

# Six Topics in Mobile Vision

**Reinhard Klette**

**Gabriel Hartmann**

**Simon Herrmann**

**Sandino Morales**

**Yi Zeng**

**Hongmou Zhang**

The *.enpeda..* Project, The University of Auckland, New Zealand

Our mobile platform:  
Test vehicle HAKA1 at Tamaki campus, Auckland



## **Binocular Mobile Vision**

- 1 Stereo vision in a driver assistance context
- 2 Segmentation of dynamic scenes based on disparity maps
- 4 3D modeling of road environments from a mobile platform

## **Monocular Mobile Vision**

- 3 Optical flow calculation
- 5 Camera pose determination
- 6 Camera in a hexacopter

## **Conclusions**

# 1

## Stereo vision for driver assistance

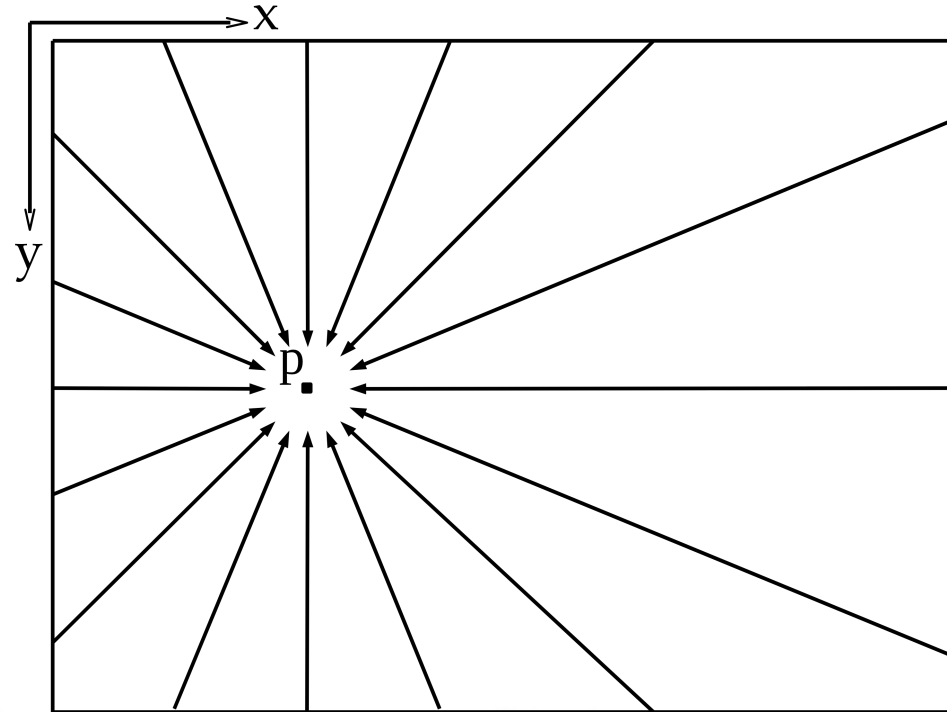
Example of rectified binocular recording



recorded at 25 Hz and 10 bit from our mobile platform HAKA1

# We recall: Semi-global matching (SGM)

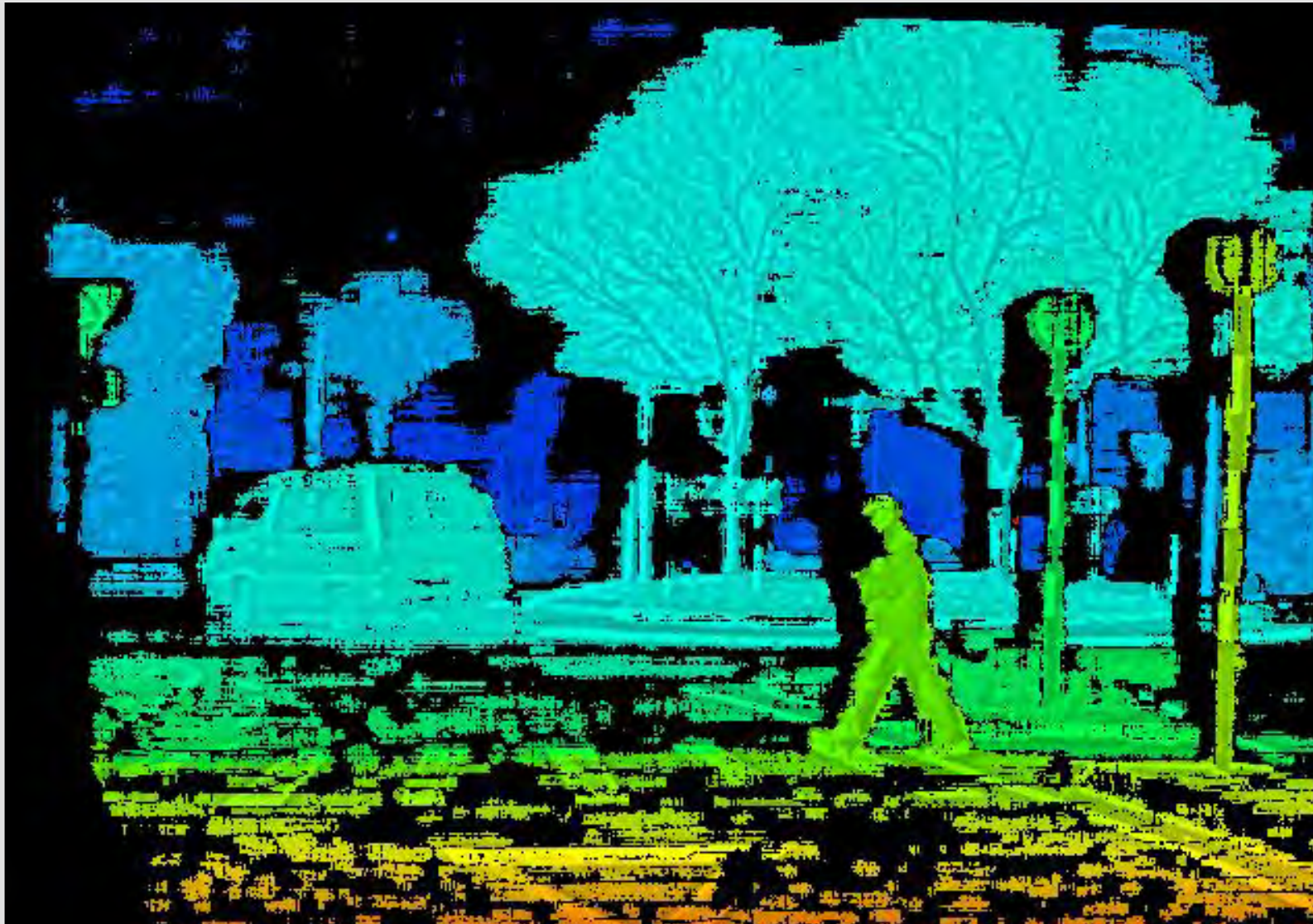
Cost accumulation  
along linear paths



$$E_a(p_i, d) = D_{p_i}(d) + \min \left\{ \begin{array}{l} E_a(p_{i-1}, d) \\ E_a(p_{i-1}, d-1) + c_1 \\ E_a(p_{i-1}, d+1) + c_1 \\ \min_{\Delta} L_a(p_{i-1}, \Delta) + c_2 \end{array} \right\} - \min_{\Delta} E_a(p_{i-1}, \Delta)$$



# SGM with mode filter on 10 bit data



Colours encode distances

Simon Hermann 2011

Black pixels: low confidence



(see, e.g., Ralf Haeusler et al., 2011...)

- iSGM is an iterative stereo matcher that extends the integration strategy of SGM
- SGM is a method often applied in industrial vision systems due to its robust performance and real-time capability. It is therefore a relevant concept for stereo matching.
- What are the benefits of iSGM over standard SGM?

# Answer 1: The KITTI Vision Benchmark Suite

## The KITTI Vision Benchmark Suite

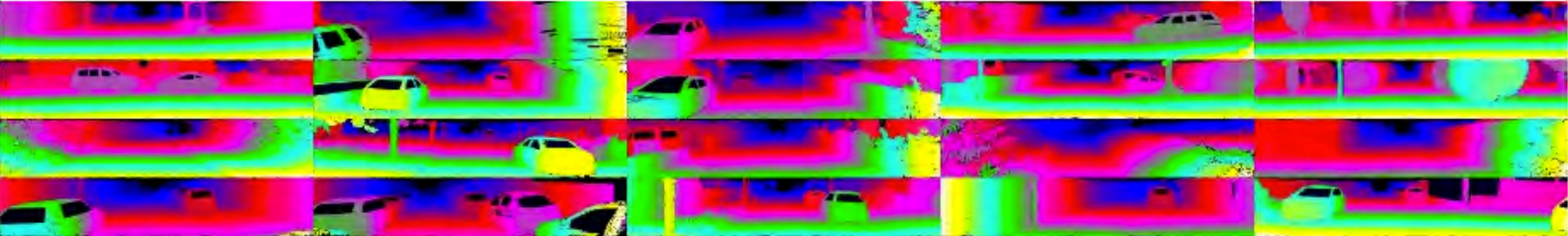
A project of Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago



[home](#) [setup](#) [stereo](#) [flow](#) [odometry](#) [object](#) [tracking](#) [raw data](#) [submit your results](#)

Andreas Geiger (KIT) | Philip Lenz (KIT) | Christoph Stiller (KIT) | Raquel Urtasun (TTI-C)

### Dataset



The stereo / flow benchmark consists of 194 training image pairs and 195 test image pairs, saved in loss less png format. Our evaluation server computes the average number of bad pixels for all non-occluded or occluded (=all groundtruth) pixels. We require that all methods use the same parameter set for all test pairs. Our development kit provides details about the data format as well as MATLAB / C++ utility functions for reading and writing disparity maps and flow fields.

- [Download stereo/optical flow data set \(720 MB\)](#)
- [Download multi-view extension \(20 frames per scene, all cameras\) \(17 GB\)](#)
- [Download stereo/optical flow development kit \(1 MB\)](#)

### Evaluation



# iSGM only marginally better on KITTI than SGM

iSGM is as fast as standard SGM (iSGM is not SSE optimized which gives a speed up factor of at least 4)

# 70 times faster than the top-ranked method

Top-ranked methods can be considered to be equally good with respect to the KITTI reference index.

KITTI is a new benchmark, but with limited challenges in given real-world data (e.g., only good weather conditions)

iSGM performs well, but does not seem to have a major benefit (compared to SGM) on the KITTI index.

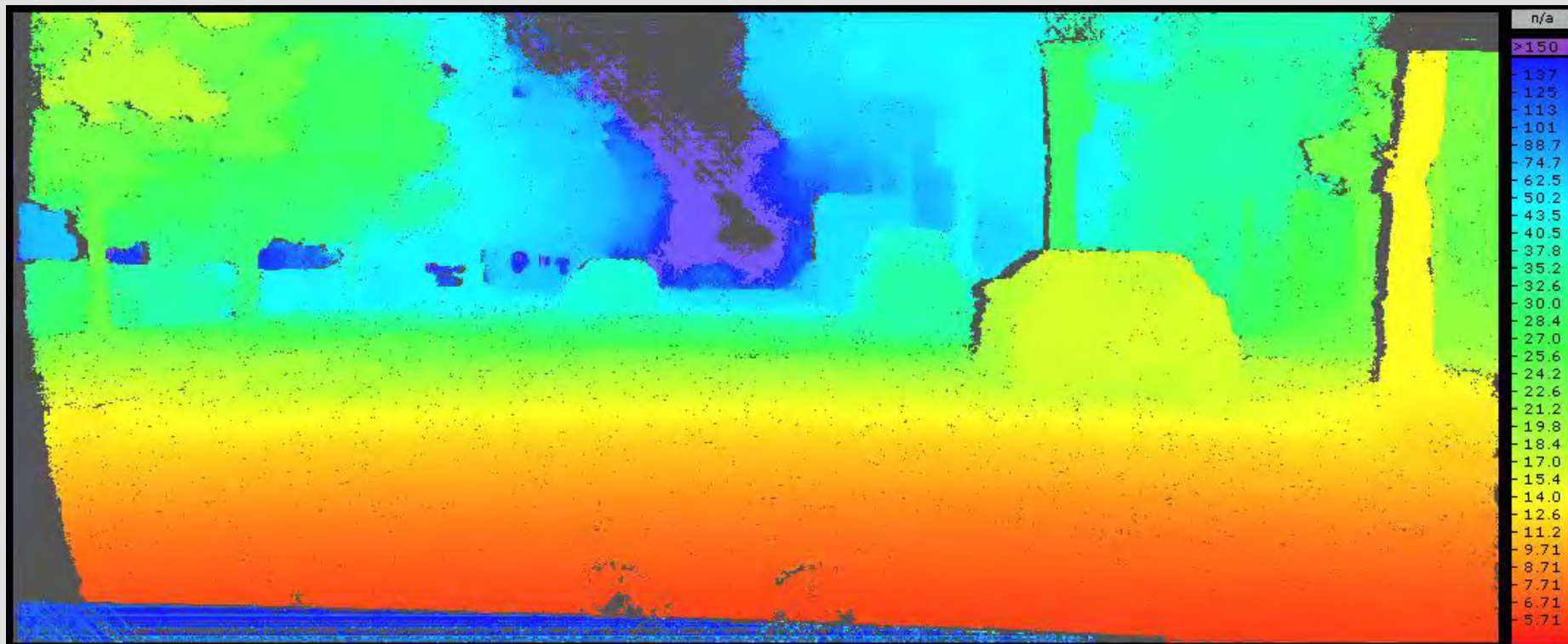
6 November 2012

Rank	Method	Setting	Out-Noc	Out-All	Avg-Noc	Avg-All	Density	Runtime	Environment
1	<a href="#">PCBP</a>		4.13 %	5.45 %	0.9 px	1.2 px	100.00 %	5 min	4 cores @ 2.5 Ghz (Matlab + C/C++)
Koichiro Yamaguchi, Tamir Hazan, David McAllester and Raquel Urtasun. <a href="#">Continuous Markov Random Fields for Robust Stereo Estimation</a> . ECCV 2012.									
2	<a href="#">iSGM</a>		5.16 %	7.19 %	1.2 px	2.1 px	94.70 %	8 s	2 cores @ 2.5 Ghz (C/C++)
Simon Hermann and Reinhard Klette. <a href="#">Iterative Semi-Global Matching for Robust Driver Assistance Systems</a> . ACCV 2012.									
3	<a href="#">SGM</a>		5.83 %	7.08 %	1.2 px	1.3 px	85.80 %	3.7 s	1 core @ 3.0 Ghz (C/C++)
Heiko Hirschmueller. <a href="#">Stereo Processing by Semi-Global Matching and Mutual Information</a> . IEEE Transactions on Pattern Analysis and Machine Intelligence 2008.									



