

Person Counting Using Stereo

David Beymer*
SRI International
333 Ravenswood Ave
Menlo Park, CA 94025
beymer@ai.sri.com

Abstract

Stores and shopping malls would like to keep track of shopper volume by employing automatic techniques for counting shoppers. Existing approaches instrument doors with infrared beams and count beam interruptions, but this approach cannot resolve groups of people well. We are applying a vision-based approach that detects and tracks people from a stereo camera mounted above a door and pointing down. After applying real-time stereo and 3D reconstruction, the system segments the scene by selecting stereo pixels falling inside a 3D volume of interest, which is placed to capture the heads and torsos of adult shoppers. The main novelties of our approach include (1) remapping the stereo disparities to an orthographic "occupancy map", which simplifies person modeling, and (2) tracking people using a Gaussian mixture model. On a test set of 900 enter/exit events in 4 hours of video, our system has achieved a net counting error rate of just 1.4%.

1 Introduction

Driven by the need for marketing information about shoppers, shopping malls and department stores are deploying more and more sensors to keep track of shoppers. Given a system that can count shoppers or track movement within a store, malls can set lease rates based on foot traffic, stores can monitor shopper volume, and the effectiveness of store displays can be evaluated.

In this paper, we focus on the problem of counting shoppers as they enter or exit a door. Allowing the doors to be double doors makes the problem difficult, as multiple people can be entering side by side, or one person can be stopped in the doorway while others pass. In addition, we want to exclude counting objects such as shopping carts or small children.

There are already a number of different types of sensing systems in place for counting shoppers. Infrared beam counters are placed so that a beam is interrupted when shoppers enter or exit. While two parallel beams can be used to give the direction of the shopper, the system will undercount two people who are side by side. Infrared motion detectors are also in use, but they suffer from similar problems with resolving groups of people.

Person tracking systems using a camera mounted overhead and pointing straight down are becoming commercially available. Orienting a camera this way alleviates the partial occlusion problem associated with more oblique mounting angles. A monocular system from ShopperTrak is being deployed by The Gap and Banana Republic. A stereo-based person tracking system from Point Grey Research, CenSys3D, was recently demonstrated at CVPR 2000.

Our approach to person counting is to track stereo blobs within a 3D volume of interest. A stereo camera is mounted in the ceiling near the door, pointing straight down. An operator first specifies a 3D volume of interest in a normally empty region by the door, but outside the door swing region. The person counter then detects and tracks the heads and upper torsos of people as they pass through the volume of interest. Because the ceilings of a typical store installation are only 8-9 feet, wide field of view lenses will be required to capture a region wide enough to monitor a double door. Since this introduces severe radial lens distortion and accentuates perspective distortion, our system remaps the stereo disparities into an orthographic, top-down view.

The main components of our system are real-time stereo, reprojection to an orthographic view, and multiple person detection and tracking. The stereo system performs real-time 3D reconstruction by employing table lookup to map from image coordinates and disparities to 3D. Based on the 3D coordinates, the scene is (1) segmented by filtering out pixels outside the volume of interest, and (2) reprojected to a top-down, orthographic view. Finally, people are detected and tracked in the orthographic reprojection using a Gaussian mixture model and Kalman filtering.

What are the advantages of a stereo-based system over non-video or monocular approaches to person counting? Compared to infrared beam counters, our system can handle multiple people moving side by side, avoiding the undercounting problem. Compared to monocular approaches, stereo provides a robust method for estimating object height, which is useful for distinguishing adult shoppers from children or shopping carts. Also, the coupling of 3D reconstruction and the volume of interest provides for an extremely robust

* Author's current affiliation: IBM Almaden Research Center, 650 Harry Rd, San Jose, CA 95120

segmentation. A popular alternative segmentation approach is background differencing, which introduces the difficult problem of maintaining a background image over time.

In this paper, we first discuss related work, and then we describe camera setup, calibration, and specification of the volume of interest. The next section describes the main tracking algorithm, including real-time stereo and reprojection. Next, we present experimental results of the system on roughly four hours of video, and we close with a discussion of future work.

