Vision-based User-interfaces for Pervasive Computing

CHI 2003 Tutorial Notes

Trevor Darrell
Vision Interface Group
MIT AI Lab

Table of contents

Biographical Sketch

Prof. Trevor Darrell leads the Vision Interface group at the MIT Artificial Intelligence Laboratory and has an appointment in the Electrical Engineering and Computer Science Department. Prior to joining the faculty of MIT in 1999, he worked as a member of research staff at the Interval Research Corp. in Palo Alto, CA. He received his PhD from the MIT Media Arts and Sciences Program in 1996. At the Media Lab he developed several interactive systems using real-time vision including the ALIVE system for interaction with virtual worlds, and systems for real-time hand gesture and facial expression recognition.

Agenda

14:00 Welcome and Overview
14:15 Responding to Faces
15:15 Tracking Hands and Gestures
16:00 Interacting with Arms
16:30 Brainstorming Activity
17:00 Privacy Issues
17:20 Conclusion

Perceptive User Interfaces

- Free users from desktop and wired interfaces
- Allow natural gesture and speech commands
- Give computers awareness of users
- Work in open and noisy environments

- Vision’s role: provide ***perceptual context***

Perceptual Context

- *Who is there? (presence, identity)*
- *Which person said that? (audiovisual grouping)*
- *Where are they? (location)*
- *What are they looking / pointing at? (pose, gaze)*
- *What are they doing? (activity)*

Perceptual context should be provided across platforms…*(PDA, Desktop, Environment)*

CV Face and Gesture Literature

- PUI Workshop
- FG Conferences
- CVPR, ICCV, ICPR….

Three metaphors

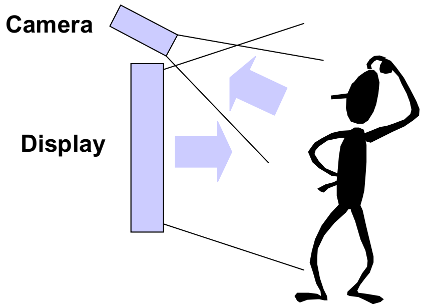
- Perceptual displays
- Gloveless hand tracking
- “Smart” environments

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

Perception-based manipulation

- “Gloveless” VR hand tracking
- Track hand position(s) in 3-D
- Usually desktop-based interface
- Control virtual character
- Navigation, etc.

e.g., Wren and Pentland, MIT Media Lab

Intelligent Environments

From PUI to Pervasive Computing…

- Integrate multiple perceptual algorithms.
- No single point of interaction (desktop/screen)

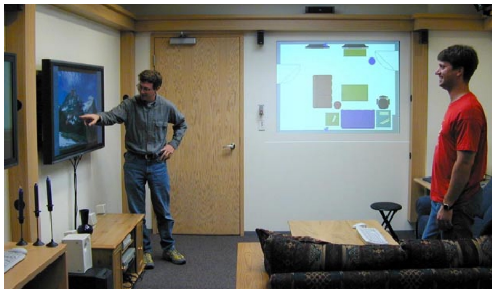
Offices & homes with

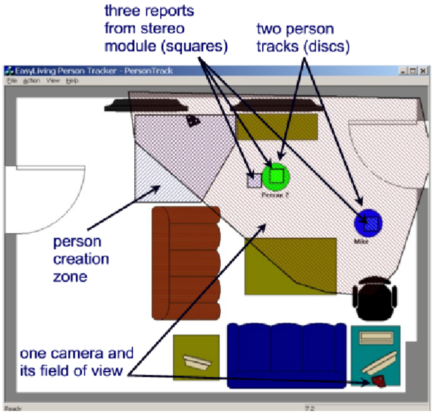
- Vision-based detection, ID, and tracking of occupants
- Speech interface to recognize commands and perform keyword indexing

Applications

- meeting recording; activity-dependent indexing
- active videoconferencing; presence; abstract messaging
- eldercare/childcare

Microsoft EasyLiving Project




three reports from stereo module (squares)
two person tracks (discs)
EasyLiving Person Tracker - PersonTrack
Person 1
Mike
person creation zone
one camera and its field of view