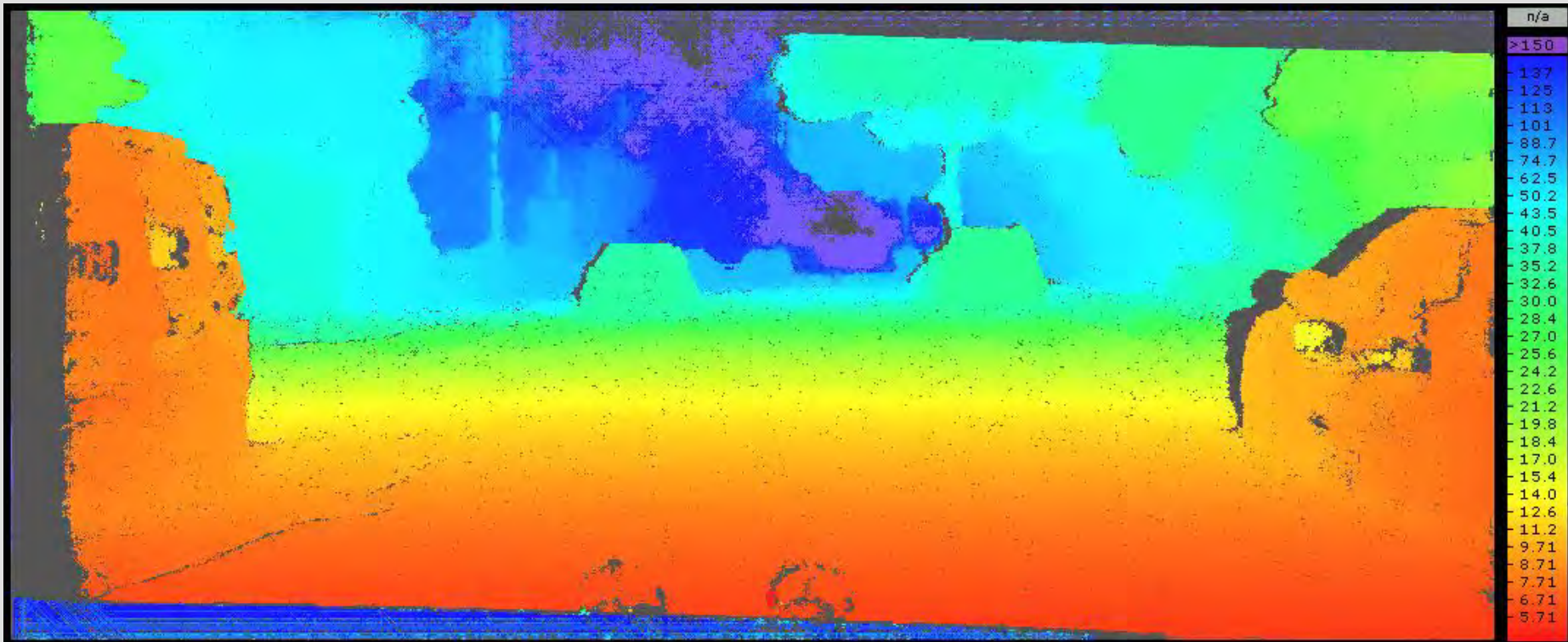


iSGM for KITTI's good weather data (Example 1)





iSGM for KITTI's good weather data (Example 2)





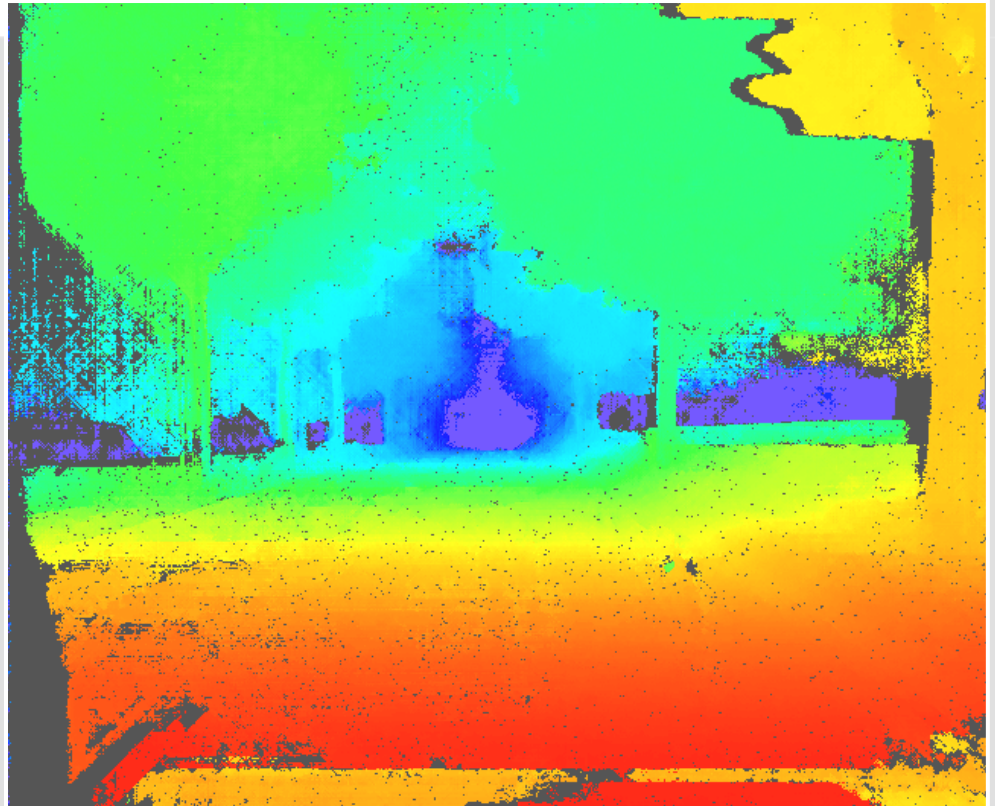
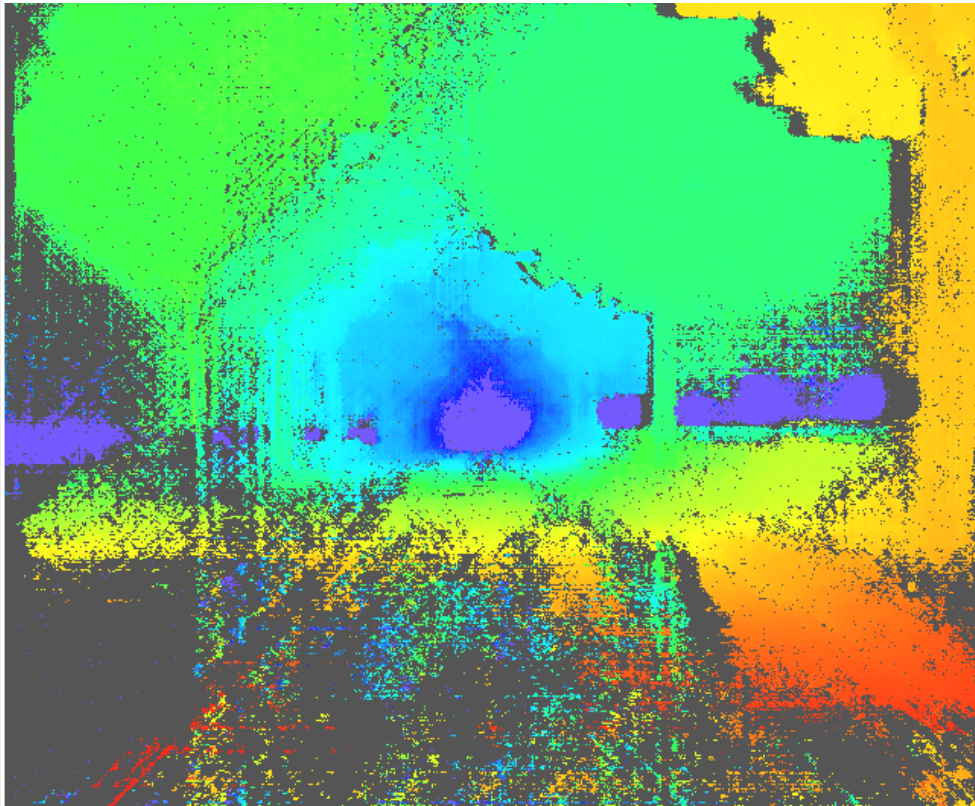
Rain (these are not KITTI data but HCI data)



SGM (standard)



iSGM



# Main benefit of iSGM in our opinion: **its robustness**

- iSGM comes with improved robustness on challenging stereo data in real-world scenarios (e.g. traffic scenes in the rain, with snow, sun strike, or night scenes)
- This has been acknowledged by the jury of the **Bosch Vision Challenge at ECCV 2012**: iSGM has been chosen to be the winner of the 2012 ECCV competition in Florence



# Answer 2: Bosch Robust Vision Challenge

## Jury at ECCV 2012

**Microsoft Research:** Simon Baker, Ph.D.

**Texas Instruments:** Goksel Dedeoglu, Ph.D.

**Volkswagen Research Driver Assistance & Integrated Safety:** Jan Effertz, Dr.

**Sony:** Oliver Erdler, Dipl.Ing. and MBA

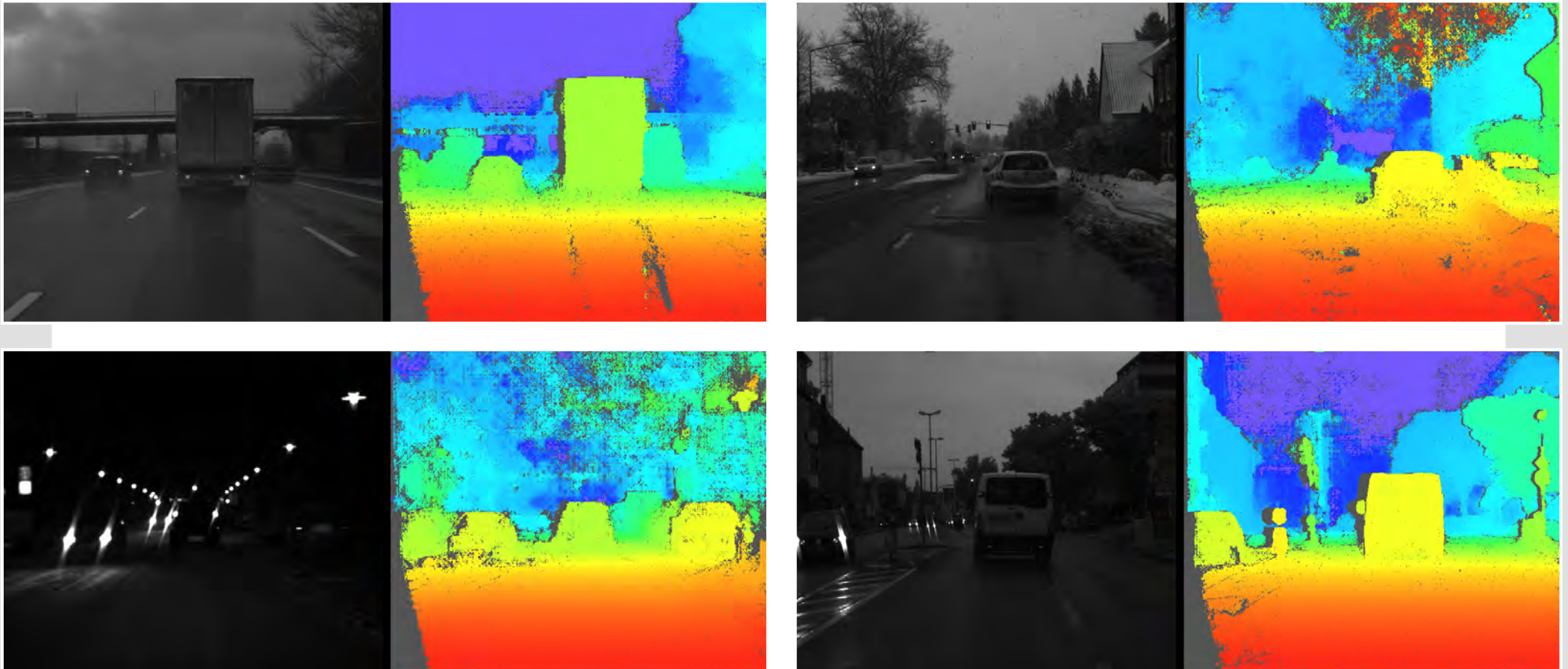
**Robert Bosch GmbH:** Wolfgang Niehsen, Dr.

