Lehrstuhl Informatik 8
Computergrafik und
Multimedia

RWTH AACHEN UNIVERSITY

Mobile Visual Scene Understanding in Highly Dynamic Environments

Bastian Leibe

Mobile Multimedia Processing
Computer Sciences 8 – Computergraphics & Multimedia
RWTH Aachen

MIRACLE Workshop, St. Augustin, 30.10.2009

UMIC

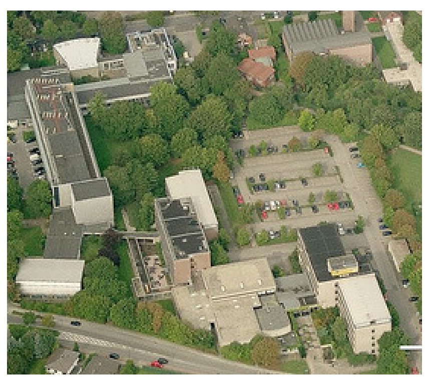
RWTH AACHEN UNIVERSITY

RWTH Computer Graphics & Multimedia Group

- **Prof. Dr. Leif Kobbelt**
  - Computer Graphics
  - Geometric Modelling

- **Prof. Dr. Bastian Leibe**
  - Computer Vision
  - Machine Learning

Lehrstuhl Informatik 8
Computergrafik und
Multimedia



UMIC Research Centre

Research Focus: Mobile Vision Applications

- Three main scenarios

Mobile phones,
Wearable computing

Mobile robotics,
Personal mobility

Intelligent vehicles

On-board computation →

← Unrestricted environment

Research Directions

- Mobile Visual Search
  - Recognition from mobile phones
  - Automatic content creation
  - *Towards mobile AR applications…*

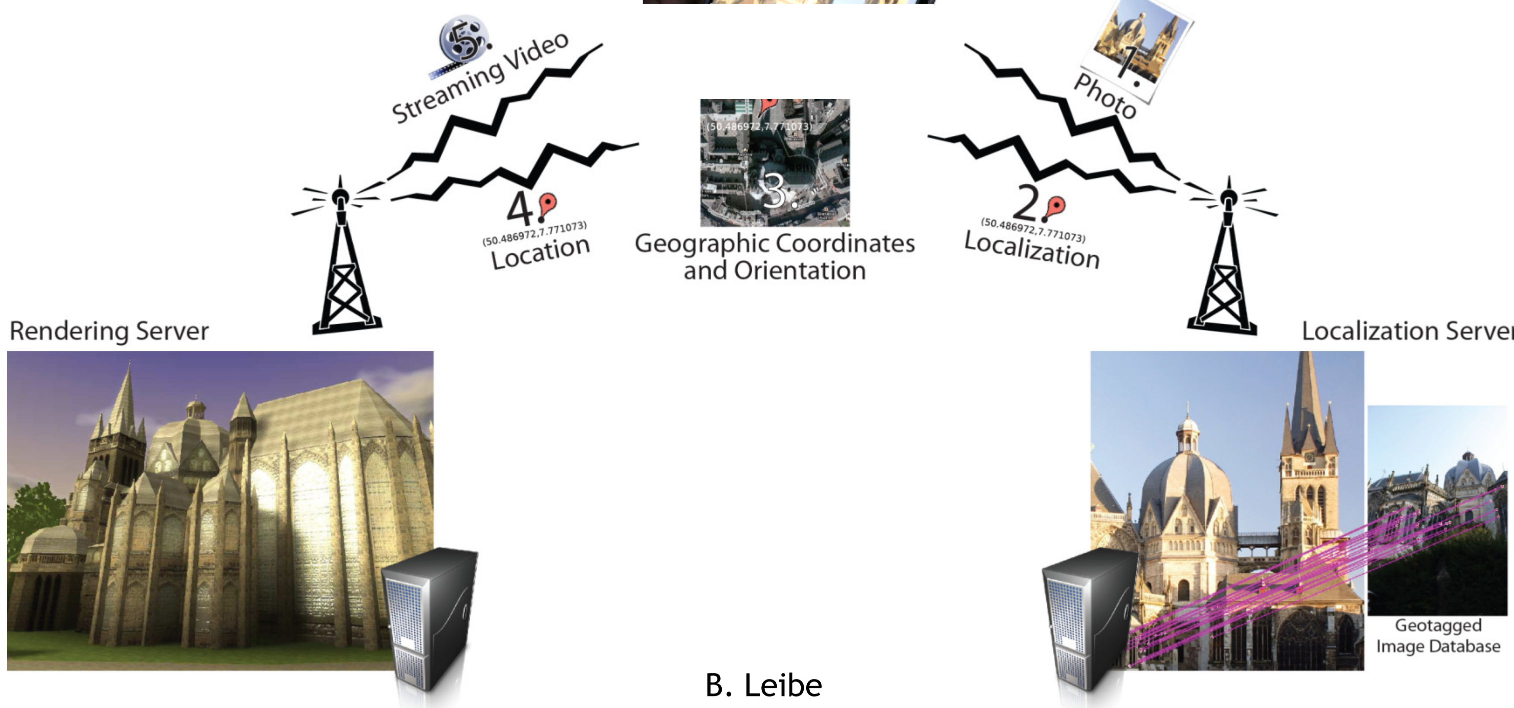
- Mobile Object Detection and Tracking
  - Object categorization
  - Scene geometry estimation
  - Multi-person tracking
  - Detailed body pose analysis

Target Scenario: Pedestrian Navigation

- **Mobile visual search**
  - ➢ **Simply point the camera to any object/building of interest.**
  - ➢ **Images are transmitted to a central server for recognition.**
  - ➢ **Object-specific information is sent back to be displayed on the mobile phone (mobile AR).**

System Overview

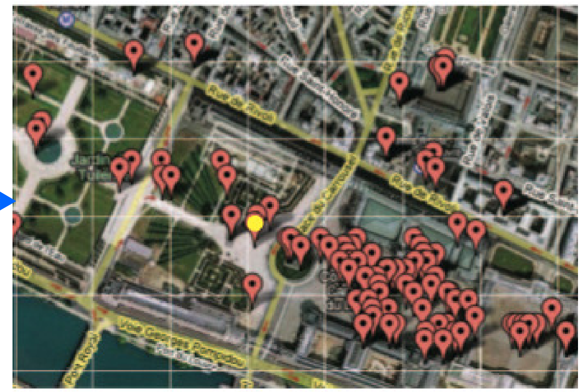
- **How can we make this scalable for an entire city?**



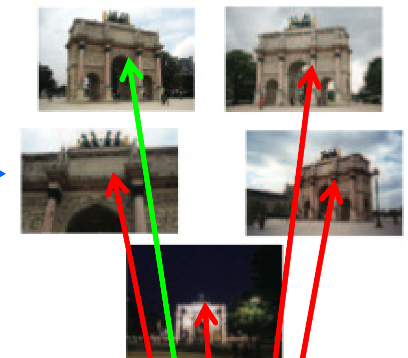
World-Scale Mining for Content Creation

flickr BETA

Mining geotagged images



Extracted Image clusters



Auto-annotation

museum
museum louvre
carrousel
carrousel triomphe
carrousel triomphe arc
carrousel triomphe arc du
carrousel triomphe du
carrousel arc
carrousel arc du

Frequent tags

You can **support Wikipedia** by making a tax-deductible donation.

article | discussion | edit this page | history

Arc de Triomphe du Carrousel

From Wikipedia, the free encyclopedia

This article does not cite any references or sources. *(January 2008)*
Please help improve this article by adding citations to reliable sources. Unverifiable material may be challenged and removed.

The **Arc de Triomphe du Carrousel** is a triumphal arch in Paris, France. It is located on the Place du Carrousel, just to the west of the Louvre.

Designed by Charles Percier and Pierre Léonard Fontatine, the arch was made between 1806-1808 by the Emperor Napoleon I on the model of the Arch of Septimius Severus in Rome. It was commissioned to commemorate France's military victories in 1805. It was originally surmounted by the famous horses of Saint Mark's Cathedral in Venice, captured by Napoleon, but these were returned to Venice in 1815. They were replaced by a quadriga sculpted by Baron François Joseph Bosio, depicting Peace riding in a triumphal chariot led by gilded Victories on both sides. The composition commemorates the Restoration of the Bourbons following Napoleon's downfall.

The highest arch is flanked by another two smaller ones. Around its exterior are eight Corinthian columns of granite, topped by eight soldiers of the Empire. In the attic between the soldiers, bas-reliefs depict:

- the Arms of the Kingdom of Italy with figures representing History and the Arts
- the Arms of the French Empire with Victory, Fame, History and Abundance
- Wisdom and Strength holding the arms of the Kingdom of Italy, accompanied by Prudence and Victory.