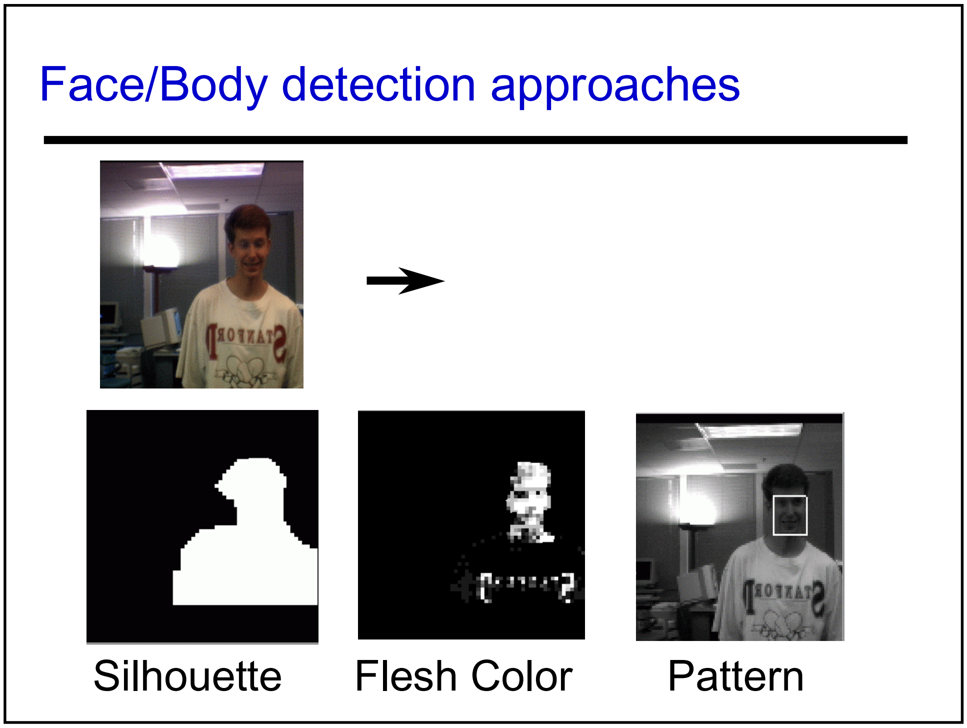
[Shafer, Brumitt, Krumm et al.]

Other smart environment projects

- GaTech AwareHome
- MIT AI Lab Intelligent Room
- MIT Media Lab Facilitator
- SRI

