Lecture-18

Monitoring Human Behavior

www.cs.ucf.edu/~ayers

Goals of the System

- Recognize human actions in a room for which **prior knowledge** is available.
- Handle multiple people
- Provide a textual description of each action
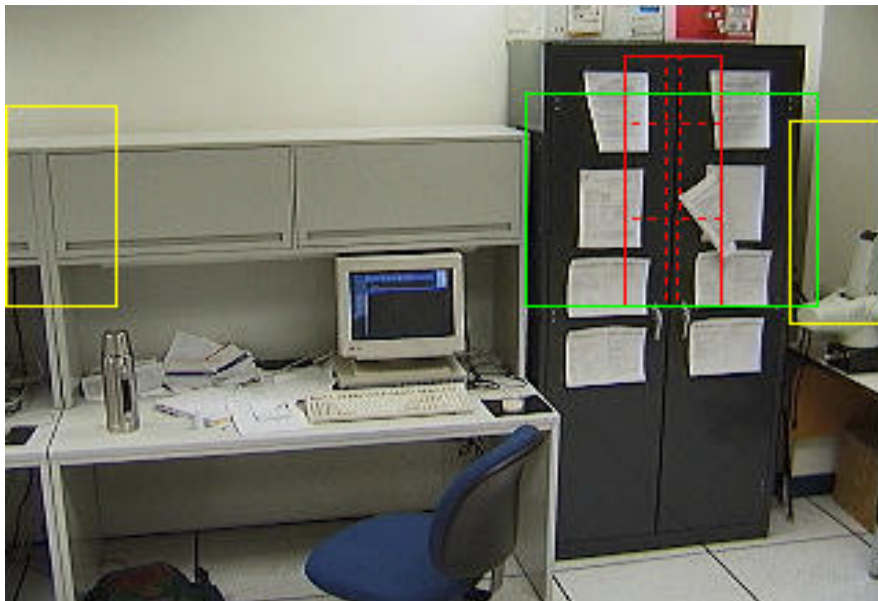- Extract "key frames" for each action

Possible Actions

- **Enter**
- **Leave**
- **Sitting** or **Standing**
- **Picking Up Object**
- **Put Down Object**
- **…..**