Napoleon's diplomatic and military victories are commemorated by bas-reliefs executed in rose marble, depicting the Peace of Pressburg, Napoleon entering Munich, Napoleon entering Vienna,

Arc de Triomphe du Carrousel

http://en.wikipedia.org/wiki/Arc_de_Triomphe_du_Carrousel

[Quack, Leibe, Van Gool, CIVR'08]

Example Results: Famous Tourist Sights

http://en.wikipedia.org/wiki/Basilica_of_the_Sacr%C3%A9_C%C5%93ur
426 Elements, 233 users, 287 days. Precision: 100%

http://en.wikipedia.org/wiki/Tour_Montparnasse
40 elements, 10 users, 11 days. Precision: 100%

http://en.wikipedia.org/wiki/Notre_Dame_de_Paris
588 elements, 287 users, 334 days. Precision: 100%

[Quack, Leibe, Van Gool, CIVR'08]

Example Results: Matching Under Occlusion

http://en.wikipedia.org/wiki/Old_Town_Square_(Prague)
262 elements, 122 users, 195 days. Precision: 98%.

http://en.wikipedia.org/wiki/Colosseum
582 elements, 190 users, 252 days. Precision: 100%

[Quack, Leibe, Van Gool, CIVR'08]

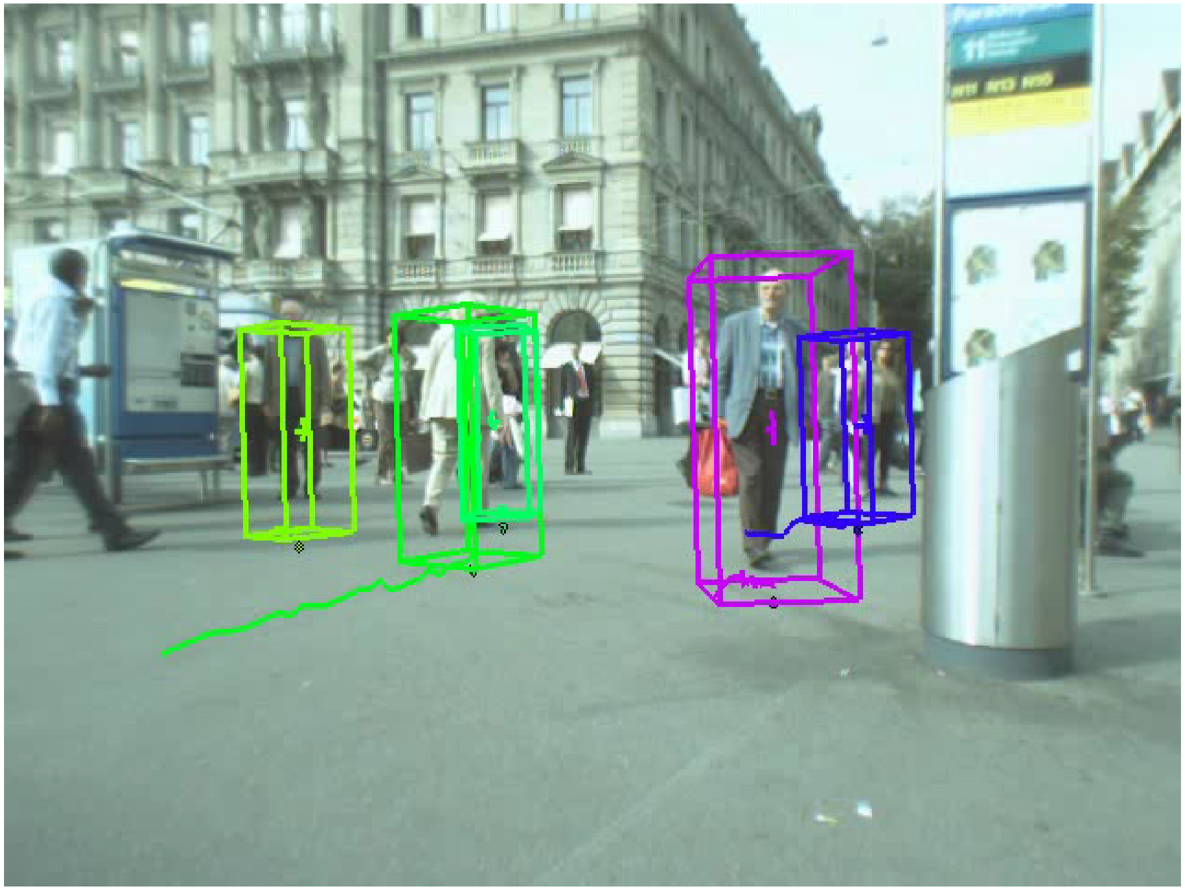
Research Directions

- Mobile Visual Search
  - Recognition from mobile phones
  - Automatic content creation
  - *Towards mobile AR applications…*

- Mobile Object Detection and Tracking
  - Object categorization
  - Scene geometry estimation
  - Multi-person tracking
  - Detailed body pose analysis

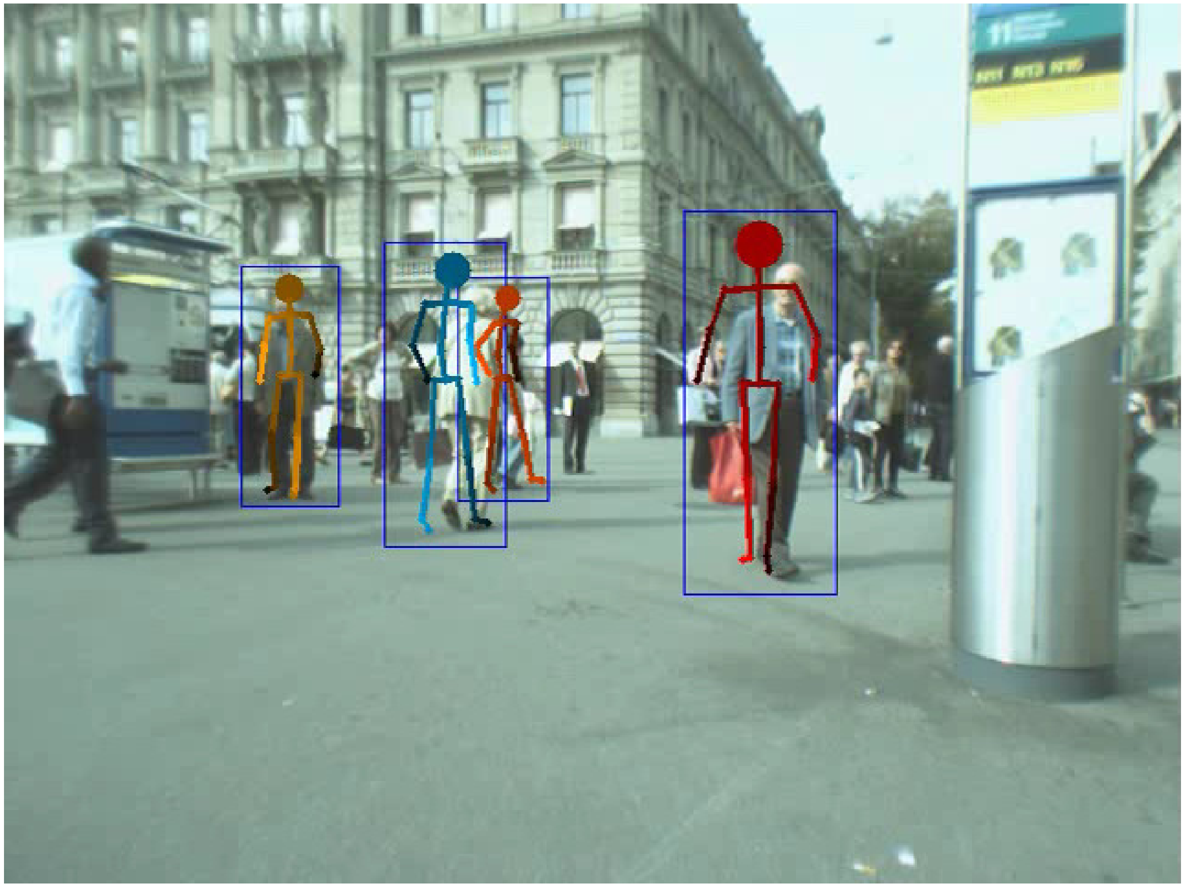
Joint work with: Andreas Ess, Stephan Gammeter, Luc Van Gool (ETH Zurich); Konrad Schindler (TU Darmstadt)

Towards Visual Scene Understanding



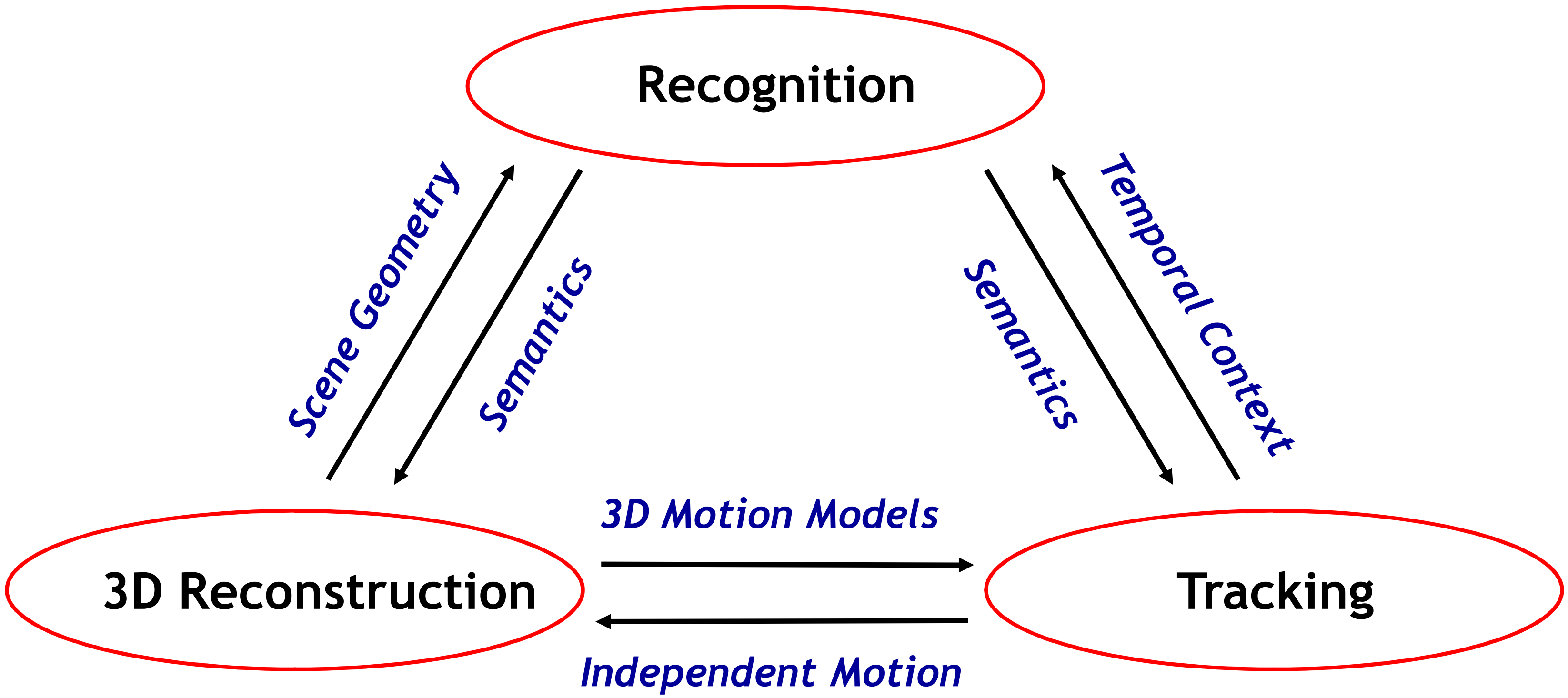
- **Objectives**
  - Detect & track people in environment
  - Interpret their motion
  - Predict their future behavior
- **Challenges**
  - We are moving
  - Objects are moving
  - Significant occlusions