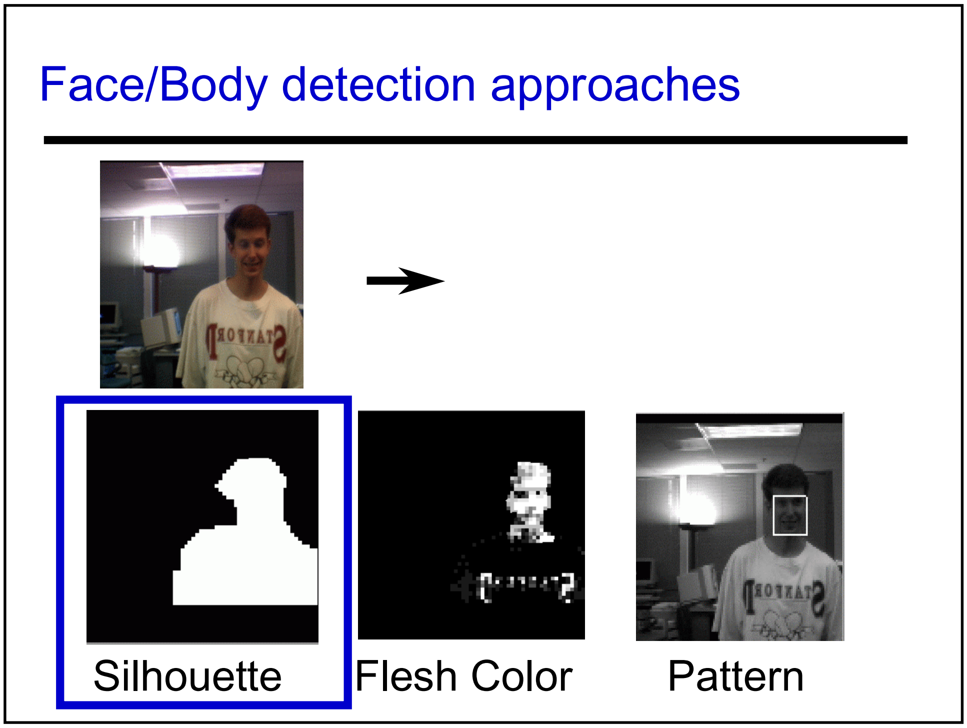
Today’s Topics

- Face Detection and Recognition
- Head Pose Estimation
- Eye Gaze Tracking
- Face Expression
- Hand Tracking
- Gesture Recognition
- Privacy Issues



Face/Body detection approaches

Silhouette Flesh Color Pattern



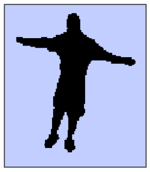
Face/Body detection approaches

Silhouette Flesh Color Pattern

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

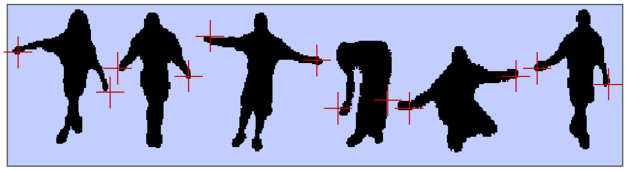
Given new image, generate silhouette by marking those pixels that are significantly different from the “background” value.



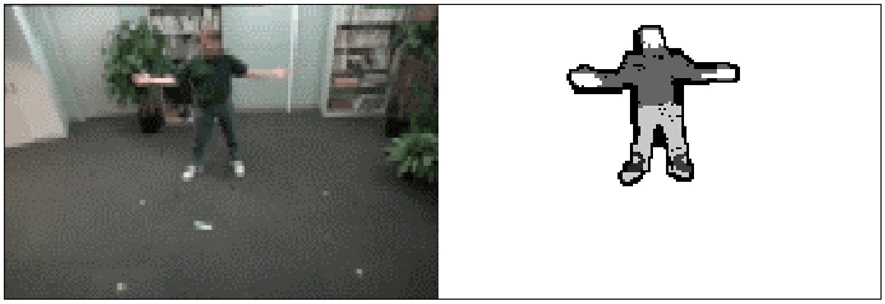
Finding Features

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)

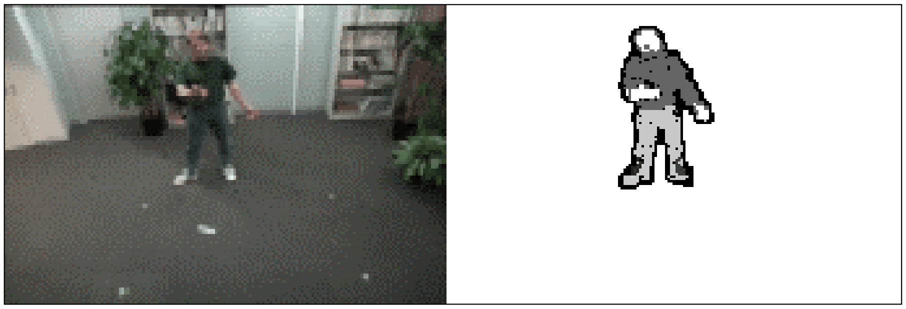


Static Background Modeling Examples



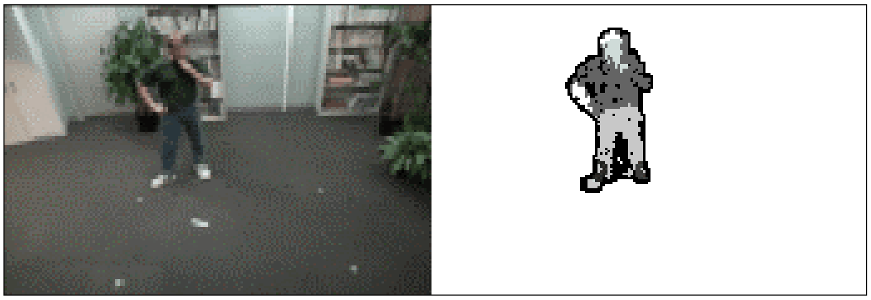
[MIT Media Lab Pfinder / ALIVE System]

