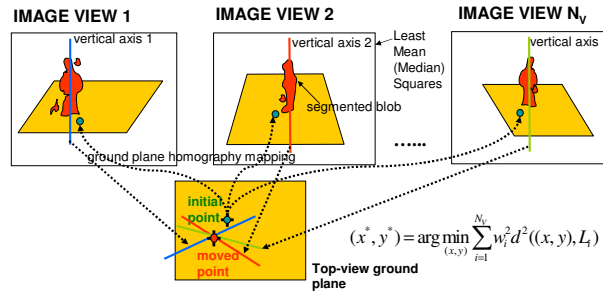




Our approach and motivation

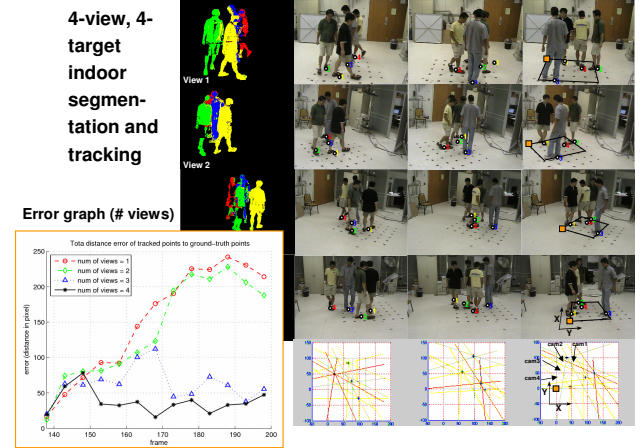
- **Previous work** (multi-target or multi-view)
 - Elgammal, Davis (ICCV'01): segmentation using MLE and occlusion info., increasing # hypotheses.
 - Zhao, Nevatia (CVPR'03): model-based segmentation in crowded scene.
 - Javed et al. (ICCV'03): tracking across non-overlapping views.
 - Mittal, Davis (IJCV'03): region-based stereo (calibrated cameras), good for indoors only.
- **Our approach & Motivation**
 - Tracking people's ground points and Segmenting blobs using overlapping views, ground-plane homography, occlusion analysis.
 - Robust multi-view integration from imperfect segmentation
 - Exponential increase of state space # targets, # views → Searching the space by several iterations of multi-view 'segmentation'
 - Monitoring crowded spaces: building entrance, store, casino, etc.

Multi-view integration



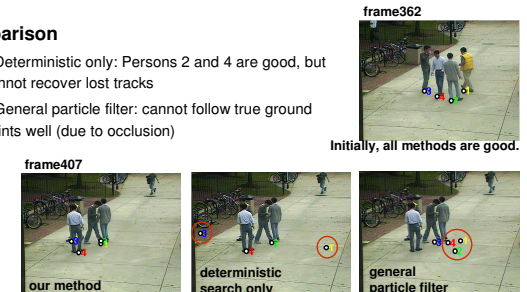
- **Interesting fact:** the homographic images of all the vertical axes of a person across different views intersect at (or are very close to) a single point (the location of that person on the ground) when mapped to the top-view.
- Severely occluded blobs does not account for integration
- Works even when the ground point of a person from some view is occluded.

Experiments



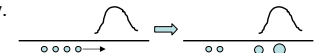
Comparison

- Deterministic only: Persons 2 and 4 are good, but cannot recover lost tracks
- General particle filter: cannot follow true ground points well (due to occlusion)



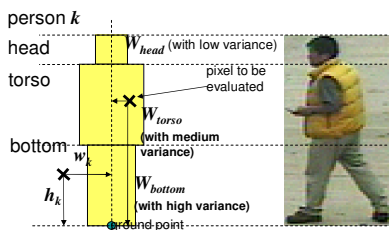
Contributions

- Center vertical axes of the person across views are mapped to the top-view - robust to poor background subtraction and invisible ground points (occlusion).
- Iterative segmentation-searching is incorporated into a particle filtering framework - a few good particles identified in low computational costs.
- In addition, even if all the particles are away from the true ground point, some of them move towards the true one as long as they are located nearby.



Foreground segmentation

- **Bayesian pixel classification**
 - k : class, x : pixel
 - $P(k | x) = \frac{P(k)P(x | k)}{P(x)}$
- **Occupancy probability**
 - $O_k(h_k(x), w_k(x)) = P[w_k(x) < W(h_k(x))]$
 - $= 1 - \text{cdf}_{W(h_k(x))}(w_k(x))$
- The probability that l occludes k depends on their relative position and l 's probabilistic width
 - $P(k) = O_k(h_k, w_k) \prod_{g_y(k) < g_y(l)} (1 - O_l(h_l, w_l))$
- The best class k^* by maximum a posteriori estimation
 - $k^* = \arg \max_k P(k)P(x | k)$



$$p_M(c) = \frac{1}{N_C} \sum_{i=1}^{N_C} \prod_{j=1}^d \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2} \left(\frac{c_j - c_{j,i}}{\sigma_j} \right)^2}$$

Extension to particle filtering framework

- The ground location of the object in the top-view: $s=(x,y)$
- A random walk dynamics (zero mean Gaussian)
- For multi-person tracking: a combination of N_p single-person states.
 - $s_i = (s_{1,d}, \dots, s_{N_p,d})$
- Observation: histogram technique [Perez ECCV02]

