



Multiple-Sensor Indoor Surveillance System

Valery A. Petrushin ¹⁾, Omer Shakil ²⁾, Damian Roqueiro ³⁾, Gang Wei ¹⁾,
Anatole V. Gershman ¹⁾

Accenture Technology Labs

¹⁾ Accenture Technology Labs, Chicago, IL, USA

²⁾ University of Texas at Austin, Austin, TX, USA

³⁾ University of Illinois at Chicago, Chicago, IL, USA
valery.a.petrushin@accenture.com

Environment

- The whole ATL floor
 - Area covered ~18,000 ft² (1,670 m²)
- 32 AXIS-2100 webcams
 - ~ 3-7 Hz sampling rate
- Infrared badge system
 - 91 IR readers
- PTZ camera
- Fingerprint reader



Multiple Sensor Indoor Surveillance Project

- MSIS Objectives:
 - Search and browsing the Event Repository database
 - Event Classification, Clustering and Visualization
 - How to visualize events captured by 32 cameras for 24 hours?
 - How to find rare events?
 - People Counting
 - **People Localization -> Awareness Map**
 - Real-time People Tracking
 - People recognition -> Interaction
 - Behavior and Activity Understanding

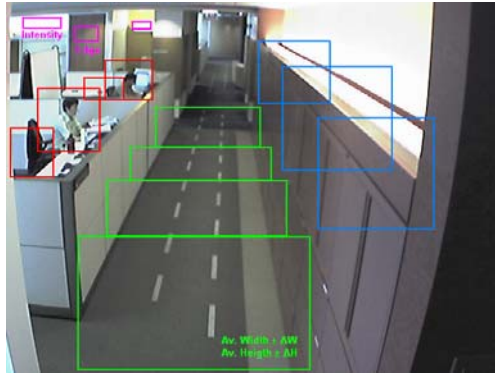


What does a camera see?



Camera Specification

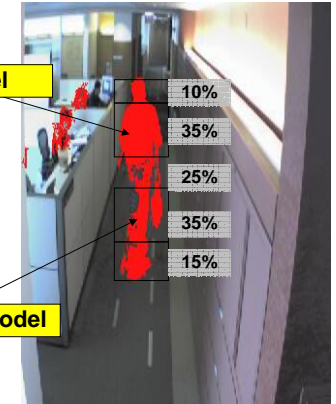
- Operating Zone
- BG Model Type
 - Single Frame
 - Median Filter
- Important Areas
- Unimportant Areas
- Camera Calibration
 - Walking Zones
- Indicators
 - Edge/Intensity/Color



People Localization: Feature Extraction

- Features
 - Geometric features
 - Color-based Appearance Model
 - Color Histograms (RGB, YCrCb, etc.)
 - Gaussian (Mixture) Models
- Blob Extraction
 - Merging vertically
 - Separating horizontally

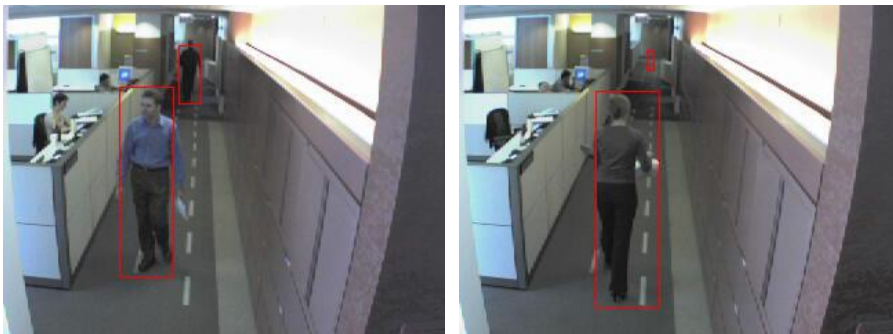
Top Model



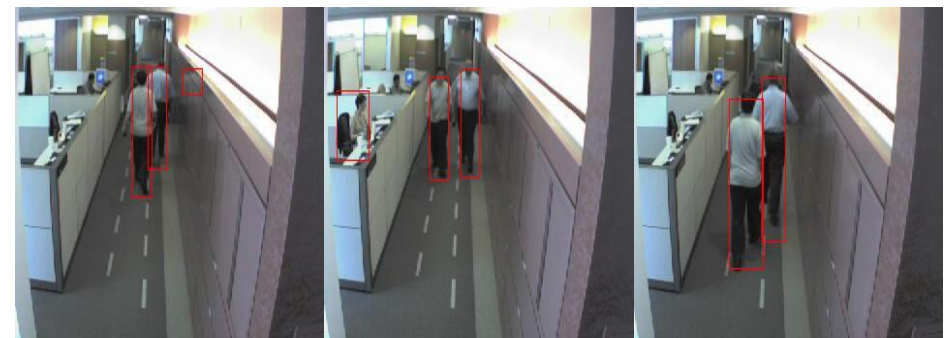
Bottom Model

People Localization: Blob extraction

BG Model → FG Pixels → Merge Vertically

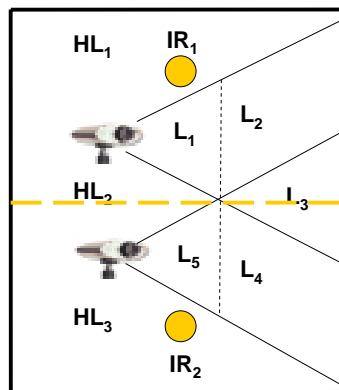


People Localization: Blob separation



Sources of Sensorial Data

- Video cameras
 - Cover large area
 - Relatively reliable
- IR badge system
 - Covers large area
 - Rather precise
 - Unreliable
- Fingerprint station
 - Very precise
 - Rare available
- Face recognition
 - Moderately precise
 - Rare available
- Human intervention
 - Precise
 - Expensive