Prior Knowledge

- Spatial layout of the scene:
  - Location of **entrances** and **exits**
  - Location of **objects** and some information about how they are use
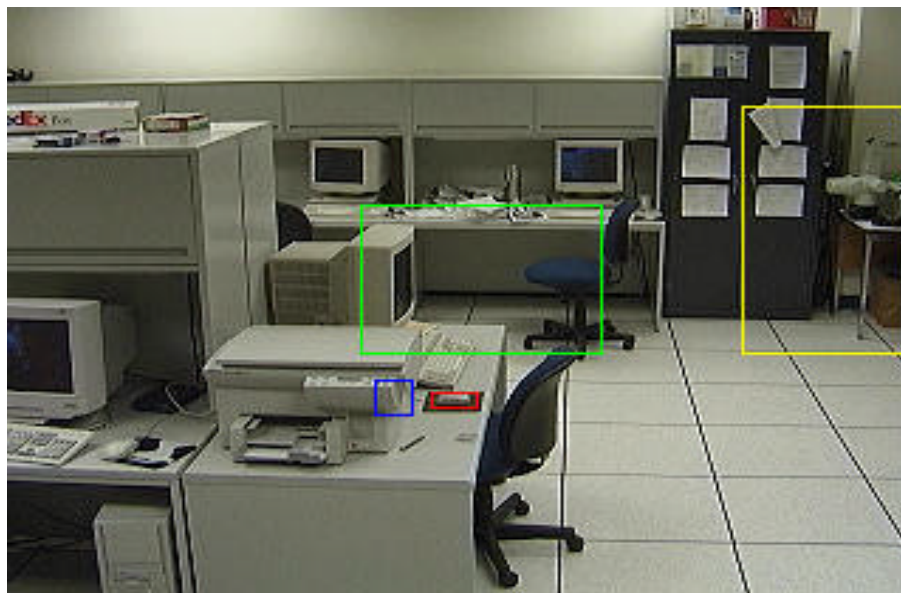- Context can then be used to improve recognition and save computation
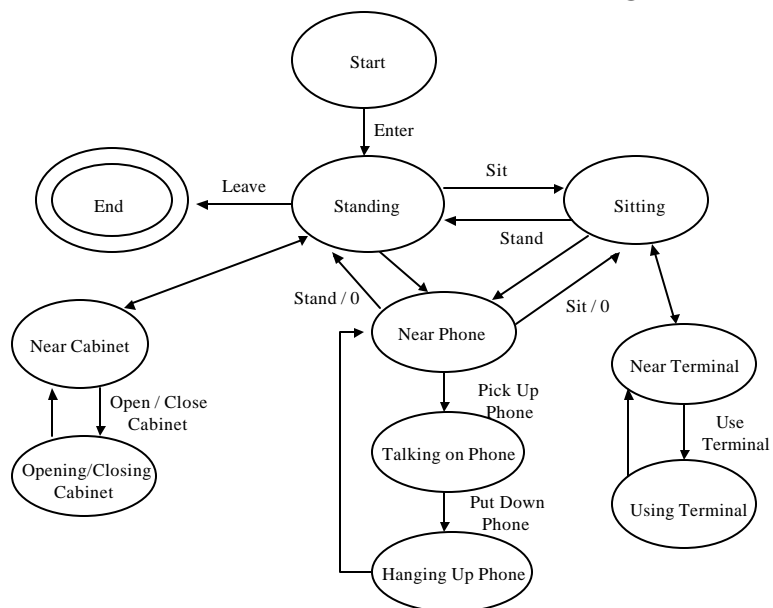
Layout of Scene 1
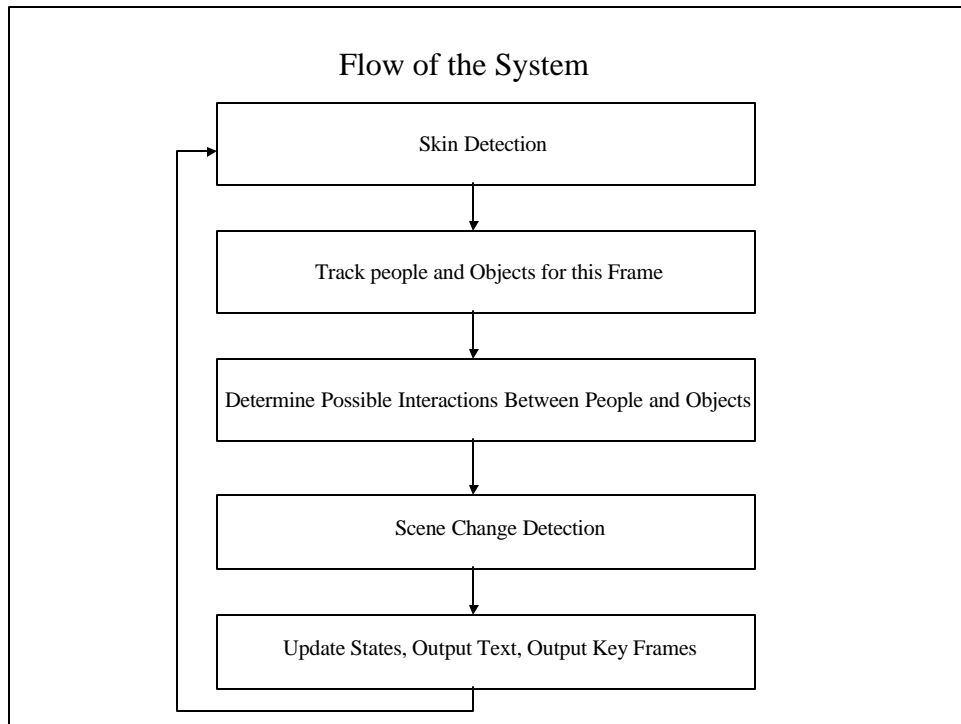
Layout of Scene 2

Layout of Scene 4

Major Components

- Skin Detection
- Tracking
- Scene Change Detection
- Action Recognition

State Model For Action Recognition

Flow of the System

```
          ┌─────────────────────────────────────────────────────────┐
      ┌──→ │                     Skin Detection                      │
      │    └─────────────────────────────────────────────────────────┘
      │                               │
      │                               ▼
      │    ┌─────────────────────────────────────────────────────────┐
      │    │            Track people and Objects for this Frame       │
      │    └─────────────────────────────────────────────────────────┘
      │                               │
      │                               ▼
      │    ┌─────────────────────────────────────────────────────────┐
      │    │   Determine Possible Interactions Between People and Objects │
      │    └─────────────────────────────────────────────────────────┘
      │                               │
      │                               ▼
      │    ┌─────────────────────────────────────────────────────────┐
      │    │                  Scene Change Detection                 │
      │    └─────────────────────────────────────────────────────────┘
      │                               │
      │                               ▼
      │    ┌─────────────────────────────────────────────────────────┐
      │    │       Update States, Output Text, Output Key Frames     │
      │    └─────────────────────────────────────────────────────────┘
      │                               │
      └───────────────────────────────┘
```

Key Frames

- Why get key frames?
  - Key frames take less space to store
  - Key frames take less time to transmit
  - Key frames can be viewed more quickly
- We use heuristics to determine when key frames are taken
  - Some are taken before the action occurs
  - Some are taken after the action occurs

Key Frames

- "Enter" key frames: as the person leaves the entrance/exit area
- "Leave" key frames: as the person enters the entrance/exit area
- "Standing/Sitting" key frames: after the tracking box has stopped moving up or down respectively
- "Open/Close" key frames: when the % of changed pixels stabilizes

Key Frames Sequence 1 (350 frames), Part 1
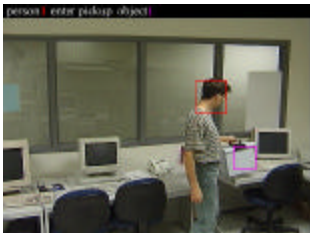