Towards Visual Scene Understanding



- **Objectives**
  - Detect & track objects in environment
  - Interpret their motion
  - Predict their future behavior
- **Challenges**
  - We are moving
  - Objects are moving
  - Significant occlusions

Integration Principle: Cognitive Loops



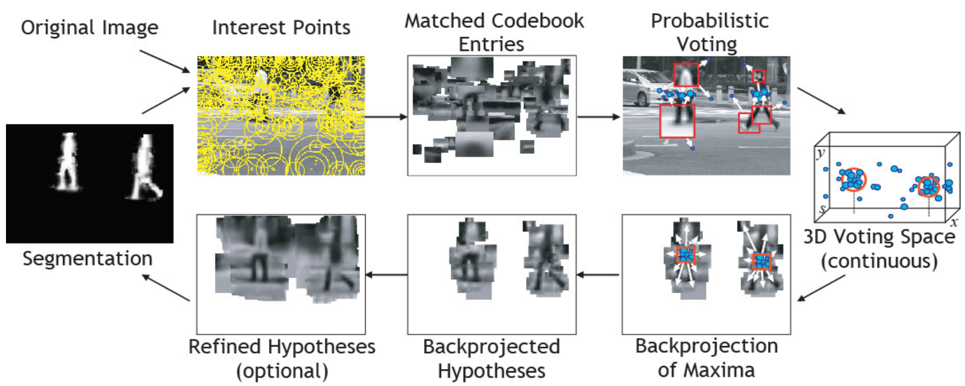
Integrate different vision modalities through
Cognitive Feedback

Outline

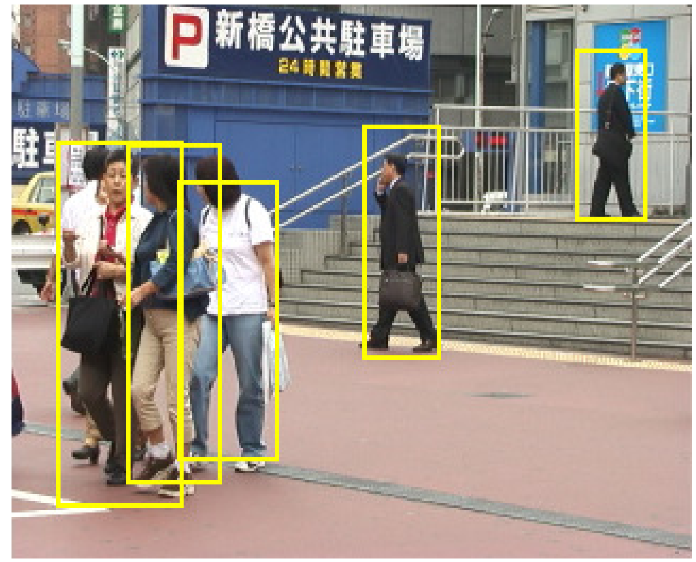
- Object Recognition
  - Implicit Shape Model (ISM) approach
- Integration with Scene Geometry
  - Coupled object detection & 3D estimation
- Temporal Integration
  - Multi-hypothesis tracking-by-detection
- Visual Odometry
  - Feedback from detection and tracking
- Putting It All Together…
  - Mobile pedestrian tracking
  - Articulated tracking under egomotion

Object Categorization & Detection

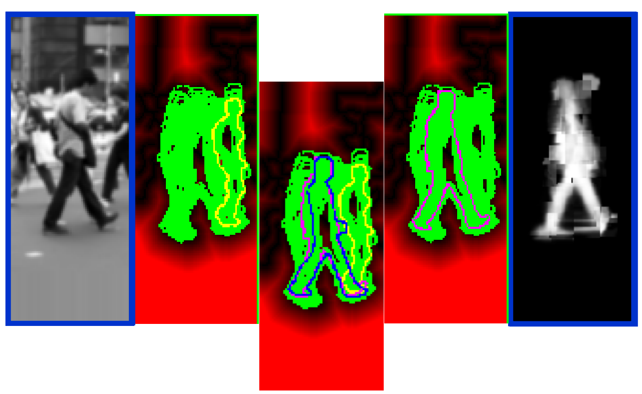
- ISM Object Detection



[Leibe, Leonardis, Schiele, IJCV’07]



- Pedestrian Detection in Crowds



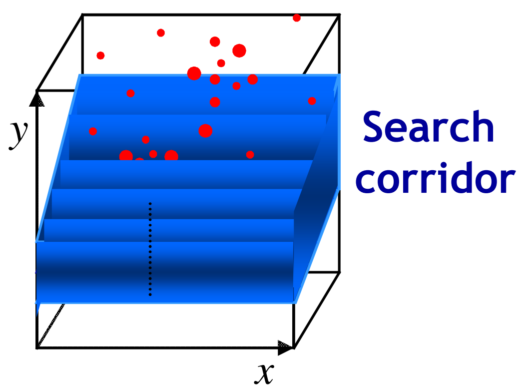
[Leibe, Seemann, Schiele, CVPR’05]

Outline

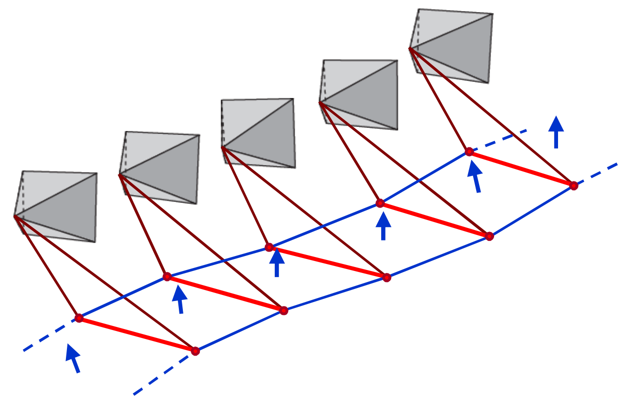
- Object Recognition
  - Implicit Shape Model (ISM) approach
- **Integration with Scene Geometry**
  - **Coupled object detection & 3D estimation**
- Temporal Integration
  - Multi-hypothesis tracking-by-detection
- Visual Odometry
  - Feedback from detection and tracking
- Putting It All Together…
  - Mobile Pedestrian Tracking
  - Articulated tracking under egomotion

Scene Geometry Estimation

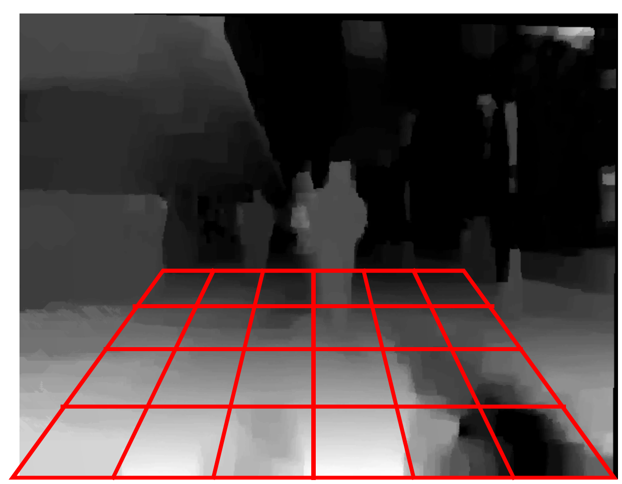
- Goal: Find the ground plane
  - Restrict object location
  - Assume Gaussian size prior
  - ⇒ Significantly reduced search space



Search Space

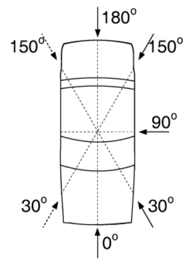


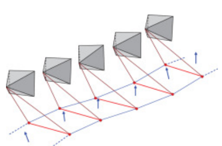
Structure-from-Motion



Dense stereo

Detections Using Ground Plane Constraints

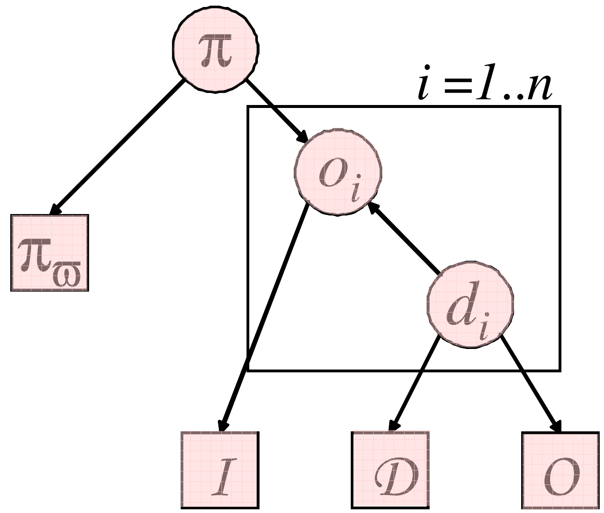




left camera
1175 frames

Object & Ground-plane Reasoning

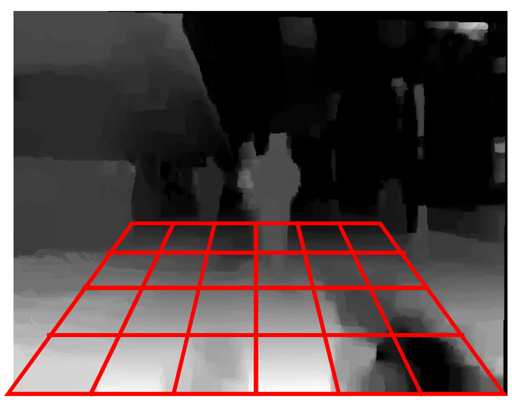
- **Probabilistic combination in Bayesian network**