Static Background Modeling Examples



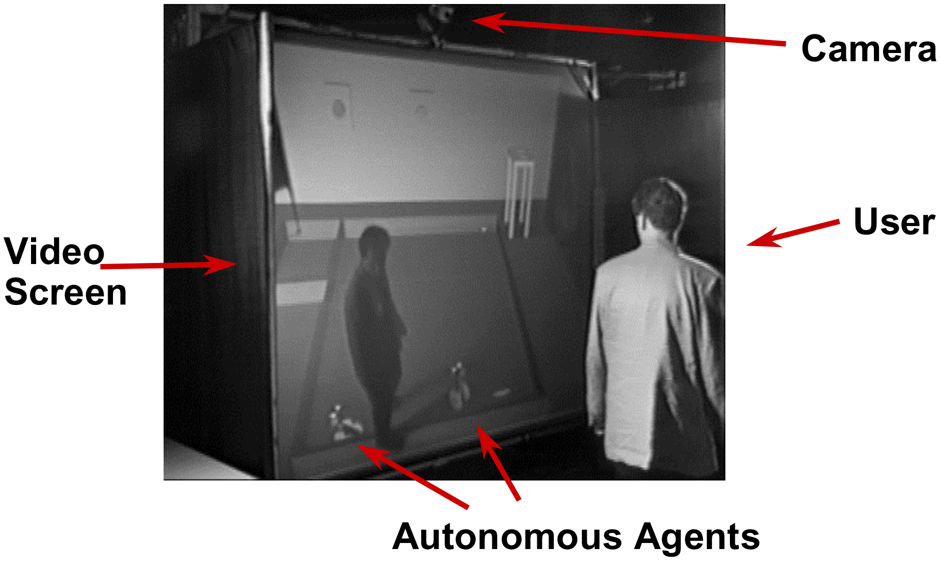
[MIT Media Lab Pfinder / ALIVE System]

Static Background Modeling Examples



[MIT Media Lab Pfinder / ALIVE System]

The ALIVE System



Camera

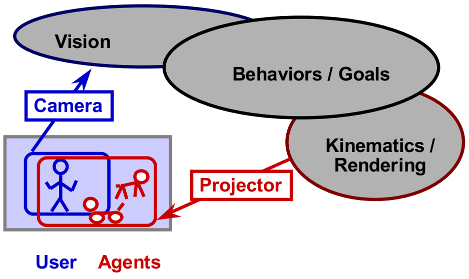
User

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]

General Background modeling

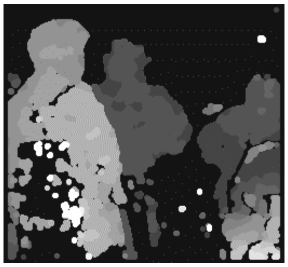
- Outdoor analysis; richer model of per-pixel background variation
  - MIT AI Lab VSAM project [Grimson and Stauffer]
  - UMD W4 project [Davis]
- Key assumption: static background
- How to deal with crowded environments and dynamic backgrounds?

Video-Rate Stereo

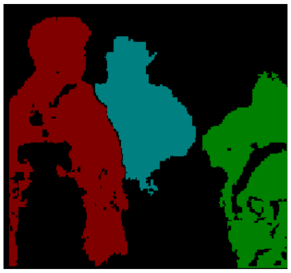
- Two cameras –> stereo range estimation; disparity proportional to depth
- Depth makes tracking people easy
  - segmentation
  - shape characterization
  - pose tracking
- Real-time implementations becoming commercially available