**The Foundry:** Phil Parsonage, M.Sc.

**Robert Bosch GmbH:** Stephan Simon, Dr.

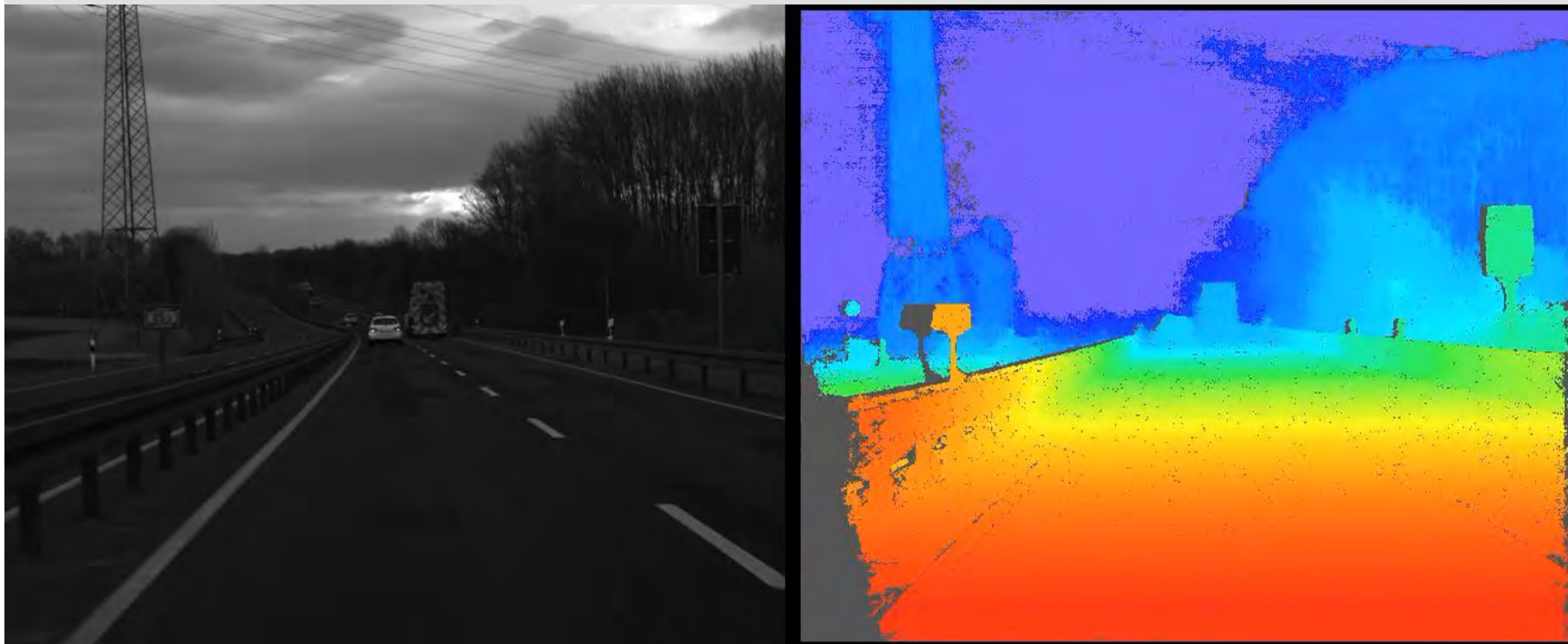
**BMW Group:** Christian Unger, M.Sc.

# iSGM wins the 'Bosch Robust Vision Challenge' at ECCV 2012 ([hci.iwr.uni-heidelberg.de/Static/challenge2012](http://hci.iwr.uni-heidelberg.de/Static/challenge2012))



Meister, S., Jähne, B., Kondermann, D.: Outdoor stereo camera system for the generation of real world benchmark data sets. Optical Engineering **51** (2012) paper 021107, 6 pages

# This was the iSGM submission for this competition



No ground truth, decision based on subjective evaluation only

# Differences between SGM and iSGM

Standard SGM works as follows:

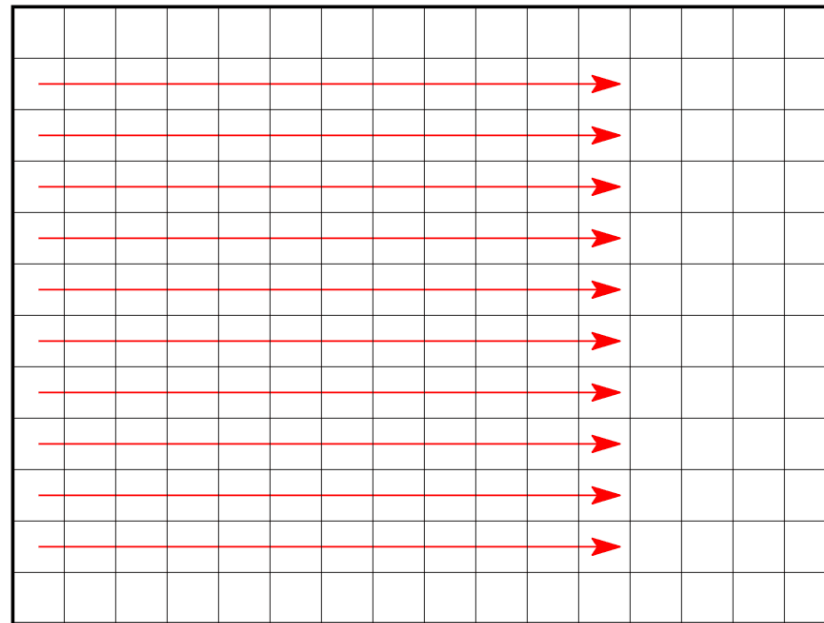
- Based on dynamic programming (DP)
- 8-path integration
- Order of integration paths does not matter



# Standard SGM

Integrates and minimizes along 1D energy paths in 8 different directions.

Read cost matrix  
(local stereo) →



→ Save  
integration matrix

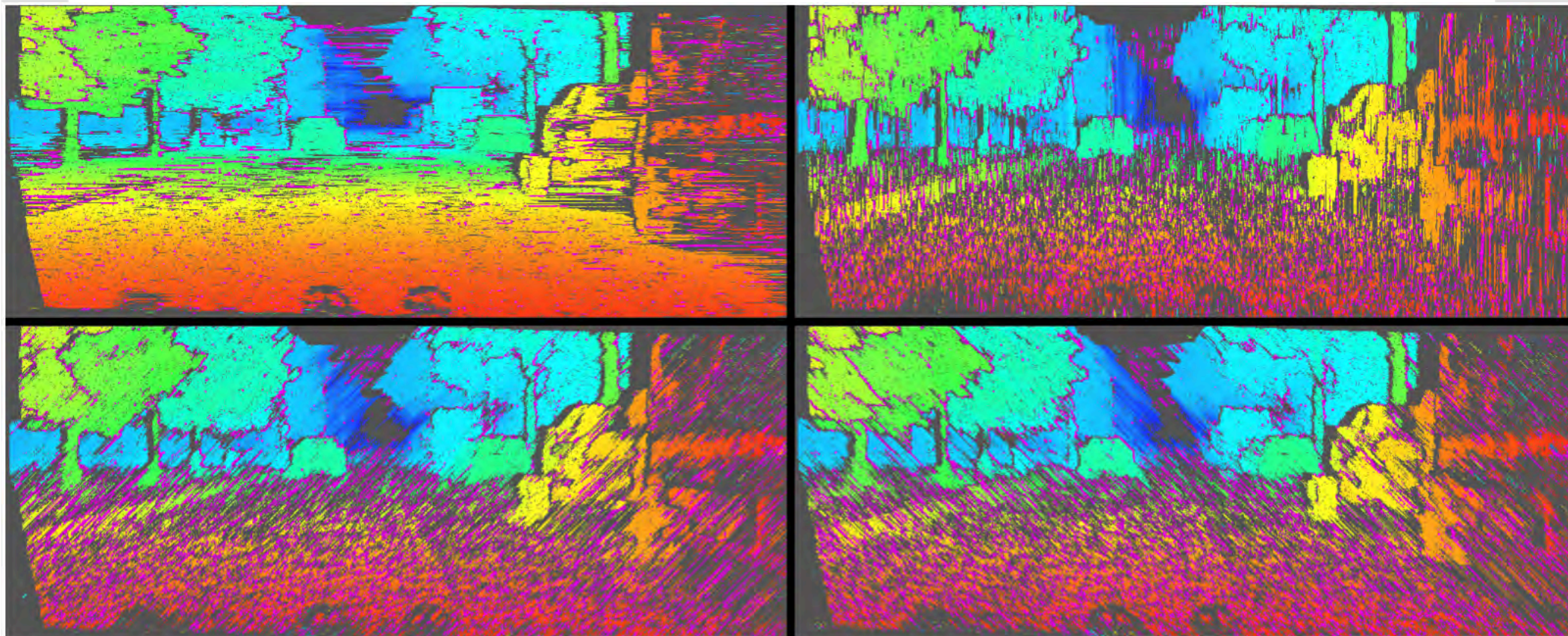
H. Hirschmüller. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In *CVPR*, pages 807--814, 2005.

# iSGM comes with the following differences

- iSGM introduces an order for the integration paths
- It also defines and applies *homogeneous disparity maps* (HDMs)
- See the ACCV (main conference) paper by Hermann&Klette for the definition of HDMs

# DP along individual directions

Invalidated pixels of HDMs are shown in pink.



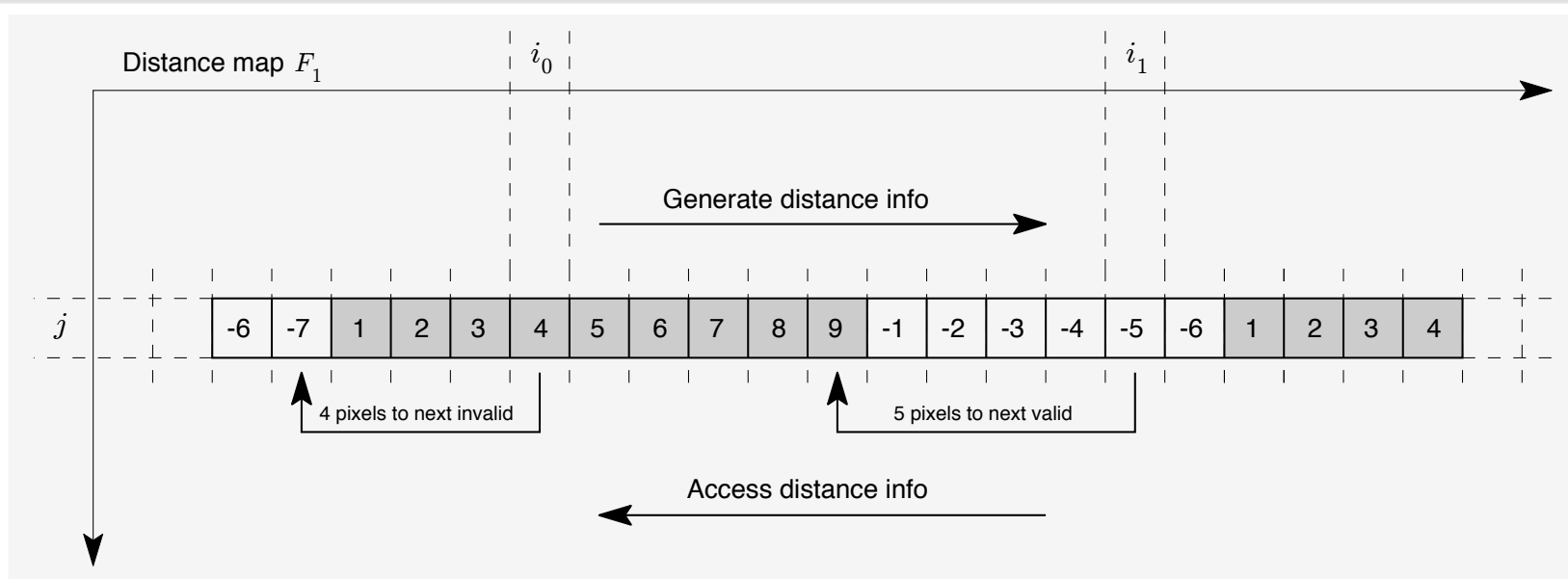
Results from horizontal cost accumulation provide the most robust information especially on the road surface. Therefore we chose the order: horizontal first, then vertical, then diagonal directions.

# Intermediate Evaluations

- Evaluate the accumulation buffer after horizontal integration
- Adjacent scan lines (i.e. above and below) are regularized independently
- We therefore look for non-horizontal consistency (whatever this means) and derive a reliable disparity estimate.
- HOW?



# Semi-Global Distance Maps (SGDMs)



A semi-global distance map is a family of eight 1D distance maps, each in one of the directions of the used integration paths. At each pixel location, a 1D distance map encodes the distance to SG neighbor = the nearest pixel in the opposite category (i.e. valid or invalid).