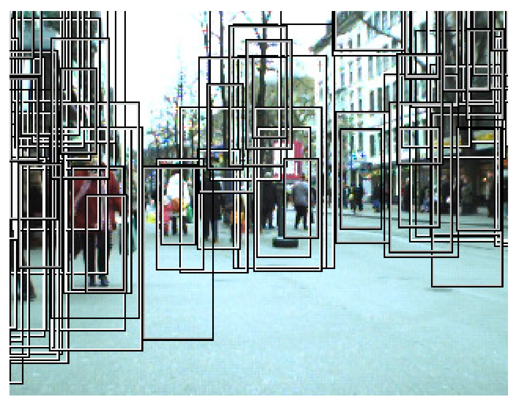
Groundplane π

Object detections o_i

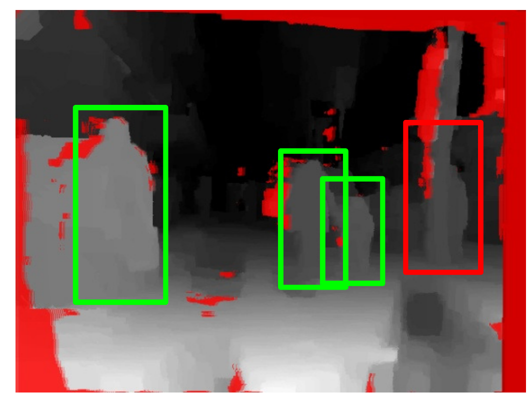
Depth measurements d_i



Groundplane measurements



Object detections (ISM)



Depth verification

[Ess, Leibe, Van Gool, ICCV'07]

Object & Ground-plane Reasoning

recording setup

- **Effect:**
  - **Reliable detections from scene context**
  - **Accurate 3D positioning from depth map**

[Ess, Leibe, Van Gool, ICCV'07]

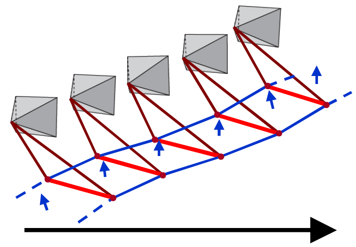
RWTH AACHEN UNIVERSITY

Outline

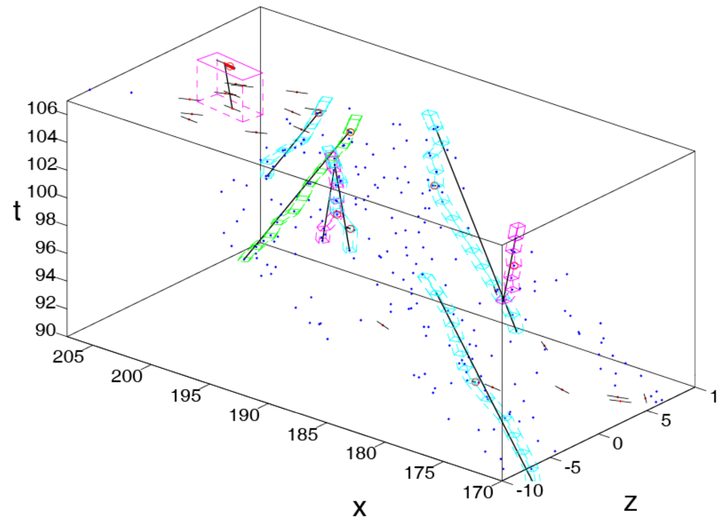
- Object Recognition
  - Implicit Shape Model (ISM) approach
- Integration with Scene Geometry
  - Coupled object detection & 3D estimation
- **Temporal Integration**
  - **Multi-hypothesis tracking-by-detection**
- Visual Odometry
  - Feedback from detection and tracking
- Putting It All Together...
  - Mobile Pedestrian Tracking
  - Articulated tracking under egomotion

Coupled Detection and Tracking

Object detections



3D through SfM

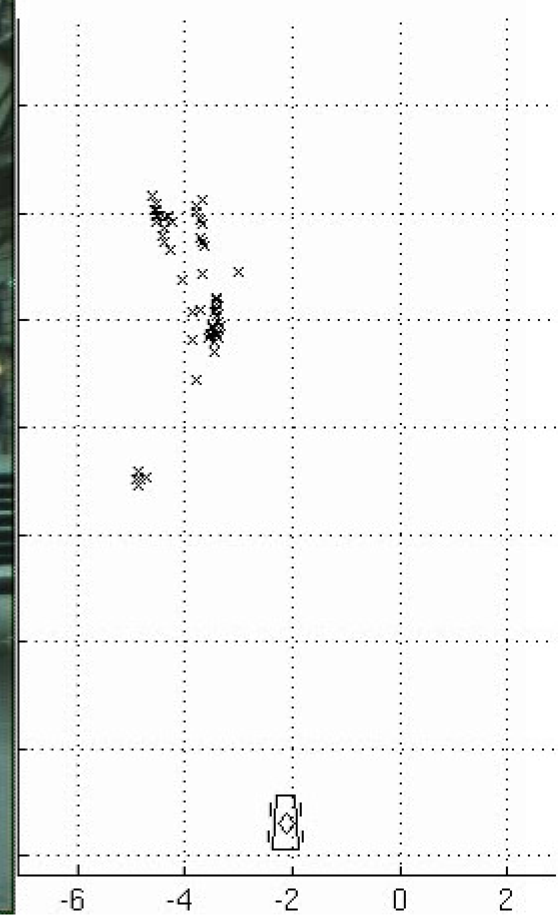


Spacetime trajectories

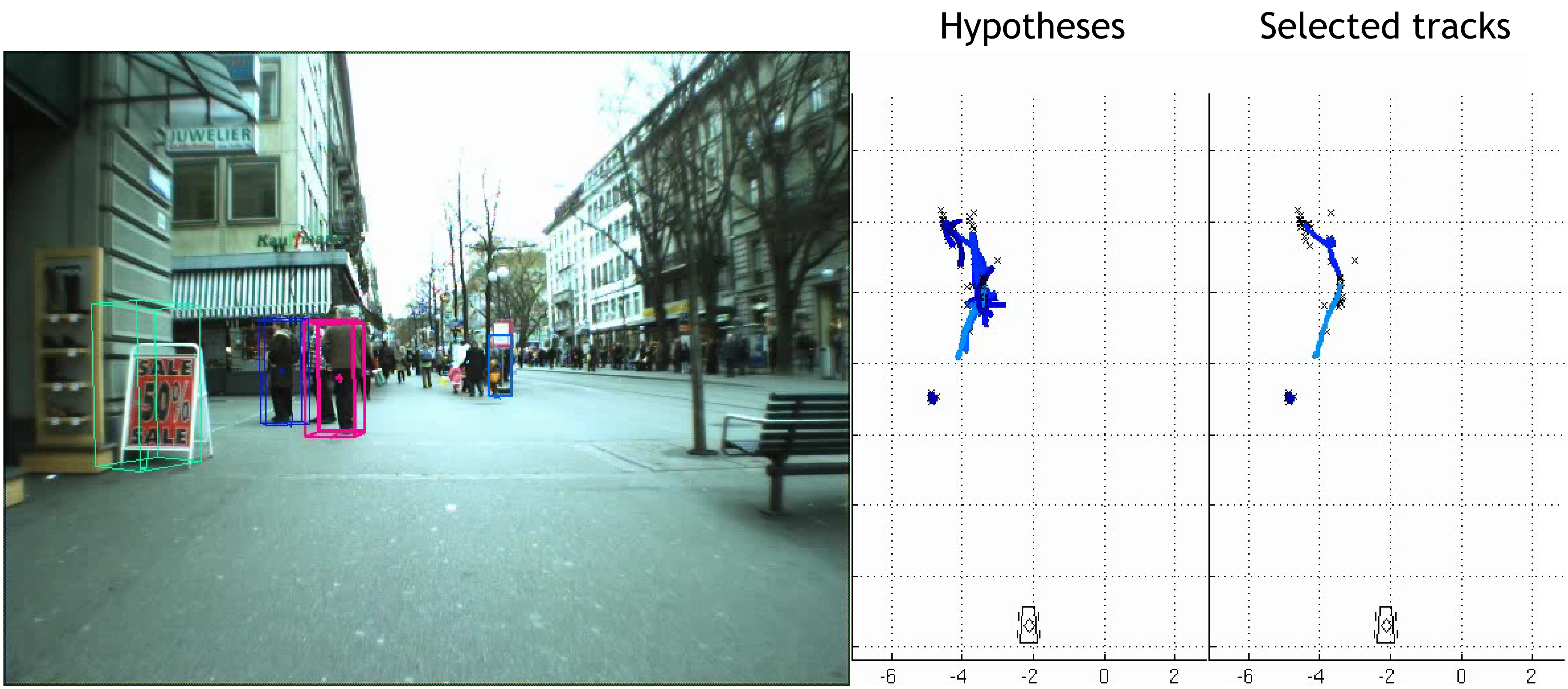
Model selection

(Quadratic Boolean Optimization)

Multi-Object Tracking by Detection

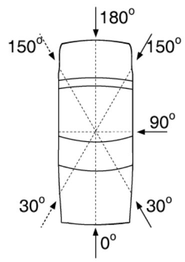


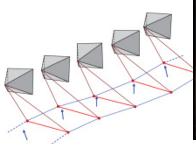
Multi-Object Tracking by Detection



- Multi-hypothesis tracking with model selection in each frame
- Ability to recover temporarily lost tracks

Dynamic Scene Analysis



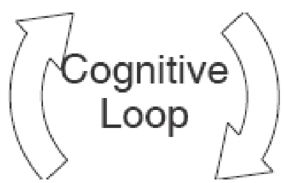


[Leibe, Cornelis, Cornelis, Van Gool, CVPR'07]

Application: Augmented 3D City Model

Enhancing your driving experience…

Original **3D Reconstruction**



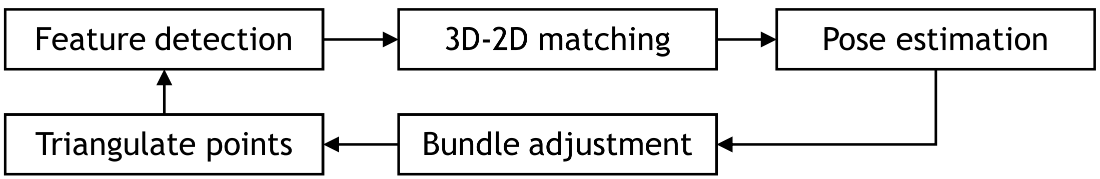
[Cornelis, Leibe, Cornelis, Van Gool, IJCV'08]

