Advances in Smart Camera Networks

Stanislava Soro

SENSORS Group Meeting

November 27. 2007

Smart Camera Networks vs Visual Sensor Networks

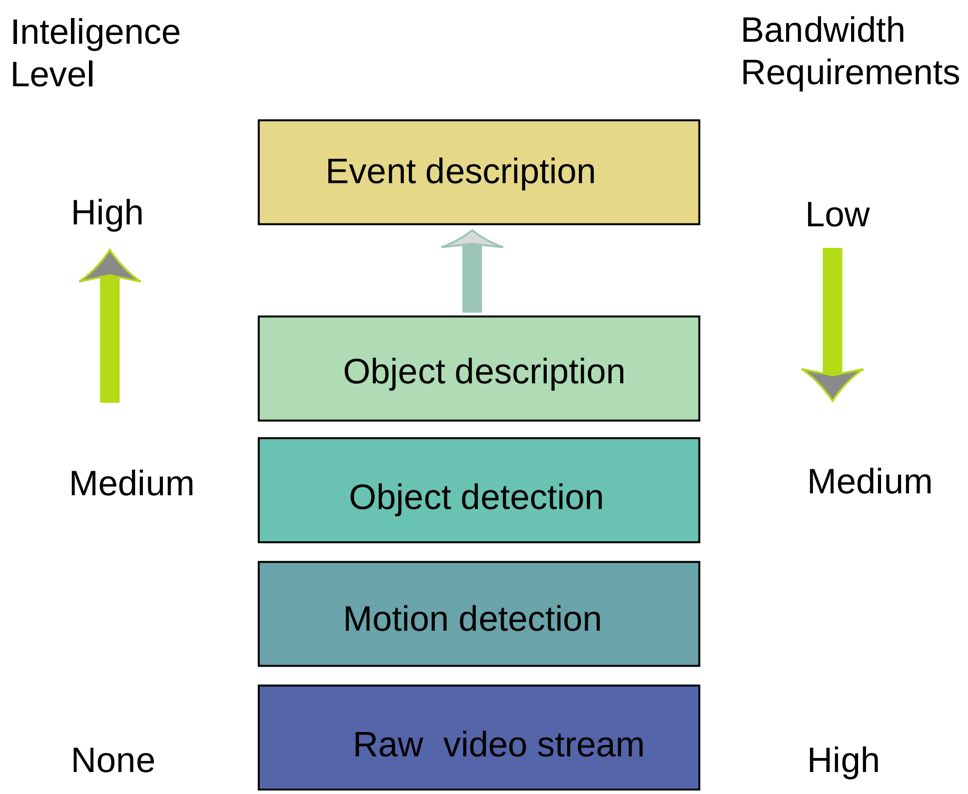
Smart Camera Networks

- Embedded processing
- Complex algorithms
- Wireless or wired cameras
- Energy consumption is not the main concern
- Performance is the main concern

Visual Sensor Networks

- Energy and bandwidth constraints
- Low {power, processing, resolution}
- Wireless data transmission

Design trade-offs for Smart Camera Node

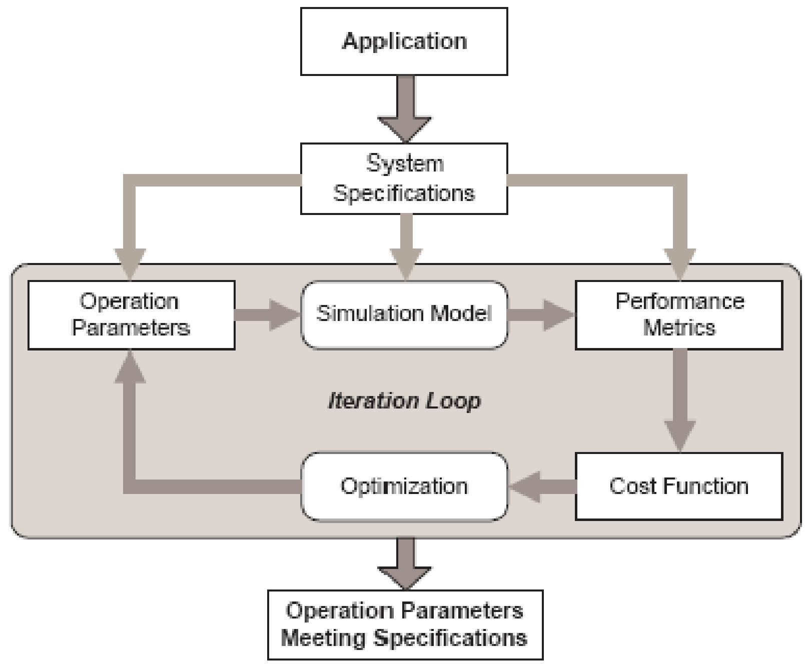


Smart Camera Networks – Research Directions

- New hardware smart camera architectures
- Application oriented design of smart cameras
- Scene analysis
  - face orientation estimation
  - determining dominant motion in crowded scenes
  - vehicle classification
  - people counting
- Power-aware distributed camera management
- Camera network architectures
- Tracking and surveillance
  - object reacquisition, object reoccurrence…
  - audio-visual tracking
  - object tracking with camera clusters
- Middleware for smart camera network
- Behavior analysis (detection of abnormal people trajectory/action)
- Distributed coding