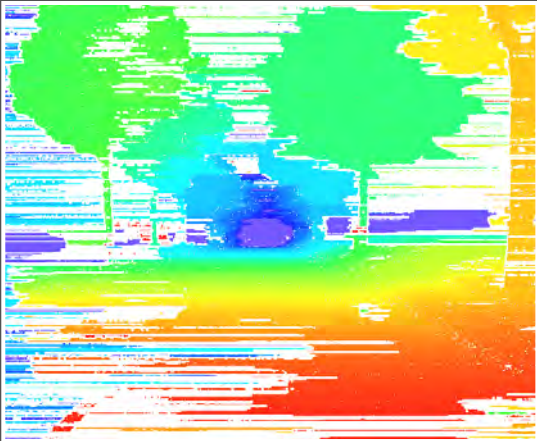
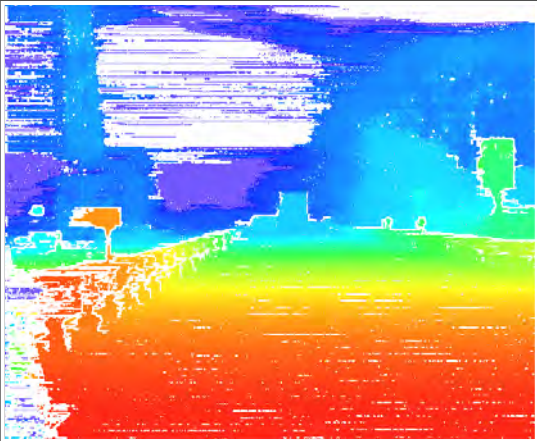
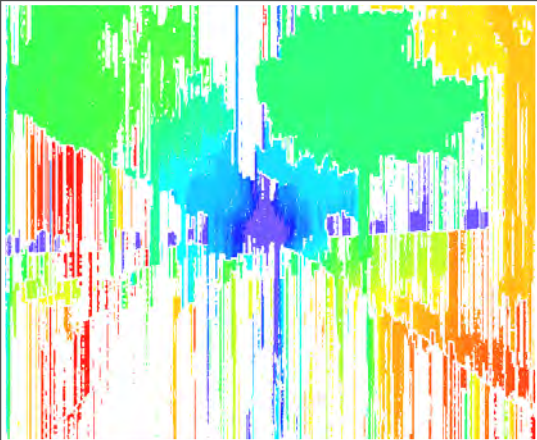
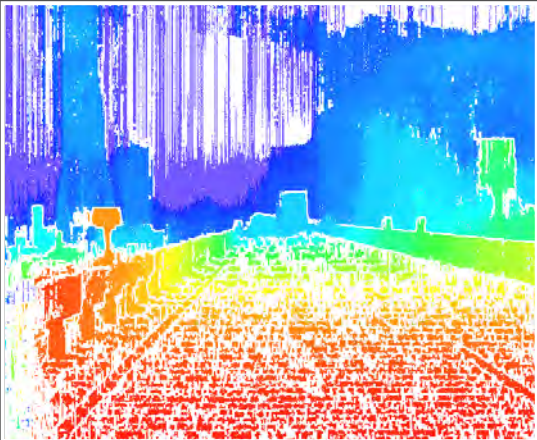
# Search Space Reduction

- Calculate on the SGDM the maximum of these minimum distances to all the semi-global neighbors.
- If the identified maximum is above a threshold → lock this disparity as an intermediate result and reduce the search space at that pixel
- This also reduces the memory requirement for the used hardware.

# Search Space Reduction for Invalid Pixels

- Also possible for invalid pixels.
- The assumption is that the search space can be locked to be between the minimum and maximum disparity of all the semi-global neighbors.

# Horizontal vs. vertical again

	Challenging	Easy data
Horizontal	 A landscape image with horizontal lines of color, showing a road and trees. The image is heavily distorted with horizontal streaks and noise, making it difficult to recognize.	 A landscape image with horizontal lines of color, showing a road and trees. The image is clear and easy to recognize.
Vertical	 A landscape image with vertical lines of color, showing a road and trees. The image is heavily distorted with vertical streaks and noise, making it difficult to recognize.	 A landscape image with vertical lines of color, showing a road and trees. The image is clear and easy to recognize.



# Problems of over-regularization

To ensure that thin vertical structures remain in the disparity map, we initialize the accumulation buffer by costs (cost priors) from a low resolution SGM.

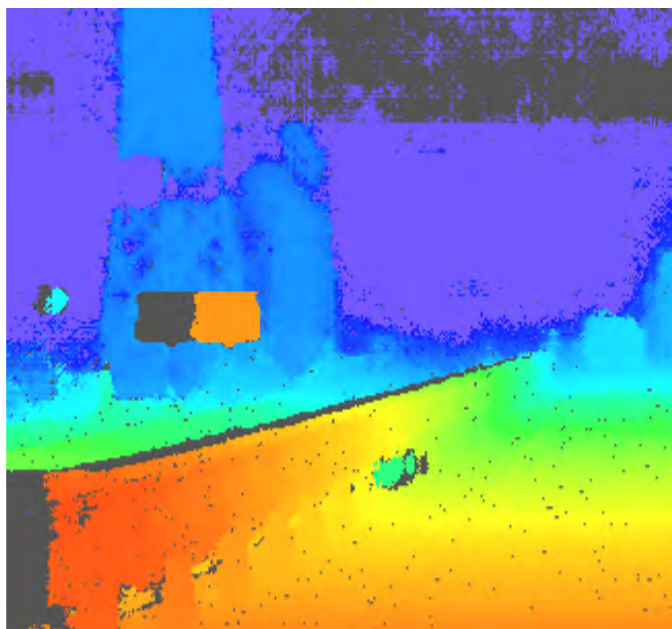
Thus we also allow contributions by vertical integration.

# Example

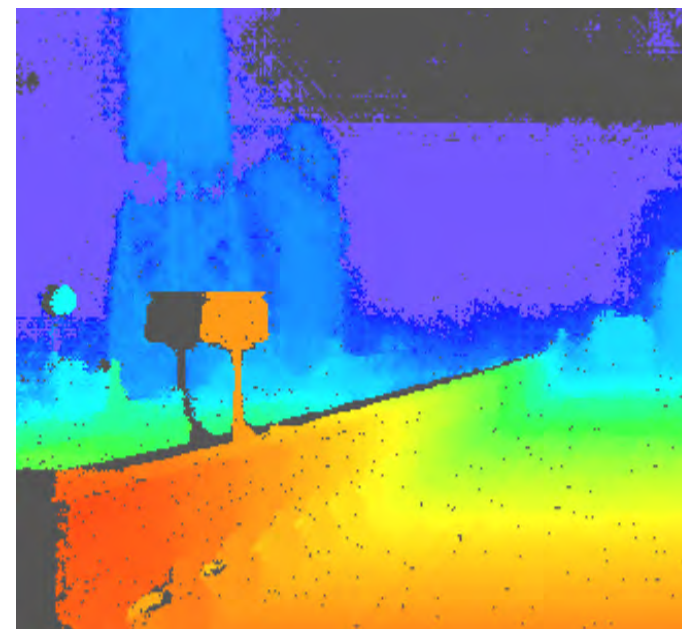
Vertical structures are preserved



Reference image



Without cost prior



With cost prior

# iSGM Summary

improved robustness of stereo processing in the context of SGM

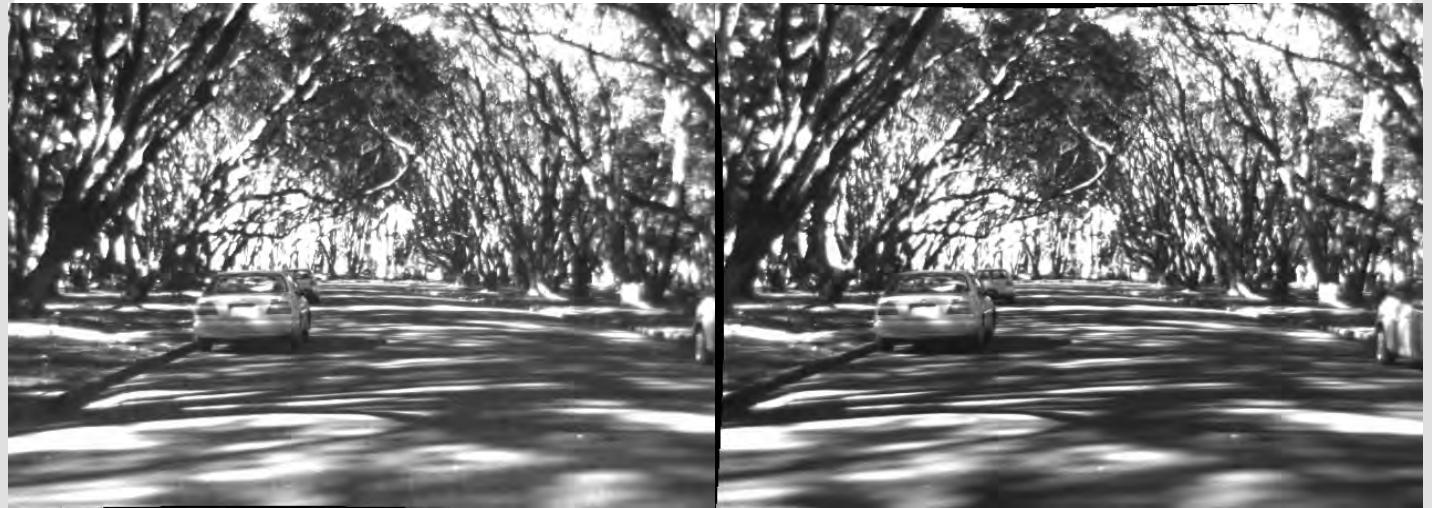
translates a conceptually parallel integration strategy into an iterative scheme

**SGDM** – a novel data structure for

- effective and efficient spatial evaluation
- iterative reduction of the search space by locking reliable disparities derived from a pre-evaluated disparity prior
- promoting horizontally accumulated costs (e.g. stabilizes road surfaces especially for challenging traffic video data)

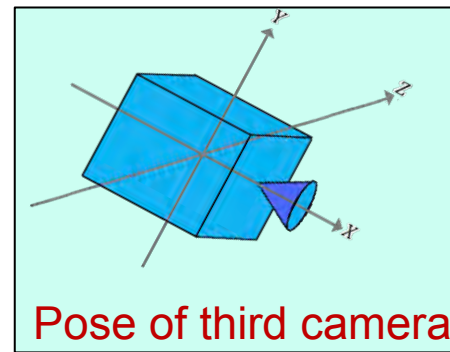
**More of this ...**

Stereo matching  
is practically  
solved for easy  
data; we need  
to focus on  
robustness for  
challenging data

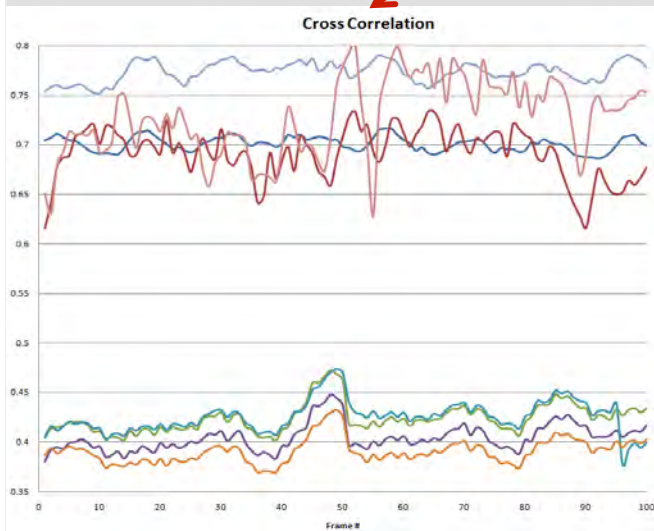




# How to evaluate on real-world data ?



Third image



Virtual Image



Depth Map

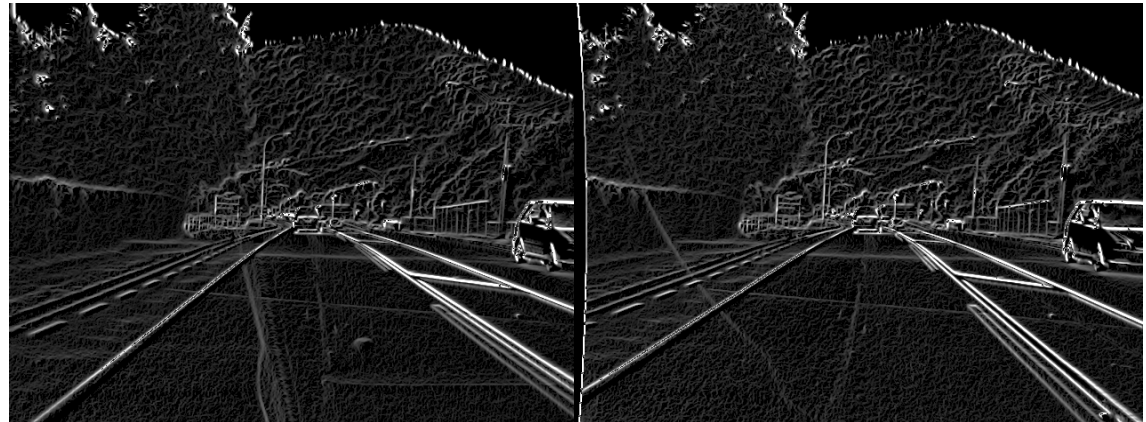
Sandino Morales and  
Reinhard Klette 2009