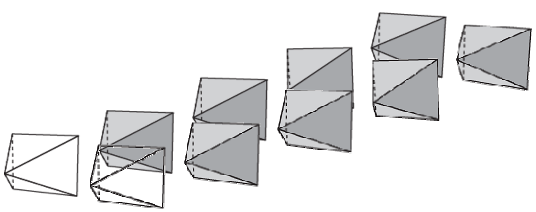
Outline

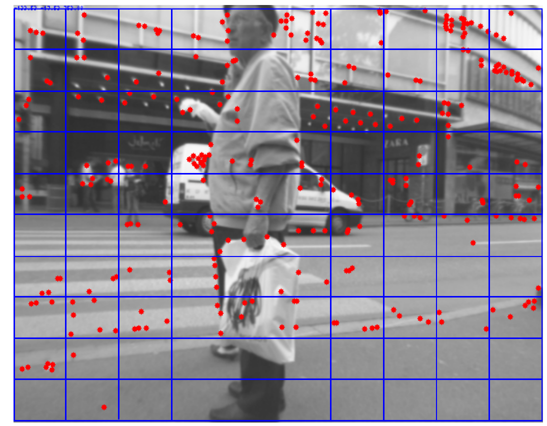
- Object Recognition
  - Implicit Shape Model (ISM) approach
- Integration with Scene Geometry
  - Coupled object detection & 3D estimation
- Temporal Integration
  - Multi-hypothesis tracking-by-detection
- **Visual Odometry**
  - **Feedback from detection and tracking**
- Putting It All Together...
  - Mobile Pedestrian Tracking
  - Articulated tracking under egomotion

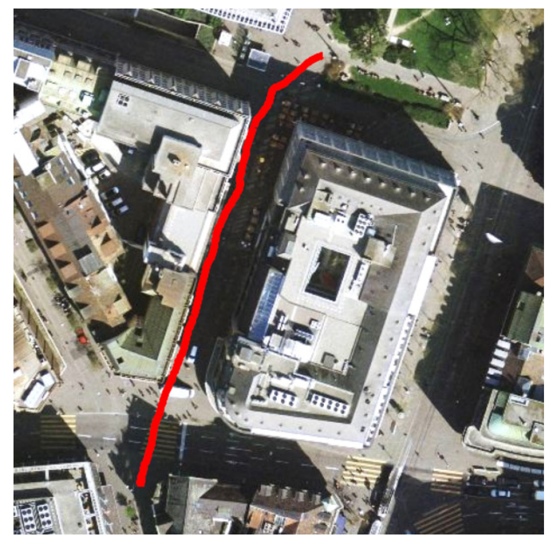
Visual Odometry

- **Defines common coordinate frame**
  - Basis for tracking-by-detection in 3D
- **Stereo-based Structure-from-Motion**
  - Similar to Nister *et al.*, CVPR'04



