Exploiting Semantics for Scheduling Real-Time Data Collection from Sensors to Maximize Event Detection

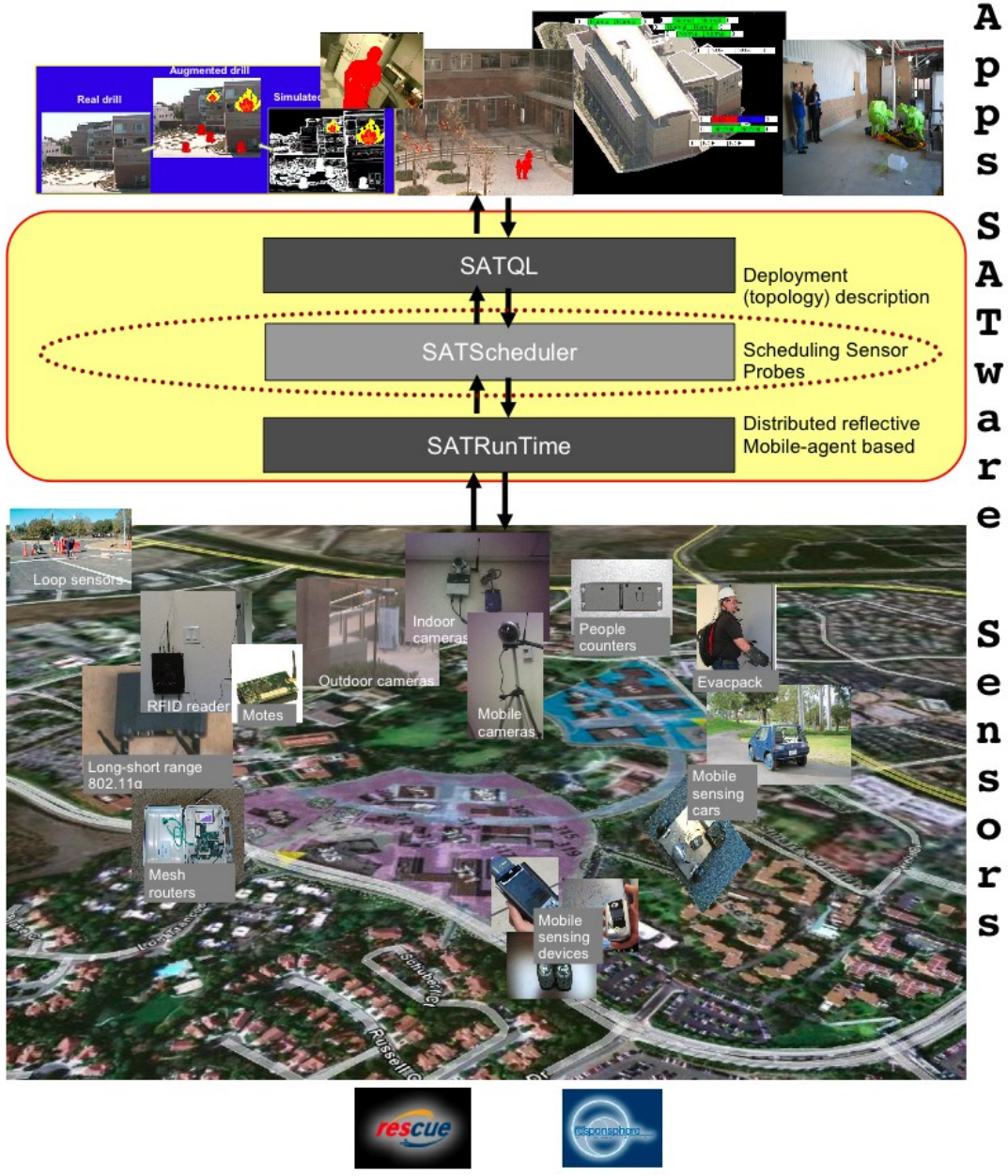
Ronen Vaisenberg, Sharad Mehrotra and Deva Ramanan

School of Information and Computer Sciences
University of California at Irvine

Multimedia and Computer Networks, 2009

UCIrvine
University of California, Irvine

Motivation



UCIrvine
University of California, Irvine

Challenge

- Resource Constraint
  - E.g., Limited network bandwidth, CPU.
- Implication: only a subset of sensors can be probed at a given time.
- Which sensors to probe and when.