Application Oriented Design of Smart Camera Networks [1]

Operation Parameters	Performance Metrics
Network topology	Detection accuracy
Vision system (single, stereo camera)	Detection delay
Camera resolution	Detection reliability
Camera poll rate	Network lifetime
Computation resources	Network throughput
Network bandwidth	
Calibration error	
Energy resources	



1 Application-Oriented Design of Smart Camera Networks, Stephan Hengstler (Stanford University, US); Hamid Aghajan (Stanford University, US)

Probabilistic Observation Model

- Single object tracking in camera network with N cameras
- Distributed Bayesian filtering
- MMSE Tracker

$$\hat{x}^{(t)}_{MMSE} = E[x^{(t)}|\overline{\underline{z}^{(t)}}] = \int x^{(t)} p(x^{(t)}|\overline{\underline{z}^{(t)}}) dx^{(t)},$$

- minimize the avg. distance between location estimate $x_{MMSE}(t)$ and the true location x(t) at discrete time t from z(t) – measurement history in time

- Using Bayesian sequential filtering, for every new measurement z(t+1) updated belief

Joint likelihood

Next believe state

$$p(x^{(t+1)}|\overline{\underline{z}^{(t+1)}}) \propto p(\underline{z}^{(t+1)}|x^{(t+1)}) \cdot \int p(x^{(t+1)}|x^{(t)}) \cdot p(x^{(t)}|\overline{\underline{z}^{(t)}}) dx^{(t)}$$

Object's dynamics

Current believe state

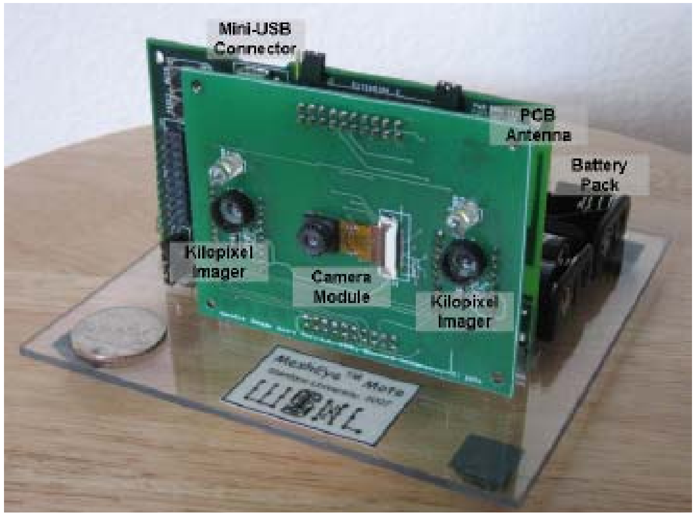
- Mutually independent measurements from the cameras =>

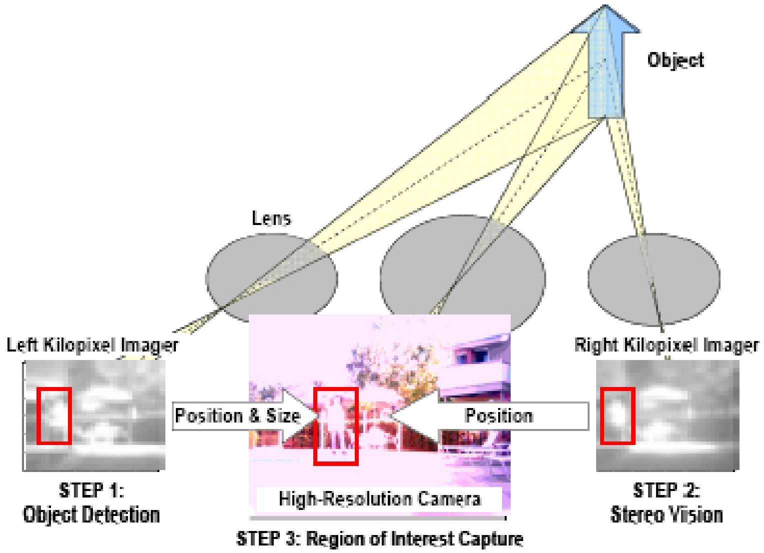
$$p(\underline{z}^{(t+1)}|x^{(t+1)}) = \prod_{n \in \mathbf{S}^{(t+1)}} p(z_n^{(t+1)}|x^{(t+1)}).$$