2 Related work

A large number of person tracking systems have been developed over the last several years, and the majority of them use color[1-5], background subtraction[3, 6-9], or contour modeling[10, 11]. In addition, a number of systems can track multiple people[3, 4, 6, 7, 9, 12], which adds the burden of maintaining person ID and handling occlusion events. In this section on related work, we concentrate on systems that either use stereo or Gaussian modeling.

The main advantages of adding stereo to a person tracking system include (1) segmentation, (2) locating objects in 3D, and (3) handling occlusion events. In [3] and [13], large baseline stereo is used to triangulate the location of person blobs in 3D. In [4], dense stereo is computed in real-time and used to segment the scene into patches of slowly varying disparity. Background subtraction on dense stereo disparities is performed in [5, 6, 12], where [6] includes image intensities and [5] includes color. [6] also uses stereo to help disambiguate occlusion events. Finally, [14] and [15] have started fitting parametric human models to dense stereo maps, with the hope that stereo can help with self-occlusion of body parts. In our use of stereo for segmentation, we reconstruct the scene in 3D and select disparity pixels in a volume of interest.

Gaussian functions have been used in person tracking systems for modeling multimodal distributions and articulated human models. [1] and [14] model human bodies as connected 3D Gaussian blobs. Gaussian mixture models have been used to model the multimodality of background images[16, 17], making background subtraction more robust to shadows or blowing leaves. In the area of multi-hypothesis tracking, [18] has used a "piecewise" Gaussian model to represent multimodality in tracking state space, allowing the system to maintain a number of competing tracking states. This is related to Condensation tracking[11], which represents multimodality non-parametrically with a set of samples in state space. In our system, we are using Gaussian mixtures to model multiple people, with one Gaussian per person.

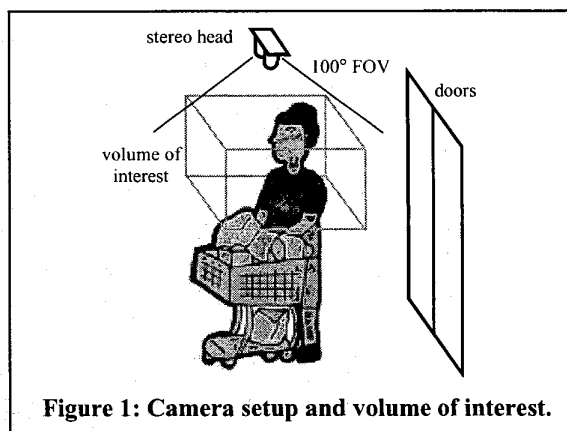


Figure 1: Camera setup and volume of interest.

3 Hardware setup

In this section, we describe the hardware setup for our person counter, including the camera setup, calibration, and specifying the volume of interest.

The stereo head we use is Videre Design's STH-V1, an inexpensive, low power stereo head with analog CMOS cameras[19]. As shown in Figure 1, the stereo head is ceiling mounted a few feet from the doors in order to minimize door visibility. Since the ceilings near the doors are typically only 8-9 feet, a wide field of view (FOV) is required to cover a set of double doors; our 2.1mm lenses give the system a horizontal FOV of approximately 100°. The choice of high FOV lenses

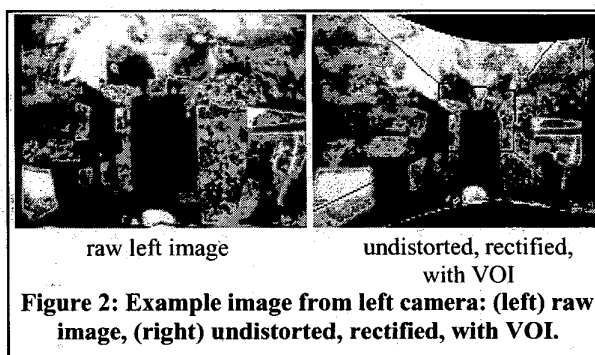


Figure 2: Example image from left camera: (left) raw image, (right) undistorted, rectified, with VOI.

increases the effect of radial lens distortion, as shown in the example left image in Figure 2.

