# Stereo Analysis for Driver Assistance Systems

Reinhard Klette

Tobi Vaudrey, Sandino Morales, Jorge Sanchez, Zhifeng Liu, Ali Al-Sarraf, Simon Hermann, James Milburn, Je Ahn, Shushi Guan, and Xuan Guo

The University of Auckland, Tamaki campus

.*enpeda*.. Project

# Vision-based Driver Assistance Systems (DAS)

Test vehicle (A class) at UoA, 2008

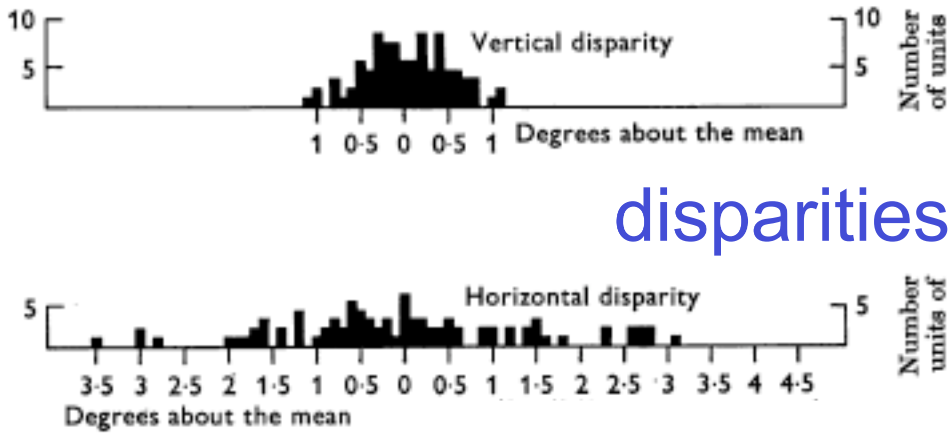

Test vehicle (S class) at Daimler AG, 2007

# Vision-Based DAS in the Market

- Mitsubishi Diamante (1995-1996): Camera for lane recognition and Radar for ACC
- Mercedes Truck (since 2000): Lane Departure Warner
- Subaru Legacy (1998-2004): Stereo-based ACC
- Cadillac (2001-2004): FIR Night Vision System
- Toyota (since 2004): Night Vision System, Parking Guide, Lane Monitoring
- Nissan (since 2004): Lane Keeping System
- Honda (since 2004): Lane Keeping System,
- Since 2005 every major car manufacturer offers camera-based driver assistance

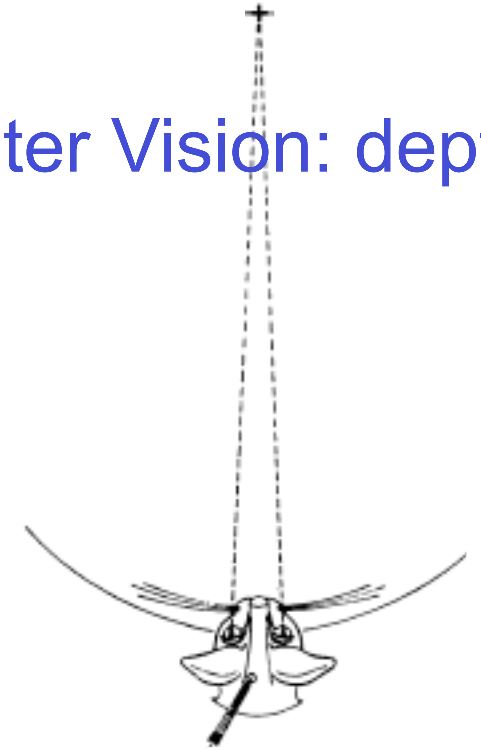
# Stereopsis

= visual perception leading to the sensation of depth




disparities

Computer Vision: depth from detected horizontal disparities



THE NEURAL MECHANISM OF BINOCULAR DEPTH DISCRIMINATION


BY H. B. BARLOW, C. BLAKEMORE* AND J. D. PETTIGREW†

*From the Neurosensory Laboratory, School of Optometry, University of California, Berkeley, California 94720, U.S.A.*



(*Received* 17 *April* 1967)

Possibly the cat, a hunting animal, surveys a wide range of depth at low accuracy, whereas man, a sophisticated toolmaker, surveys a narrow band at high accuracy, varying the position of the band with his convergence movements.

THE NEURAL MECHANISM OF BINOCULAR DEPTH DISCRIMINATION

By H. B. BARLOW, C. BLAKEMORE* and J. D. PETTIGREW†

We have a "cat" at Tamaki campus, without wide-angle vision, just with two 640 x 480 gray level cameras, but with 10 bit per pixel.