Individual likelihood f-ns

Camera Vision Model

- Likelihood functions are derived for single camera and stereo camera models
- MeshEye Mote [2] -- stereo camera consisted of:
  - 2 kilopixel imagers 30x30 pixels grayscale
  - VGA, color camera



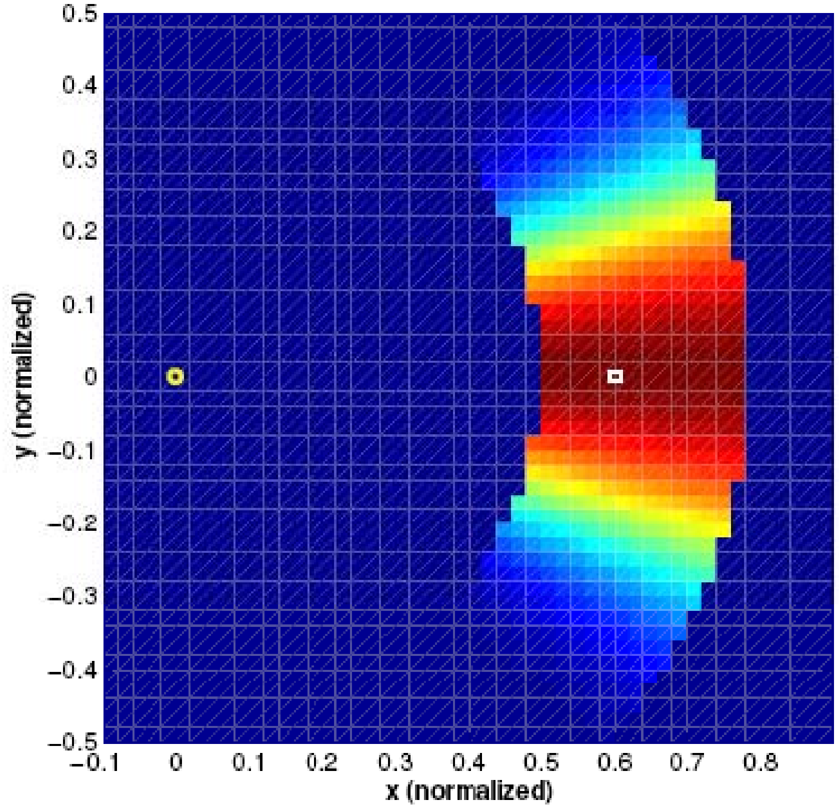


2 S. Hengstler, D. Prashanth, S. Fong, H. Aghajan, "MeshEye: A Hybrid-Resolution Smart Camera Mote for Applications in Distributed Intelligent Surveillance," Proc. of IPSN 2007.