Raking the Benefit from a Probe

Certain frames are more important than others from application perspective, which is determined by a Benefit Function (BF).

- Application 1: Real-Time building occupancy. BF = frame without presence ranked lower than all others.
- Application 2: Real-Time person location monitoring. BF = frontal image ranked higher.
- Application 3: Alert potential stolen items BF = person with back-bag ranked higher.

(a) No Presence

(b) Carrying a Bag

(c) Facial

UCIrvine
University of California, Irvine

Architecture

Two possible architectures:

- Push: Perform processing at sensor and send frame to server.
- Pull: Based on global system view, select where to probe.
- We follow the pull model for the following reasons:
  1. BF functions in general are complicated.
  2. Push is not adequate for a real-time application.

The Problem Statement

- Let $S_1, .., S_n$ be n sensors.
- Let **benefit function (BF)** be an application-dependant function based on which probes can be ranked.
- Let **cost function (CF)** be the cost of a probe plan

Specify a probe schedule that maximizes application benefit (BF) without exceeding a cost (CF) limit.

The Problem Statement

Assumptions

(d) BF=0,CF=1

(e) BF=1,CF=1

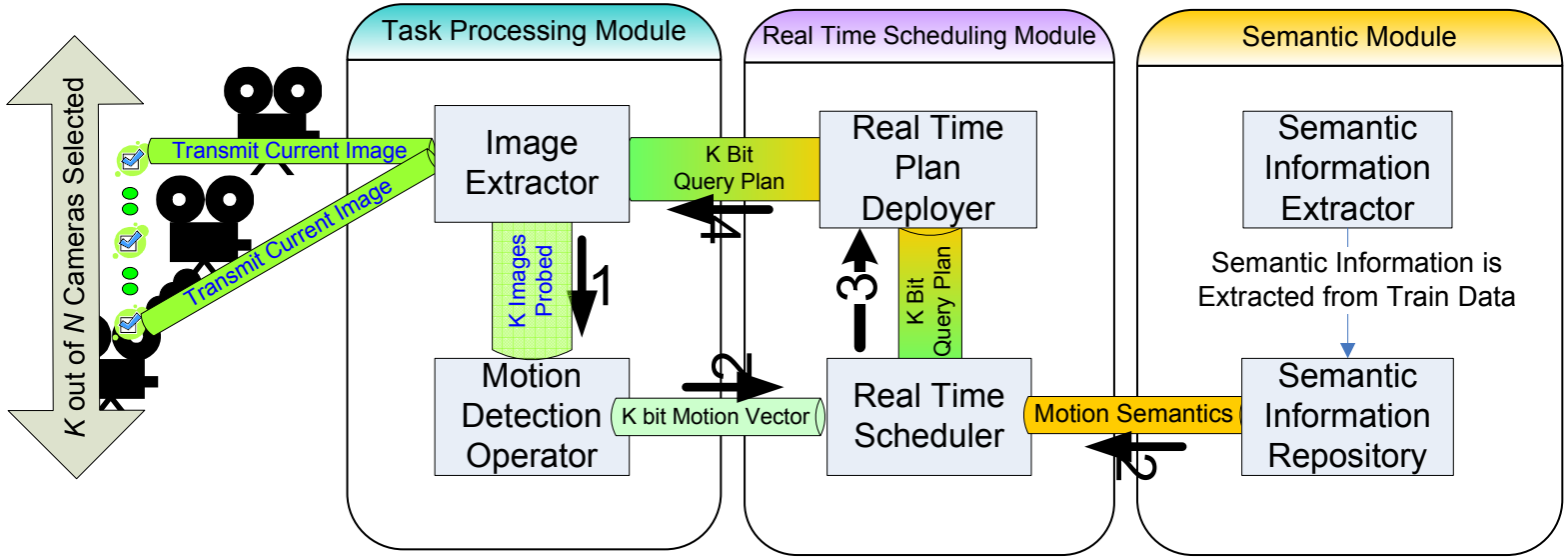
(f) BF=1,CF=1

(g) BF=1,CF=1