During stereo head calibration, we estimate the intrinsic parameters of the left and right cameras and the rigid transform between the two cameras. During the procedure, five views of a planar calibration pattern are presented to the stereo head, where the pattern is the 4x4 grid of squares used by Tsai[20]. Camera intrinsics are modeled after Tsai[20], using one radial distortion parameter K_1 , and all the parameters are estimated using a bundle adjustment procedure applied to the five stereo pairs. From the calibration, the system computes an

image warp for (1) removing radial distortion, and (2) rectifying the left and right images to align epipolar lines for stereo matching (see Figure 2).

The power of a calibrated stereo head is the ability to perform 3D reconstruction. In our person counting setup, this allows us to segment the scene by focusing on objects falling inside the 3D volume of interest (VOI). Proper placement of the VOI by a system operator allows the system to focus on the head and upper torso of adult shoppers, ignoring shopping carts and small children. The VOI is currently a 3D box (see Figures 1 and 2), although the shape can be easily customized, if necessary. Reprojection of stereo disparities into an orthographic view will be performed inside the VOI.

4 Tracking in occupancy maps

As mentioned in the introduction, tracking is done using stereo primarily because the segmentation problem is simplified. Beyond using stereo, though, we remap disparities to what we call an *occupancy map*, an orthographic reprojection of the scene disparities inside the VOI. In this section, we discuss the creation of occupancy maps and our use of Gaussian mixture models to track people in them.

4.1 Occupancy maps

Perspective distortion tends to elongate stereo blobs near the boundaries of the image, as people present more of a side view to the camera. As shown in Figure 3a, the elongation is towards the optical center of the camera. If we can resample the image from a vertical projection, then the stereo blobs should be more compact and simpler to model mathematically. 3D reconstruction from stereo enables our system to perform the orthographic resampling. Furthermore, the resampled image can be computed in real-time through the use of lookup tables.

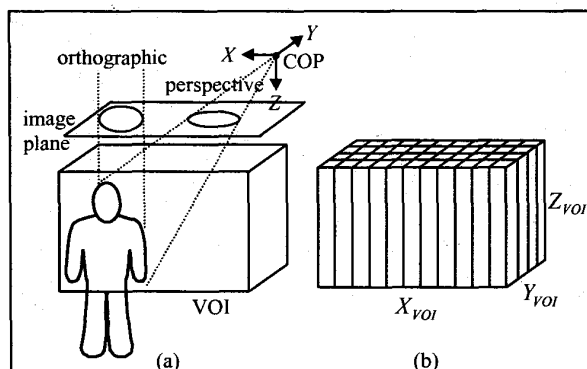


Figure 3: (a) A person's projection is more compact under orthographic projection than perspective, and (b) the division of the VOI into buckets.

We call the resampled image an *occupancy map*. Occupancy maps divide the X - Y plane of the VOI into a set of discrete vertical buckets, as shown in Figure 3b. At location $occ(X_{VOI}, Y_{VOI})$, we accumulate all disparity pixels that land in the corresponding bucket. We can also work forward from the disparity image to see how pixels map into the occupancy map. Given an image disparity at $(x, y, disp)$, we first reconstruct the point in 3D (X, Y, Z) and then we map to (X_{VOI}, Y_{VOI}) using a simple windowing transform on (X, Y) . Finally, we increment the value of $occ(X_{VOI}, Y_{VOI})$.

Since closer objects appear larger in the image, having each disparity pixel $(x, y, disp)$ contribute a fixed amount to $occ(X_{VOI}, Y_{VOI})$ favors closer objects. Thus, a technique is needed to compensate for the depth of a pixel. Using the stereo equation $Z = bf / disp$, where b and f are the baseline and focal length, one can easily show that incrementing $occ(X_{VOI}, Y_{VOI})$ by $disp_{nom} / disp$ compensates for range, where the accumulation is 1 for pixels at disparity $disp_{nom}$. An example of an occupancy map is shown in Figure 4. We always smooth the occupancy map to compensate for image noise and stereo matching errors.