Likelihood Function

- P(z/x) for stereo camera

Circle – stereo camera location
Square – object true location



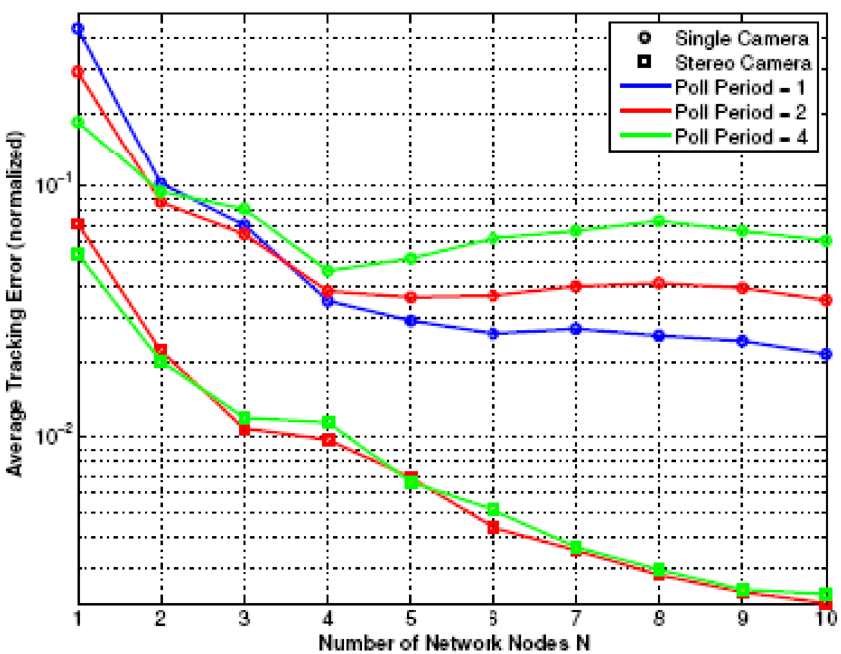
Tracking Performance

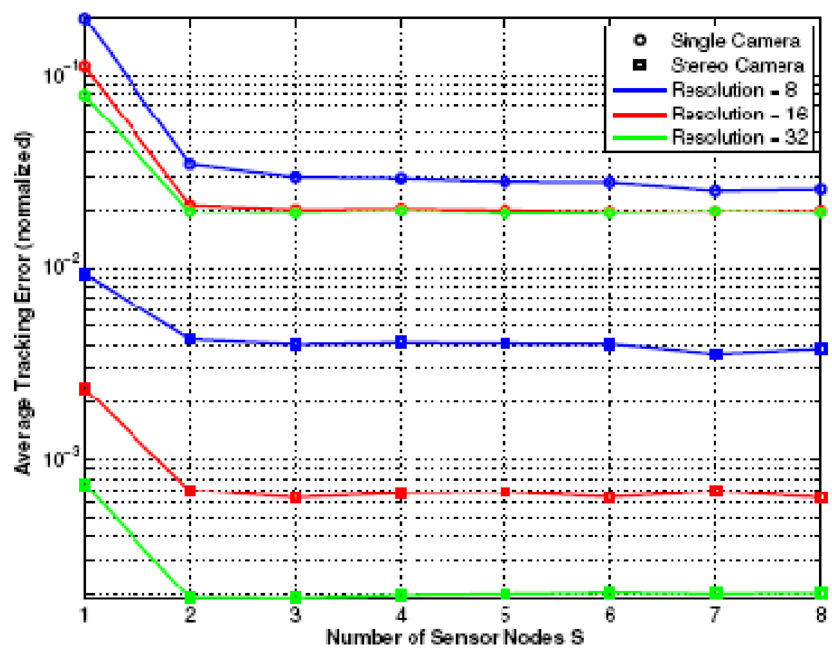
Tracking performance for different polling rate

- Single Camera:
  - Error minimized for N=5
  - Error doubles with increasing polling period
- Stereo Cameras:
  - Insensitive to poling
  - Single stereo camera good as 3-4 single cameras

Tracking performance for different camera resolution

- Collaboration between two cameras improves tracking accuracy for 54%
- Stereo cameras
  - Doubling cameras resolution results in ¼ of tracking error





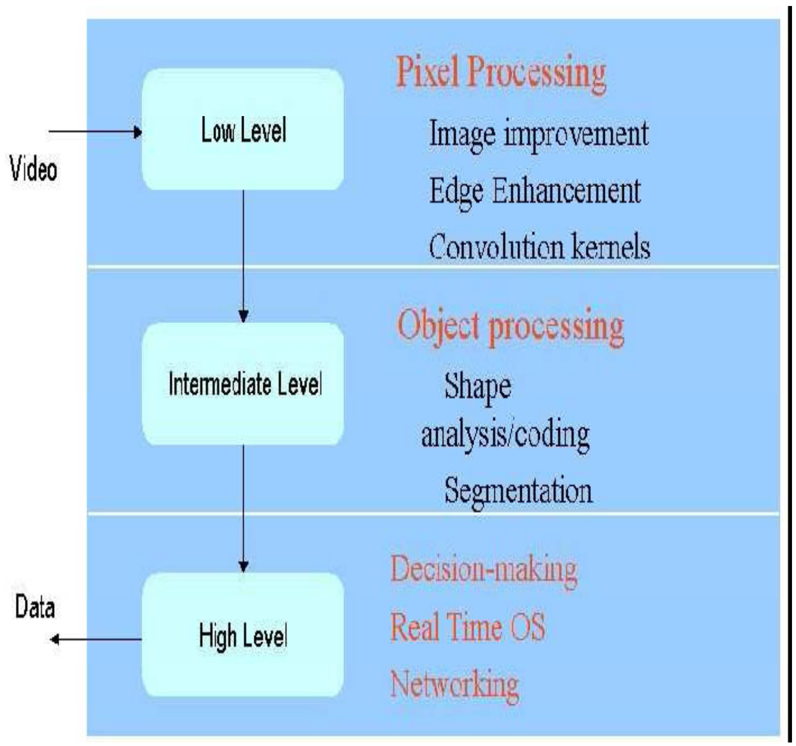
SIMD Architecture of Smart Cameras [3]

- Support for real-time vision applications
- Advanced video processing: extensive computational power, increasing complexity of algorithms, high pixel rates
- Explore *processors with parallel processing architecture*
  => Single Instruction Multiple Data (SIMD) processors
  - Uses vector data type to represent 320 element line
- Challenge: Mapping existing vision algorithms to embedded processor with parallel processing architectures

3 Mapping Vision Algorithms on SIMD Architecture Smart Cameras, Chen Wu (Stanford University, US); Richard Kleihorst (NXP Semiconductors, NL); Hamid Aghajan (Stanford University, US)

