Depth and Appearance for Mobile Scene Analysis

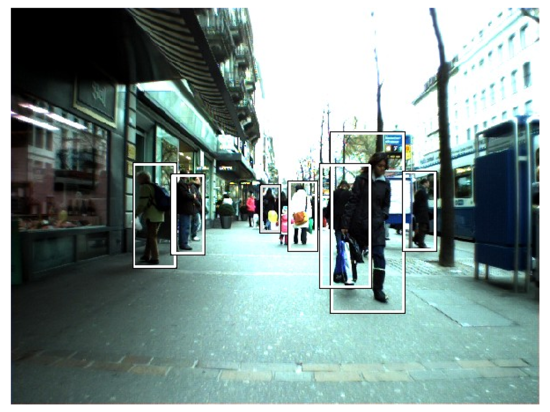
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Andreas Ess[1], Bastian Leibe[1], and Luc Van Gool[1,2]

[1]ETH Zurich, Switzerland, [2]Katholieke Universiteit Leuven, Belgium
{aess, leibe, vangool}@vision.ee.ethz.ch, Luc.VanGool@esat.kuleuven.be

Motivation and Challenges

- Detect pedestrians from a moving platform
- Available cues using pair of cameras
  - Pedestrian detection (bounding box, top-down segmentation)
  - Depth map
  - Additional geometric information
- Use combination of cues to increase reliability
  - Coupled detection and ground-plane estimation
- Causal system: use only current/previous frames
- Challenges
  - Unconstrained video data
  - Large number of moving objects
  - Frequent partial occlusions
  - Motion and bayering artefacts
  - Large range of scales
  - Suboptimal camera placement

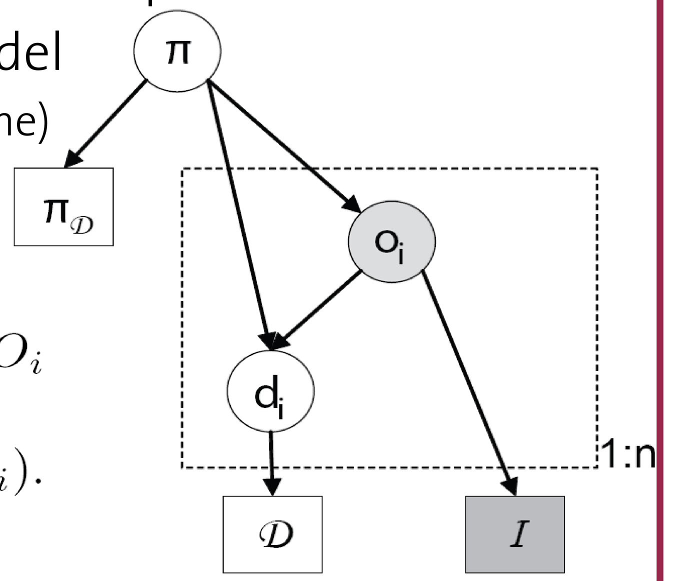


Our approach

- Combine object detections, ground plane, and depth cues
- Formulate dependencies in a graphical model
  - Find mutually best explanation of scene (per-frame)
  - Result can be plugged into tracking later
- Decomposes into

$$P(\boldsymbol{\pi}, o_i, d_i, \boldsymbol{\pi}_{\mathcal{D}}, \mathcal{D}, \mathcal{I}) = P(\boldsymbol{\pi}) P(\boldsymbol{\pi}_{\mathcal{D}}|\boldsymbol{\pi}) \prod_i^n O_i$$

$$O_i = P(o_i|\boldsymbol{\pi}) P(\mathcal{I}|o_i) P(d_i|o_i, \boldsymbol{\pi}) P(\mathcal{D}|d_i).$$



Formulation (pt. 1)

Ground plane $\boldsymbol{\pi}$

- Useful means for constraining object detection
- Based on prior, depth maps and object detection
- Discretized into 6 x 6 x 20 bins, with prior from training sequence

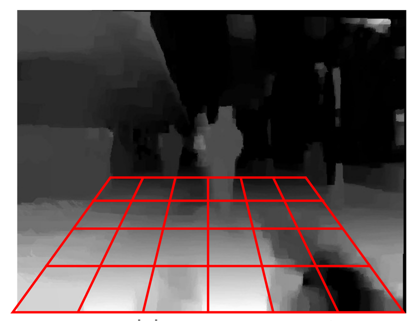
Ground plane measurements

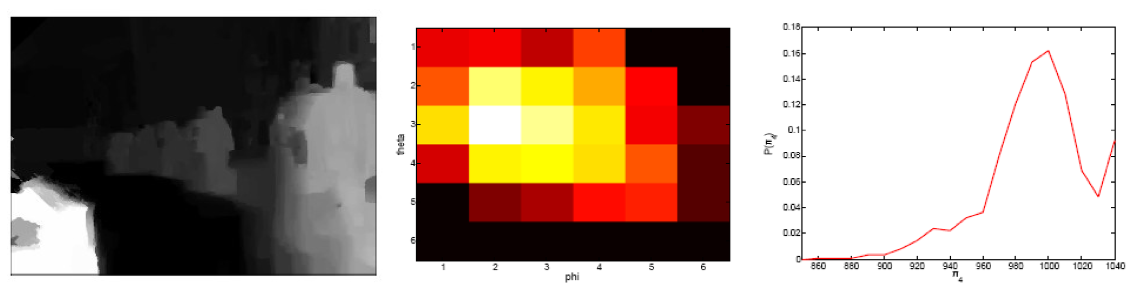
- Quality assessed using robust measure on $\mathcal{D}$