HAKA1

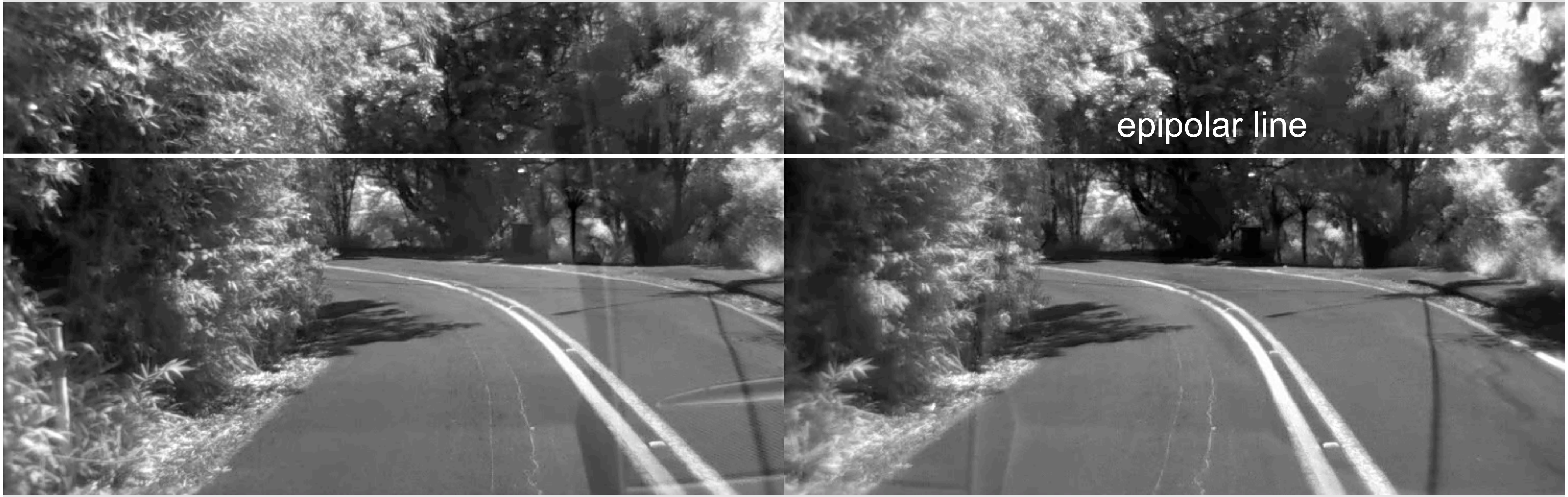
Sponsors: Mercedes Benz New Zealand & Coutts Cars North Shore

# .*enpeda*.., HAKA1 and Tobi Vaudrey on TV

also showing Dr. Uwe Franke (Daimler AG) and Ali Al-Sarraf on TV3 (Campbell Live), New Zealand, in February 2008


(for the clip, visit the TV3 website, search for “smart cars”)

# Preparing the Car and Preprocessing of Recorded Sequences




# HAKA1 in the Waitakeres: geometrically rectified stereo images

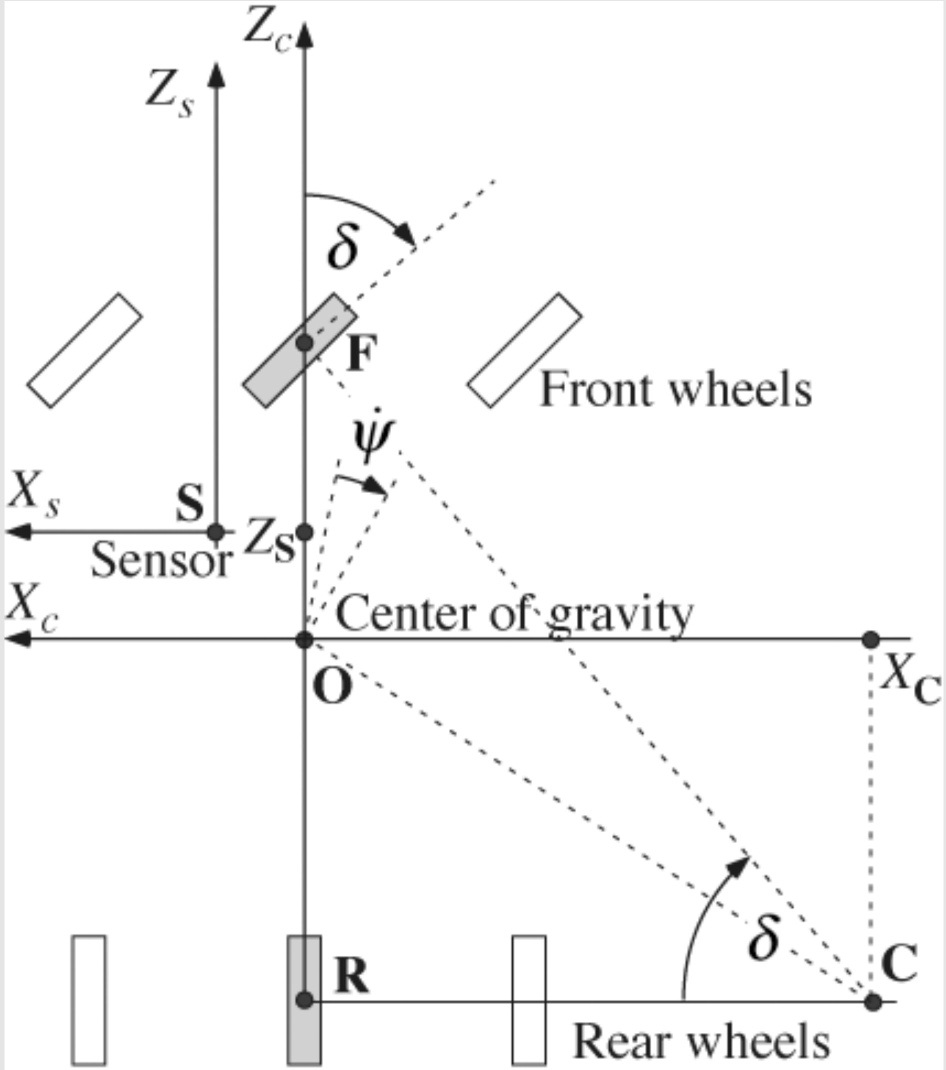
High
Awareness
Kinematic
Automobile
no. 1

(see standard references in computer vision)


## with prior calibration at Tamaki campus

(see J.-Y. Bouguet, Calibration Toolbox)

enpeda..