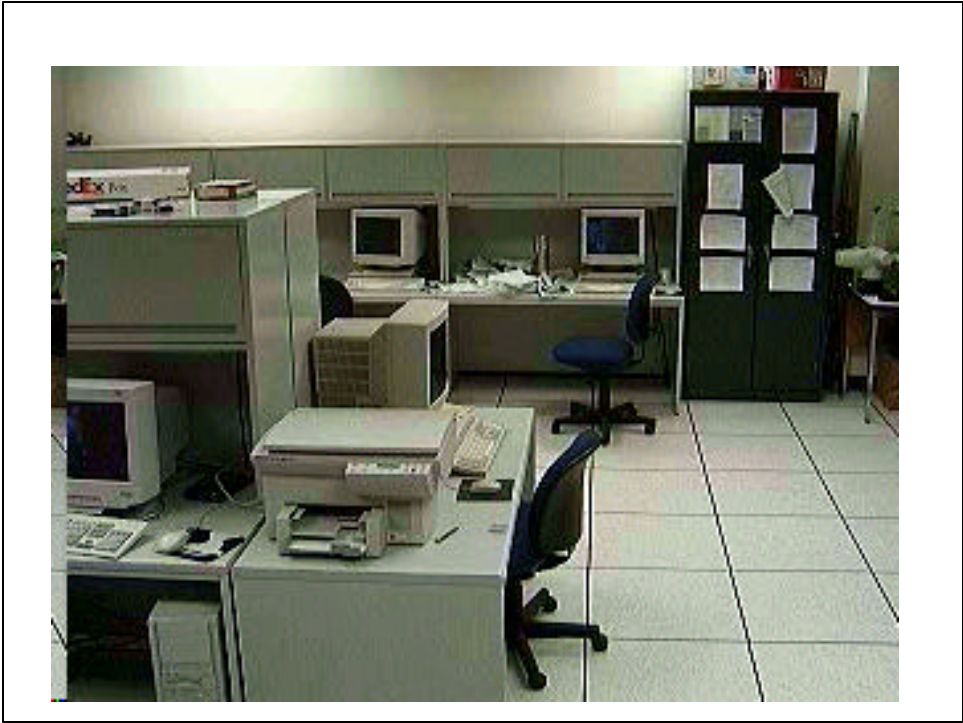# Key Frames Sequence 1 (350 frames), Part 2

Key Frames Sequence 2 (200 frames)
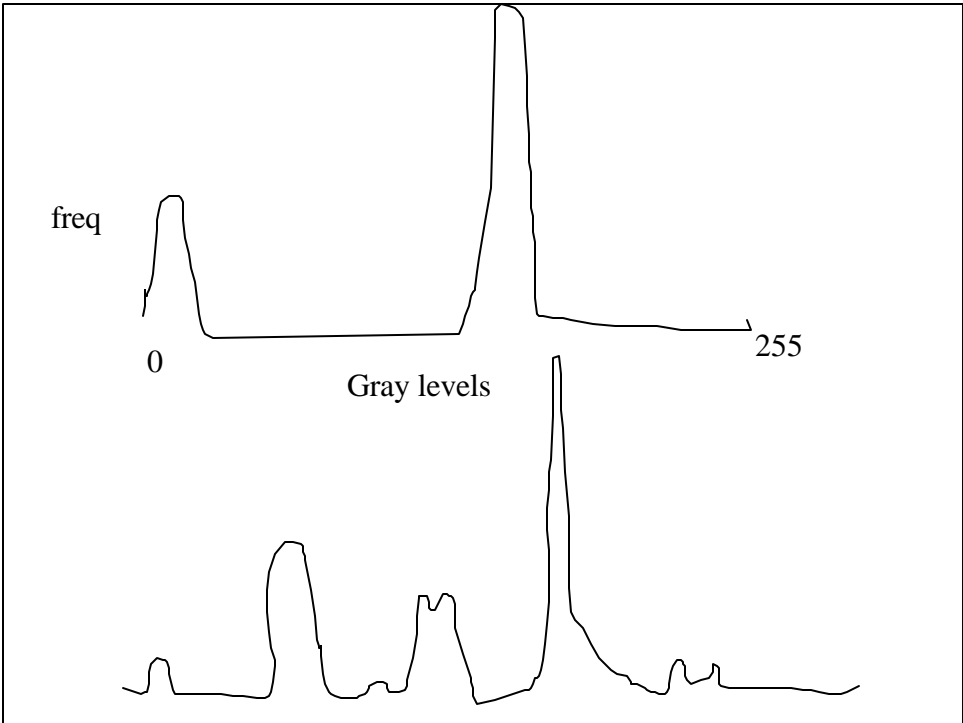
Key Frames Sequence 3 (200 frames)

Key Frames Sequence 4 (399 frames), Part 1

Key Frames Sequence 4 (399 frames), Part 2

freq

0

Gray levels

255

13

Results of skin detection

http://www.cs.ucf.edu/~rcen/icarcv_Rao_Shah.pdf

Detecting Fire