Camera Hardware Platform

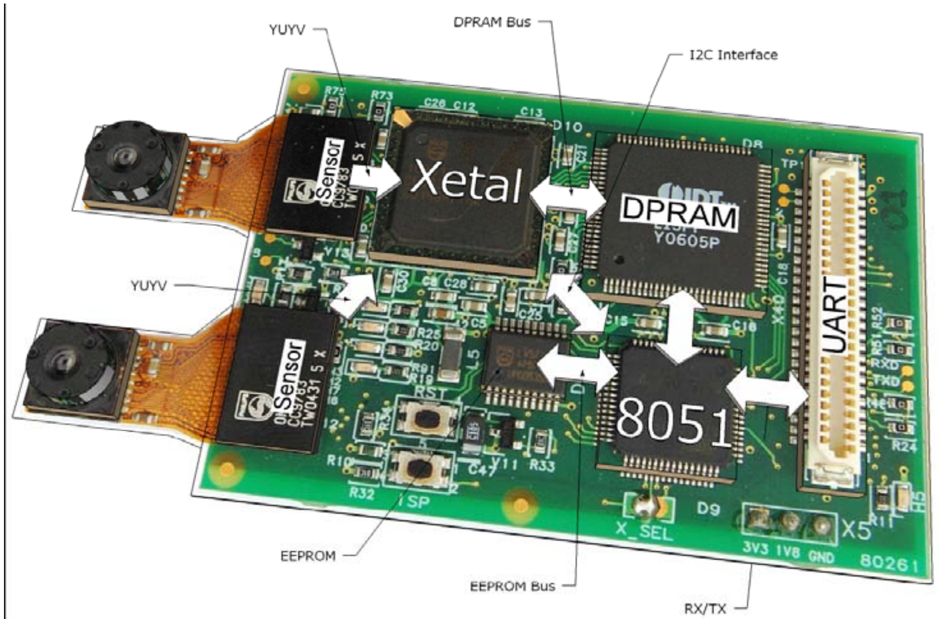
- Two types of processors:
  - SIMD parallel processor
  - General purpose DSP



- Low level processing
  - The algorithms per pixel are essentially the same
  - Processing of up to billion pixels/sec
  - use parallelism to operate on more pixels per cycle
- High- and intermediate level processing
  - Complex software tasks
  - Running OS
  - networking

WiCa Camera Mote [4]

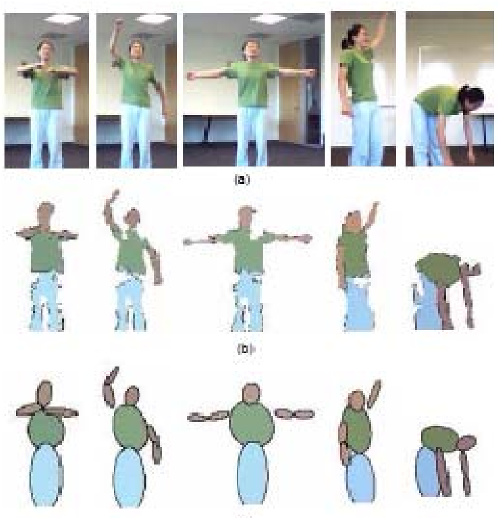
- Real time frame based video processing
- Xetal-II processor
- 80 MHz
- 600 mW at peak performance
- 10 Mb of internal memory



4 Camera Mote with a High-Performance Parallel Processor for Real-Time Frame-Based Video Processing, Richard Kleihorst (NXP Semiconductors, NL); Anteneh Abbo (Philips, NL); Ben Schueler (NXP Semiconductors, NL); Alexander Danilin (NXP Semiconductors, NL)

Reconstruction of Person’s 3D Model

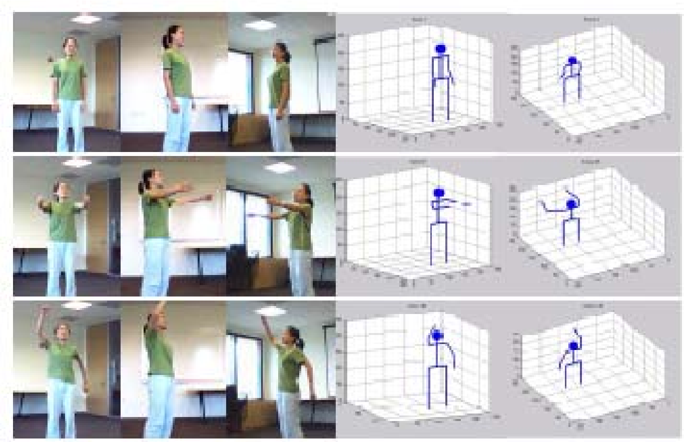
- Consists of the following steps:
  - Detect geometric configuration
  - Detect color of body parts
  - Detect motion of body parts
- Algorithm with two steps:
  - Local processing part
    - Segmentation (k-means clustering), ellipse fitting
  - Collaboration between cameras
    - Ellipses from all cameras are merged to find configuration of 3D skeleton model



Original images

Segments

Fitted ellipses



Coordinated Distributed Power Management in VSN [5]

