People detection and tracking using stereo vision and color

Rafael Munoz-Salinas, Eugenio Aguirre, Miguel Garcia-Silvente.

In *Image and Vision Computing* Volume 25 Issue 6 (2007) 995-1007.

Presented by Reid Sawtell

Introduction

The system presented by this paper is designed to take advantage of newer stereo cameras to improve person detection and tracking. Additionally, color information is used to create a more robust tracking system. Finally, all of this must be done quickly so the system can be used in real time applications.

Overview

- Design Specifications
- Building the Environment
  - Stereo Processing
  - Background
  - Foreground
- Detecting and Tracking People
  - Division of Tasks
  - Color Modeling
  - Detecting People
  - Tracking People
- Experimental Results
- Future Work

Design Specifications

- Stereo camera system (instead of the popular single camera system)
- Color information is used to improve tracking reliability
- Under-head camera position

Why Stereo?

- The price has decreased and availability has increased
- The added depth information allows for more accurate detection and tracking
  - Foreground objects can be separated from background objects
  - Can filter objects to reduce false positives
  - Added filtering also reduces the need to apply costly face detection

Why use color?

- Systems based only on position can easily confuse objects in close proximity
- Luminance can vary greatly from one location to another

Why use an under-head camera?

- Camera position is problem dependent
- It can be mounted on mobile platforms
- It is ideal for gesture recognition
- Smaller robots are less intimidating