People Localization: Algorithm

- $P(H_i | L_j)$ - probability H_i located at L_j
- $T_{ij}(H_k)$ – transition probability from L_i to L_j for H_k
- $M(H_k, L_i, C_m)$ – model for H_k at L_i and camera C_m
- Step 1. Map regions (blobs) into locations for each camera
 - $R_q \rightarrow \langle C, x, y, L, e_q = \{P(R_q | M(H_k, L, C))\} \rangle$
- Step 2. Merge blobs from different cameras
 - $E(R_q) = \{e_{q1}, e_{q2}, \dots, e_{qQ}\}$
- Step 3. Get related IR data, estimate likelihood function, map IR location into camera-based locations, add evidence to related objects
 - $O = \{O_r\}, W = \{w_{kr} \mid k=1, K, r=1, M, O_r\}$
 - $S_r = \{P(R_j, L_k, C_q | H_i)\}$

People Localization: Algorithm (cont.)

- Step 4. Motion estimation

$$\tilde{P}(H_i | L_j) = \frac{\left[\sum_{k=1}^L P(H_i | L_k) \cdot t_{kj}(H_i) \right] \cdot P(H_i | L_j)}{\sum_{l=1}^L \left[\sum_{k=1}^L P(H_i | L_k) \cdot t_{kl}(H_i) \right] \cdot P(H_i | L_l)}$$
- Step 5. Calculate posterior probability

$$P(H_i | O_r, L_k) = \frac{\tilde{P}(H_i | L_k) \cdot w_{kr} \cdot \prod_{P(R_j, L_k, C_q) \in S_r} P(R_j, L_k, C_q | H_i)}{P(O_r)}$$

$$P(O_r) = \sum_{i=1}^N \tilde{P}(H_i | L_k) \cdot w_{kr} \cdot \prod_{P(R_j, L_k, C_q) \in S_r} P(R_j, L_k, C_q | H_i)$$
- Step 6. Normalize over locations

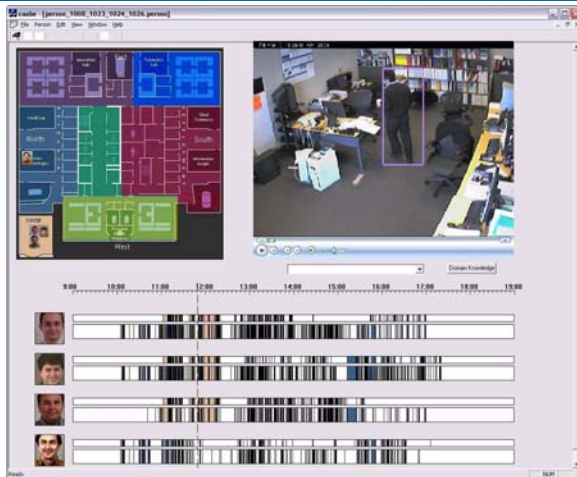
$$P(H_i | L_k) = \frac{P(H_i | O_r, L_k)}{\sum_{k=1}^L P(H_i | O_r, L_k)}$$

People Localization: Experimental Results

- Experiments
 - 4 hours
 - 15 cameras
 - 44 IR sensors
 - Close set
 - 15 “known” people
 - 7 with IR badges
 - Open Set
 - 15 “known” (7 with IR)
 - 10 “unknown”

Person ID	Closed Set Accuracy			Open Set Accuracy		
	Single Camera	Cameras and IR badge	Difference	Single Camera	Cameras and IR badge	Difference
1000	82.14%	87.56%	5.42%	71.39%	77.31%	5.92%
1002	99.27%	99.51%	0.24%	94.54%	95.81%	1.27%
*1003	86.88%	91.07%	4.19%	81.91%	86.10%	4.19%
1005	74.32%	81.69%	7.37%	62.73%	70.59%	7.86%
*1006	45.98%	69.68%	23.70%	43.78%	58.58%	14.80%
*1015	44.08%	41.12%	-2.96%	26.32%	26.58%	0.26%
1020	67.03%	76.71%	9.68%	60.75%	69.24%	8.49%
*1023	64.43%	60.77%	-3.66%	59.82%	58.49%	-1.33%
1024	41.26%	50.72%	9.46%	25.08%	32.78%	7.70%
*1025	69.26%	78.29%	9.03%	57.88%	66.81%	8.93%
1026	71.06%	73.66%	2.60%	41.34%	46.19%	4.85%
*1027	62.04%	73.41%	11.37%	58.13%	67.08%	8.95%
*1029	51.21%	57.88%	6.67%	51.50%	57.21%	5.71%
1064	77.81%	83.42%	5.61%	44.69%	51.69%	7.00%
1072	66.07%	74.49%	8.42%	52.53%	59.67%	7.14%
Average	66.86%	73.33%	6.47%	55.49%	61.61%	6.12%

People Localization: Visualization Tool



Summary

- People Localization
 - Multiple sensor integration
 - Video cameras
 - IR badge ID
 - Fingerprint reader
 - Camera specification
 - Interactive tool
 - Bayesian framework
 - Visualization tool



MSIS Future Work

- Counting people on the floor
- People Tracking
- Merging different sensors for people localization
- People Recognition and Interaction
- Behavior and Activity understanding
 - Ontology of video events
 - Rare Behavior detection
 - Working place statistics
 - Instant Event Reminder



Any questions?

- Eh???
- **Merci beaucoup!**