Stereo range estimation

Left and right images

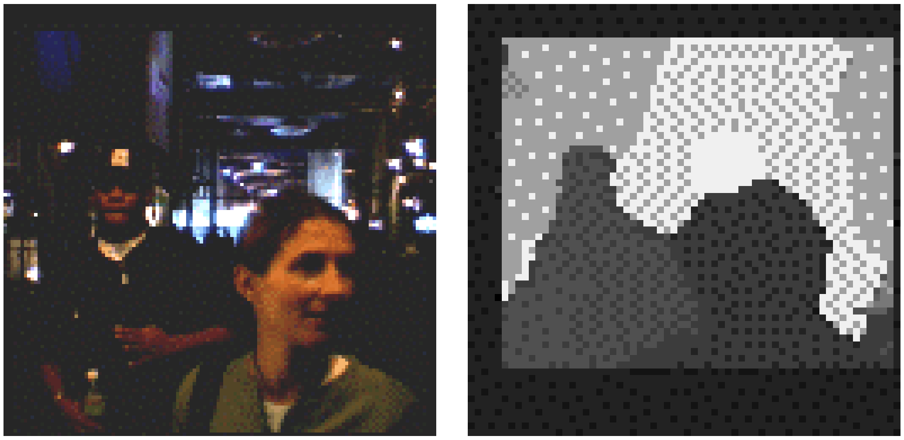


Computed disparity

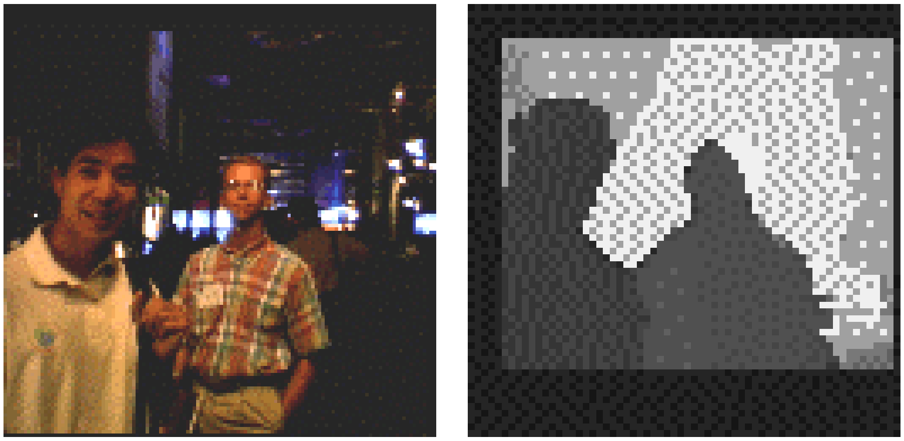


Grouping by local connectivity

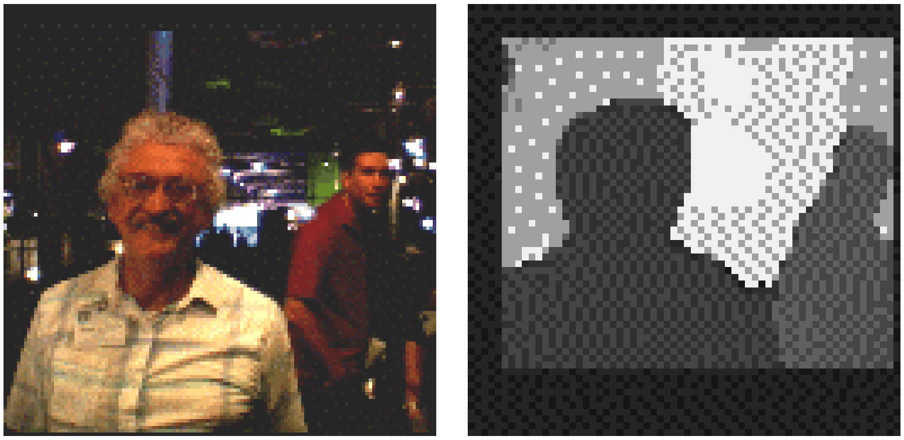
RGBZ input



RGBZ input

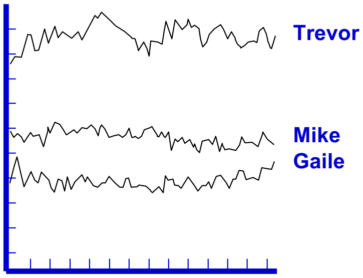


RGBZ input



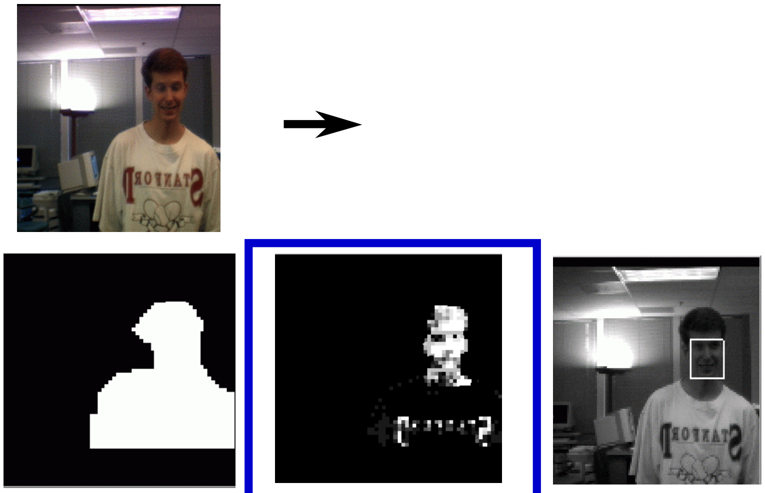
Range feature for ID

- 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.

Face/Body detection approaches



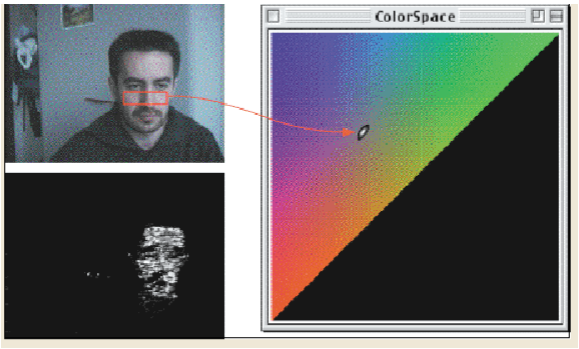