# Dealing with variations in brightness

**Cost functions:** census, zero-mean SAD, ..

**Preprocessing:**

Sobel edge images



Residuals w.r.t.  
smoothing



e.g. Tobi Vaudrey et al., 2009 ... 2011

# Quality measure

Root-Mean Squared (RMS) error:

$$R(t) = \frac{1}{|\Omega_t|} \left( \sum_{(x,y) \in \Omega_t} [I_t(x,y) - I_v(x,y)]^2 \right)^{1/2}$$

Normalized Cross Correlation (NCC):

$$N(t) = \frac{1}{|\Omega_t|} \sum_{(x,y) \in \Omega_t} \frac{[I_t(x,y) - \mu_t] \times [I_v(x,y) - \mu_v]}{\sigma_t \cdot \sigma_v}$$

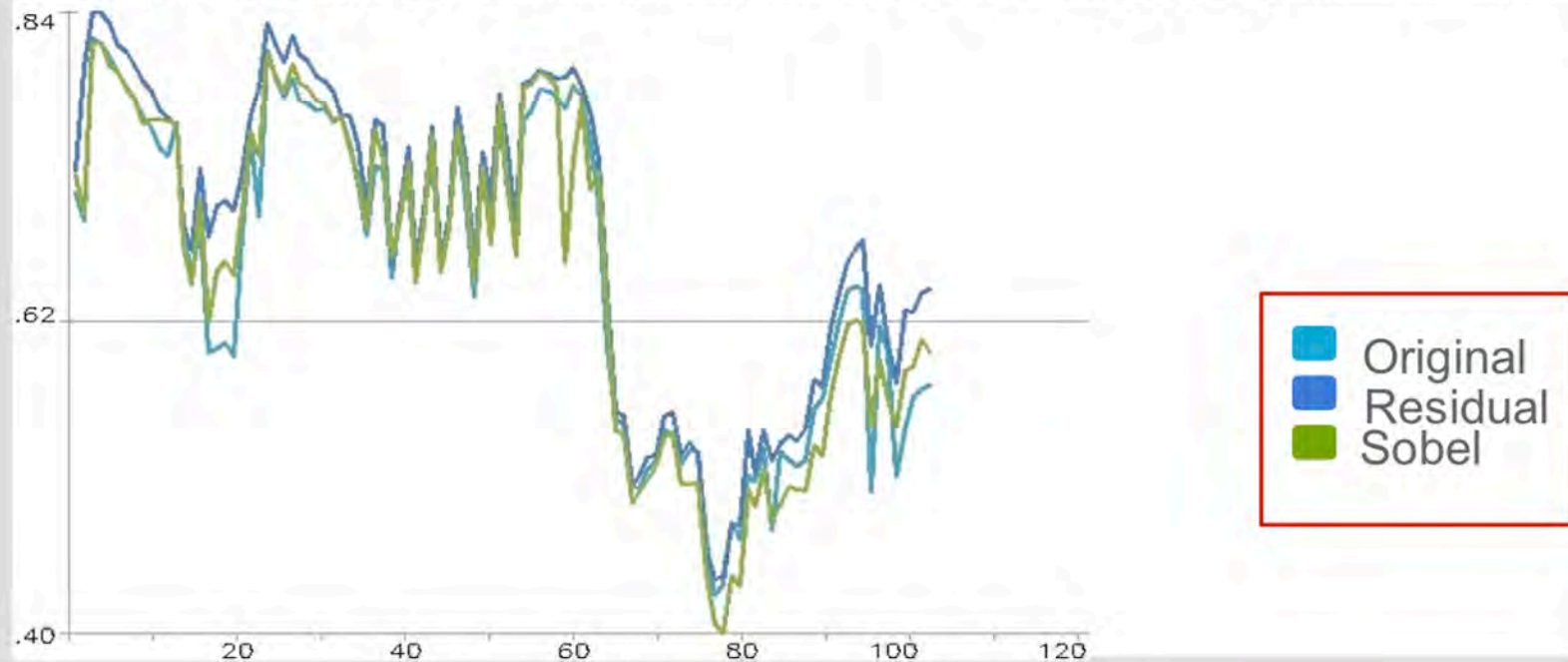


# Example: Belief propagation stereo on 120 frames

BP on original data



BP on residuals w.r.t. smoothing

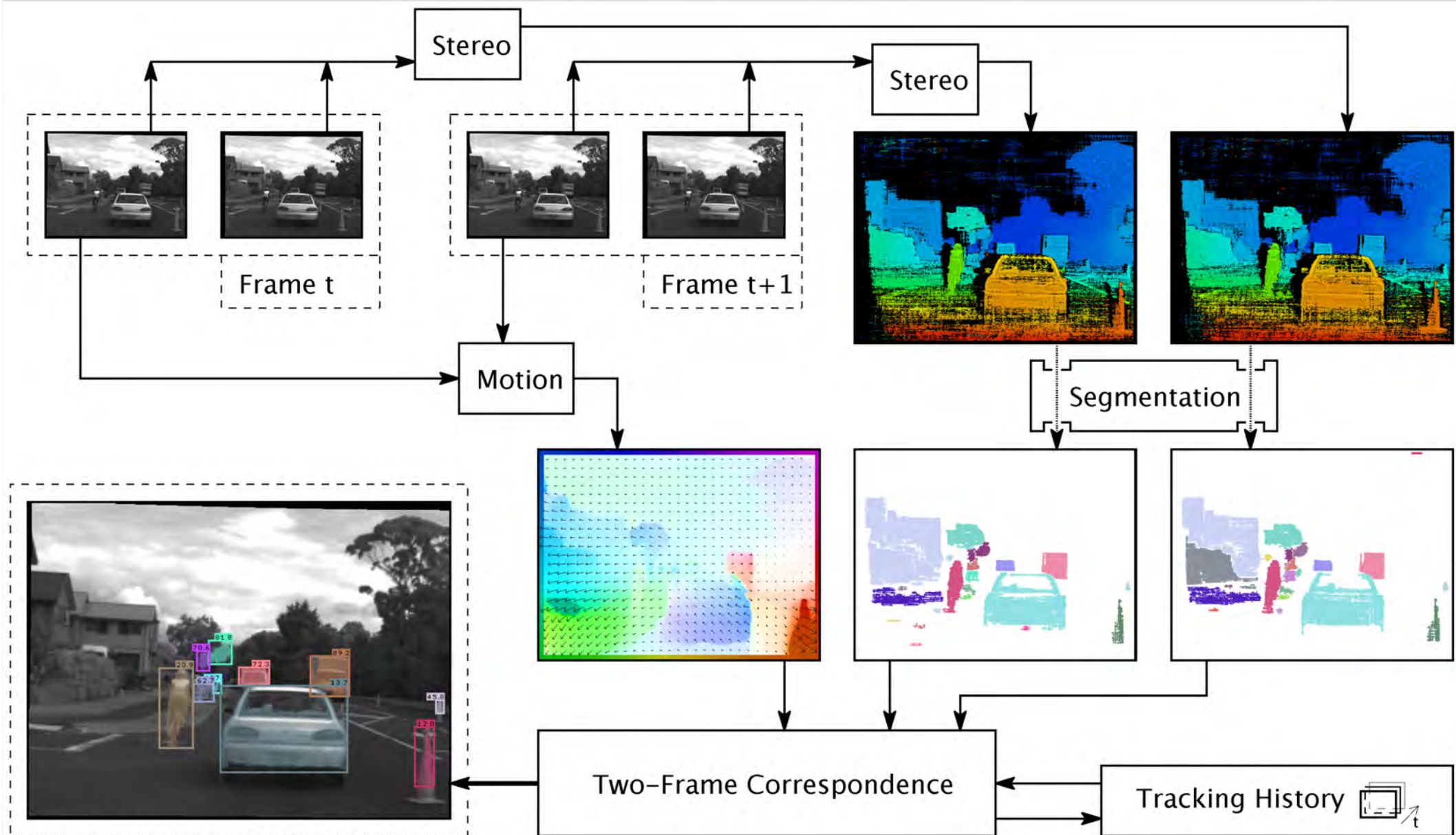




# Summary for evaluations on real-world data

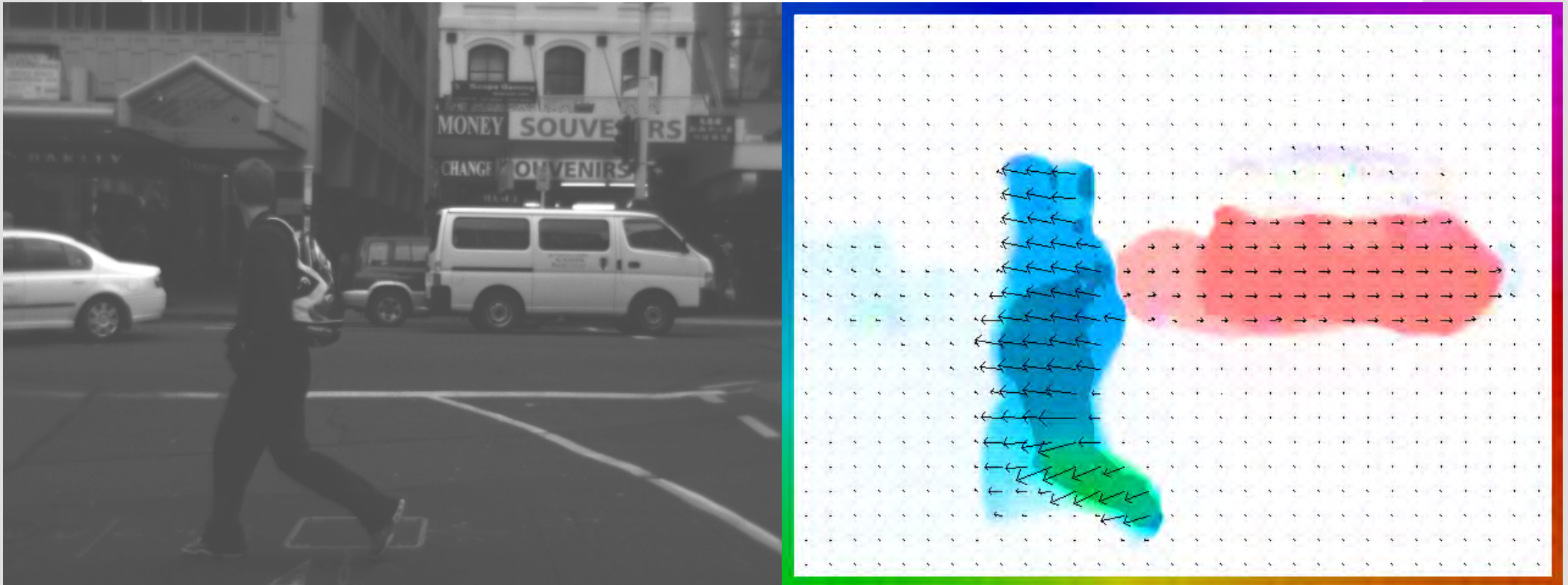
- There is no all-time winner.
- Conclusions can be drawn when evaluating on data of hours or days of recording, under various conditions (different scenarios or situations)
- Evaluations based on a few frames only (= less than one second) are meaningless as NCC curves show sudden changes in rankings on recorded data
- Different stereo matchers appear to have issues at the same frame – those situations need to be identified and studied

## 2 Segmentation based on disparity maps only



# Motion of patches in depth map

from time  $t$  to  $t+1$  is estimated by optic flow



A set metric is used to identify corresponding patch at time  $t+1$ , selected from patches in depth map at time  $t+1$



# Distance to objects using SGM with mode filter

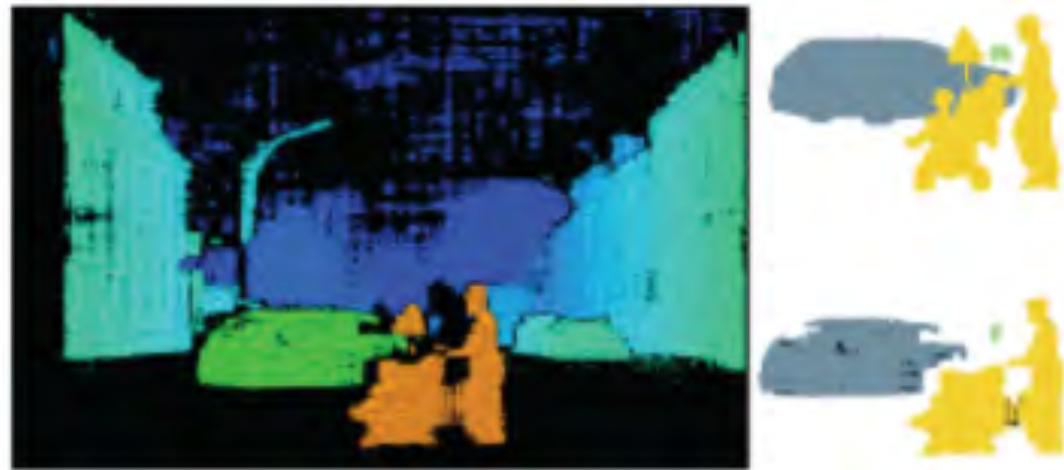




# Summary for segmentation in depth maps only

There is a possibility of **reliable segmentation and tracking in disparity domain**.

It is difficult to compare results; more ground-truth data is needed for mid-level tasks like this.



