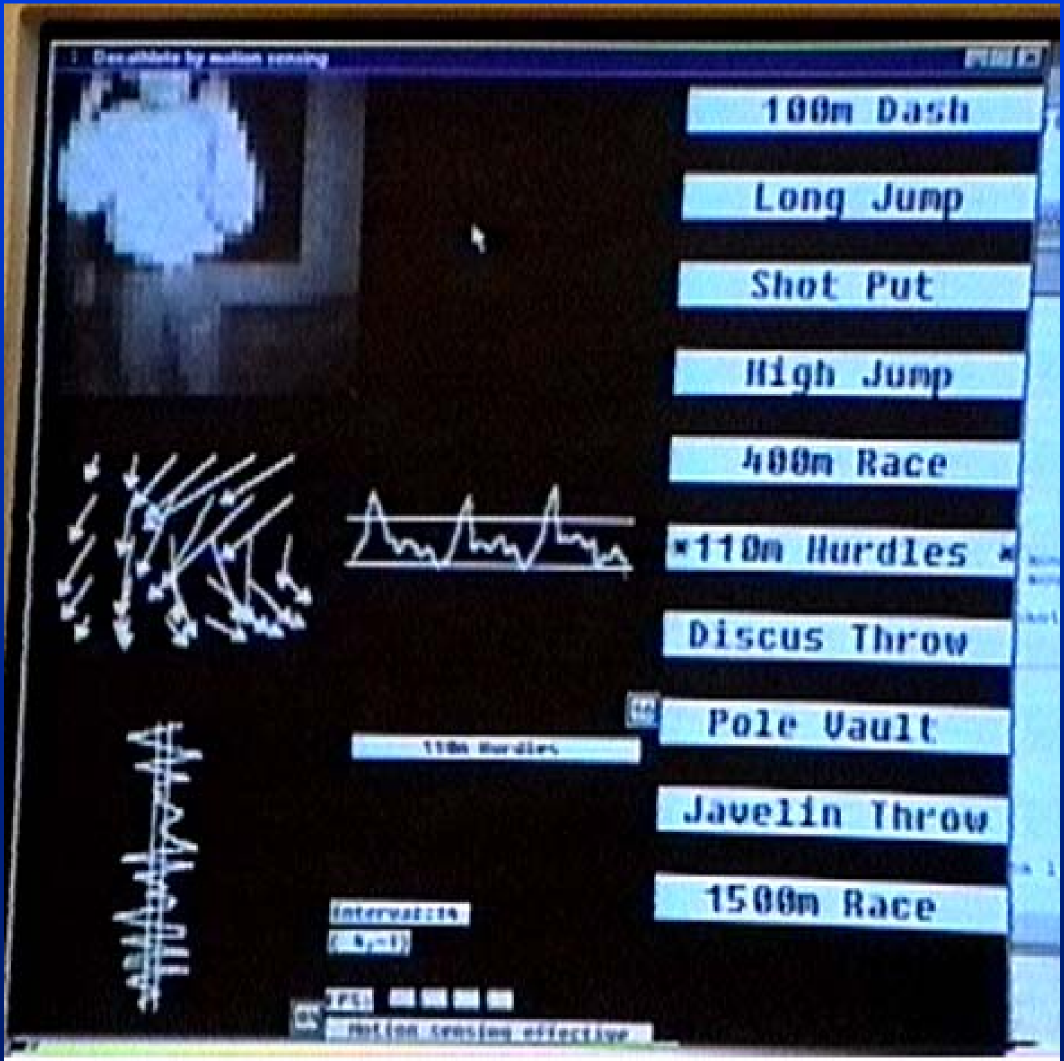


# Game:  Decathlete

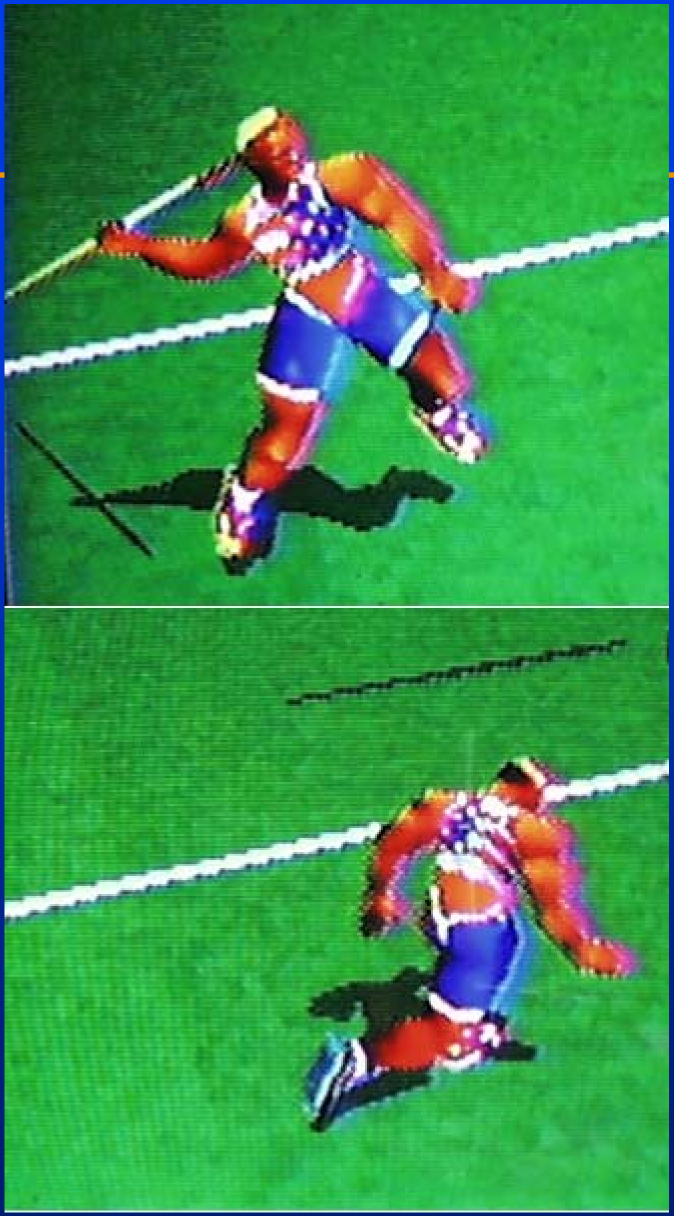


# Optical-flow-based Decathlete figure motion analysis



# Decathlete 100m hurdles

# Decathlete javelin throw



# Decathlete javelin throw