4.2 Tracking using Gaussian mixture models

Since the head and torso pixels map to Gaussian blobs in the occupancy map, we have explored using a Gaussian mixture model to approximate the occupancy map $occ(x, y)$. (In this section, we use lower case x and

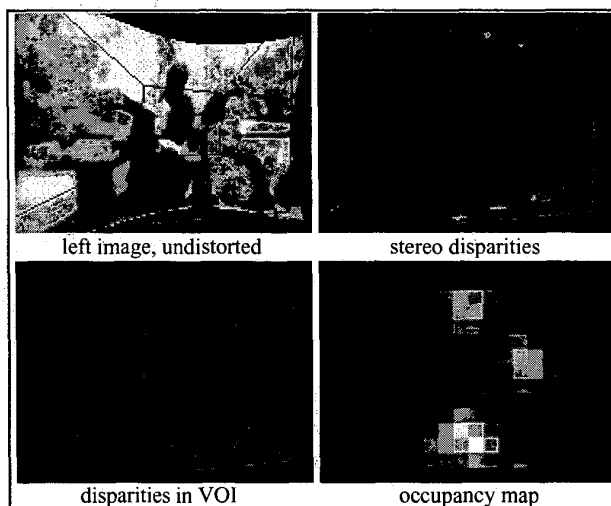


Figure 4: Stereo disparities are first filtered using the VOI and then orthographically reprojected to produce the occupancy map.

y to index the occupancy map). Each individual being tracked will be represented by a single Gaussian in the mixture. The tracking procedure is hence driven by the function approximation of $occ(x, y)$.

Two mixture models have been explored for representing occupancy. Let the mixture model G be composed of n Gaussians G_i , where

$$G_i(x, y; x_i, y_i, s_i, \sigma_i) = s_i e^{-\frac{[(x-x_i)^2 + (y-y_i)^2]}{\sigma_i^2}}$$

For the sake of brevity we will often omit the parameters x_i, y_i, s_i , and σ_i and write G_i as $G_i(x, y)$. We have explored:

1. *Additive mixture model*: $G(x, y) = \sum_{i=1}^n G_i(x, y)$

2. *Competitive mixture model*: $G(x, y) = \text{Max}_{i=1}^n \tilde{G}_i(x, y)$

The main difference between the models is that the competitive model explicitly incorporates segmentation of the occupancy map. That is, a Gaussian G_i (i.e. person i being tracked) claims the pixel set

$$\{(x, y) | G_i(x, y) > G_j(x, y), j \neq i\}.$$

The additive model, on the other hand, is more liberal and allows multiple Gaussians to explain occupancy at a given (x, y) . We found that when two people are close to one another, this additive mixing reduced the Gaussians' ability to lock onto individual modes of $occ(x, y)$, although the overall approximation was good. The competitive model was better able to track the modes of $occ(x, y)$ for groups of people.

The "measurement" step of the tracker fits the competitive mixture model to the occupancy map by iteratively performing segmentation and parameter update steps. For n people, the measurement process is:

Iterate until Gaussian parameters settle
 For each person i :
segment:
 $mask_i(x, y) = \{(x, y) | G_i(x, y) > G_j(x, y), j \neq i\}$
update G_i : minimize
 $\sum_{mask_i} (occ(x, y) - G_i(x, y))^2$ using one step of
 Newton-Raphson iteration.

An example of this fitting process is shown in Figure 5.

In this scheme, the system detects new people by looking for local maxima in $(occ(x, y) - G(x, y))$ that are above a threshold. More details are given in the next section, which provides a complete description of the system.

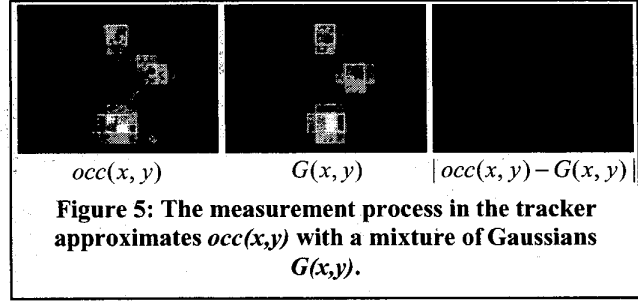


Figure 5: The measurement process in the tracker approximates $occ(x, y)$ with a mixture of Gaussians $G(x, y)$.