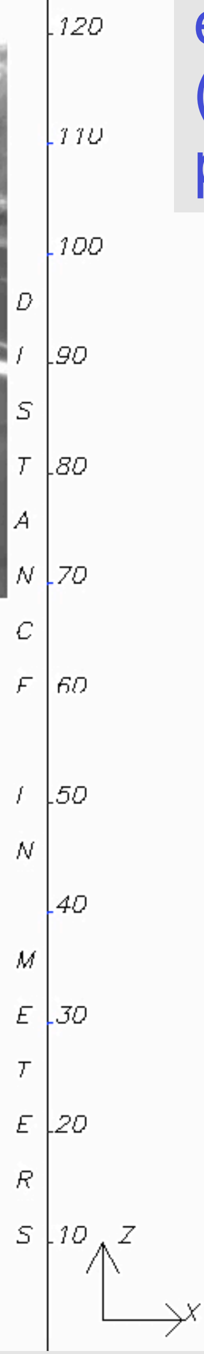

Computer installation, thanks to MIT

Sponsor: Blackhawk Tracking Devices Ltd

We read yaw rate and velocity at 25 Hz from the car's computer, thanks to the research department at

Daimler AG Germany

enpeda..




Trajectory of the ego-vehicle (assumed plane)

## *.enpeda..* Project

**En**vironment

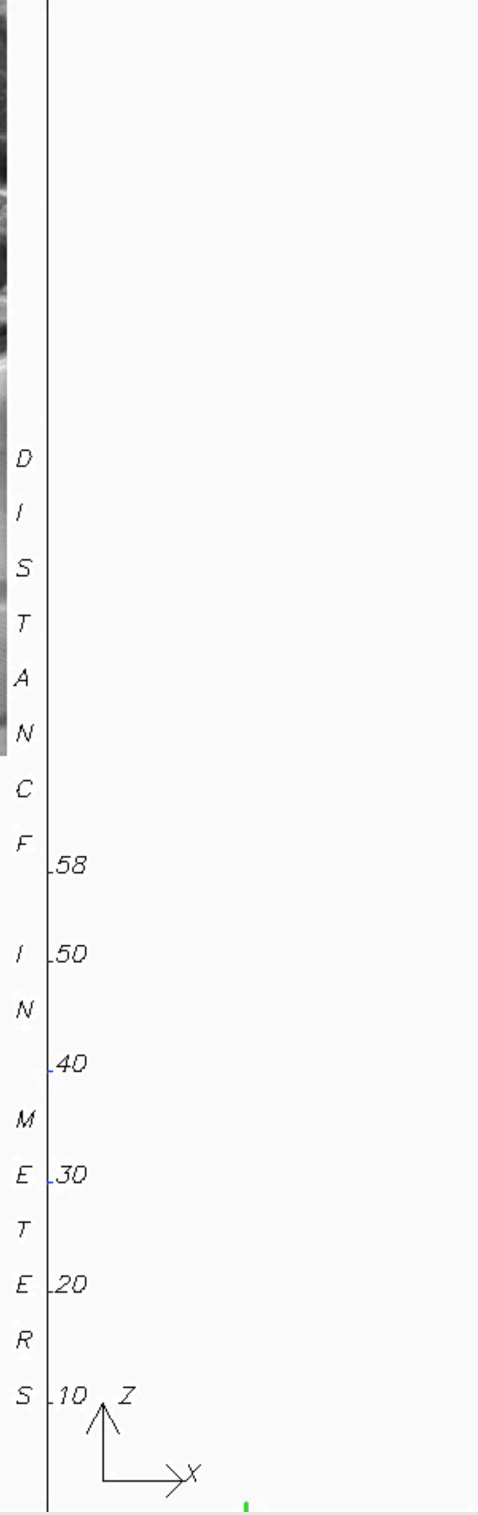
**Pe**rception and

**D**river

**A**ssistance

# Actually: roads are not planar, and there is ego-vehicle motion (pitch or tilt, and roll)

Demonstration (in 2006) of distortion in detected motion due to ego-motion (pitch), by courtesy of Uwe Franke, Daimler AG