Building the Environment

- Stereo Processing
- Background Modeling
- Foreground Extraction

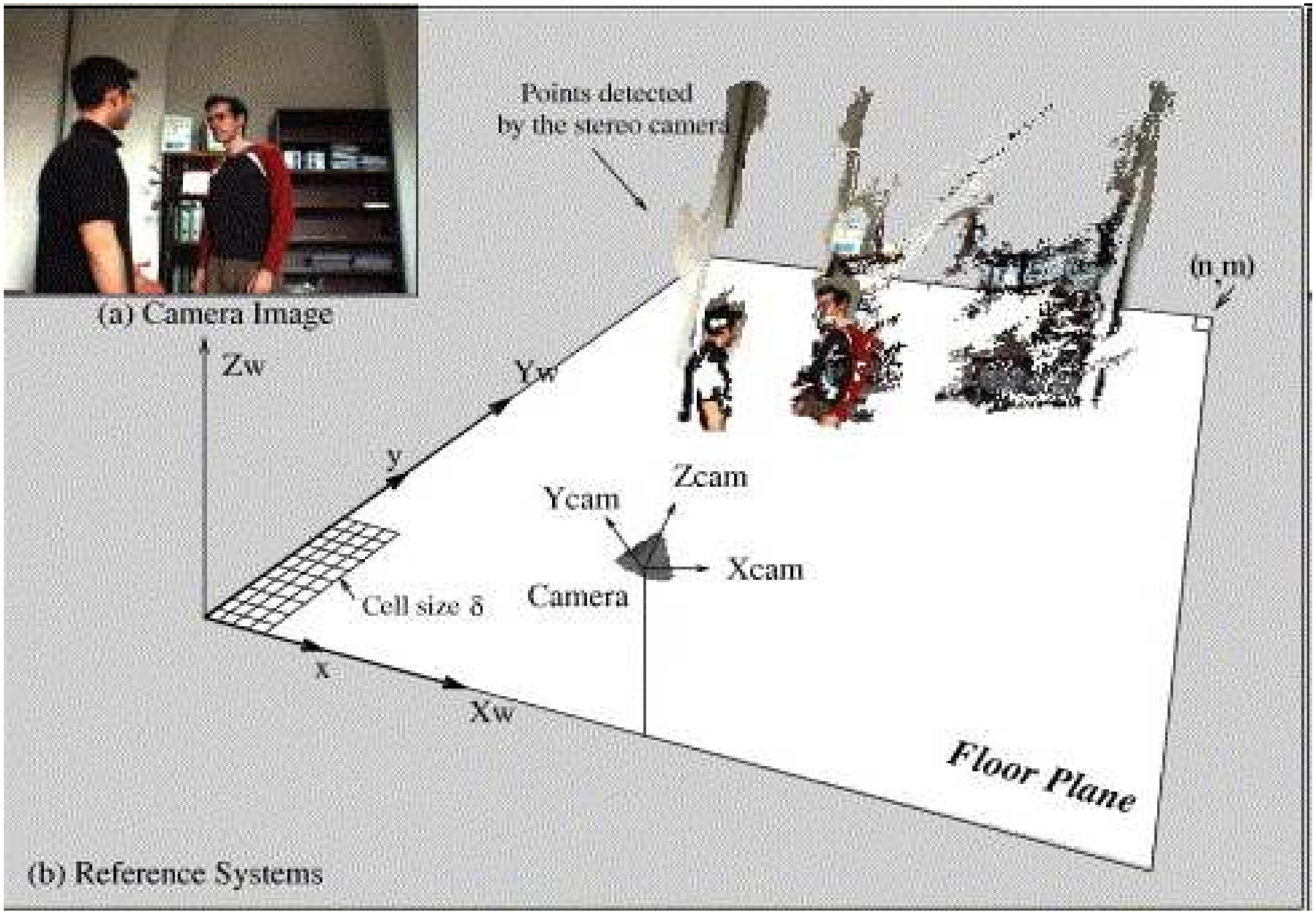
Stereo Processing

- Stereo System provides a disparity map
- Pixel location is calculated using known parameters
- Points are translated into a world model, which is likely to contain large volumes of data

Simplifying the Model

- 3D data is projected onto a 2D surface composed of uniform cells with height information
- Cell size must be carefully selected to balance memory/performance and reliability
- Filter points based on height to exclude ceiling or floor.