5 Person counting algorithm

The goal of the person counting system is to count the number of people entering and exiting a doorway. Figure 6 shows a flowchart of the entire system, beginning with stereo and ending with person track classification. In this section, we give the details of the overall counting system.

5.1 Stereo and occupancy map

For stereo, we use SRI's Small Vision System, a real-time, area correlation-based stereo implementation that runs on standard PC hardware [19]. See Figure 4 for an example disparity map, where lighter grey levels correspond to larger disparities and closer objects. To assure stereo operation at frame rate with extra cycles to spare for the tracker, we capture images at 160x120 and compute stereo over 16 disparities.

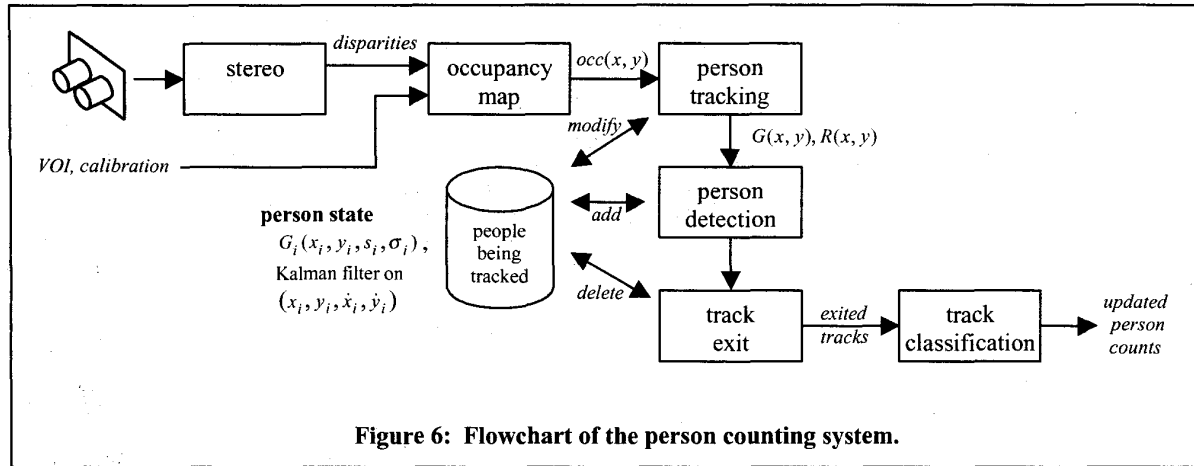
Segmentation of pixels inside the VOI and reprojection to form the occupancy map is performed using table lookup. For each possible x, y , and disparity $disp$, we need to store (1) an "inside VOI" segmentation bit, and (2) the coordinates (X_{voi}, Y_{voi}) if the pixel is indeed inside the VOI. Thus, the tables for segmentation and VOI coordinates are 3D, indexed by x, y , and $disp$. Our tables are reasonable in size (2.6 MB total), however, because of our reduced image size.

The tables are generated offline once the camera calibration and VOI are known. For each $x, y, disp$ triplet, we reconstruct the point (X, Y, Z) in 3D, and set the segmentation bit if the 3D point is inside the VOI. For interior points, we map the X and Y coordinates to the proper bucket (Figure 3b) and store the VOI coordinates in the second table.

At run time, we can easily test to see if a disparity is in the VOI and immediately map it to the occupancy map. The occupancy map is passed on to the tracking and detection modules.

5.2 Tracking and detection

As mentioned in the previous section, each person is represented by a Gaussian that tracks their movement in the occupancy map. Thus, the main part of a person's state vector is the vector $(x_i, y_i, s_i, \sigma_i)$ parameterizing



the Gaussian. In addition, the person's position (x_i, y_i) is tracked with a Kalman filter; the Kalman filter maintains position and velocity $(x_i, y_i, \dot{x}_i, \dot{y}_i)$ and assumes a simple constant velocity motion model.

