Smart Rooms

Presented by Paul DeWitt

April 22, 2003

About this paper

- Author: Alex P. Pentland, Head of Perceptual Computing Section at the Media Lab of MIT
- Published in Scientific American, April 1996

Why Smart Rooms?

- Currently, computers are deaf and blind, they only experience multimedia through strings of data without context
- To be truly helpful, computers need to see and hear, and to be able to recognize people in order to make sense of what we are thinking, so the computer can react as desired
- This paper presents a combination of technologies to create an environment to do just that. The system is called a “smart room”

Components of a Smart Room

- Cameras
- Microphones
- Network of Computers
  - A separate computer for each analysis needed
  - Each module uses maximum likelihood analysis to match observed input to stored models and takes the closest match

Locating People

- Pfinder is used to track a person by:
  1. Recording the motions with a camera
  2. Segmenting the image into blobs
  3. Using known data to build a model

When Pfinder sees a new picture

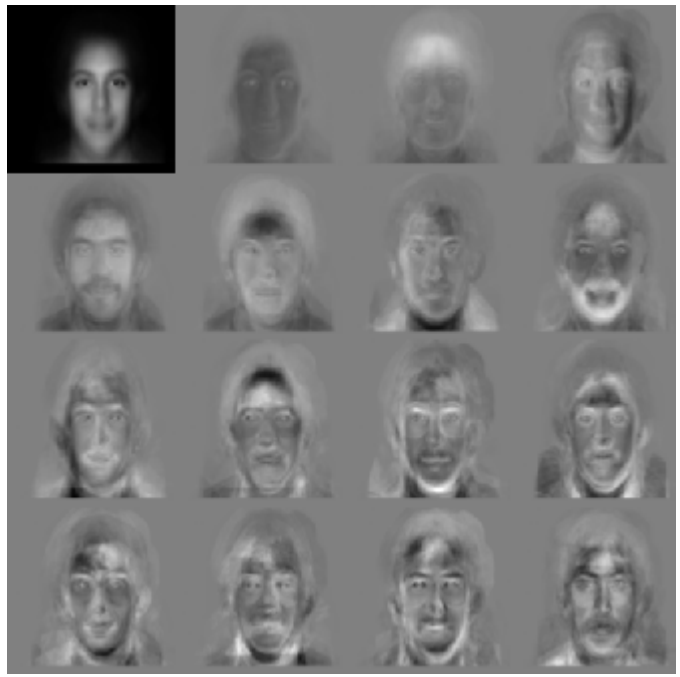
- First, Pfinder guesses what the blob model should look like, using previous motion and typical movement patterns
- Next Pfinder measures the chance that each pixel belongs to each blob by comparing the pixel's color and brightness to the blob's average color and brightness
- Pfinder compensates for changes in the room by continuously updating the background picture

Understanding Speech

- Typical speech recognition algorithms require the user to wear a microphone or to sit near one
- This doesn't make the room seem very smart
- To compensate, the smart room uses the Pfinder's location information to direct the microphones to pick up the sounds from the user's mouth

Identifying Users

- To recognize a user, the Smart Room uses Eigenfaces with 99% accuracy
- The system uses Pfinder to extract the face and normalizes it
- Models the face based on its similarity to each Eigenface
- Compares this to the simularity of known faces
- If it is close, assume the face is recognized



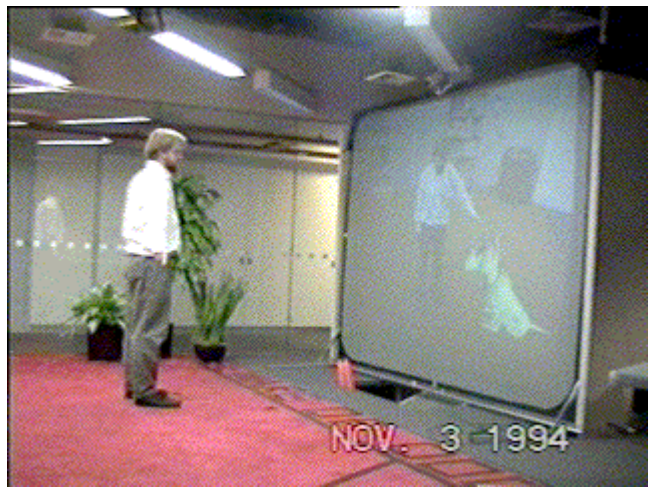
Typical Eigenfaces

Identifying Expressions

- The Smart Room system also identifies expressions by analyzing facial muscle movement with 98% accuracy

Adding Context

- It is useful to place motions into context
- To determine intentions, Smart Rooms use a generalization of the phoneme technique for voice recognition
  - Ex. Artificial Life Interactive Video Environment (ALIVE)



ALIVE uses the smart room's ability to recognize intent to allow the user to play with a virtual dog. The user can throw a virtual ball and the dog will retrieve it.

Uses for smart-room technology

- Controls for a virtual reality environment
- Has been used to translate a 40-word subset of American Sign Language in real time with 99.2% accuracy, could be extended to more words
- Smart Cars could monitor drivers to provide feedback
  - A test of this concept monitored the actions taken when passing, following, turning, stopping, accelerating, and changing lanes, and was able to identify the action within .5 seconds from the start 86% of the time and in 2 seconds, 97% accurate

Uses for smart-room technology(cont.)

- Potentially usable in smart clothing, such as eyeglasses that could recognize people encountered and whisper their names in the user's ear
- Could be used in credit cards that can recognize their owners and so could report themselves stolen
- Further study into human actions would allow a system to identify when students or drivers stop paying attention
- Smart-room technology could be used in sports to allow the camera to identify a play and follow the action

Strengths and Weaknesses

- Pros
  - Uses existing technology, including typical computers, to achieve real time analysis
  - Modular design allows new capabilities to be added as they are developed, and allows them to take advantage of information obtained through the other modules

- Cons
  - The system (particularly Pfinder) seems to only support a single user in the room
  - Scalability? The examples presented in the paper typically only dealt with a small selection of motions to analyze. As more motions are added it would get harder to differentiate between them, and it is not totally obvious that this system can be extended indefinitely.

Summary

- Smart rooms consisting of Pfinder, voice recognition, Eigenface based face recognition, methods of expression and context determination, all based on maximum likelihood analysis can be used to provide a mechanism for allowing a user to interact with a computer without relying on a mouse or keyboard.

Related Works

- http://www-white.media.mit.edu/vismod/demos/smartroom/
- http://www-white.media.mit.edu/vismod/demos/pfinder/
- http://www.media.mit.edu/wearables/
- http://www-white.media.mit.edu/vismod/demos/facerec/index.html

Any Questions?