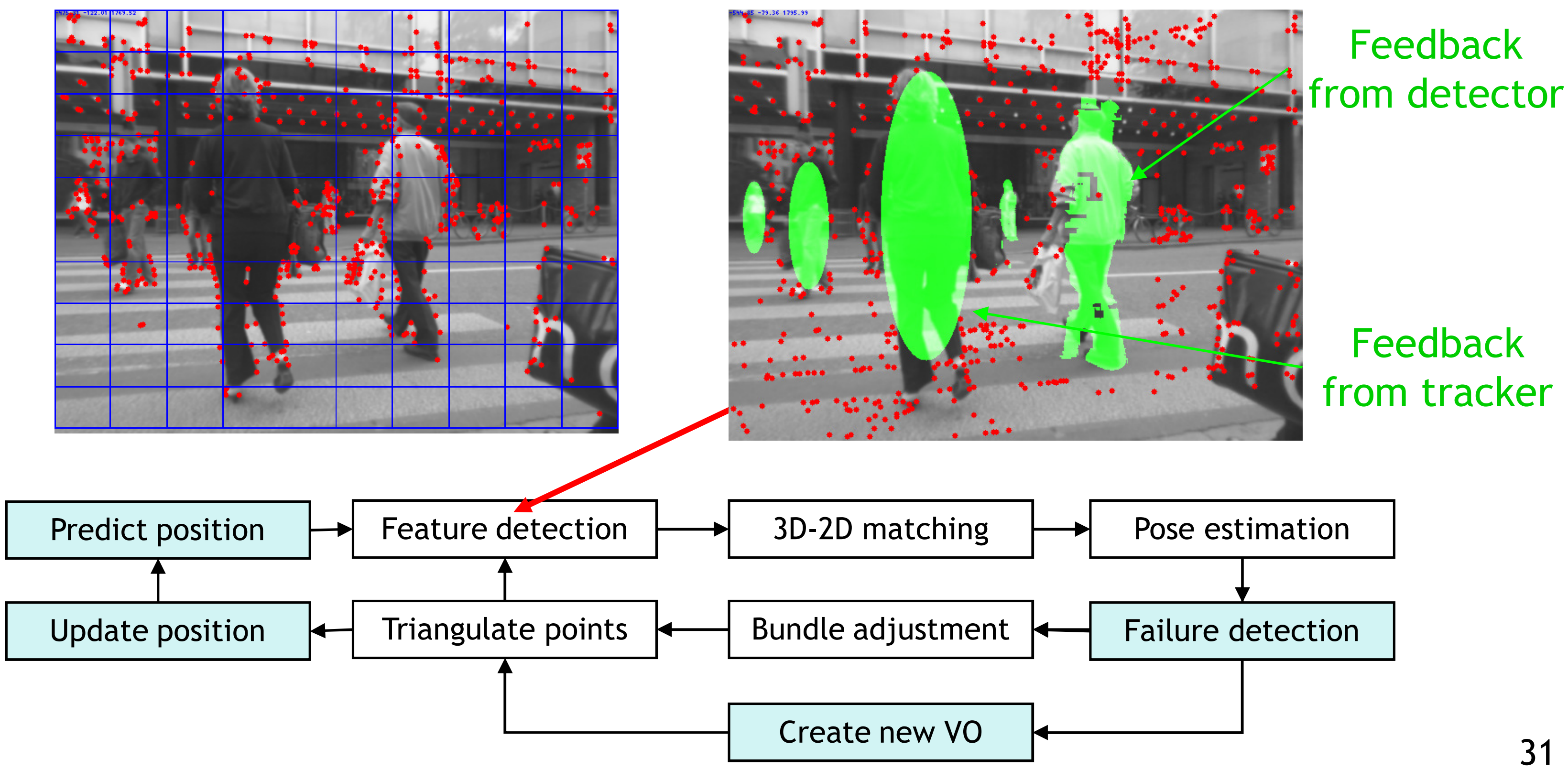




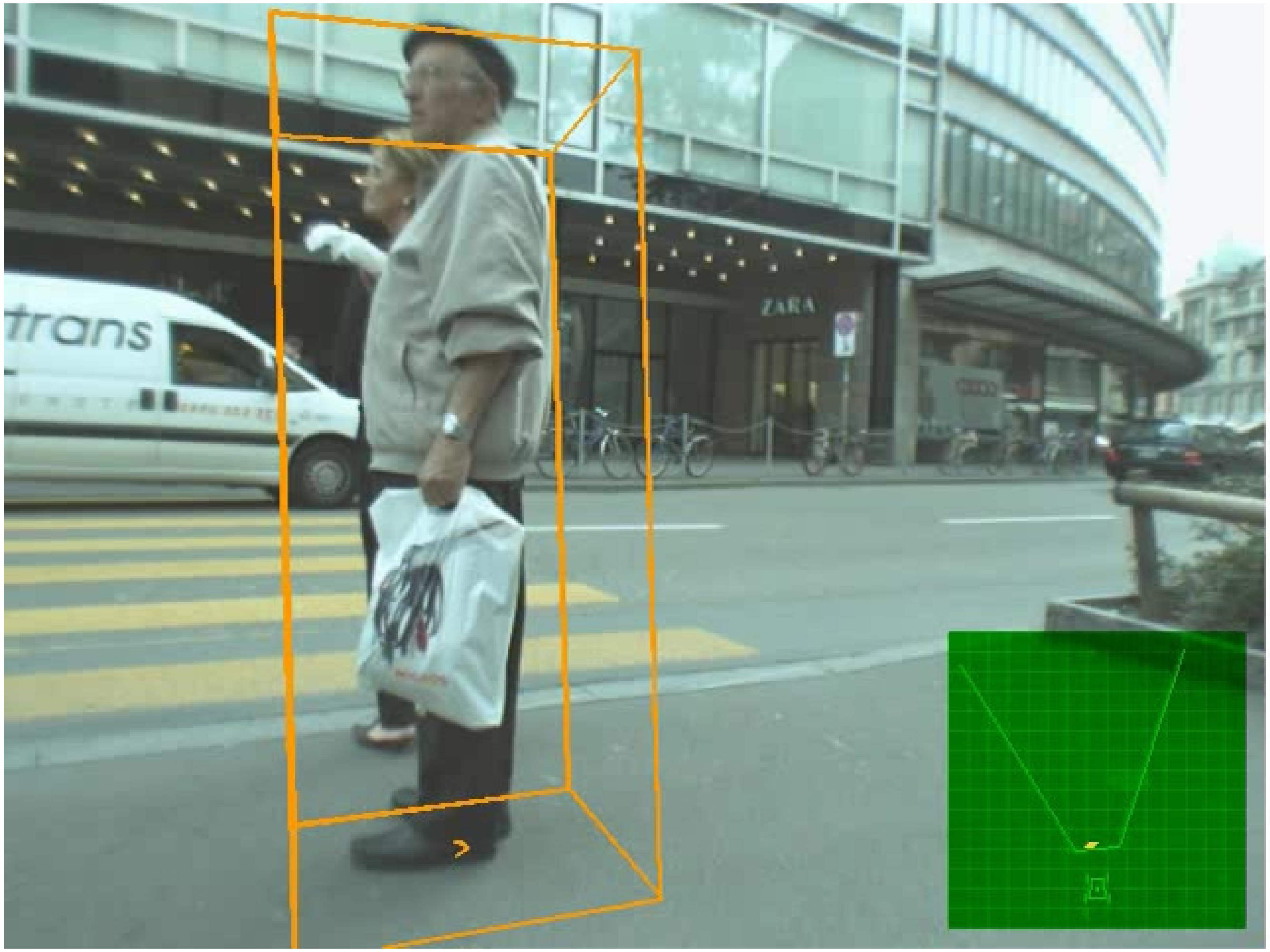


Feedback to Visual Odometry

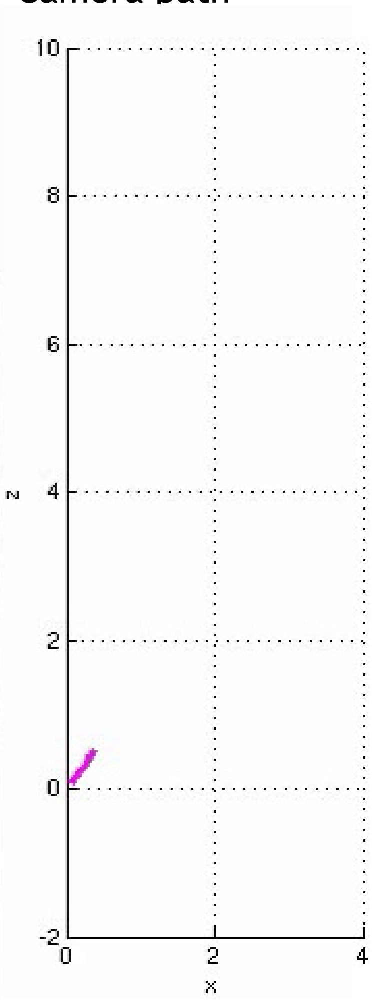
- Not all parts of scene are static
  - Detector / Tracker give semantic information
  - Mask out moving parts
    $\Rightarrow$ Restrict localization efforts to static parts



Feedback to Visual Odometry



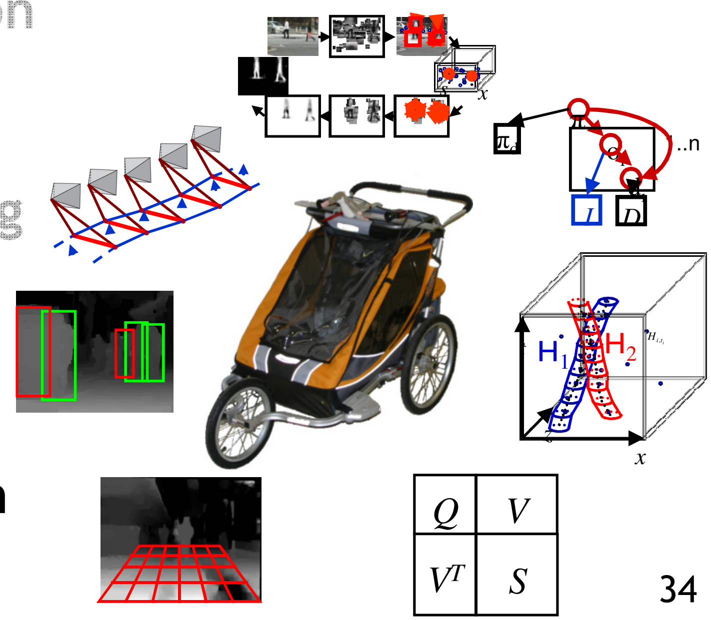
Camera path



Standard VO
Failure detection, no masking
Failure detection + masking

Outline

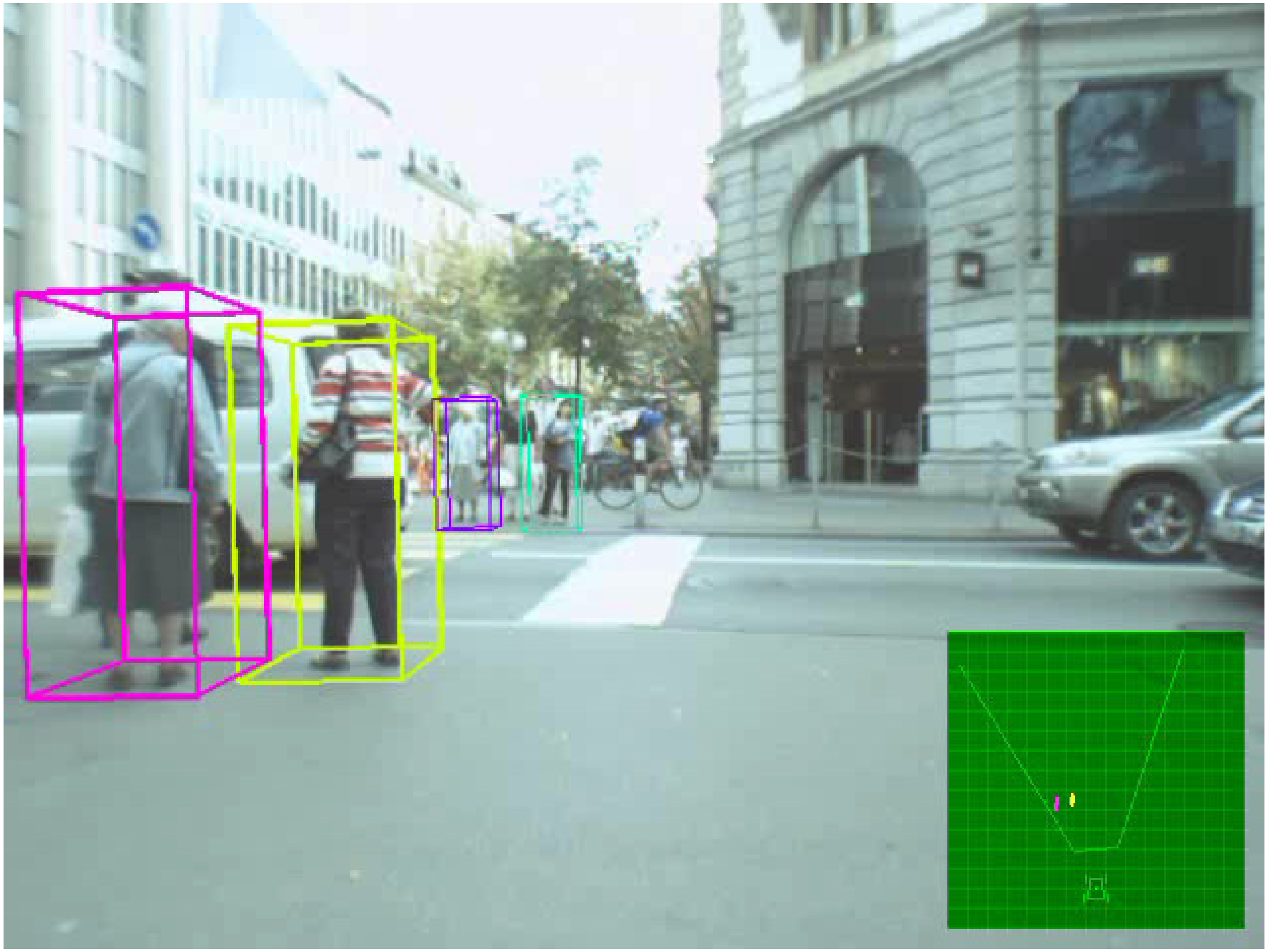
- Object Recognition
  - Implicit Shape Model (ISM) approach
- Integration with Scene Geometry
  - Coupled object detection & 3D estimation
- Temporal Integration
  - Multi-hypothesis tracking-by-detection
- Visual Odometry
  - Feedback from detection and tracking
- **Putting It All Together…**
  - **Mobile Pedestrian Tracking**
  - **Articulated tracking under egomotion**



Mobile Tracking Through Crowds

[Ess, Leibe, Schindler, Van Gool, CVPR’08]

An Extreme Case…



[Ess, Leibe, Schindler, Van Gool, CVPR'08]

Predicting Behavior of Dynamic Obstacles

(Cooperation with Toyota Motor Corporation)

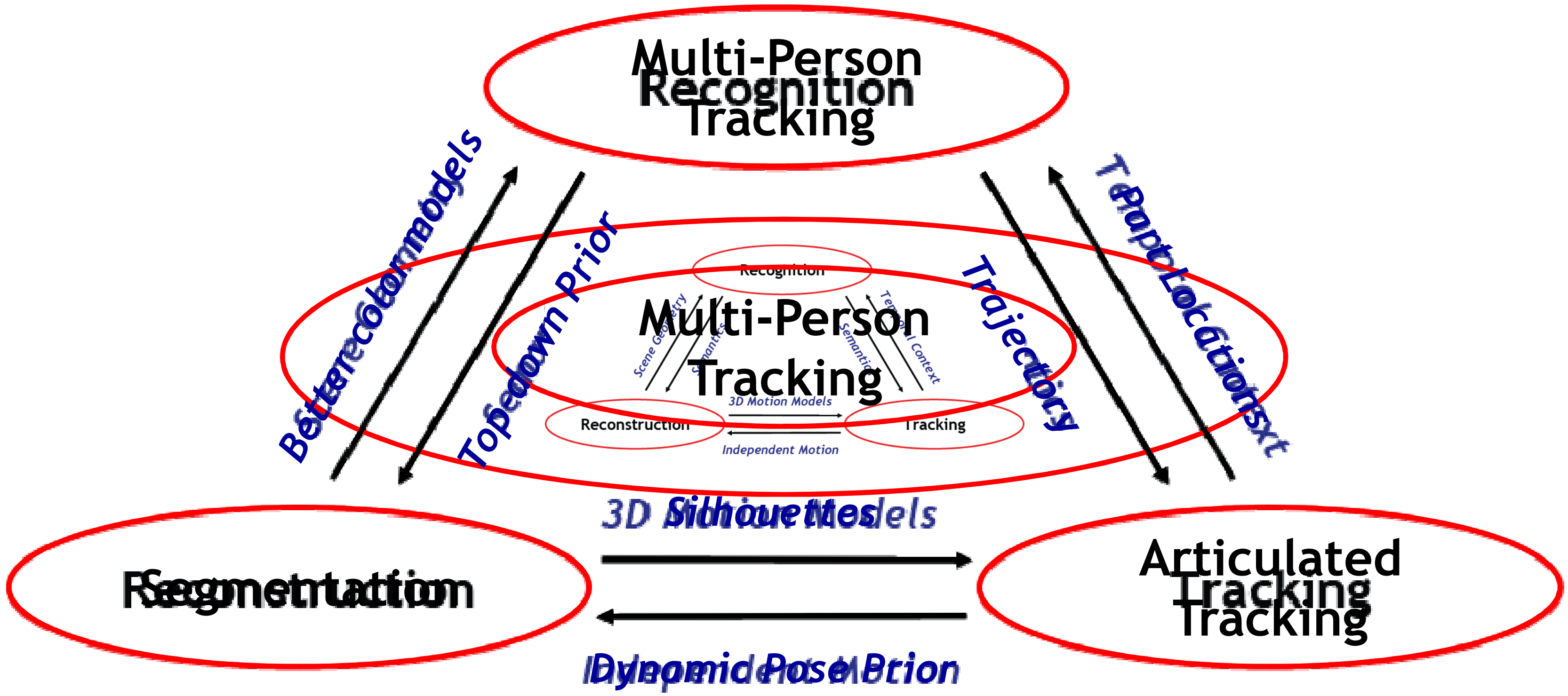
[Ess, Leibe, Schindler, Van Gool, ICRA'09]