Top: ground truth as provided in Set 7 of the EISATS online test data

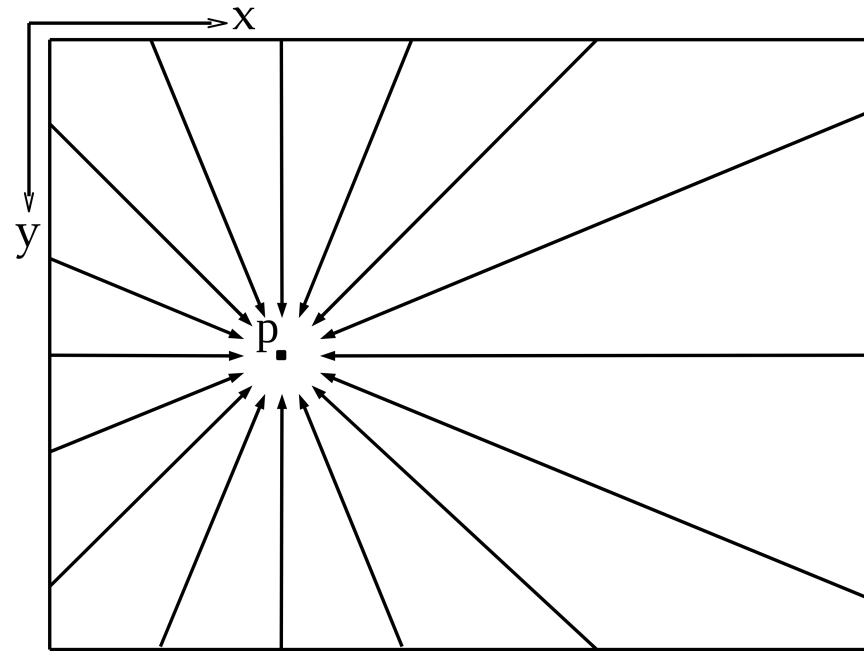
Bottom: our obtained segments

How to measure the difference?

## 3

## Flow by semi-global matching (fSGM)

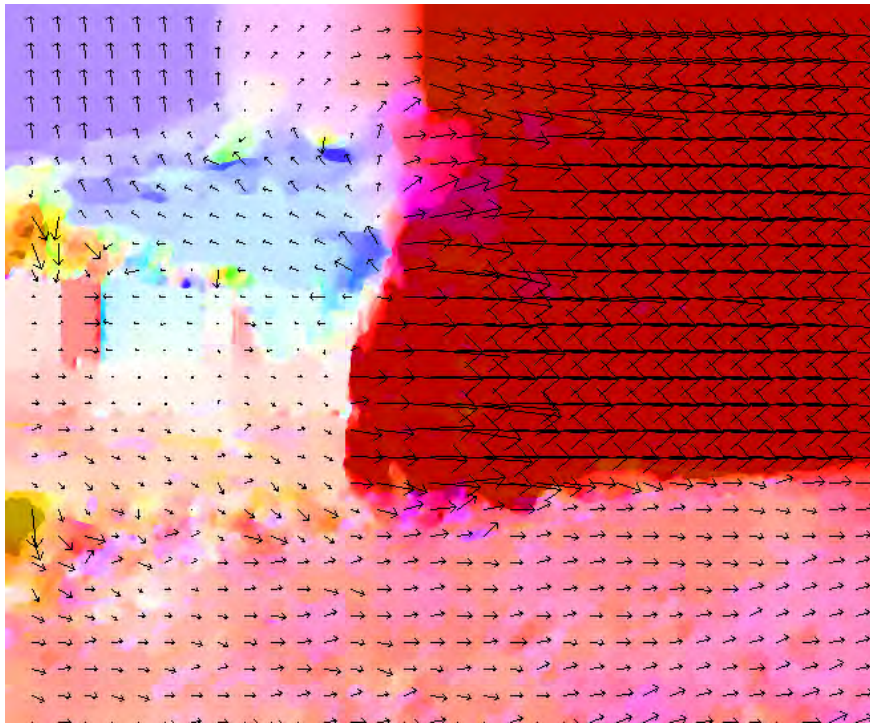
- fSGM is a discrete optic flow estimator
- It uses dynamic programming in combination with the SGM integration strategy
- Maps (1D disparity) labels 1-1 on 2D flow vectors
- Combines mid-scale flow analysis with large-scale analysis



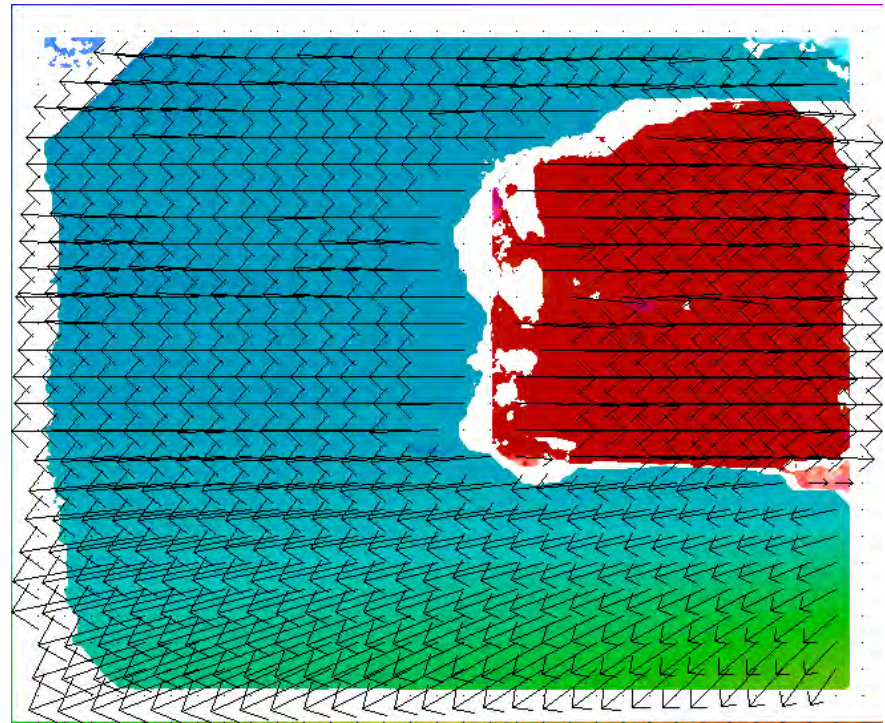
# fSGM on HCI data

Example of a frame of HCI data on which fSGM outperforms variational methods

TV-L1

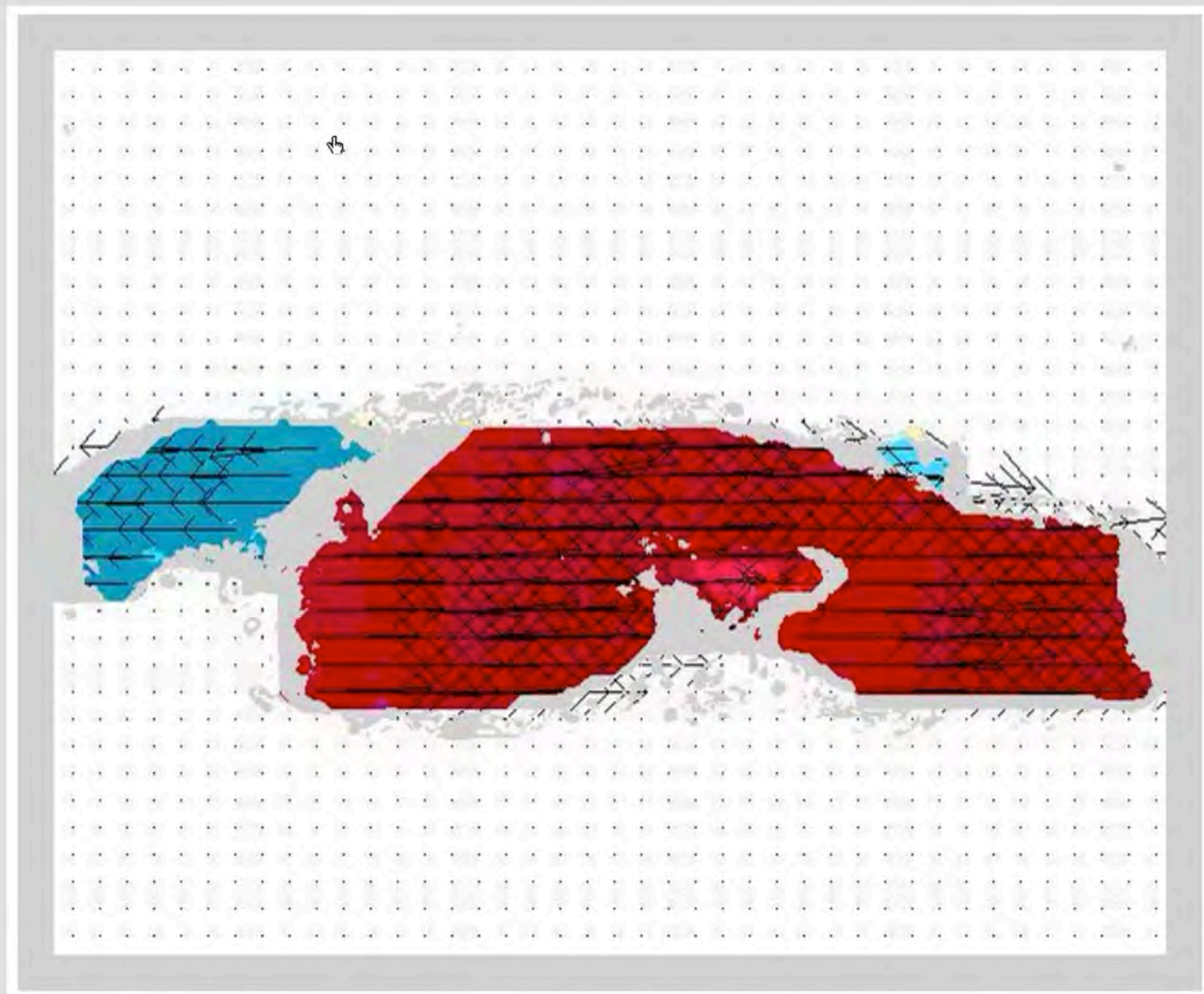


fSGM





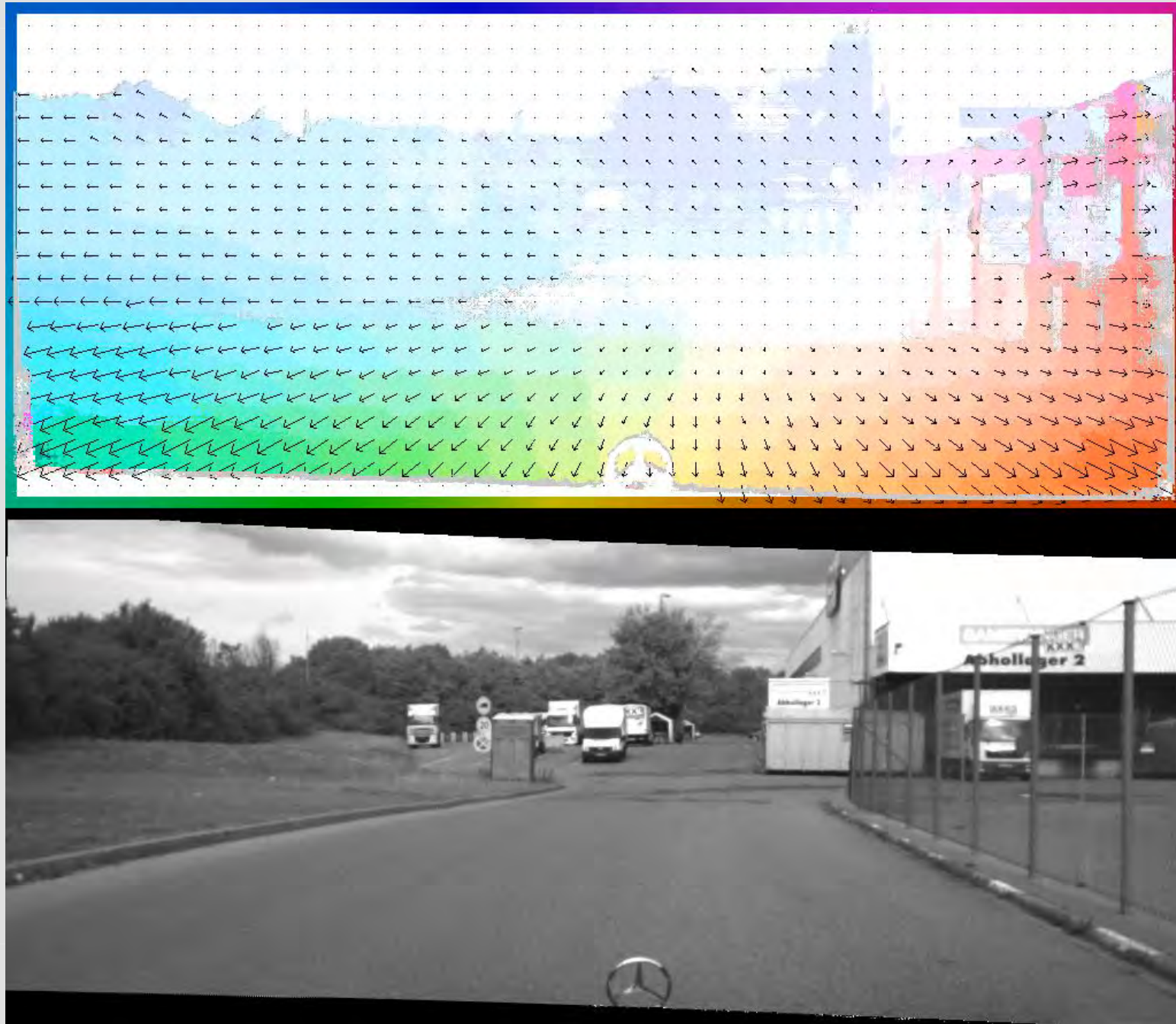
# fSGM and iSGM win the Robust Vision Challenge at ECCV 2012



Meister, S., Jähne, B., Kondermann, D.: Outdoor stereo camera system for the generation of real world benchmark data sets. Optical Engineering **51** (2012) paper 021107, 6 pages



# fSGM submission to ACCV 2012 (10 fps)



# A brief fSGM summary

fSGM is not better than variational methods in general; ranking depends on input data.