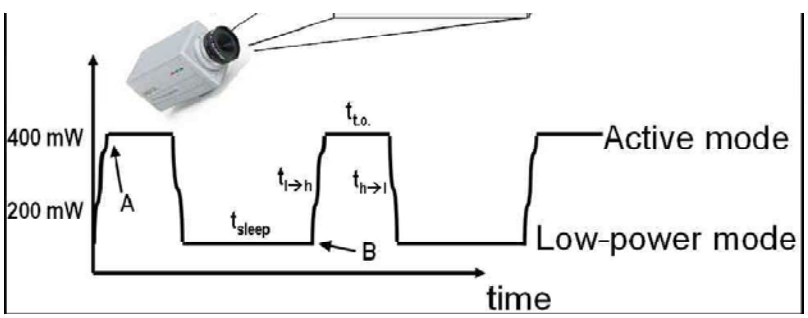
- Nodes consider the content of image data sensed locally and by neighboring nodes in order to decide about working mode (sleep, active)
- Motion detection and estimation
  - Node detects motion using sum of absolute differences (SAD) for two successive frames
  - Mean shift algorithm to identify objects and estimate motion
- Prototype VSN
  - Zigbee IEEE 802.15.4 transceiver
  - 30 frames/s
  - TI DSP
  - Mesh-like networking

5 Coordinated Distributed Power Management with Video Sensor Networks: Analysis, Simulation, and Prototyping, Nicholas Zamora (Carnegie Mellon University, US); Radu Marculescu (Carnegie Mellon University, US)

Power Management Polices

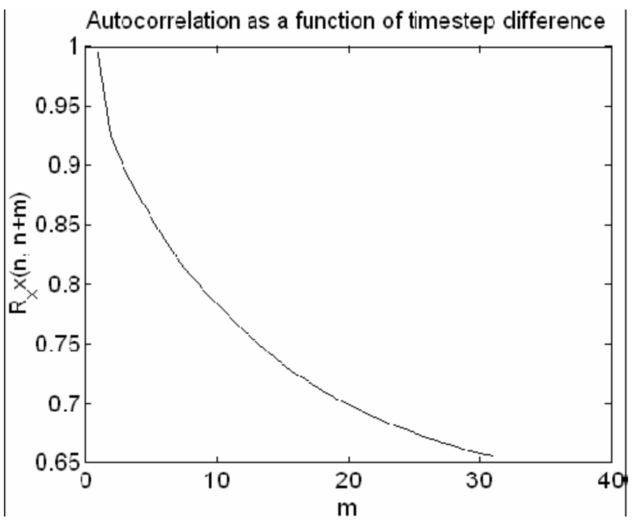
Each video node makes its own decision while considering the decisions from neighboring nodes

- Coordinated timeout :
  - After entering active mode, broadcast this info to neighbors
  - If the motion is not detected for some threshold time and all neighbors are in time out, the node enters sleep mode
- Voting policy:
  - Each node periodically broadcast the summary of motion detected
  - If enough neighbors decide that there is no motion, then node enters sleep mode

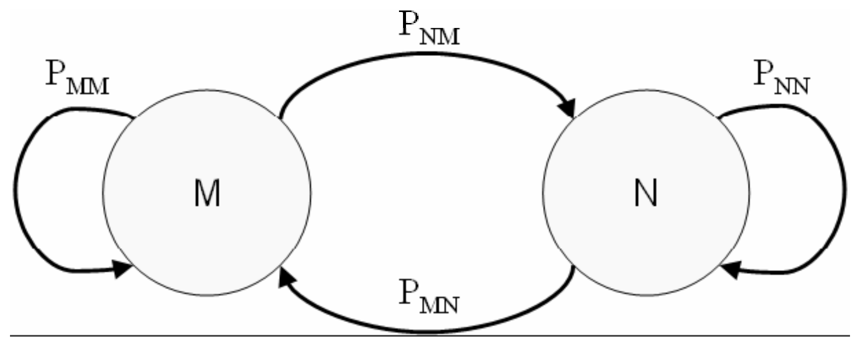


Model to capture burst motion

- Video is highly correlated – motion in one frame implies the high probability of motion in the next frame



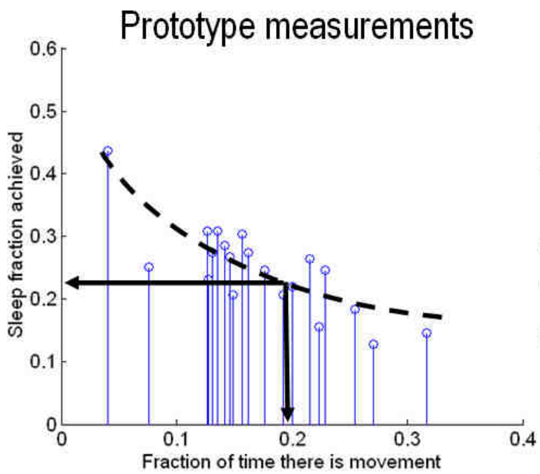
- Likelihood of movement in next m video frames, if the current video frame contains the movement (m- number of frame)