Application in Cars

(Cooperation with Toyota Motor Corporation)

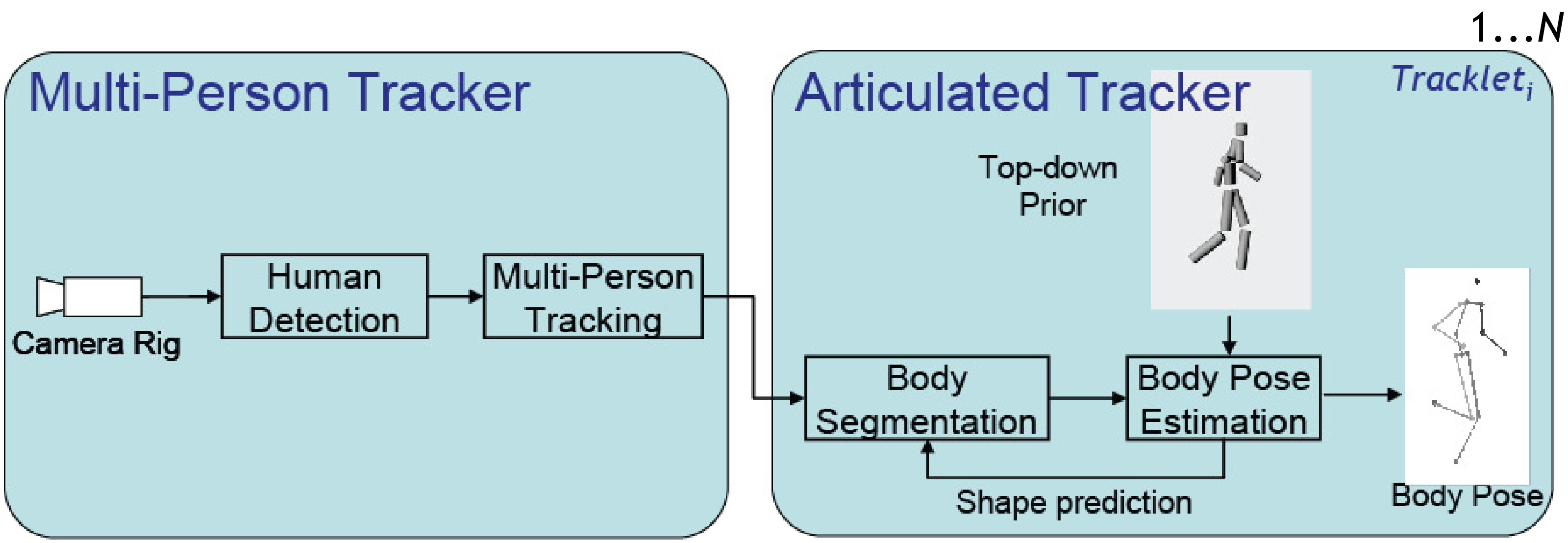
[Ess, Leibe, Schindler, Van Gool, PAMI'09]

Multi-Person Tracking as new Basic Unit



- **Many interesting ways to go on from here…**

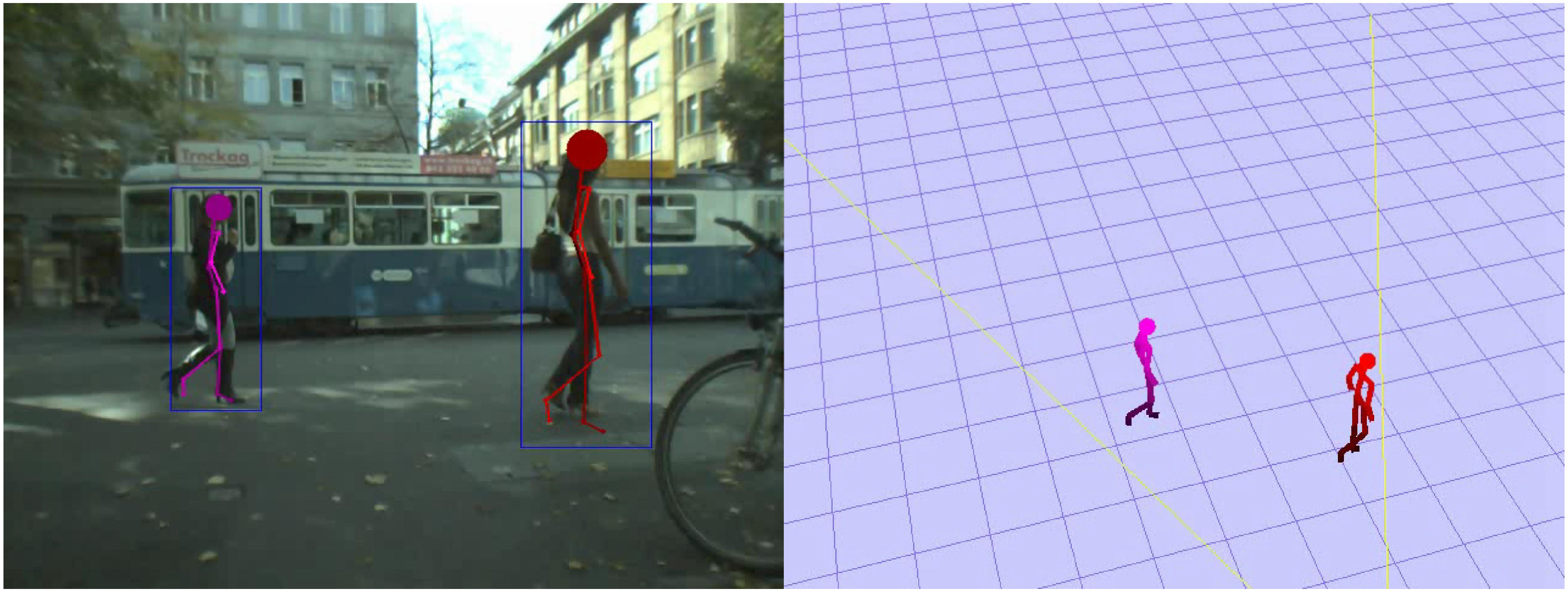
Recovering Articulations



- **Idea: Only perform articulated tracking where it’s easy!**
- **Multi-person tracking**
  - **Solves hard data association problem**
- **Articulated tracking**
  - **Only on individual “tracklets” between occlusions**

[Gammeter, Ess, Jaeggli, Schindler, Leibe, Van Gool, ECCV’08]

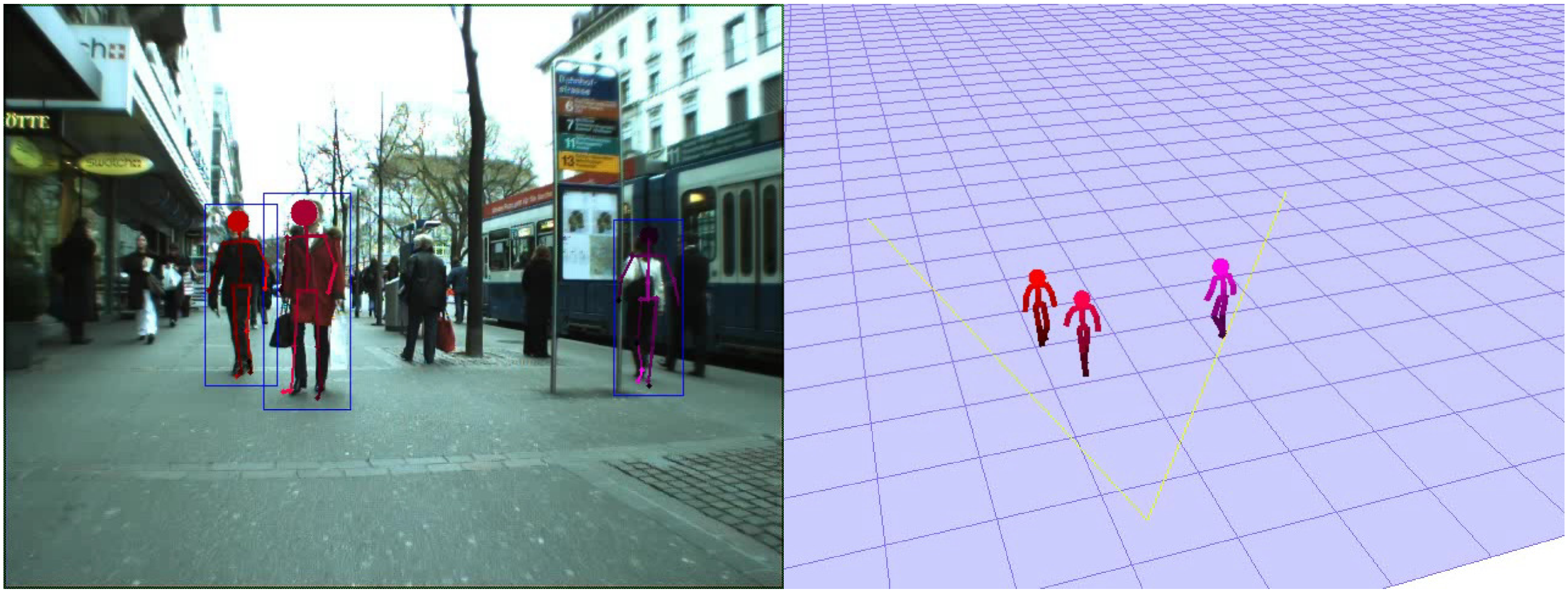
Articulated Multi-Person Tracking



- Multi-Person tracking
  - Recovers trajectories and solves data association
- Articulated Tracking
  - Estimates detailed body pose for each tracked person