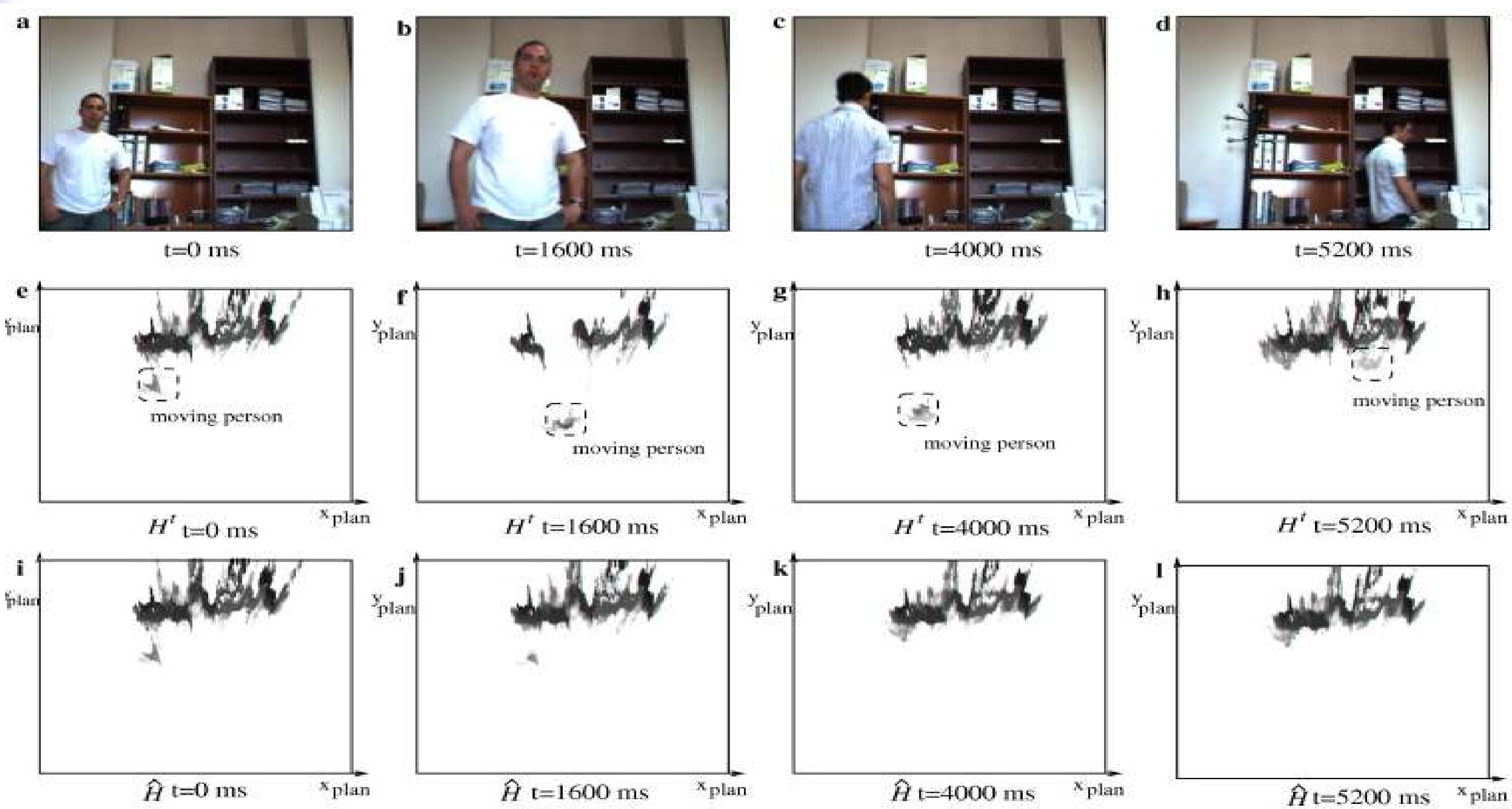
Scene reconstruction



Background Modeling

- People move, so separating them from the background is advantageous
- The heightmap is a model of the environment people move through
- Generating a reliable heightmap cannot be done instantaneously
- Must be updated more frequently than people are detected and tracked

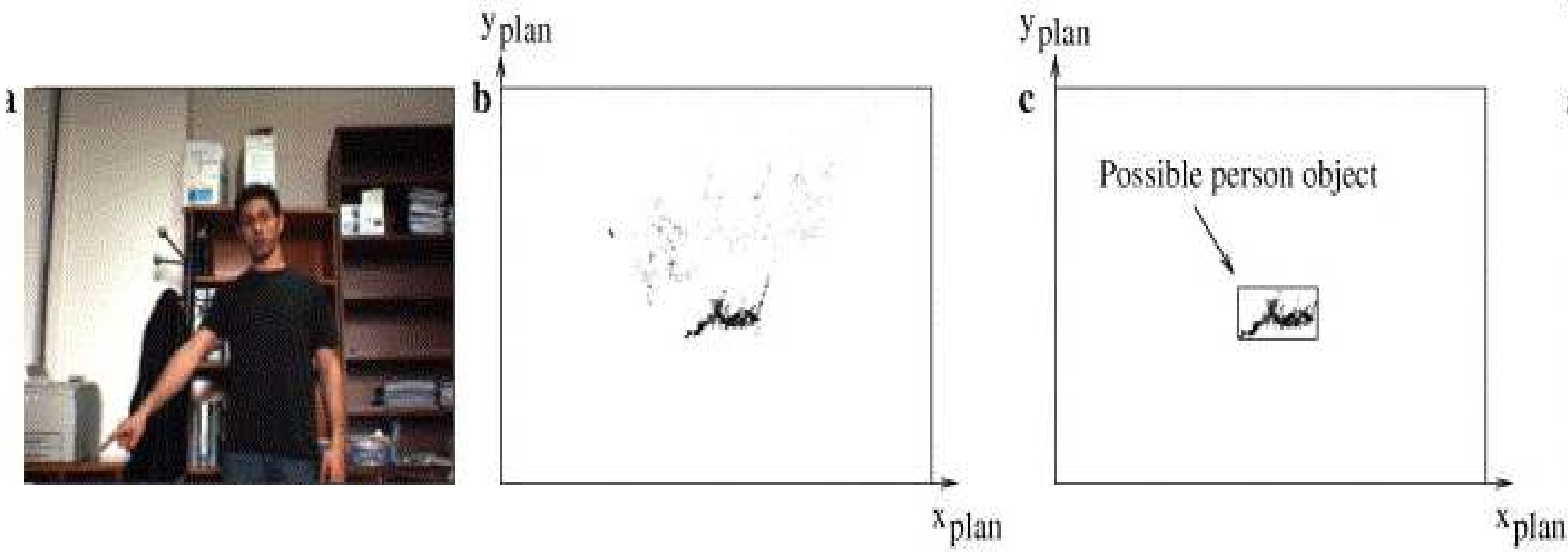
Generating the heightmap over time



Detecting Foreground Objects

- Foreground points are easily isolated using the heightmap
- Occupancy map shows the distribution of foreground points over the scene
- Points are scaled based on proximity to camera
- A person is likely to be found in areas with high occupancy