- While the framework is more general, in the paper we make two simplifications:
  - **(BF)=1** if frame contains motion, 0 otherwise.
  - **(CF)=1** if sensor probed, 0 otherwise.
- Under assumptions: specify a probe schedule that maximizes $BF(plan)$ while $CF(plan) \leq k$

UCIrvine
University of California, Irvine

Framework



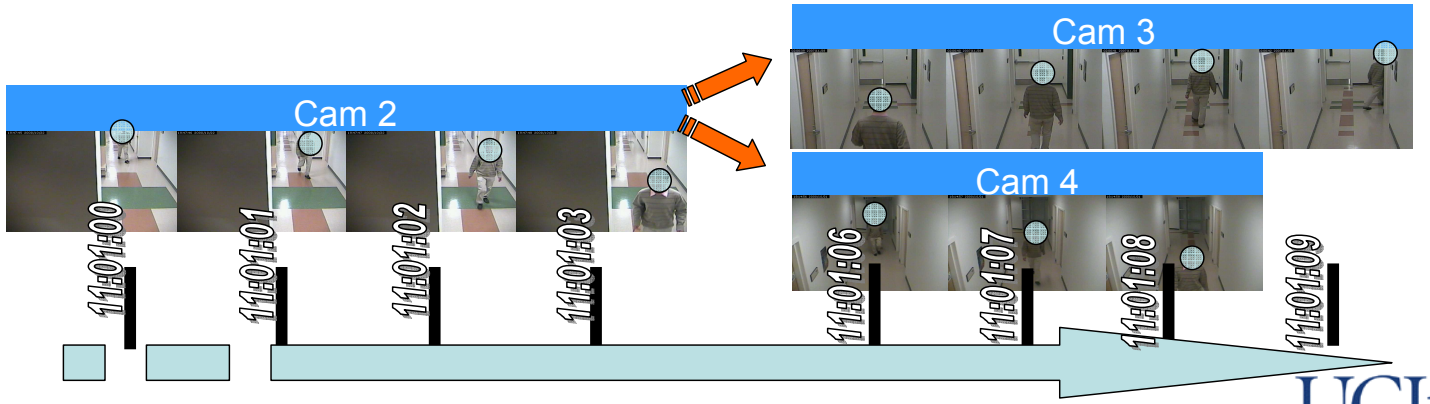
UCIrvine
University of California, Irvine

The approach we adopt is to exploit semantics

- 1. **Apriori** probability of where motion is more likely to initiate.

- 2. **Self correlation** of camera stream over time.

- 3. **Cross Correlations** over time between cameras.



UCIrvine
University of California, Irvine

Key Contributions and Challenges

- 1. A Probabilistic semantic model for unscripted human activity.
- 2. Predicting system state based on constrained, partial current state.
- 3. A fast and light weight real-time scheduler.
  - A delayed probe will miss the action.

UCIrvine
University of California, Irvine

Deterministic Alternatives

1. **RR** A based line algorithm, Round Robin.
2. **RR + Stay** Perform RR, Stay as long as motion is detected.
3. **RR + Stay + Go to Nearest Neighbor** Perform RR, Stay as long as motion is detected when it ends, wait (60 time units) to probe motion at nearest neighbor.