$$r(\boldsymbol{\pi}, \mathcal{D}) = \text{med}_{\mathbf{x} \in \mathcal{D}} \| d_{\perp}(\boldsymbol{\pi}, \mathbf{x}) \|_{c_d}$$

- Probability that real $\boldsymbol{\pi}$ generated evidence

$$P(\boldsymbol{\pi}_{\mathcal{D}}|\boldsymbol{\pi}) \propto \mathcal{N}(r(\boldsymbol{\pi}, \mathcal{D}); 0, \sigma^2_{\mathcal{D}})$$


groundplane measurements

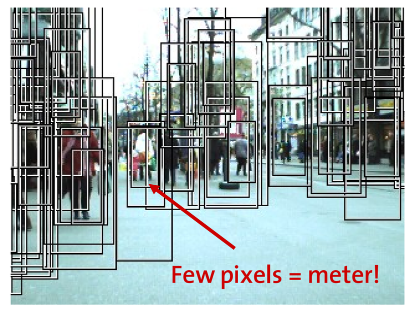


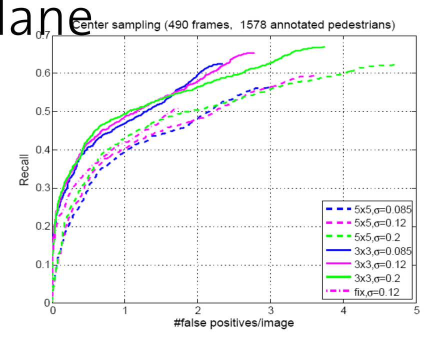
Formulation (pt. 2)

Object detection

- ISM detector yields n object hypotheses
- Object modeled as $o_i = \{ v_i, \mathbf{c}_i \}$
  - Validity flag v_i
  - Object center and scale $\mathbf{c}_i$ (yields bounding box)
  - Overlapping objects create loop over pixels!
- Bounding box transferred into world via ground plane
- Optimize bounding box placement
  - Allow slight changes in center/scale
  - Distribution learnt from training sequence
  - Solves problems with depth estimates

$$P(o_i|\boldsymbol{\pi}) = P(v_i|\mathbf{c}_i, \boldsymbol{\pi}) P(\mathbf{c}_i|\boldsymbol{\pi})$$

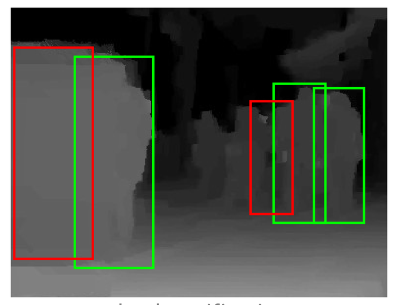




Depth cues

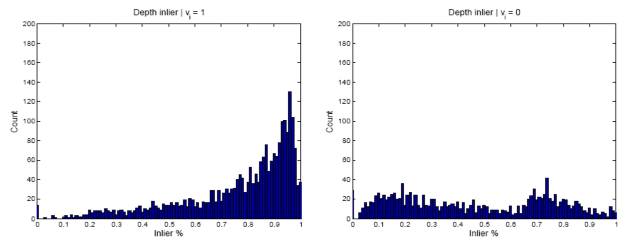
- Generated using BP-based stereo
  - Mostly accurate, sometimes excessive smoothing
  - Use robust statistical measures for inference

$$P(d_i|o_i, \boldsymbol{\pi}) \propto P(v_i|\mathbf{c}_i, d_i, \boldsymbol{\pi}) P(\mathbf{c}_i|d_i, \boldsymbol{\pi}) P(d_i)$$


depth verification

- Two cues
  - Median depth should coincide with depth predicted by bounding box
  - Depth distribution should be (robustly) uniform

$$q_i = \frac{\left| \{ x \in [LQ, UQ] \mid -\sigma_{di} < x < \sigma_{di} \} \right|}{UQ - LQ}$$



Inference

- Practical problem: inference?
- Graphical model contains implicit loop over image pixels!
- Modeling on pixel level infeasible
- Find best explanation in 2-stage approach
  1. Loopy BP disregarding implicit loops over images
     - Executed first to give geometric meaning to detections
  2. QBP-based optimization on result, handling overlap on pixel level using top-down segmentation
     - Selects best image explanation using hard constraints $\max_{\mathbf{m}} \mathbf{m}^T \mathbf{Q} \mathbf{m}$
     - cmp. Leibe et al., CVPR 2005
  - Iterate if necessary