Occupancy Map



Detecting and Tracking People

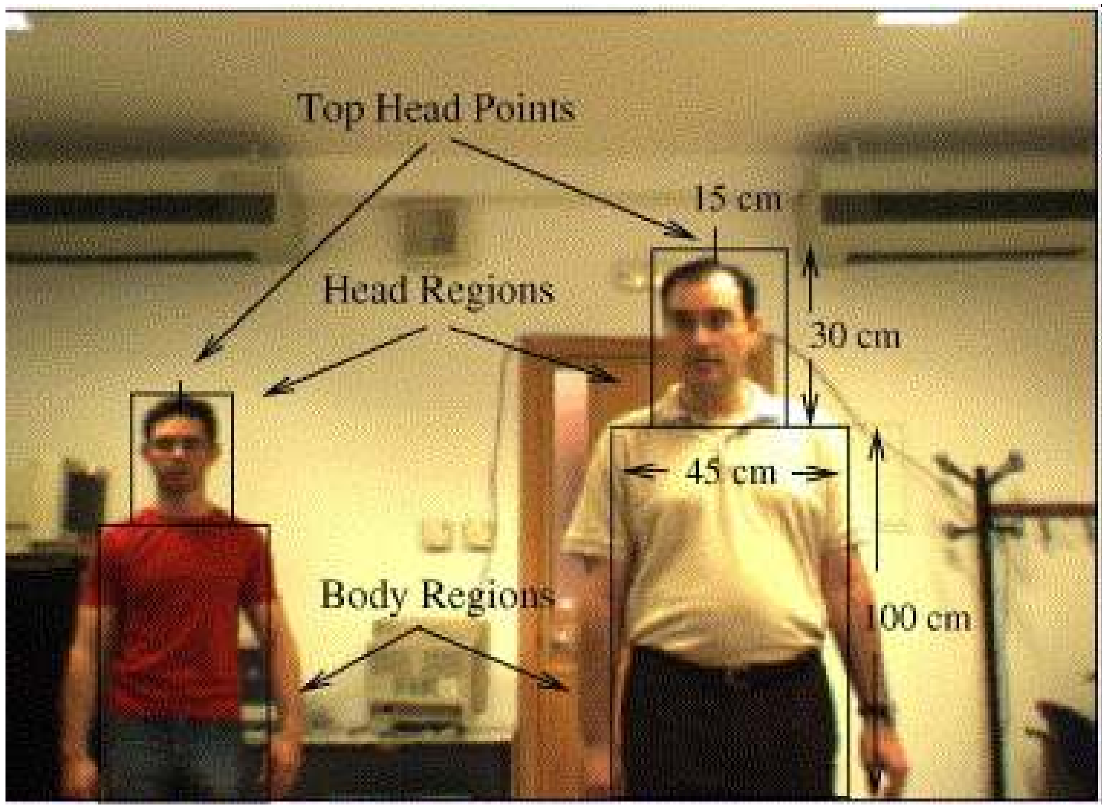
- Division of Tasks
- Color Modeling
- Detecting People
- Tracking People

Division of Tasks

- Face detection is computationally costly
- Tracked people can first be assigned to objects detected in a scene
- Remaining objects can then be studied as possible candidates for new people

How objects are Classified

- Generate a ‘head’ region based on highest point in occupancy map
- Place a ‘body’ region underneath the head region