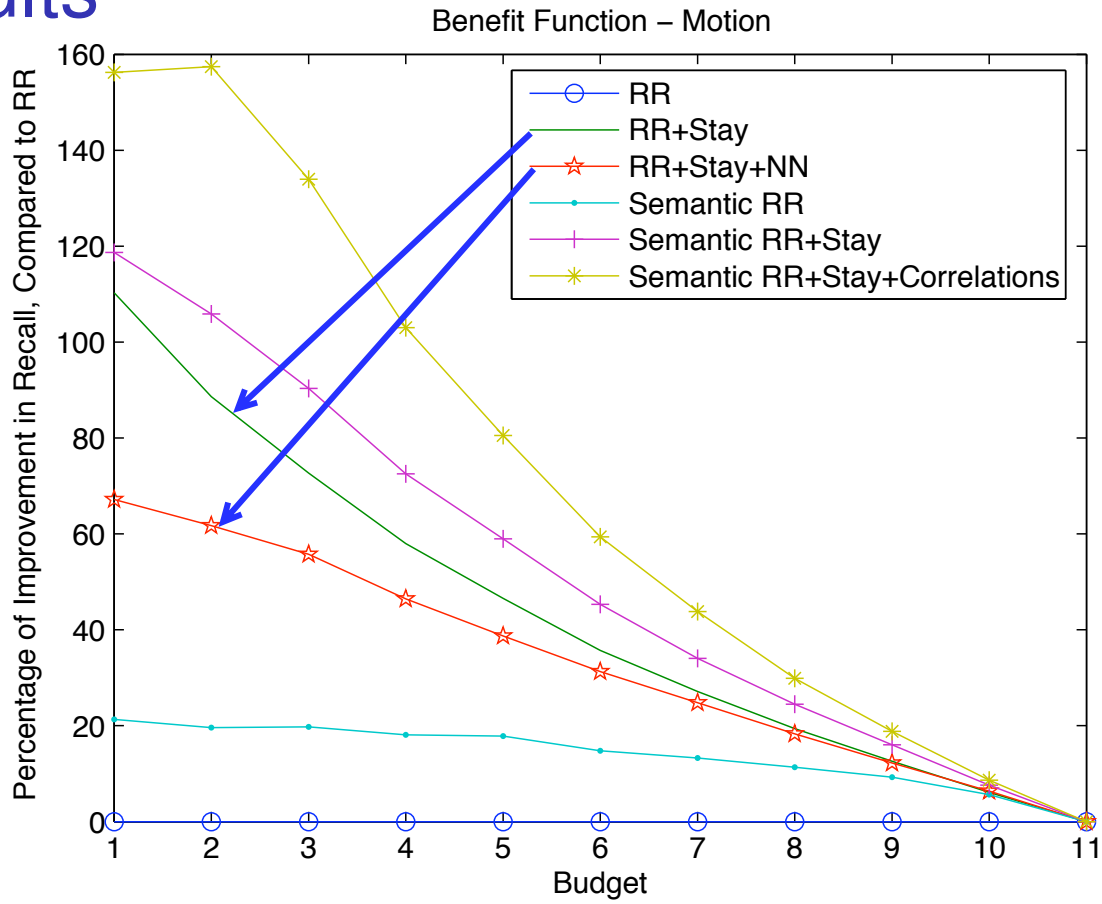
1. Not trivial to do better than RR+Stay, following a deterministic algorithm.

Evaluation Setup

- We evaluated our algorithms on a network of 11 cameras distributed on two floors in the CS building at UC Irvine.
- The Data we collected was over a period of two weeks (the first and second weeks of the Spring quarter).
- All cameras were located in hallways.
- We then used the first week to train for optimal collection of the other week.

UCIrvine
University of California, Irvine

Some Results



- "RR+Stay+NN" performed *much worse* than "RR+Stay". – an ad-hoc approach will prove to perform much worse than a simple alternative("RR+Stay").
- The probabilistic approach "Semantic RR" proved to work better then RR but worse than all other semantic alternatives.

UCIrvine
University of California, Irvine

Semantic RR

- Offline: from train data,compute for each sensor:

$$\frac{cases\ with\ motion}{cases\ with\ or\ without\ motion}$$

- Online: Regardless of the actual probes, choose k nodes based on probability distribution.