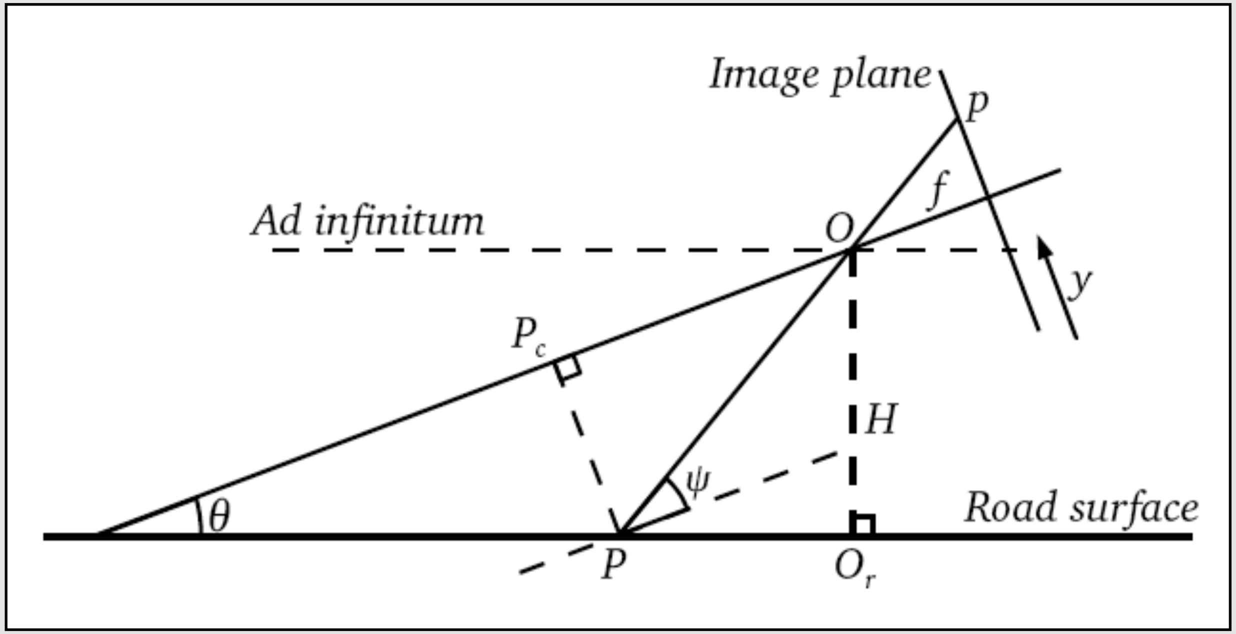

Right: trajectory of the ego-vehicle (still on an assumed plane)

BUT

There are (various) ways to correct for ego-motion.

# One option:

# Ego-motion correction based on calculated disparities

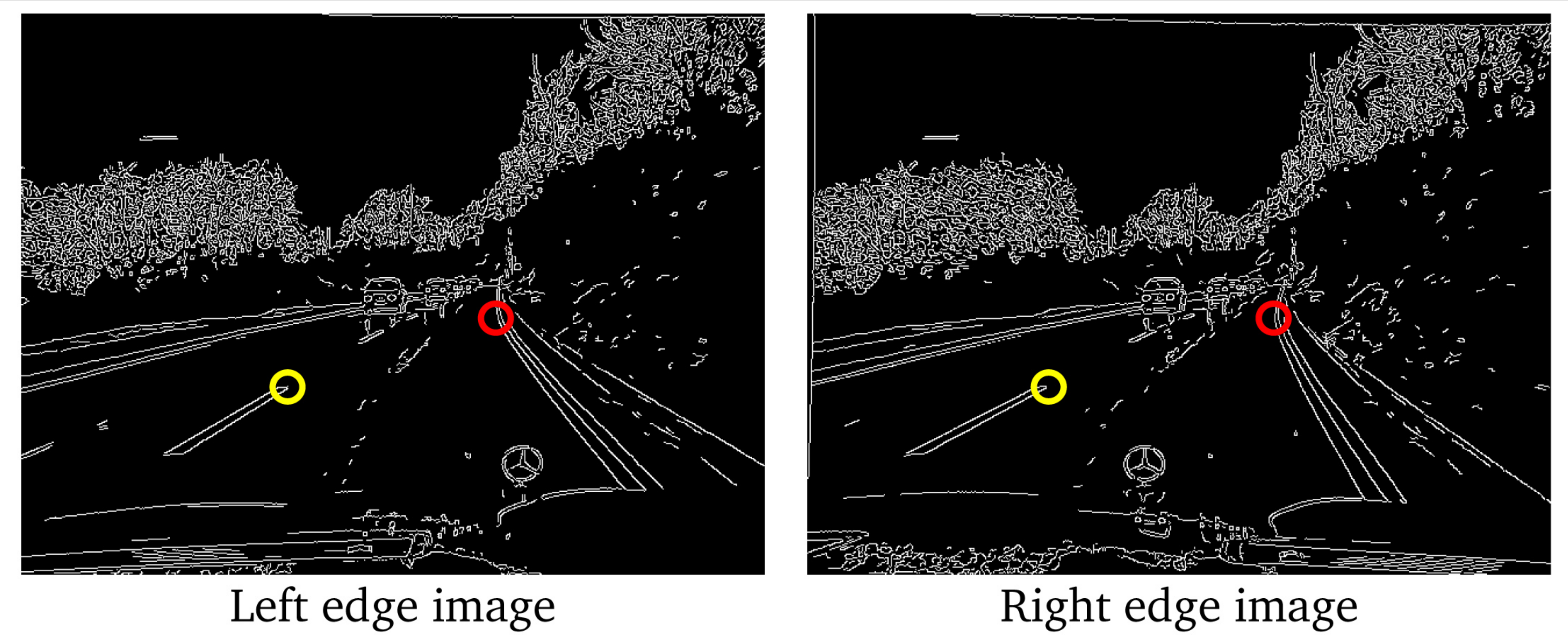



*d* is the disparity at *P*
*b* is the base distance
*θ* tilt angle (pitch)