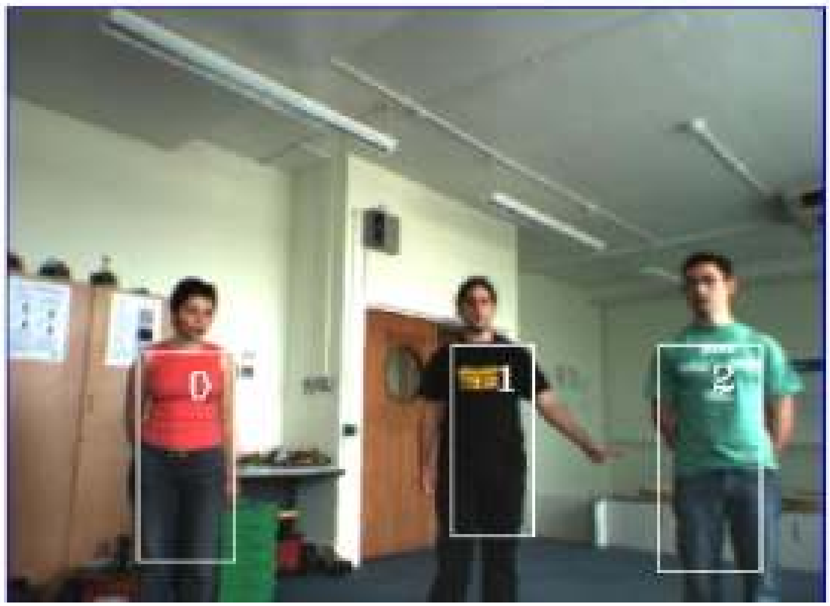


Modeling Color

- The color of each object is modeled as a histogram
- Some illuminance information is held since color is not reliable at extreme values
- Pixels near the center are given more weight to reduce error
- Color models can be compared to compute a similarity value
- Color model for a tracked object is updated over time

Similarity of Body Regions

a



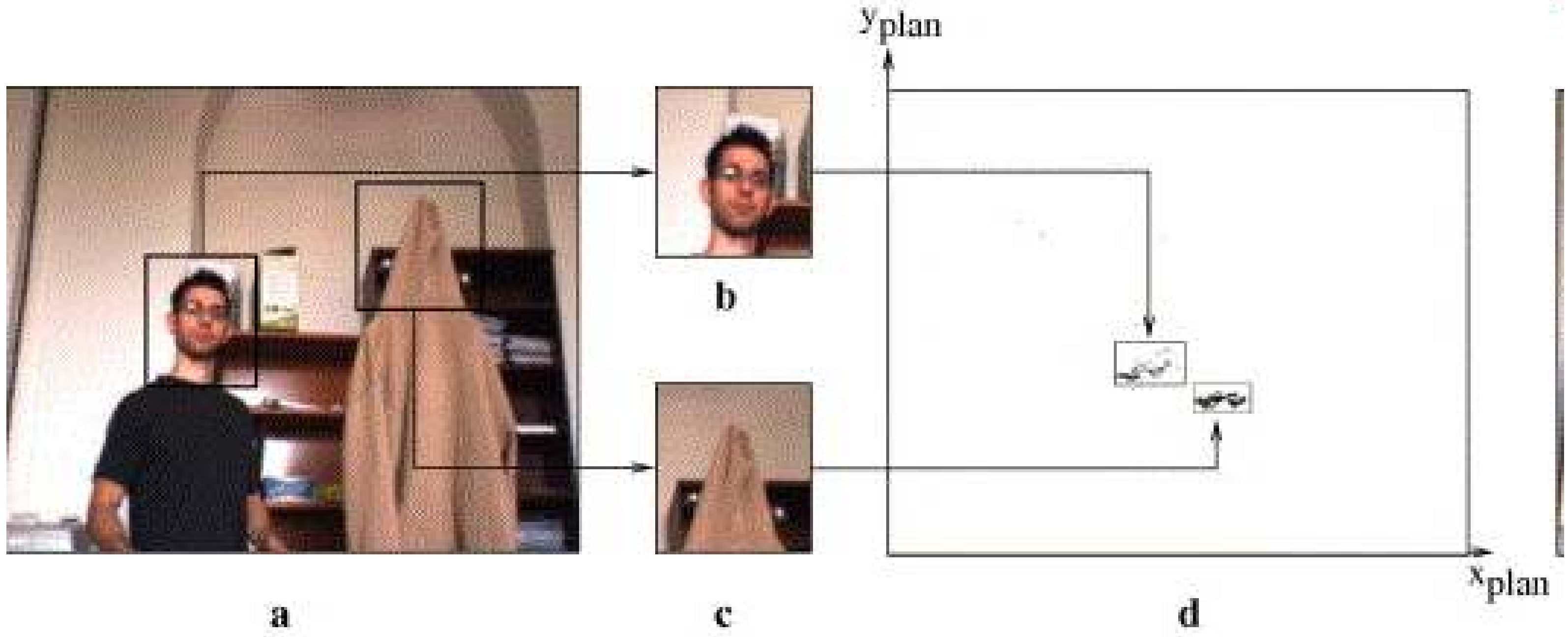
b

	0	1	2
0	1	0.492	0.655
1	0.492	1	0.282
2	0.655	0.282	1

Detecting People

- Detected object not associated with tracked people may be new people in the scene
- Each head region is analyzed to determine if it is similar to a human head
  - If it is, apply face detection based on common library routines, ignore it otherwise
  - Pruning of regions reduces computational cost and occurrence of false positives