Silhouette

Flesh Color

Pattern

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.]

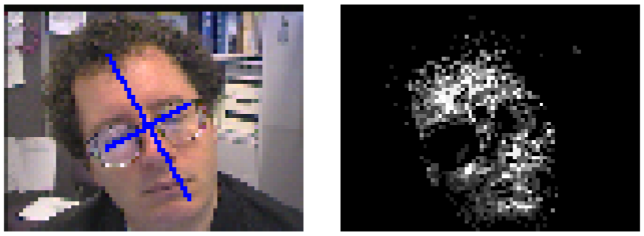
(log(r) - log(g), log(b) - log((r+g)/2))

Hue & Saturation

Flesh color tracking

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

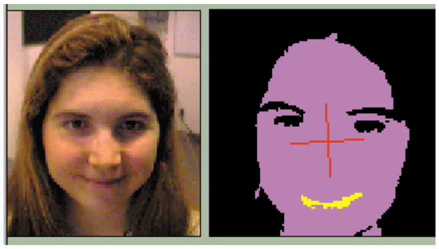
(open source impl. avail.!)



[Bradsky, Intel]

Flesh color tracking

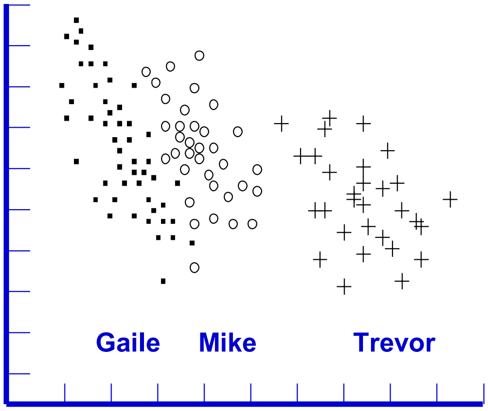
MIT Media Lab's Lafter--simultaneous face and lip hue tracking.



[Oliver and Pentland]