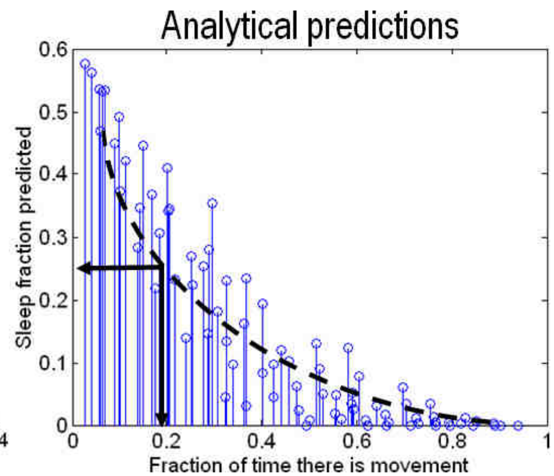


- Two state model to describe burst motion (Gilbert Elliot Model)
- This model predicts the performances of PM polices
- Node analyzes each frame:
  - Frames with the detected motion are in the state M
  - Frames with no motion are represented with N

Results

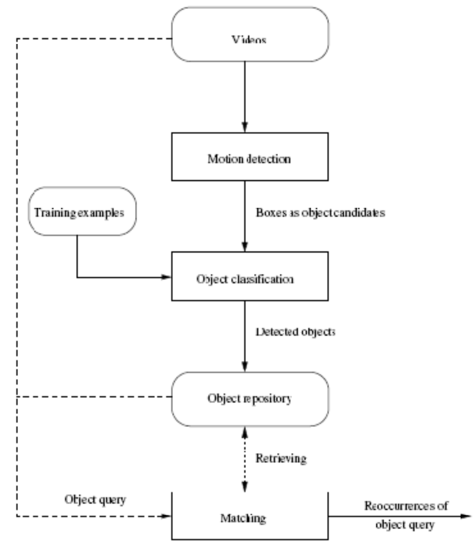
- Measure the fraction of *sleep time* obtained analytically (described model) and through the experiments
- Analytical prediction of sleep time – vary Pmm and Pnn (0.92-0.995)
- Prototype based results – the nodes execute coordinated timeout policy
- Consistency of analytical and practical model





- Better results in the case when additional states are used to capture movement within 2 previous frames

Tracking and Surveillance with Smart Camera Network



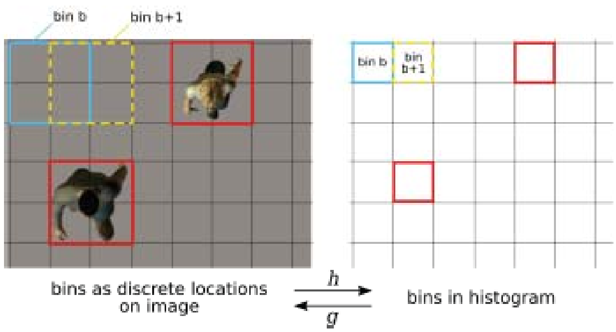
- Object reacquisition and re-identification
  - process of matching objects between images taken with different cameras
- Object reoccurrence in multi camera network
  - after event scenarios
  - provide help to police to quickly identify the occurrence of the subject of interest
- Audio Visual Tracking using STAC sensors
  - STAC sensor is composed of a single camera mounted between two microphones
  - estimation of time difference of arrival (TDOA) between the two microphones
  - Target localization (performed using color-based change detection) is integrated with audio estimates

Camera Network Architectures

- Distributed tracking and activity recognition in assisted living applications
  - By 2040 the population over age 85 will triple (from 4.5 to 14.2 millions)
  - After 22 minutes an operator misses up to 95% of scene activities
  - Tracking and activity recognition are embedded at the camera
  - Currently supports detection of activity such as falling person
  - Activity detection requires training sequences
  - Embedded activity recognition using classification with Support Vector Machines (SVM)

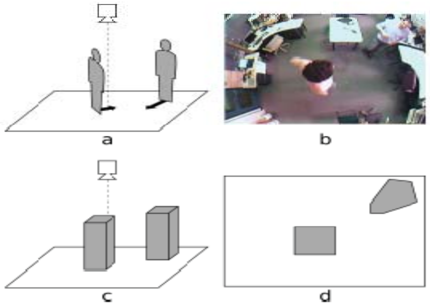
SmartClassySurv -- A Smart Camera Network for Distributed Tracking and Activity Recognition and its Application to Assisted Living, S. Fleck, R. Loy, C. Vollrath, F. Walter, W. Strasser (University of Tübingen, DE)