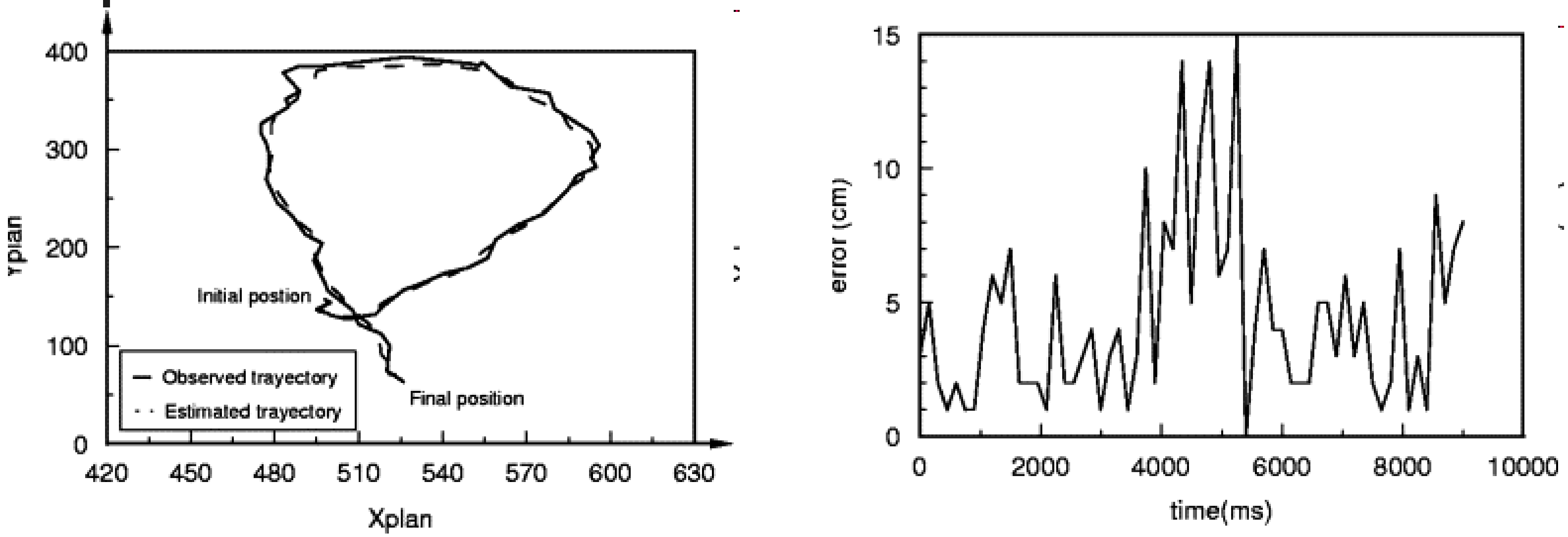
Identifying People



Tracking People

- Once a person is detected, they must be tracked
- Assign tracked people to detected objects using Kuhn's Hungarian Method
- Color similarity and estimated position from the Kalman Filter are used to compute how likely each object is a given person
  - Color similarity is important when people are in close proximity
- If a person is not detected for some time, they are removed from the list

Estimating Position with the Kalman Filter



- Even though the filter uses a linear movement model, it works well because the time interval is small
- Worst error occurs when people turn