See [Liu and Klette, ICONIP 2008]

$$\theta = \arcsin\left(\frac{H \cos\psi \cdot d}{b \cdot f}\right) - \psi$$

$$\psi = \arctan\left(\frac{(y_p - y_0)s_y}{f}\right)$$




Estimate a few (dozens) of accurate disparities at feature points

(use of a feature-based stereo algorithm for sparse but accurate disparities)

## Our "top 4" in edge detectors [Al-Sarraf, Vaudrey, Klette, Woo, IVCNZ 2008]:

# Example: estimated mean tilt angles

for (disjoint) intervals of 10 stereo frames within one stereo sequence

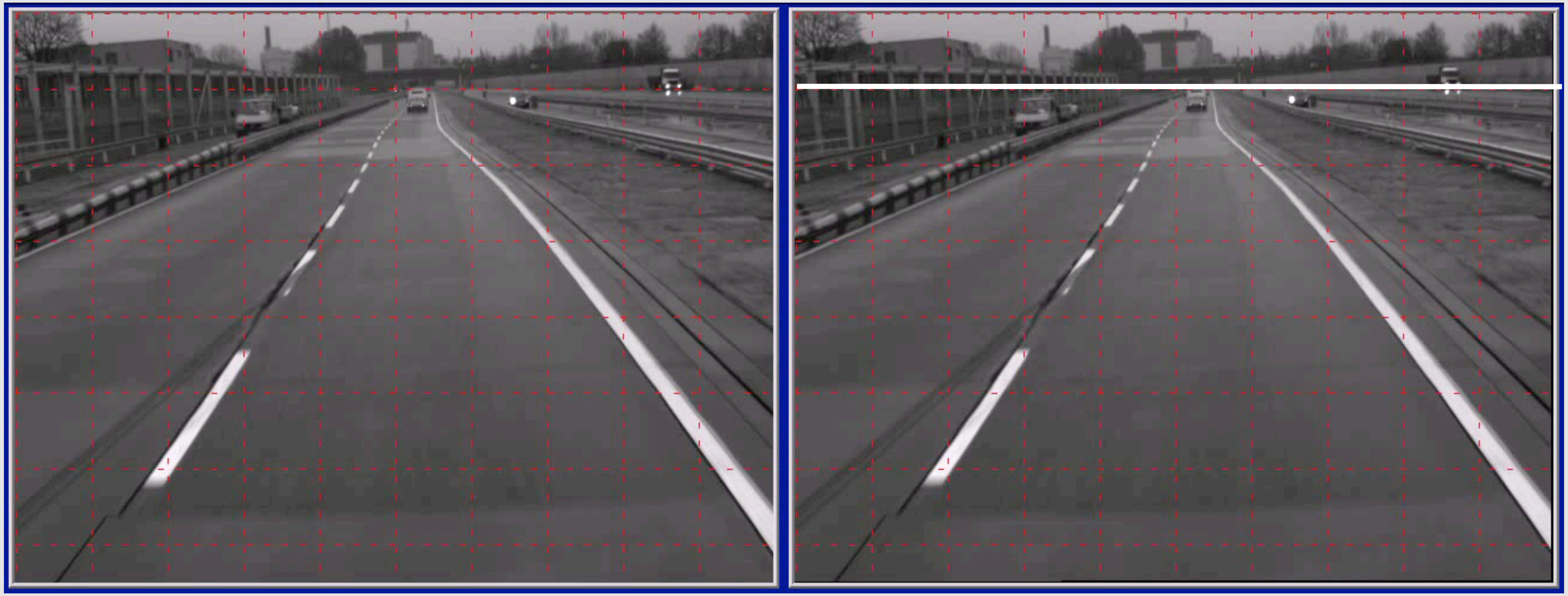
| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| First pair of frames | 1 | 11 | 21 | 31 | 41 | 51 | 61 | 71 | 81 | 91 | 101 | 111 |
| Tilt angle ($10^{-3}$ of a radian) | 80 | 71 | 60 | 60 | 62 | 63 | 65 | 70 | 77 | 71 | 63 | 66 |
| First pair of frames | 121 | 131 | 141 | 151 | 161 | 171 | 181 | 191 | 201 | 211 | 221 | 231 |
| Tilt angle ($10^{-3}$ of a radian) | 60 | 50 | 50 | 59 | 58 | 54 | 55 | 56 | 58 | 53 | 53 | 42 |

See [Liu and Klette, PSIVT 2009]

Demonstration (in 2006) of ego-motion (i.e., tilt an roll) correction.

Left: original sequence. Right: corrected.