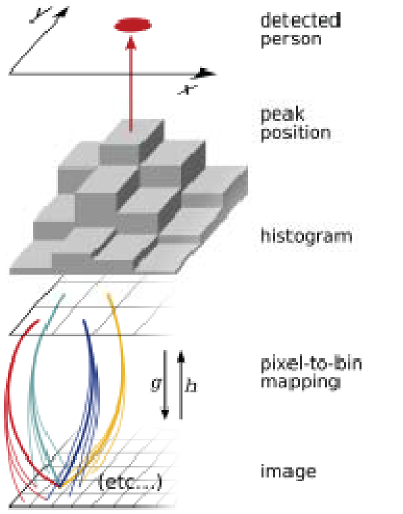
People Counting and Localizing in Indoor Spaces



- Applications: assisted living, safety, entertainment…
- Set of wide-angle cameras mounted at the ceiling
- Image divided into overlapped bins (histogram bin)
- Bin size found from the expected size of the human
- Value of each histogram bin equals to the number of foreground pixels in this bin
- Motion detection – frame differencing

- Optimizing the histogram – wide angle consideration
  - Due perspective and lens distortion
  - Human model – 3D cuboid
  - Motion histogram is calculated for the bin that is the projection of 3D cuboid on image plane





Lightweight People Counting and Localizing in Indoor Spaces using Camera Sensor Nodes, Thiago Teixeira (Yale Univesity, US); Andreas Savvides (Yale University, US)

Histogram-based Counting

- The algorithm does not uniquely identify the persons
- Persons are identified in the small periods of time, until some ambiguity happens
- Histogram can have more than one peak – depends on the number of people and of bin's size
- Each person is presented by the feature vector $\langle \mathbf{s}_{ti}, h_{ti}, b_{ti}, \mathbf{v}_{ti}, \mathbf{a}_{ti} \rangle$ (peak position, peak height, peak breadth, velocity and acceleration)
- People's positions are predicted using the current values for position, velocity and acceleration

$$\mathbf{s}_{t+1} = \mathbf{s}_t + \mathbf{v}_t t + (\mathbf{a} t^2)/2$$

- For each person find the matching peak:
  - Difference measure between the predicted values and values from the histogram

$$d_{ij} = \beta_s \|\hat{\mathbf{s}}_{ti} - \mathbf{s}_{tj}\| + \beta_h |\hat{h}_{ti} - h_{tj}| + \beta_b |\hat{b}_{ti} - b_{tj}|$$

Counting Results

- Testbed with Intel iMote2 sensor nodes coupled with omniVision OV7649 cameras (320x240 downsampled to 80x60 pixels)
- Processing and sending requires about 110ms (~ 8frames/s)

- Example:
  - person(s) walk and stands still => the network correctly counts the number of people in the room for 89.1% (1 person), 82% (two persons) and 79.8% (three persons)
  - This is followed by the increase on the packet drop rate, since all cameras want to tx at the same time

  - Demo at:

  http://www.eng.yale.edu/enalab/behaviorscope/counting.htm

Clustering in Wireless Camera Networks

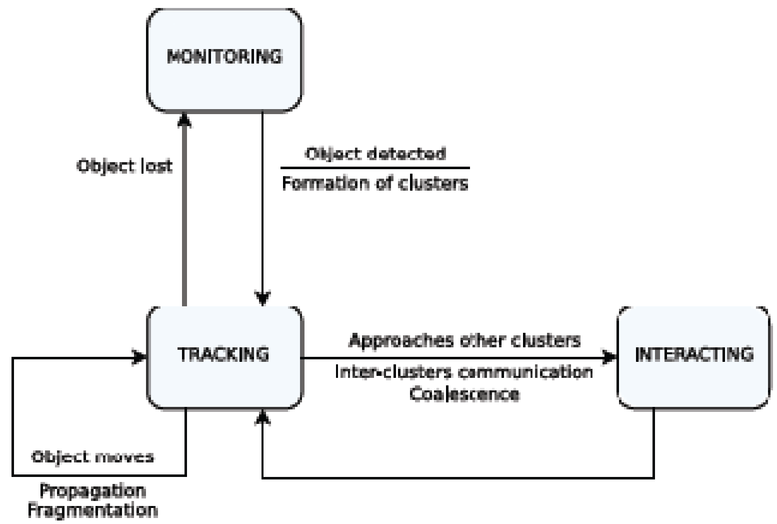
- Object tracking
- Clustering triggered by detection of object

- Cluster head election
- Clustering propagation

- Experiments with Micaz nodes
with attached Cyclopes cams

How the clustering is done in the
Case of multiple targets?

What the cluster members do?



A Light-Weight Event-Driven Protocol for Sensor Clustering in Wireless Camera Networks, Henry Medeiros (Purdue University, US); Johnny Park (Purdue University, US); Avinash Kak (Purdue University, US)

Resources

- ACM/IEEE 1st International Conference on Distributed Smart Cameras 2007

http://www.icdsc.org/

- 2007 IEEE International Conference on Advanced Video and Signal based Surveillance

http://www.elec.qmul.ac.uk/staffinfo/andrea/avss2007.html