**But we see it as a robust alternative.** fSGM indicates that discrete methods are also (or: still) of relevance for optic flow calculations.

More research in the field of DP and optic flow is required to understand effects (for particular situations or scenarios) even better.

For more details, see my talk later at the IMV workshop.

4

## 3D incremental ( $t$ , $t+1$ , ...) reconstruction (5 fps)

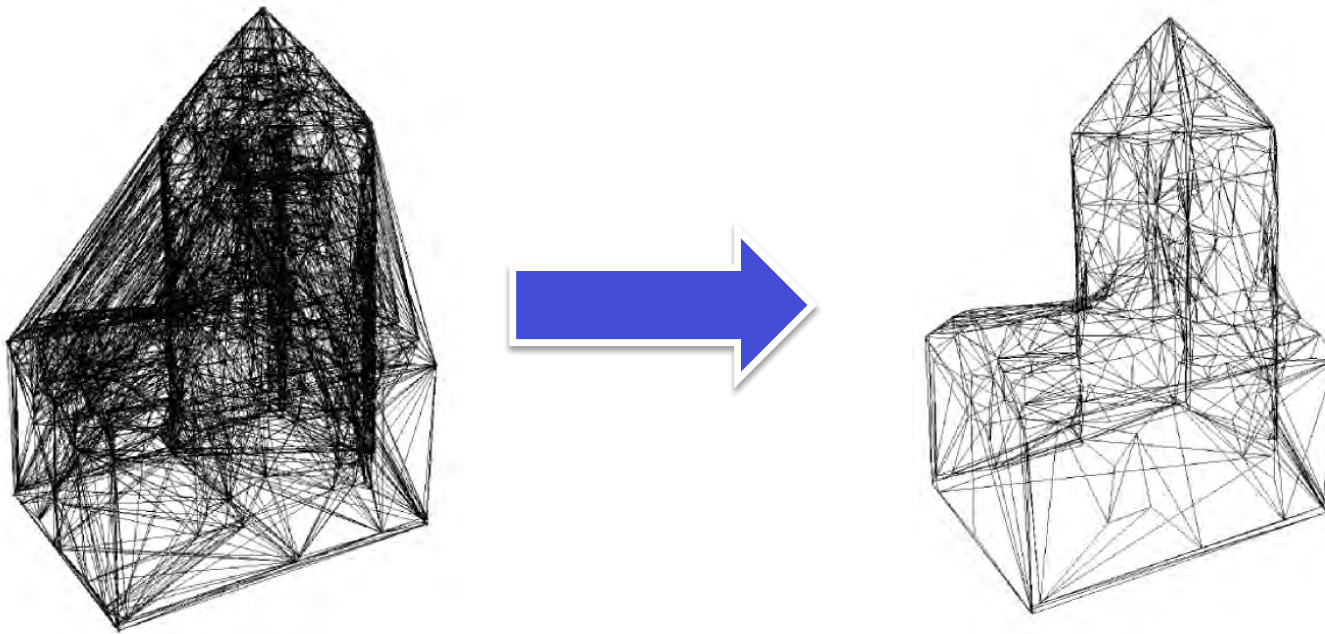


OpenCV block matcher, Yi Zeng 2012



# Surface Reconstruction

The shape formed by a set of 3D points:  
Tetrahedra and alpha shape (volume carving)



[http://mi.eng.cam.ac.uk/~qp202/my\\_papers/BMVC09/](http://mi.eng.cam.ac.uk/~qp202/my_papers/BMVC09/)



# Landmark Tracking

Use of a simple visual odometry algorithm only

SURF (Speeded Up Robust Feature)

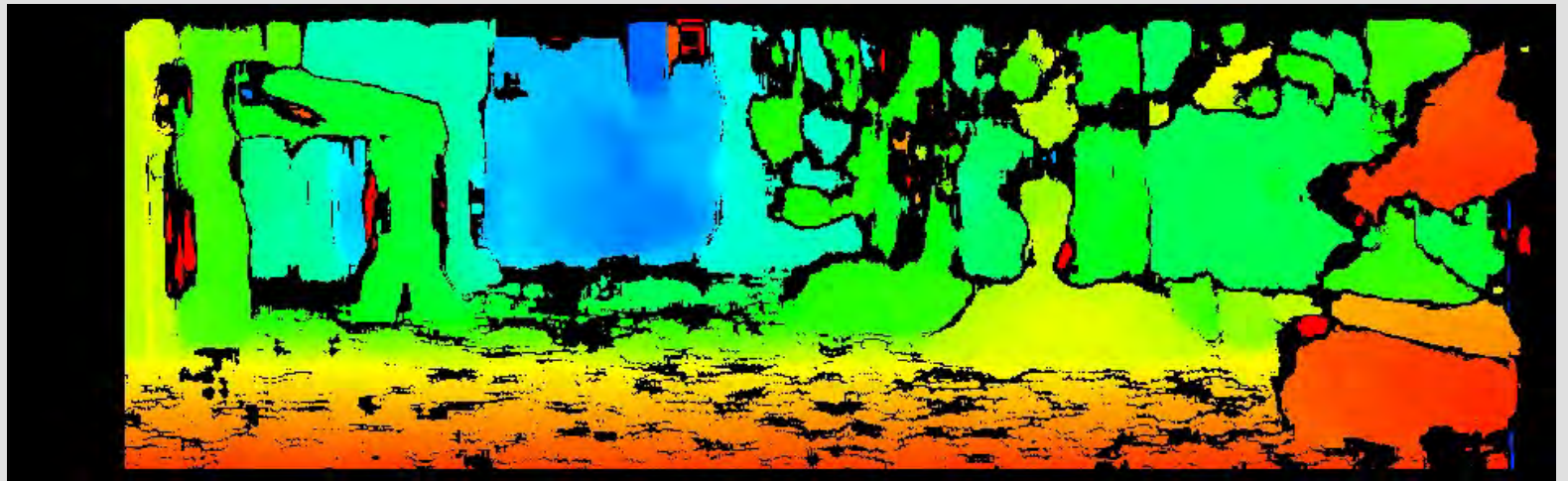
Tracking: Lucas-Kanade algorithm



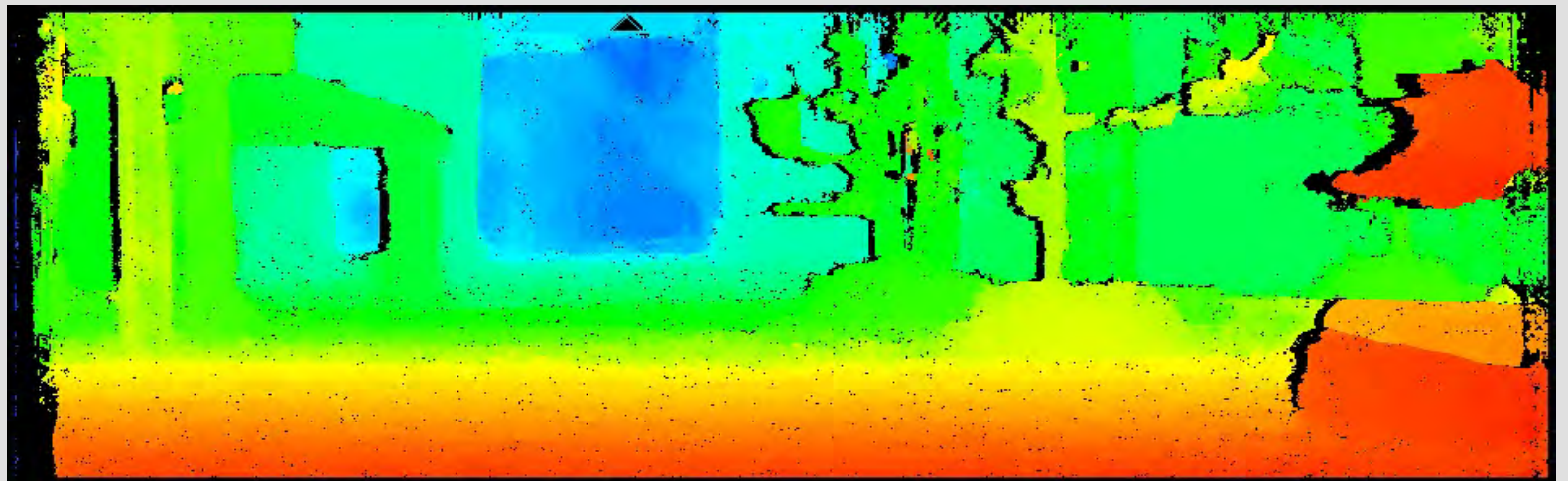
Input



OCV\_BM



iSGM





# 3D reconstruction using iSGM (5 fps)



## The problem here ...

is that ego-motion needs to be determined at a very high level of accuracy; errors for camera pose in centimeters or one degree already lead to unsatisfactory errors when merging depth data from  $t$  and  $t+1$

Depth data (use of iSGM) at a particular time are basically fine for this application

Also note that the example before was just for 5 fps due to memory limitations



How do we combine landmark observations and IMU data for high-accuracy ego-motion estimation?

Properties of these inputs:

IMU data is noisy and biased. The cheaper the IMU, the worse it is. E.g., smartphones contain the cheapest available.

Monocular landmark observation is bearing only, and also contains noise due to, at the very least, the reality of discrete pixels.

Thus: >>> Kalman filter

# Unscented Kalman filter (UKF)

UKF uses the Unscented Transform (UT) to make predictions

The UT is a statistical approach to predict states which undergo (highly) non-linear transformations:

- Select some points (sigma points) from the state space surrounding the current best estimate of state (i.e. make some smart guesses)
- Make sure they capture the mean and covariance of the state
- Push all the points through either the process function or the measurement function.
- Apply a weighted mean
- Use that mean to produce the relevant covariance matrix.

# Adding Landmarks

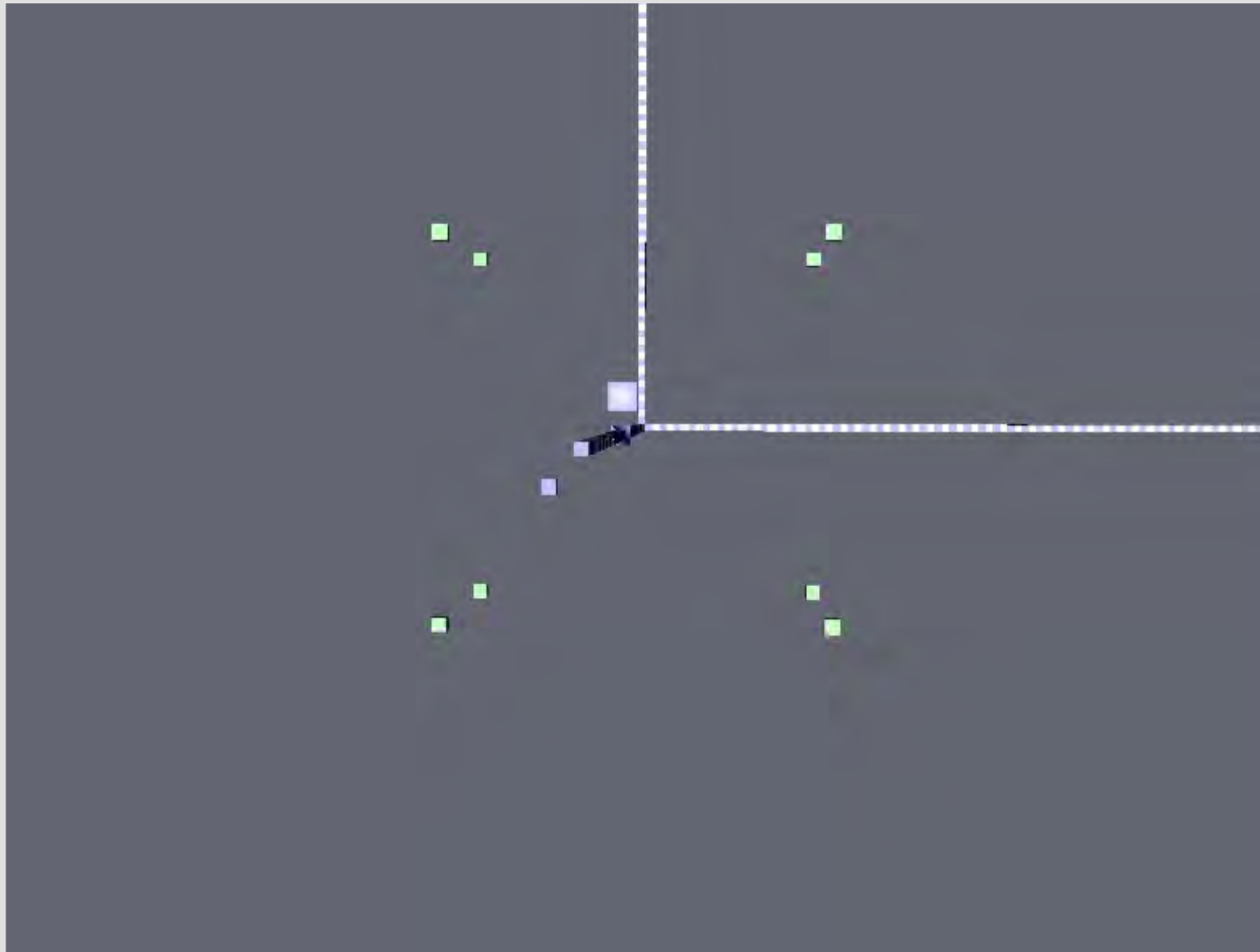
The UKF does not provide a mechanism *per se* for changing the size of the state or its covariance.