The tracker follows a simple Kalman filter framework with the Gaussian mixture modeling as the "measurement" process. For each person we predict the expected location (x_i, y_i) in the occupancy map; this location is used as initial conditions for the Gaussian fitting process. Next, we run the measurement process described in section 4.2 to update the Gaussian locations.

Finally, we perform a Kalman filter update on $(x_i, y_i, \dot{x}_i, \dot{y}_i)$. Figure 7 shows a tracking example, where (a) shows person locations plotted as circles in the occupancy map computed from (b). In Figure 7(b), person locations are plotted by projecting a 3D vertical bar through the 3D Gaussian centroid. In Figure 7(c), we show a track over time, where a vertical bar shows the present person location. Finally, in (d), a number of tracks are shown in the occupancy map.

The detection of new people entering the VOI uses the residual image $R(x, y) = occ(x, y) - G(x, y)$ from the tracker. $R(x, y)$ contains stereo data that is not explained by current tracks. The detection algorithm first smoothes $R(x, y)$ for noise immunity and then looks for local maxima above a detection threshold. If a local maximum is found, we fit a Gaussian to $R(x, y)$ and begin a new track. To avoid a problem with multiple detections when a person is slightly multimodal in $occ(x, y)$, we have an additional test that tries to link the new detection to an existing track. Given the new potential track and an existing track, this linkage test is based on (1) their proximity, and (2) the drop-off in occupancy between the two Gaussian modes.

5.3 Track exit and classification

When a person exits the VOI, their track is deleted by noticing either:

- (1) (x_i, y_i) is out of bounds,
- (2) $\sigma_i < \text{threshold}$ (Gaussian width is too narrow),
- (3) $s_i < \text{threshold}$ (occupancy is too weak).

Upon track exit, the system classifies the track in order to one of four categories, as shown in Figure 8, based on a simple geometric analysis of the track path. Categories "reenter" and "passby" do not affect the person counts, while an "enter" or "exit" event will increment the corresponding count. Tracks will be rejected if they are

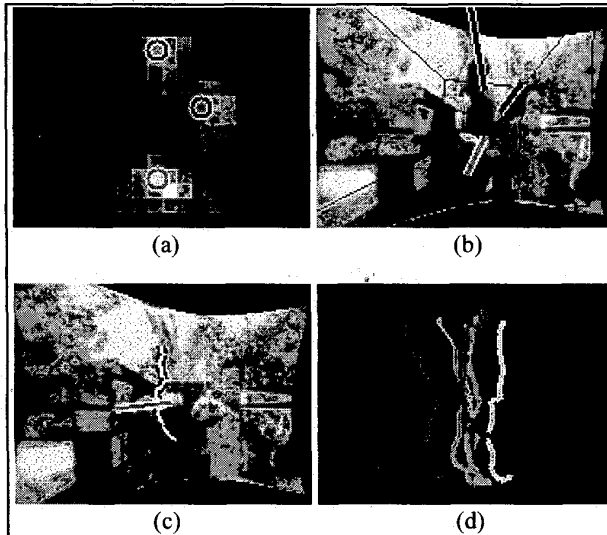
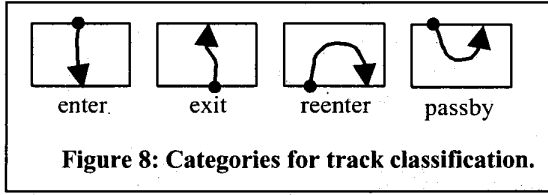


Figure 7: Tracker results: (a) the occupancy map computed from (b) with Gaussian centroids plotted as circles. In (b), person locations are plotted by projecting a 3D vertical bar. In (c), a complete track over time is shown, with a vertical bar showing the present person location. (d) shows several tracks over time in the occupancy map.



too short or if they do not fit in one of the four categories.