By courtesy of Uwe Franke, Daimler AG



# Another option:

# Ego-motion correction based on tracked features

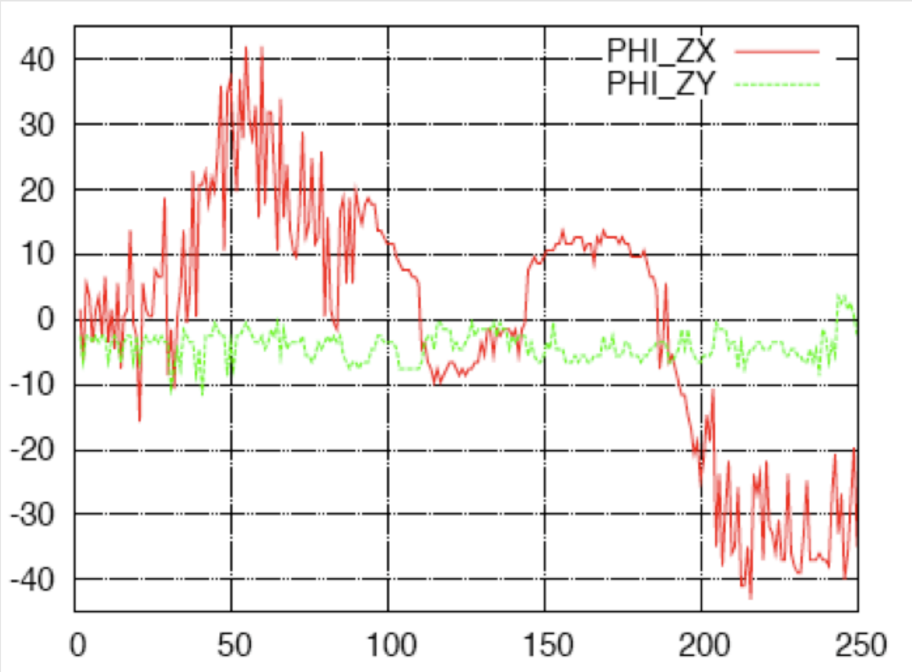
The size of shown circles corresponds to maxima of scale characteristics of tracked features in derivative scale space.

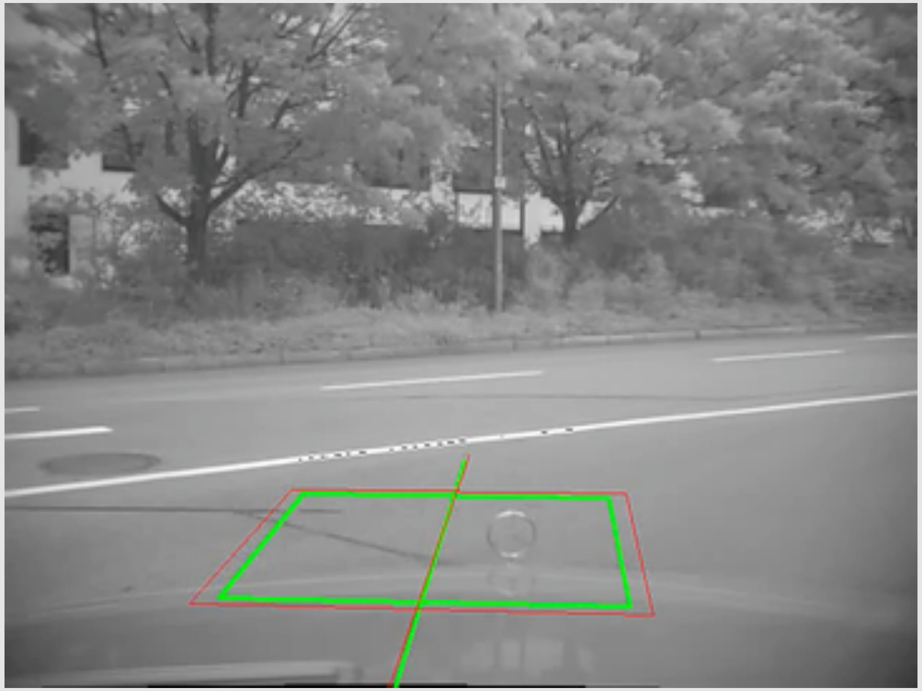
See [Sanchez, Klette, Destefanis, PSIVT 2009]

Tracked features in a test sequence provided by Daimler AG in 2007

Estimated mean navigation angles (yaw and roll) for each of the 250 stereo frames.

Illustration of those mean navigation angles (with and without Kalman filtering).