Conclusion

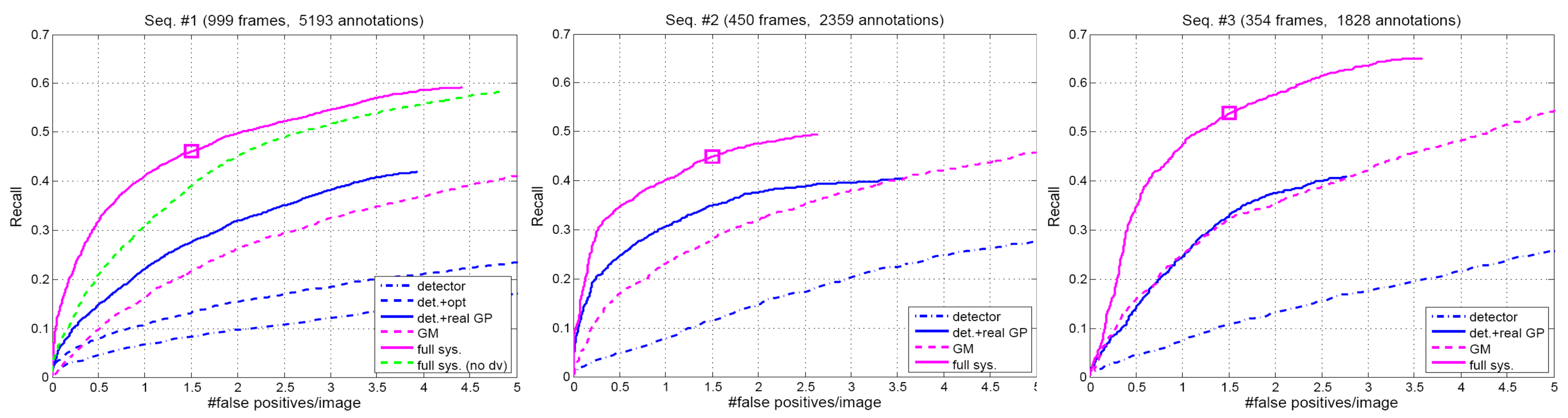
- Principled integration of depth, appearance, and ground plane
  - Improves results considerable w.r.t. various baseline systems
  - Can compensate for inaccuracies in detector
- Good detection performance in challenging real-world data
- Extension to tracking system → ask for video!

Also check out our poster:
B. Leibe, K. Schindler, and L. Van Gool "Coupled Detection and Trajectory Estimation for Multi-Object Tracking"

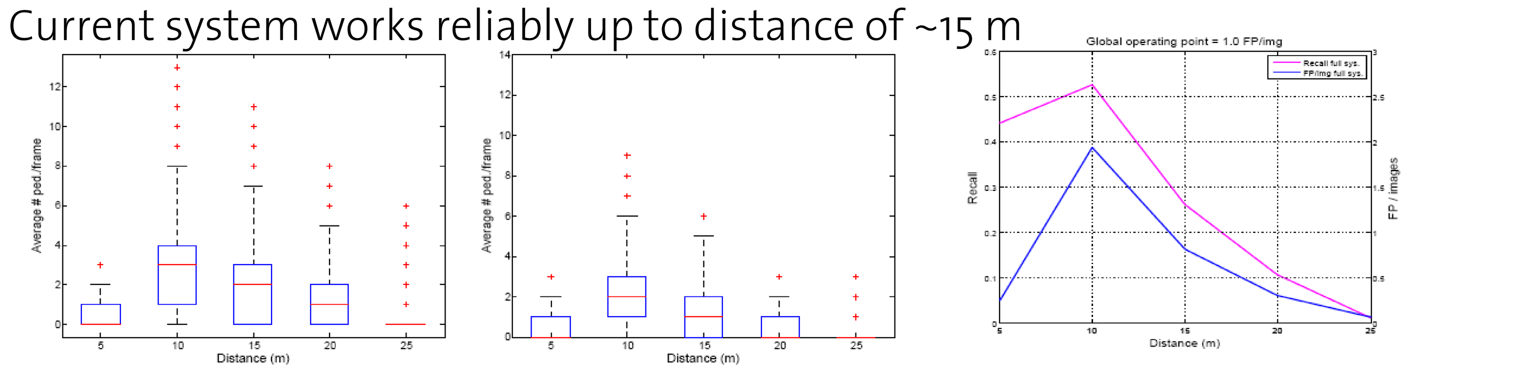
Results

Data available!

- Experimental validation
  - 3 annotated test sequences (2'293 frames, 10'958 annotations)
- Baseline performances
  - Basic detector
  - Detector with MDL optimization stage
  - Detector with MDL stage and ground-plane knowledge (Leibe *et al.*, CVPR 2007)
  - Output of graphical model
  - Full system output



- Analysis of detection performance w.r.t. depth
  - Current system works reliably up to distance of ~15 m



The authors acknowledge the support of Toyota Motor Europe and Toyota Motor Corporation, as well as the European project DIRAC (IST-027787).

Annotated data is available:
http://www.vision.ee.ethz.ch/~aess/iccv2007/