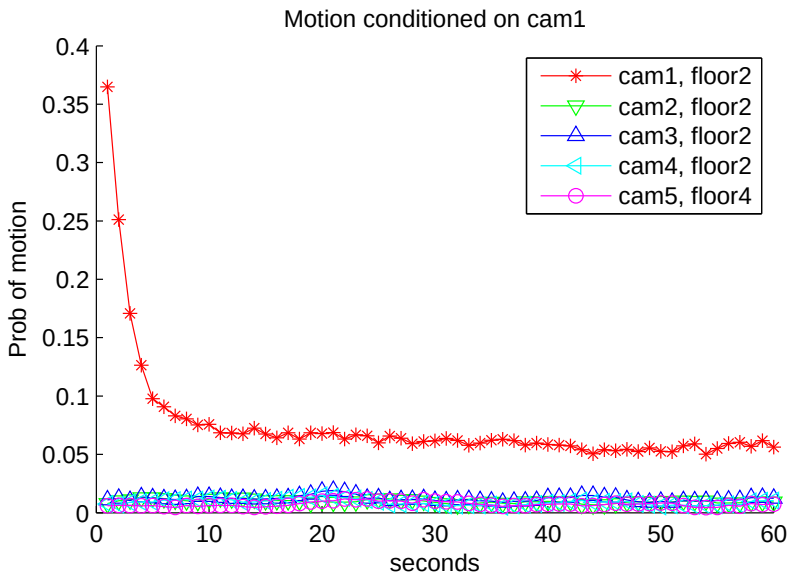
Semantic RR+Stay

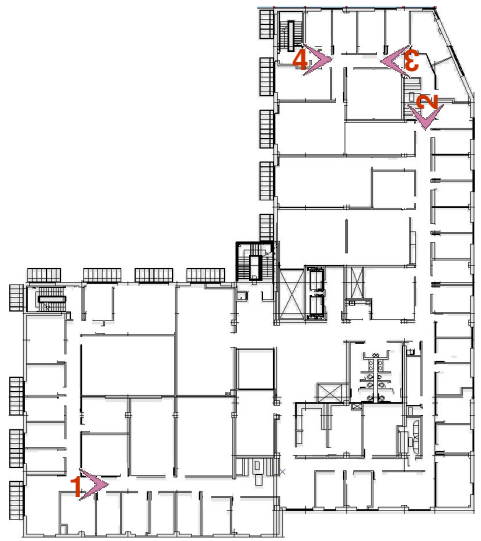
We assume a *Hidden Markov Model* (HMM) of sensor states and observations.

- Offline: from train data, compute for each sensor:

$$P(s_t|s_{t-1})$$

- Online: Based on the actual probes, dynamically compute probabilities for the next time unit and choose k nodes based on probability distribution.