But we are always encountering new landmarks which have to be added to the state (with some noise). You cannot just stick them on the end.

How do we figure out the new state covariance matrix, especially the off-diagonal elements?

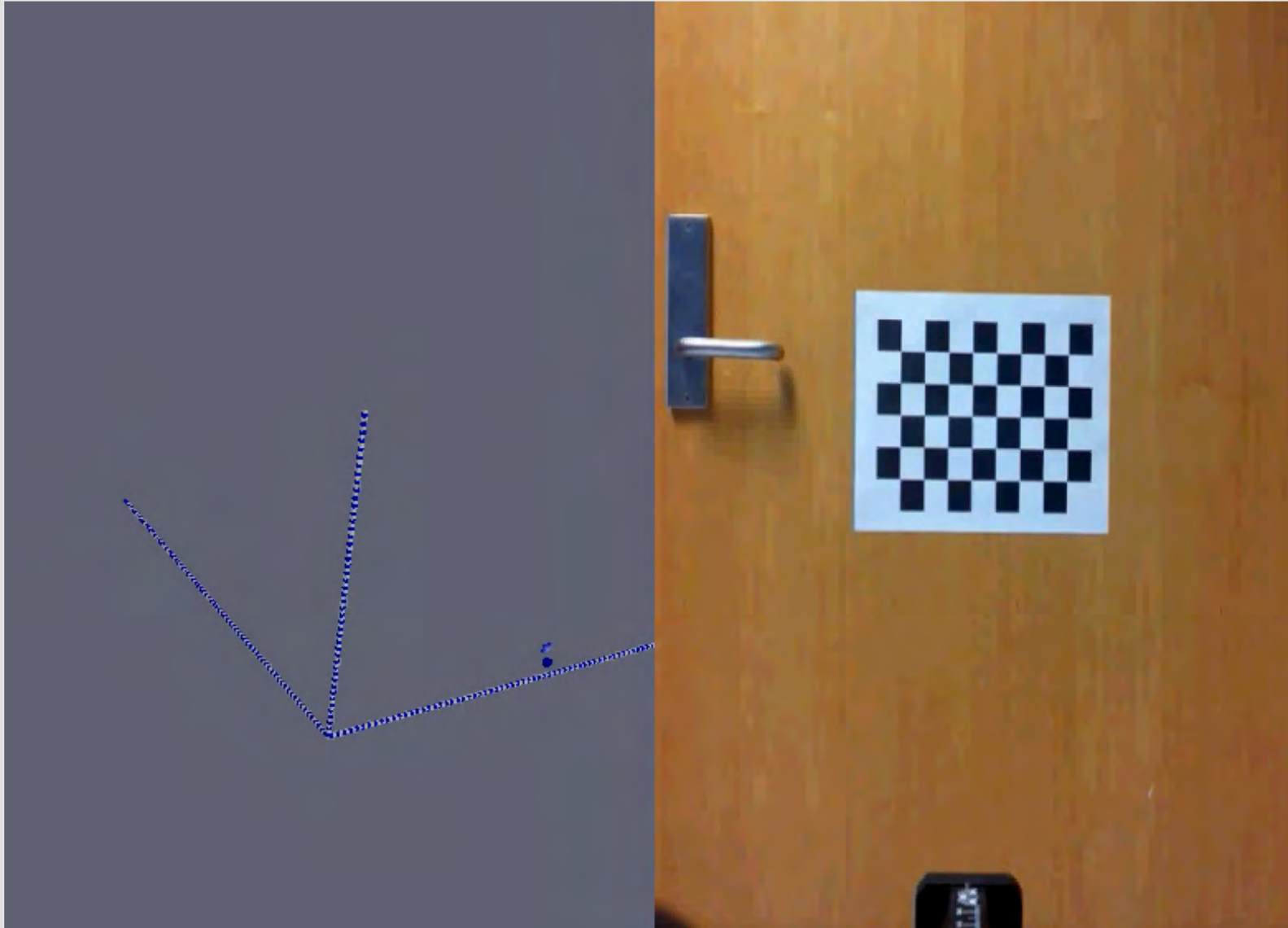
**By the UT:** it takes a state and its covariance, applies a function to these, and provides the new state and covariance.

# Simulated pose detection

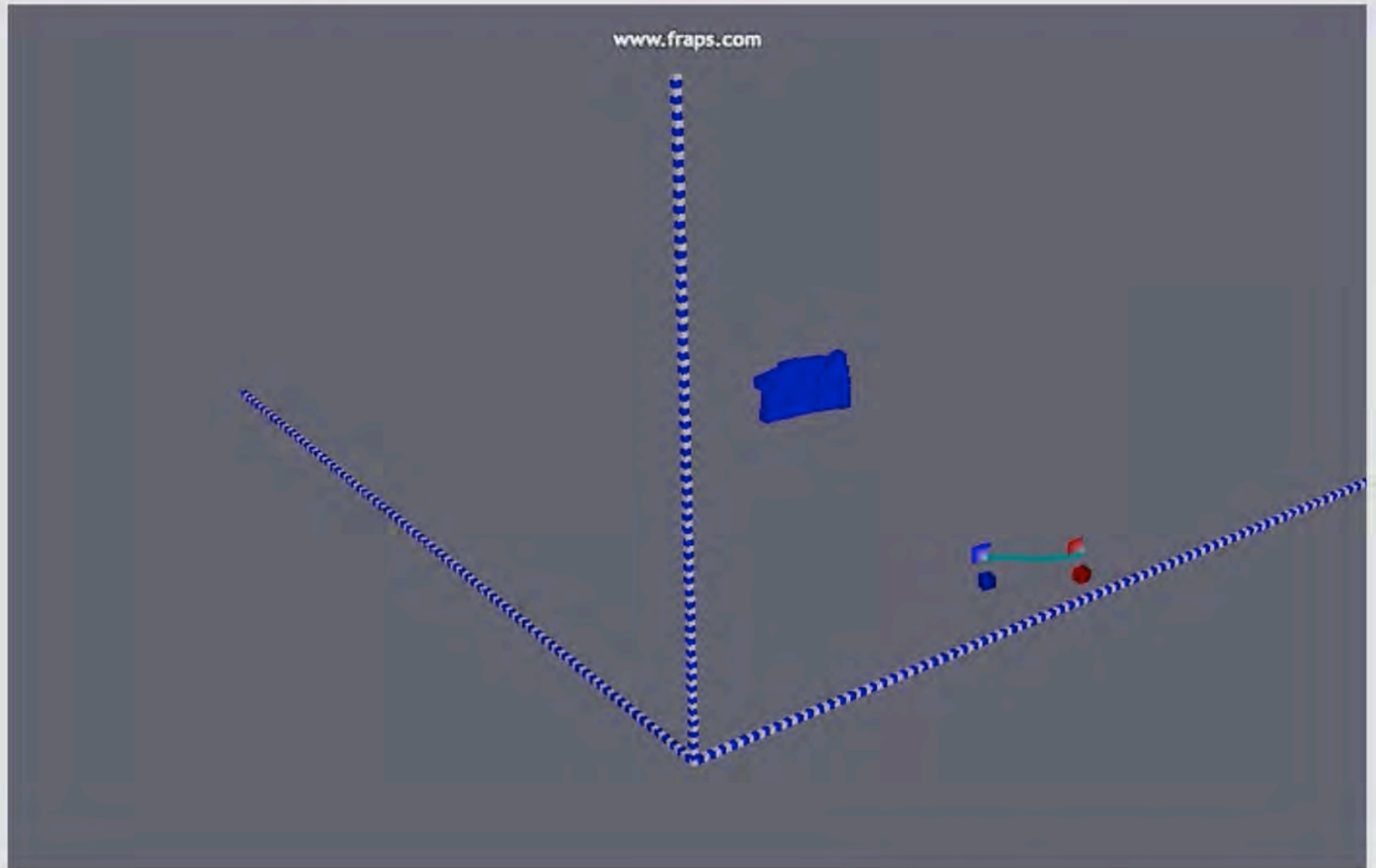




# Pose detection of handheld mobile phone



# Convergence of landmarks



6

## Increasing the complexity of camera motion



Hongmou Zhang 2012



# Recorded monocular video from a hexacopter

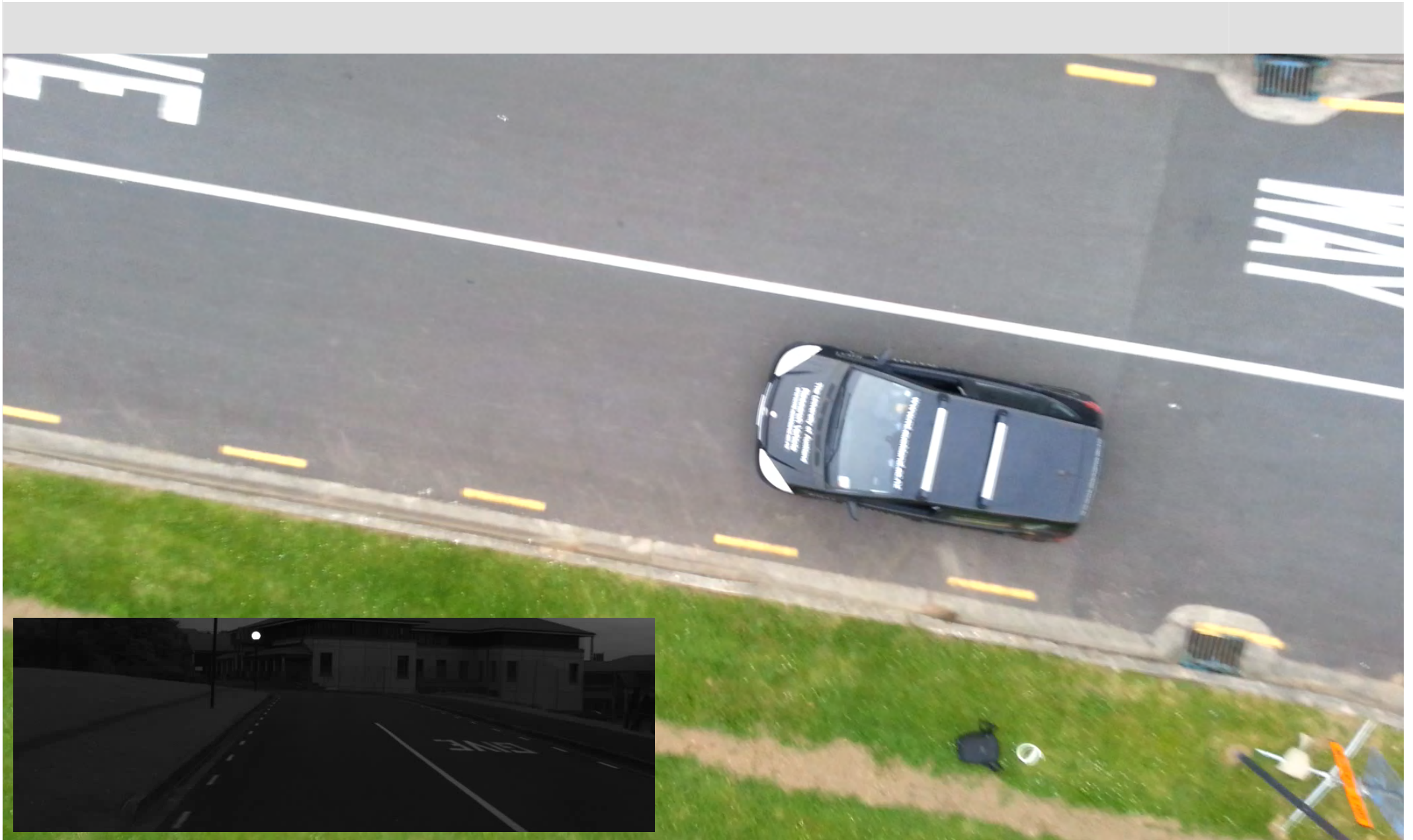




# 1st issue: vibration



# Hexacopter follows HAKA 1



Junli Tao, Hongmou Zhang, Boksuk Shin, and Dongwei Liu 2012



# More issues defined by consistency

of trajectories and camera pose, and time-synchronization



# Various interesting new tasks

**A tool for obtaining ground truth for object distances**  
“on the ground” assuming accurate camera pose data and aiming at an “alignment” with HAKA1 movements

**Complete 3D road environment model** with stereo data obtained from the top views

etc. etc.







# Conclusions for mobile outdoor vision

**Quality of stereo analysis results:** about 90% where we want to be; the challenge is difficult data; the focus should shift on robustness

**Optic flow:** about 60% only, even for easy data not yet satisfactory solved; robustness not yet the focus

Incremental 3D modelling shows: stereo accuracy is O.K. frame by frame, but **ego-motion data** (e.g. visual odometry) not yet at required level of accuracy

Unscented Kalman filter provides in principle a way for accurate visual odometry, but tracked **landmarks not yet stable enough** to provide reliable input

This is even more apparent for cameras freely moving in 3D space (e.g. X-copter applications)