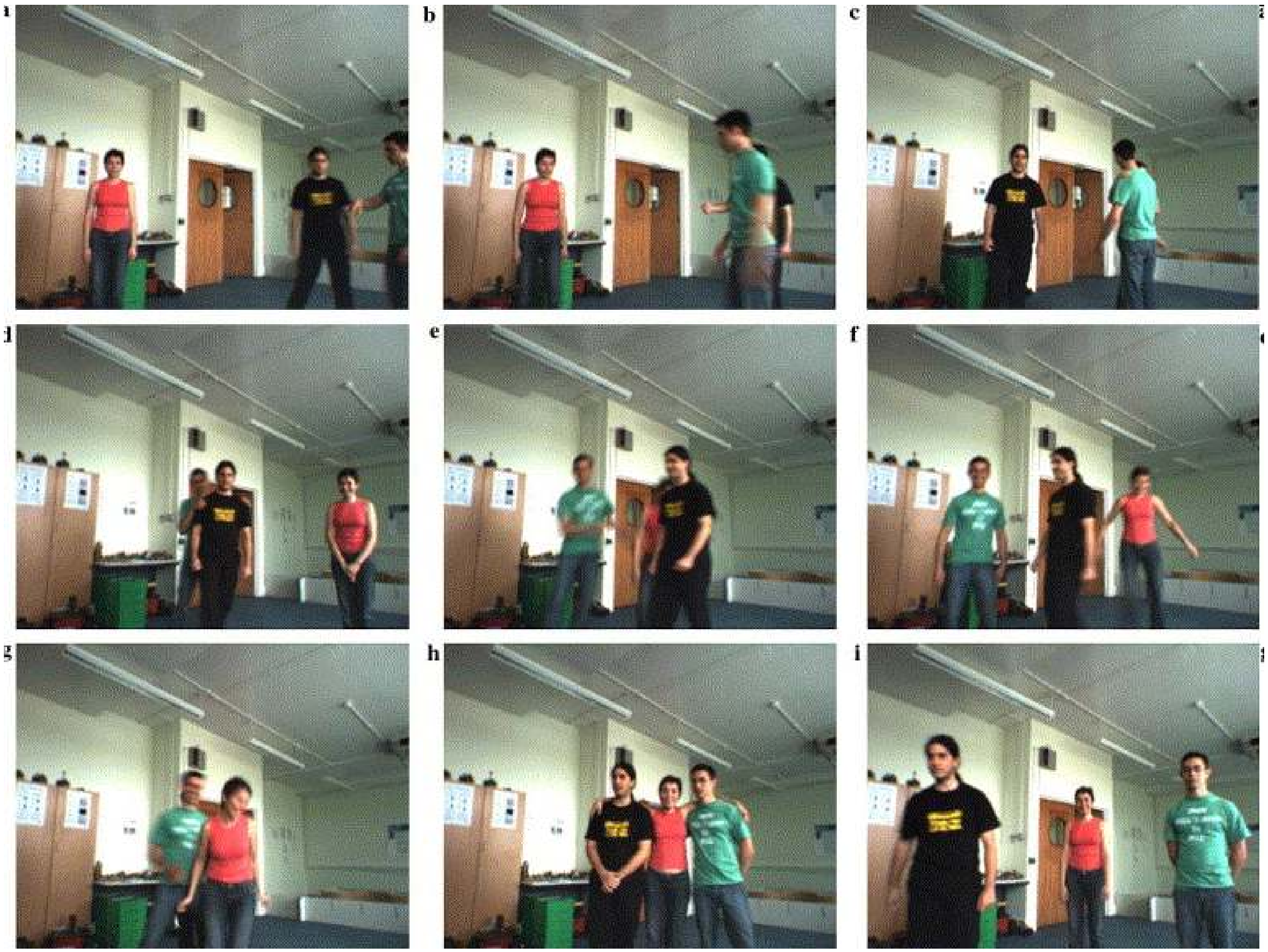
Experimental Setup

- Used 320*240 resolution images
- Operational frequency is 10Hz on a 3.2 Ghz Pentium 4 laptop
- Configured to detect people between .5 and 2.5 meters
- Cameras with 4mm and 6mm focal lengths are used
- Number of people vary from 2 to 4
- People interact in a variety of ways to try and trick the system into confusing them
- Tests were run with and without the use of color modeling

Results

- In every test case, the use of color modeling improved tracking success rate
- The 6mm camera could not track as many people as the 4mm camera
- Increasing the number of people decreases the success rate of tracking, especially when color is not used
- The use of color may allow the system to correct itself once the interacting people have separated

Example Scene



Tracking Success

#People	f(mm)	#conflicts	#NC (%)	#c (%)
2	6	30	86	100
3	6	17	58	82
3	4	52	69	100
4	4	13	50	100

Future Work

- Use multi-scale techniques when creating the height and occupancy maps
- Avoid the use of a unique cell size in hopes of improving success without hurting performance
- Use additional features aside from color and position to distinguish objects, such as face identification
- Test the system on a mobile platform such as an autonomous robot