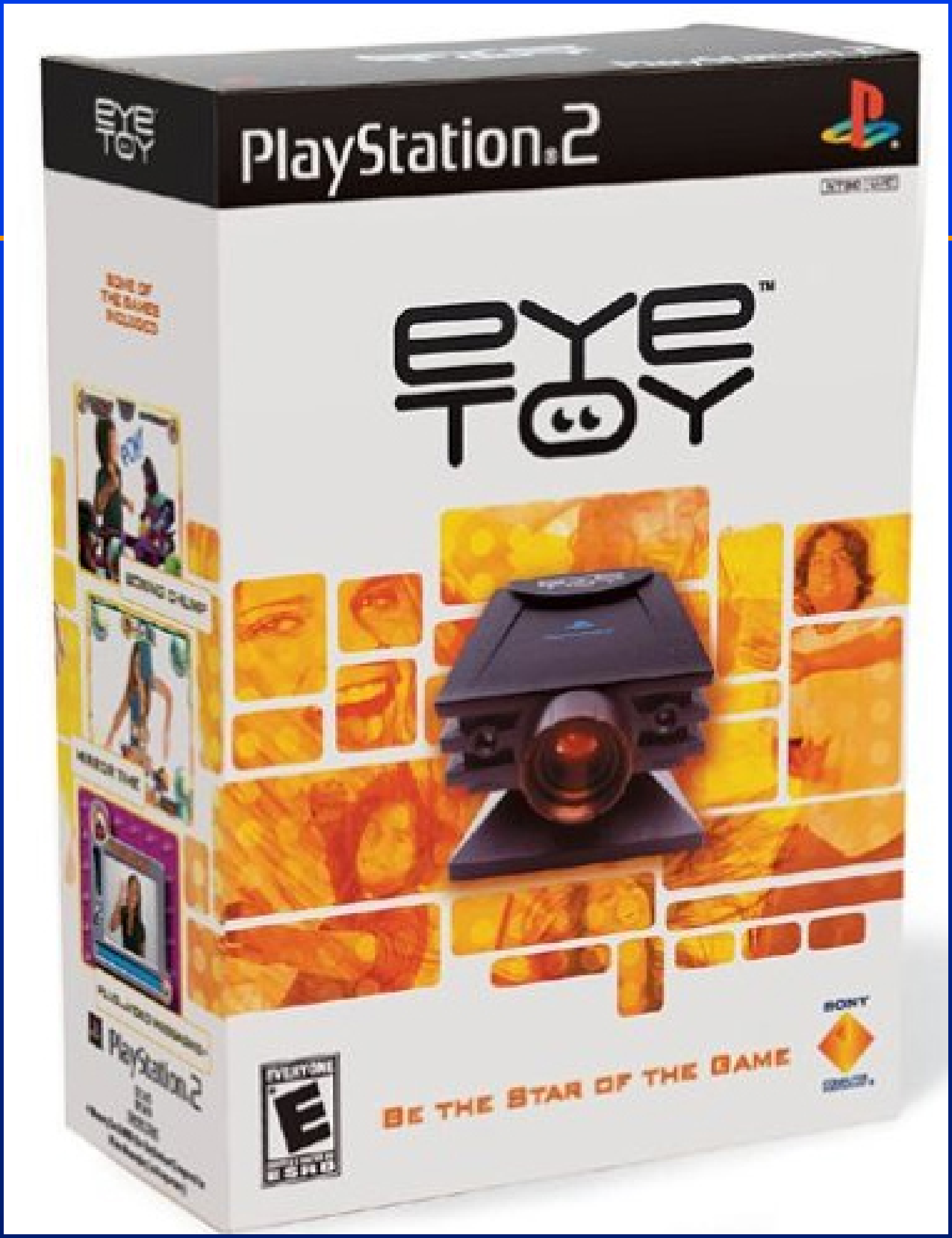
# video

# Nintendo Game Boy Camera

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

# video

# Sony ITOY



# Sony ITOY

# Sony ITOY

# Sony ITOY

## 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…