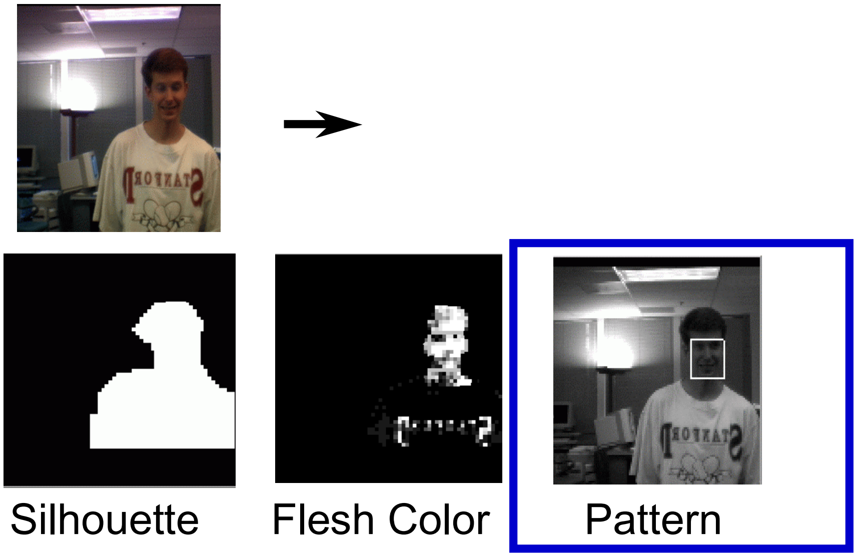
Color feature for ID

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

Face/Body detection approaches



Silhouette

Flesh Color

Pattern

Pattern Recognition

Face Detection:

Determine location and size of any human face in input image given greyscale patch. (2-class)

[Sung and Poggio; Rowley and Kanade]

Face Recognition:

Compare input face image against models in library, report best match. (n-class)

[Turk and Pentland; Cootes and Taylor; and many others…]

Pattern Recognition

Face Detection:

Determine location and size of any human face in input image given greyscale patch. (2-class)

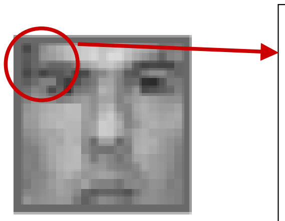
[Sung and Poggio; Rowley and Kanade]

Face Recognition:

Compare input face image against models in library, report best match. (n-class)

[Turk and Pentland; Cootes and Taylor; and many others…]

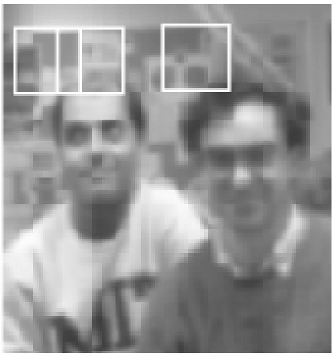
Image Basics



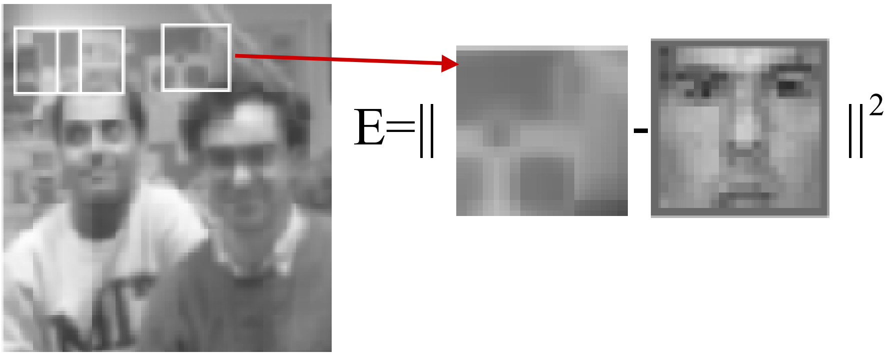
```
35 39 45 68 88 …

36 43 62 43 55 …

33 43 55 52 51 …

…
```

Template Matching

- Classic approach
- Given input image, compare template image at various offsets
- Various distance metrics
  - MSE
  - Correllation

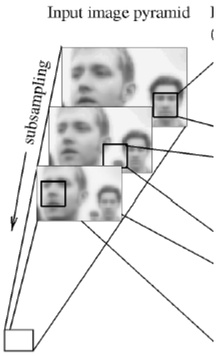


Template Matching



Multi-scale search

- Search at multiple scales (and pose)
- Multiple templates
- Single template, multiple scales
- Image Pyramid
  - decimate image by constant factor
  - efficient search



Template Matching

- Works for single (or similar) individuals, cannonical pose and lighting.

- Common extensions
  - prenormalization
  - Multiple templates
  - Subfeatures
  - How to choose?