[Gammeter, Ess, Jaeggli, Schindler, Leibe, Van Gool, ECCV'08]

Articulated Tracking under Egomotion



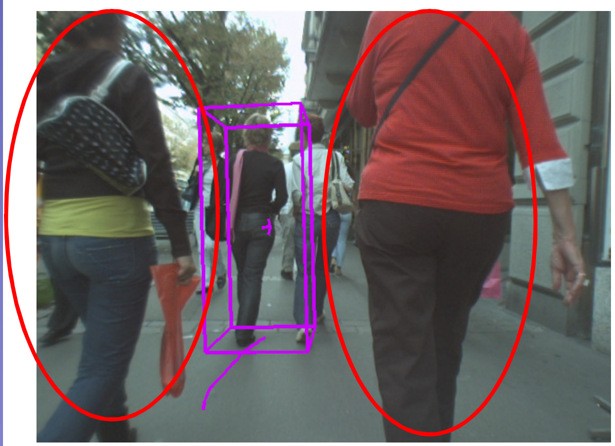
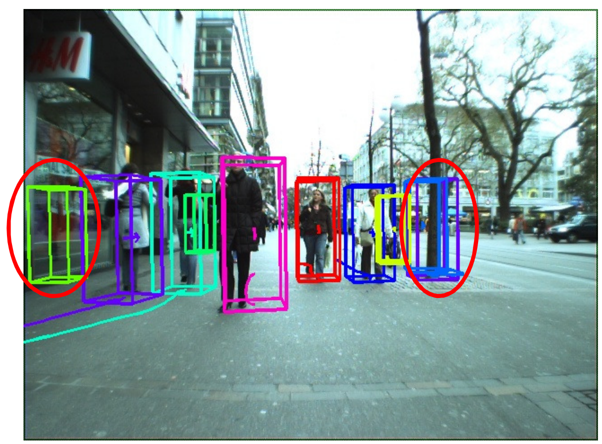
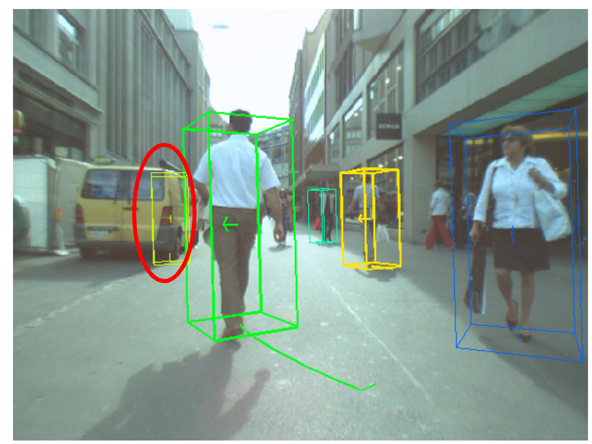
- **Guided segmentation for each frame**
  - No reliance on background modeling
  - Approach applicable to scenarios with moving camera
  - Feedback from body pose estimate to improve segmentation

[Gammeter, Ess, Jaeggli, Schindler, Leibe, Van Gool, ECCV'08]

RWTH AACHEN UNIVERSITY

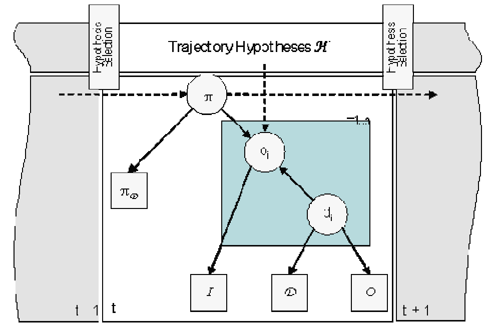
Typical Failure Cases

- Too big pedestrians
  - Not completely visible
  - Separate detector necessary
- False positives on reflections, trees, trashcans, …
  - Multi-class detector?
  - OK for path planning: in most cases, still an obstacle

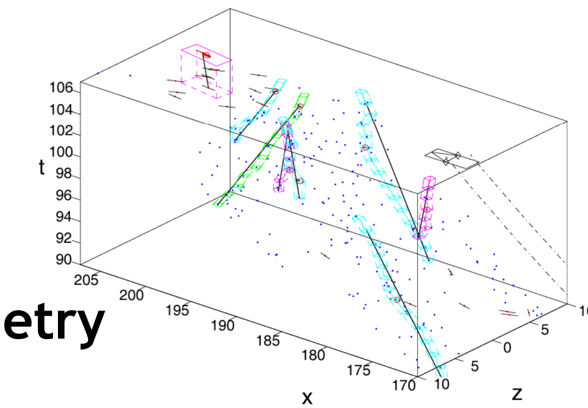




Keys for Success

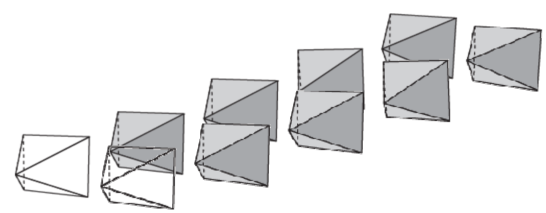
- Bayesian network for detection
  - Allow modification of bounding boxes
  - Use of confident depth information
  - No hard decision about ground plane
  - Spatial prior from tracker



- Multi-hypothesis tracking
  - Per-frame model selection
  - Depth information for localization
  - World-coordinate frame from visual odometry



- Visual Odometry
  - Feedback from Tracking/Detection
  - Failure detection



Conclusion

- Visual scene understanding
  - Vision is becoming feasible in the real world.
  - Many individual components are getting sufficiently mature.
  - Robust performance possible through combination.
- Perspective for Augmented/Mixed Reality
  - Currently still restricted to high-power hardware…
  - But real-time reachable within next 2 years.
  - Novel capabilities for AR applications?
    - Object categorization
    - Reaction to people
    - Augmenting categorical objects
- What use can we make of this for AR/MR applications?

RWTH AACHEN UNIVERSITY

Thank you very much!

Collaborators

Local Features
K. Mikolajczyk
N. Cornelis
T. Quack

Obj. Detection
E. Seemann
M. Fritz
A. Thomas
A. Lehmann
K. Mikolajczyk
A. Leonardis
B. Schiele

Tracking
K. Schindler
A. Ess

Body Pose Est.
T. Jaeggli
S. Gammeter

SfM & Stereo
N. Cornelis
K. Cornelis
A. Ess
T. Weise

System Integr.
A. Ess
K. Schindler
L. Van Gool

http://mmp.rwth-aachen.de

RWTH AACHEN UNIVERSITY

Timing

- C/C++ implementation
  - Without detector currently 3-4 fps
- Many calculations can be cached
- Detector current bottleneck
  - Faster detectors can be plugged in (e.g. fastHOG on GPU)
  - Parallelization possible

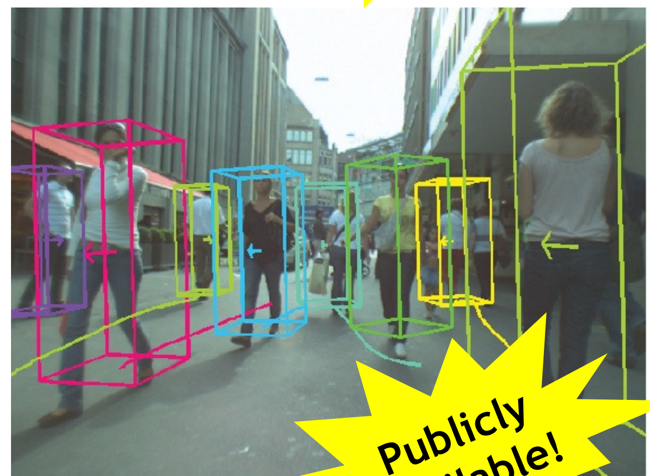
Component	CPU	GPU	Time
Detector	×		2 x 15s
Depth map	×	×	15s / 0.020s
Bayesian Network	×		0.200s
Visual odometry	×	×	0.020s
Tracking	×		0.100s

(per frame)

Large Datasets Available

- ICCV'07 Data
  - 4 Sequences
  - ~2200 frame pairs total
  - ~10,900 Pedestrian annotations
  - Cameras + groundplane from SfM
  - Various baseline performances

- CVPR'08 Data
  - 3 New sequences
  - ~2750 additional frame pairs
  - Pedestrian annotations (every 4^{th} frm)
  - Camera + groundplane from SfM
  - Various baseline performances



Data available at
http://www.vision.ethz.ch/aess

Towards Mobile Scene Understanding

Lehrstuhl Informatik 8
Computergrafik und
Multimedia

RWTH AACHEN UNIVERSITY

Mobile Multi-Person Tracking in Highly Dynamic Environments

Bastian Leibe

Mobile Multimedia Processing
Computer Sciences 8 – Computergraphics & Multimedia
RWTH Aachen

MIRACLE Workshop, St. Augustin, 30.10.2009

UMIC

Lehrstuhl Informatik 8
Computergrafik und
Multimedia

Visual Scene Understanding in Highly Dynamic Environments

Bastian Leibe

Mobile Multimedia Processing
Computer Sciences 8 – Computergraphics & Multimedia
RWTH Aachen

MIRACLE Workshop, St. Augustin, 30.10.2009

RWTH AACHEN UNIVERSITY