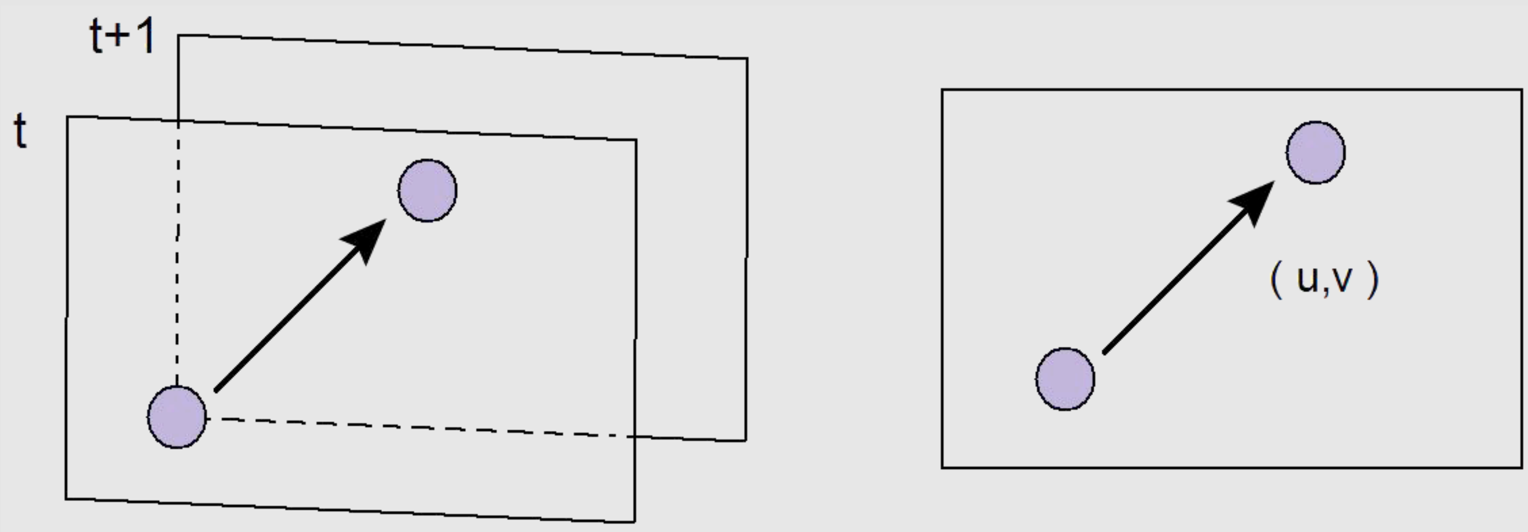
# Motion Analysis

with one eye

# Local motion (optic flow) estimation

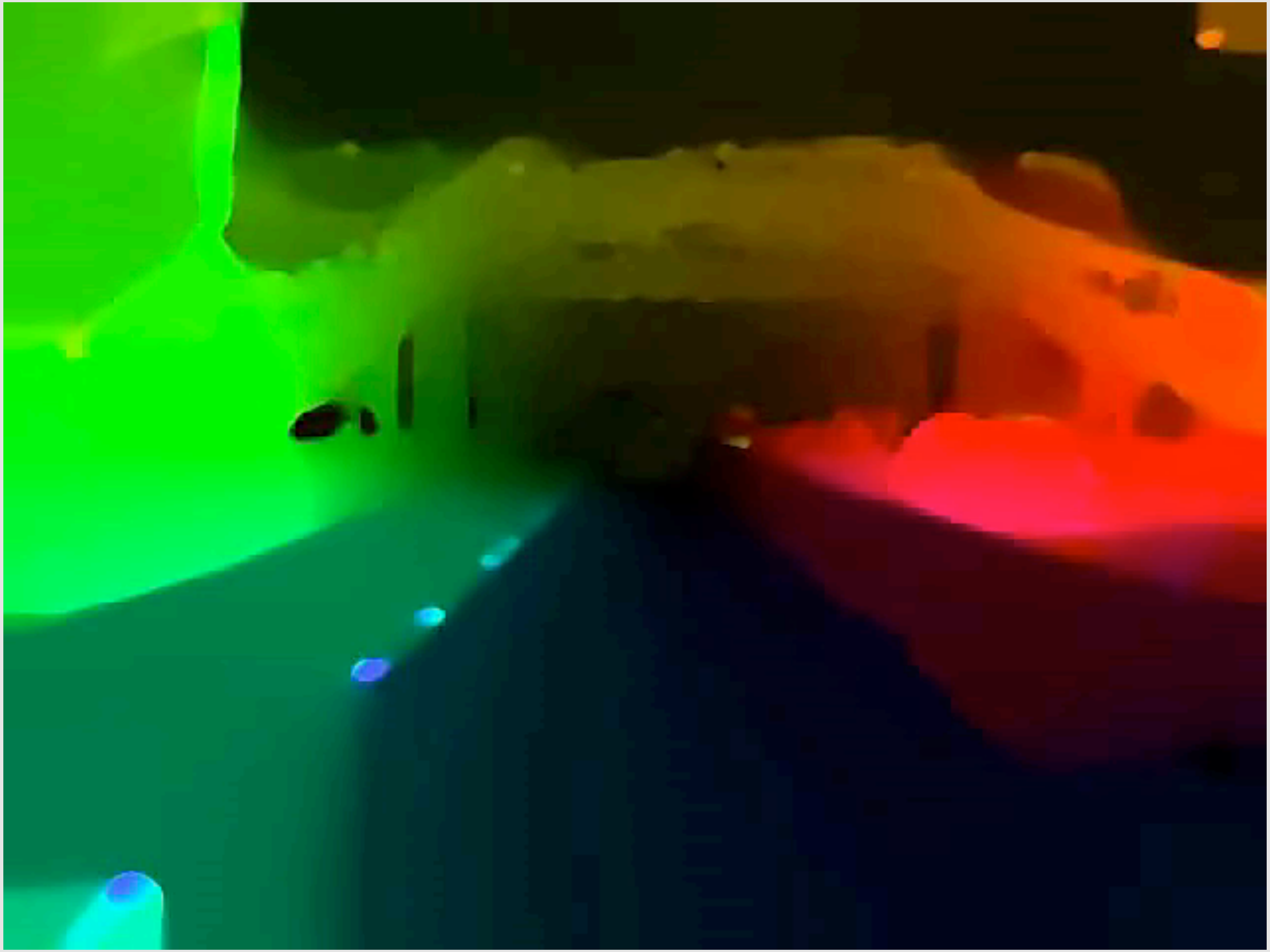
flow key

For each of the two sequences, motion vectors ($u$,$v$) are calculated by comparing frames along the time scale. Vectors at ($x$,$y$) are shown as one colored dot, representing direction by using the HSI scheme on the left; note that length is encoded by intensity (black = no motion).