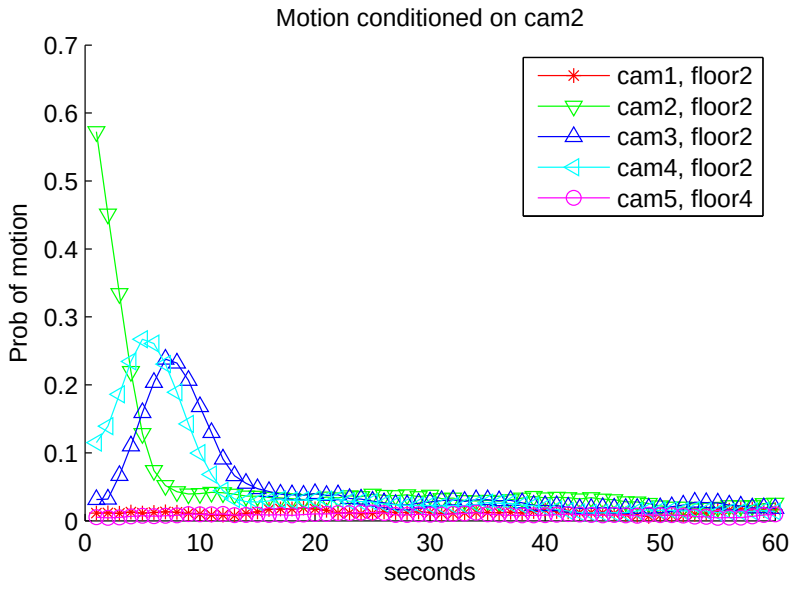


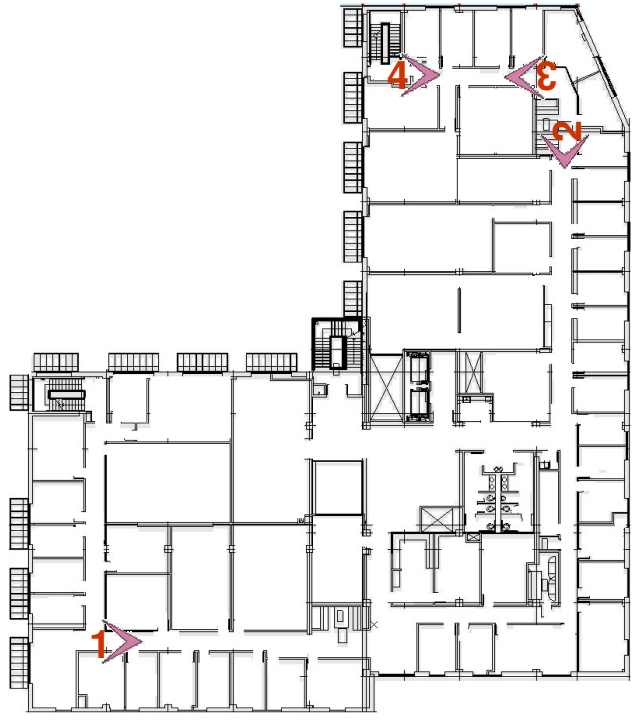


UCIrvine
University of California, Irvine

Semantic RR+Stay+Correlations(t=1)

We would like to capture interactions between N cameras:





$$P(s_t^i | s_{t-1}^{1:N}) \quad (1)$$

- Superscripts denote cameras, subscripts - time.

Exponential Blowup:

- The conditional probability table has 2^N entries.

UCIrvine
University of California, Irvine

Semantic RR+Stay+Correlations(t=1)

Making it Scale

We assume that people move independently throughout our camera network.

- We want to intuitively "add" such probabilities across i to compute $P(s_t^j)$.
- The transition model now simplifies to:

$$P(s_t^i | s_{t-1}^{1:N}) = 1 - \prod_{j=1}^{N} (1 - \alpha_{ij} s_{t-1}^j) \tag{2}$$

1. α_{ij} - the probability that a person seen at camera i moves to camera j in the next timestep.
2. The above model is usually refereed to as a *noisy-OR* model.

The above model has $O(N^2)$ parameters - scales.

UCIrvine
University of California, Irvine

Semantic RR+Stay+Correlations(t=M)

- Again, a native implementation of $P(s^i_{t+1}|s^{1:N}_{(t-M):t})$ - exponential.

With similar reasoning, the transition model simplifies to:

$$P(s^i_t|s^{1:N}_{(t-M):(t-1)}) = 1 - \prod_{o=1}^{M}\prod_{j=1}^{N}(1 - \alpha^o_{ij}s^j_{t-o}) \tag{3}$$

where α^o_{ij} - the probability that a person in camera i moves to camera j, o timesteps later.

The above model has $O(M * N^2)$ parameters - scales.

Semantic RR+Stay+Correlations(t=M)

The M^{th}-Order model states that if there is no motion activity in any camera in the past M-seconds, there can be no motion in any camera currently.

