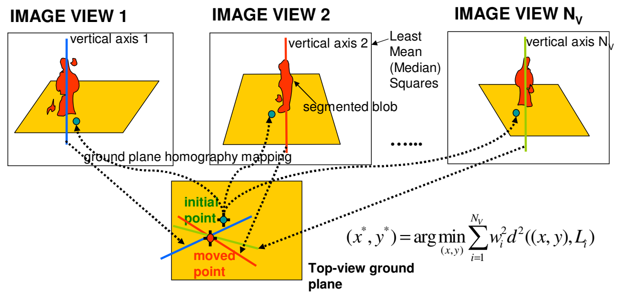
Multi-camera Tracking and Segmentation of Occluded People on Ground Plane Using Search-Guided Particle Filtering

Kyungnam Kim*, Larry Davis
CVL, University of Maryland, USA
*IPIX Corporation, USA

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

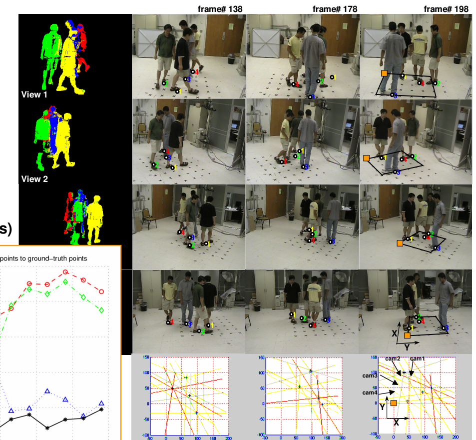


$$(x^*, y^*) = \arg\min_{(x,y)} \sum_{i=1}^{N_V} w_i^2 d^2((x,y), L_i)$$

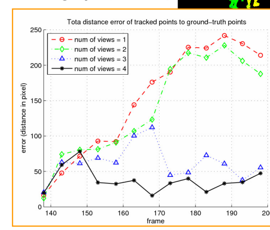
- **Interesting fact**: the homographic images of all the vertical axes of a person across different views intersect at (or are very close to) a single point (the location of that person on the ground) when mapped to the top-view.
- Severely occluded blobs does not account for integration
- Works even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

Experiments

4-view, 4-target indoor segmentation and tracking



Error graph (# views)

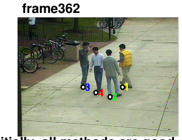


Comparison

– Deterministic only: Persons 2 and 4 are good, but cannot recover lost tracks

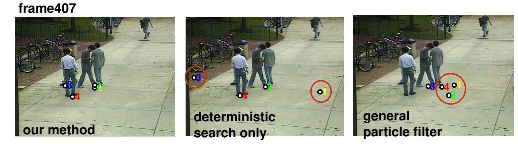
– General particle filter: cannot follow true ground points well (due to occlusion)

frame362



Initially, all methods are good.

frame407



Foreground segmentation

- Bayesian pixel classification
  - k: class, $\mathbf{x}$: pixel

$$P(k \mid \mathbf{x}) = \frac{P(k)P(\mathbf{x} \mid k)}{P(\mathbf{x})}$$

- Occupancy probability

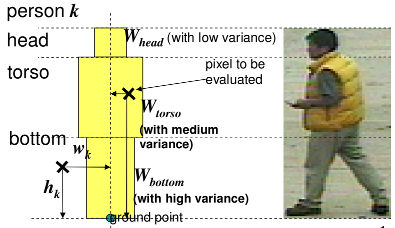
$$O_k(h_k(\mathbf{x}), w_k(\mathbf{x})) = P\left[w_k(\mathbf{x}) < W(h_k(\mathbf{x}))\right] = 1 - \mathrm{cdf}_{W_{(h_k(\mathbf{x}))}}(w_k(\mathbf{x}))$$

- 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

$$k^* = \arg\max_k P(k)P(\mathbf{x} \mid k)$$



- Human appearance model

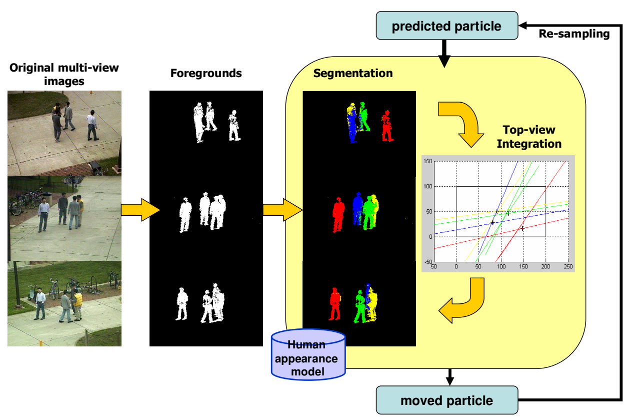
$$p_M(\mathbf{c}) = \frac{1}{N_C} \sum_{i=1}^{N_C} \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{c_j - c_{i,j}}{\sigma_j}\right)^2}$$

Extension to particle filtering framework

- The ground location of the object in the *top-view*: $\mathbf{s} = (x, y)$
- A random walk dynamics (zero mean Gaussian)
- For multi-person tracking: a combination of N_p single-person states.

$$\mathbf{s}_t = (\mathbf{s}_{1,t}, \ldots, \mathbf{s}_{N_p,t})$$

- Observation: histogram technique [Perez ECCV02]



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.