6 Experimental results

We have evaluated the person counter on roughly 4 hours of video collected at double door entrances on the SRI campus. About 3.5 hours were taken at lunchtime at the SRI cafeteria over a period of two days, and around 30 minutes was taken at a SRI auditorium during a meeting. The data was recorded to VCR (our stereo head interlaces the left and right views into a single video feed) and digitized into clips lasting from a few seconds to 20 seconds. Clips are chosen so that the track classifications are unambiguous. Finally, for each clip we manually determine ground truth counts; for the i th clip let the count for each event type be $enter_i^{gt}$, $exit_i^{gt}$, $reenter_i^{gt}$, and $passby_i^{gt}$. The net count for a clip i is $net_i^{gt} = enter_i^{gt} - exit_i^{gt}$.

Next, the person counter is run on all the clips, recording the tracker counts similarly as $enter_i^{tr}$, $exit_i^{tr}$, $reenter_i^{tr}$, and $passby_i^{tr}$. Three error rates are defined:

- (1) Net count error: $error_{NC} = \frac{\sum_i |net_i^{gt} - net_i^{tr}|}{\sum_i (enter_i^{gt} + exit_i^{gt})}$
- (2) Enter error: $error_{enter} = \frac{\sum_i |enter_i^{gt} - enter_i^{tr}|}{\sum_i enter_i^{gt}}$
- (3) Exit error: $error_{exit} = \frac{\sum_i |exit_i^{gt} - exit_i^{tr}|}{\sum_i exit_i^{gt}}$

Table 1 shows our error rates for three groups of clips, as well as the total error rate. The most important number is in the lower left: 13 net count errors for 905 enter/exit events. The two dominant sources of error are (1) tall people being "multimodal" in the occupancy map and hence overcounted, and (2) the track classifier rejecting a track because of a wild motion at track exit.

	Error rates		
	Net count	Enter	Exit
Cafe1	2.1% (10/471)	2.1% (5/241)	1.8% (4/218)
Cafe2	0.9% (3/305)	2.1% (4/193)	1.9% (3/159)
All hands	0.0% (0/79)	0.0% (0/69)	0.0% (0/4)
Total	1.4% (13/905)	1.8% (9/503)	1.8% (7/381)

Table 1: In our experimental results, the system achieved a net count error of 1.4% on 905 enter/exit events.

It is possible for the net count error to be lower than the combined enter and exit errors, as enter and exit errors can cancel each other within a clip. Also, the enter/exit error rates do not include clips with passby or reenter events since how the tracker classifies them depends on the placement of the VOI (i.e. the path of such an event can avoid the VOI; enter or exit events cannot avoid the VOI, however).

The system has achieved a frame rate of 20-25 Hz, depending on the number of people being tracked. This timing was done on a dual Pentium II running at 400 MHz (but the system is single threaded and only uses one processor).

7 Conclusion

The main contributions of this paper are the introduction of the occupancy map and the use of Gaussian mixture models for tracking multiple people. The occupancy map remaps stereo disparities in a way that simplifies camera viewpoint and subsequent modeling. In our person counting problem, it turned the problem of tracking irregularly shaped blobs with varying depth into one of tracking Gaussian shaped blobs. Second, mixture models have previously been applied for modeling multimodal distributions (e.g. background modeling) and handling ambiguous state spaces (e.g. multi-hypothesis tracking). We have applied it to the occupancy map, where the modes correspond to people, showing that it is useful for multiperson tracking.

With regards to the person counting problem, we have demonstrated a real-time system that can resolve groups of people, unlike the popular infrared beam systems commonly in use today. Our stereo-based system segments the scene by focusing on pixels in a 3D volume of interest and remaps those pixels to an orthographic view called the occupancy map. Persons are tracked in the occupancy map by using a "competitive" Gaussian mixture model. The system runs at 20-25 Hz on a 400 MHz Pentium II, and it achieves a net counting error of only 1.4% on a test set of 900 enter/exit events.

For future work, we plan to perform more field testing, especially at department stores or shopping malls. The error rate could be reduced by investigating the problem of tall people appearing to be multimodal in the occupancy map. Finally, we would like to simplify the camera calibration procedure to the point where a field technician can perform it.

8 References

- [1] Wren, C., et al., *Pfinder: Real-Time Tracking of the Human Body*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997. 19(7): p. 780-785.