- To overcome this limitation, we allow for a probability α_i that a person will spontaneously appear at camera i - this is equivalent to a ***leaky* noisy OR.**
- We add in $(1 - \alpha_i)$ as an extra term:

$$P(s_t^i | s_{(t-M):(t-1)}^{1:N}) = 1 - (1 - \alpha_i) * \prod_{o=1}^{M} \prod_{j=1}^{N} (1 - \alpha_{ij}^{o} s_{t-o}^{j}) \qquad (4)$$

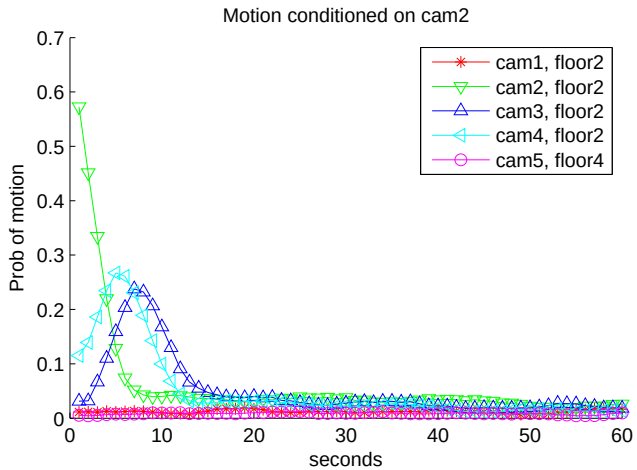
UCIrvine
University of California, Irvine

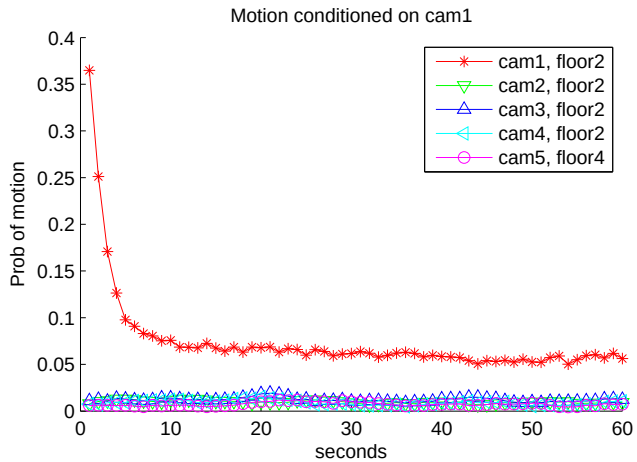
Semantic RR+Stay+Correlations(t=M)

Implementation

Further optimization:

- We observe that most of the α_{ij}^{o} dependencies are weak (far-away cameras do not effect each other).
- We zero out the α_{ij}^{o} values below some threshold.
- In our case, thresholding on 0.05 reduced table from 1815 to 165 entries.





UCIrvine
University of California, Irvine

Semantic RR+Stay+Correlations(t=M)

Implementation

While experimenting we identified two pitfalls:

1. First we choose to probe based on **sampling**, as opposed to probing based on the order of probabilities (high to low).
2. The second relates to the history table, only motion probed should be used for prediction.
   - When predicting based on probability for motion -

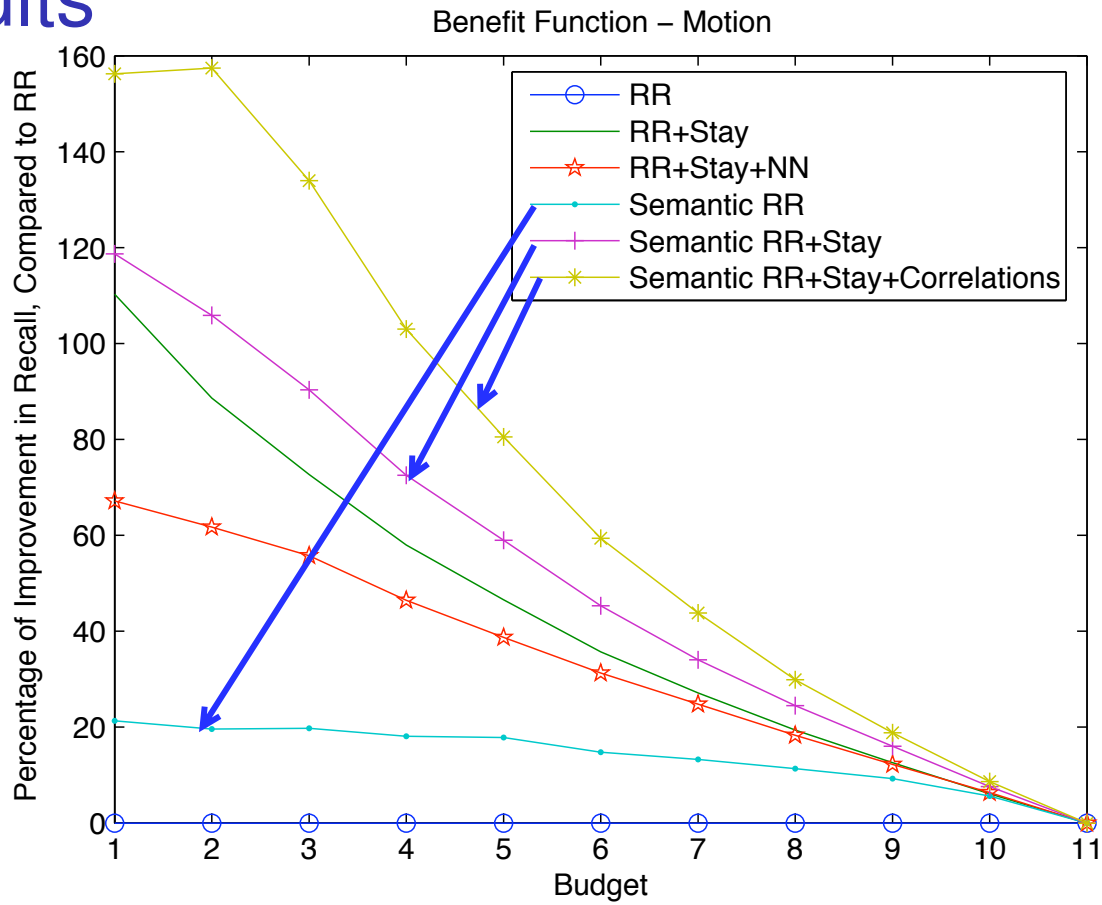
plan	C_{10}	C_{10}	C_9	C_{10}	C_{10}	C_9	C_9	C_9
Cam9 Projected Prob.	0.11	0.26	0.33	0.51	0.66	0.65	0.65	0.67
Cam10 Projected Prob.	0.45	0.66	0.69	0.79	0.875	0.88	0.9	0.916
time	1	2	3	4	5	6	7	8
result	+	+	-	+	+	-	-	-

   - When basing decision only on probed motion -

plan	C_10	C_{10}	C_9	C_{10}	C_{10}	C_9	C_9	C_5
Cam9 Projected Prob.	0.11	0.21	0.25	0.34	0.42	0.43	0.41	0.38
Cam10 Projected Prob.	0.45	0.65	0.56	0.71	0.8	0.67	0.63	0.6
time	1	2	3	4	5	6	7	8
result	+	+	-	+	+	-	-	-

UCIrvine
University of California, Irvine

Some Results



- With a budget of 1, our last semantic algorithm achieved over 150%(!) improvement over RR.
- With budget of 50%, the last semantic algorithm achieves about 90% recall (compared to about 50% of RR).
- Demo.

UCIrvine
University of California, Irvine

Summary

- We designed and implemented a fully functional system using available off-the-shelf cameras to learn motion semantics.
- Our semantic based algorithm that takes into account all available semantic information proves that even under significant resource constraints, we can detect a very large number of events.
- The scheduling algorithm is *very* light weight and is suitable for a real-time application.

UCIrvine
University of California, Irvine

Outlook

- Experimenting with other **Benefit Functions**.
  - Instead of 0/1 no motion/motion.
- Take into account different **Cost Functions**.
  - Probe in low/high resolution.
- Updating the Semantic Model while the system is in operation.
  - Relates to a problem addressed by the database communitee: estimating query selectivity based on the actual query results.

Thank You

Thank You!
.. Any questions?

UCIrvine
University of California, Irvine