- [2] Fieguth, P. and D. Terzopoulos. *Color-Based Tracking of Heads and Other Mobile Objects at Video Frame Rates*. in *IEEE Conference on Computer Vision and Pattern Recognition*. 1997. San Juan, Puerto Rico, pp. 21-27.
- [3] Rehg, J.M., M. Loughlin, and K. Waters. *Vision for a Smart Kiosk*. in *IEEE Conference on Computer Vision and Pattern Recognition*. 1997. San Juan, Puerto Rico, pp. 690-696.
- [4] Darrell, T., G. Gordon, and M. Harville. *Integrated person tracking using stereo, color, and pattern detection*. in *IEEE Conference on Computer Vision and Pattern Recognition*. 1998. Santa Barbara, California, pp. 601-608.
- [5] Krumm, J., et al. *Multi-Camera Multi-Person Tracking for EasyLiving*. in *IEEE International Workshop on Visual Surveillance*. 2000. Dublin, Ireland.
- [6] Haritaoglu, I., D. Harwood, and L.S. Davis. *W⁴S: A Real-Time System for Detecting and Tracking People in 2 1/2D*. in *European Conference on Computer Vision*. 1998. Freiburg, Germany, pp. 877-892.
- [7] Kanade, T., et al. *Advances in Cooperative Multi-Sensor Video Surveillance*. in *Proceedings of the Image Understanding Workshop*. 1998. Monterey, California, pp. 3-24.
- [8] Boulton, T.E., et al. *Frame-Rate Omnidirectional Surveillance and Tracking of Camouflaged and Occluded Targets*. in *IEEE International Workshop on Visual Surveillance*. 1999. Fort Collins, CO, pp. 48-55.
- [9] Rosales, R. and S. Sclaroff. *3D Trajectory Recovery for Tracking Multiple Objects and Trajectory Guided Recognition of Actions*. in *IEEE Conference on Computer Vision and Pattern Recognition*. 1999. Fort Collins, CO, pp. 117-123.
- [10] Baumberg, A. and D. Hogg. *Learning Flexible Models from Image Sequences*. in *European Conference on Computer Vision*. 1994. Stockholm, Sweden, pp. 299-308.
- [11] Isard, M. and A. Blake. *Contour Tracking by Stochastic Propagation of Conditional Density*. in *European Conference on Computer Vision*. 1996, pp. 343-356.
- [12] Beymer, D. and K. Konolige. *Real-Time Tracking of Multiple People Using Stereo*. in *IEEE Frame Rate Workshop*. 1999. Corfu, Greece.
- [13] Azarbayejani, A., C. Wren, and A. Pentland. *Real-Time 3-D Tracking of the Human Body*, MIT Media Lab, Technical Report No. 374, 1996
- [14] Jojic, N., M. Turk, and T.S. Huang. *Tracking Self-Occluding Articulated Objects in Dense Disparity Maps*. in *IEEE International Conference on Computer Vision*. 1999. Kerkyra, Greece, pp. 123-130.
- [15] Lin, M.H. *Tracking Articulated Objects in Real-Time Range Image Sequences*. in *IEEE International Conference on Computer Vision*. 1999. Kerkyra, Greece, pp. 648-653.
- [16] Grimson, W.E.L., et al. *Using Adaptive Tracking to Classify and Monitor Activities in a Site*. in *IEEE Conference on Computer Vision and Pattern Recognition*. 1998. Santa Barbara, California, pp. 22-29.
- [17] Friedman, N. and S. Russell. *Image Segmentation in Video Sequences: A Probabilistic Approach*. in *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*. 1997. Providence, RI: Morgan Kaufmann.
- [18] Cham, T.-J. and J.M. Rehg. *A Multiple Hypothesis Approach to Figure Tracking*. in *IEEE Conference on Computer Vision and Pattern Recognition*. 1999. Fort Collins, CO, pp. 239-245.
- [19] Konolige, K. *Small Vision Systems: Hardware and Implementation*. in *Eighth International Symposium on Robotics Research*. 1997. Hayama, Japan, pp. 111-116.
- [20] Tsai, R.Y., *A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses*. *IEEE Journal of Robotics and Automation*, 1987. **RA-3**(4): p. 323-344.