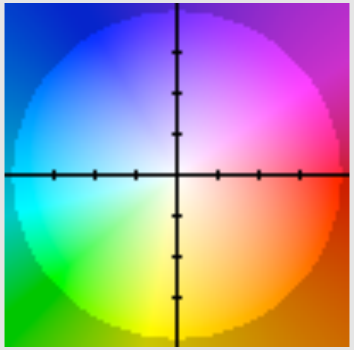
# BBPW algorithm on left and right sequence separately

# BBPW algorithm on the Auckland Harbor Bridge

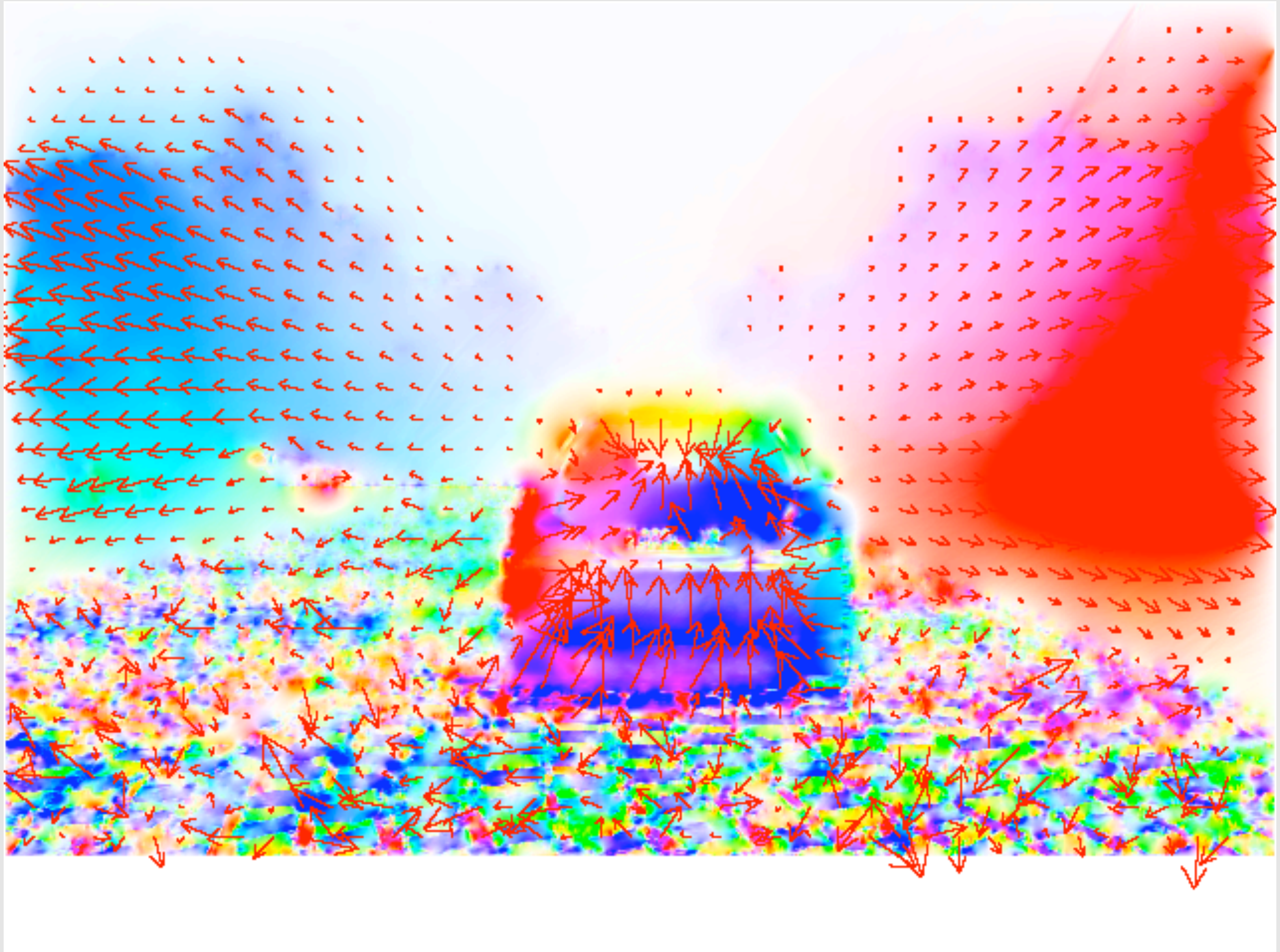


# CLG algorithm (combining local and global motion analysis)

[Bruhn, Weickert, Schnörr, IJCV 2005]



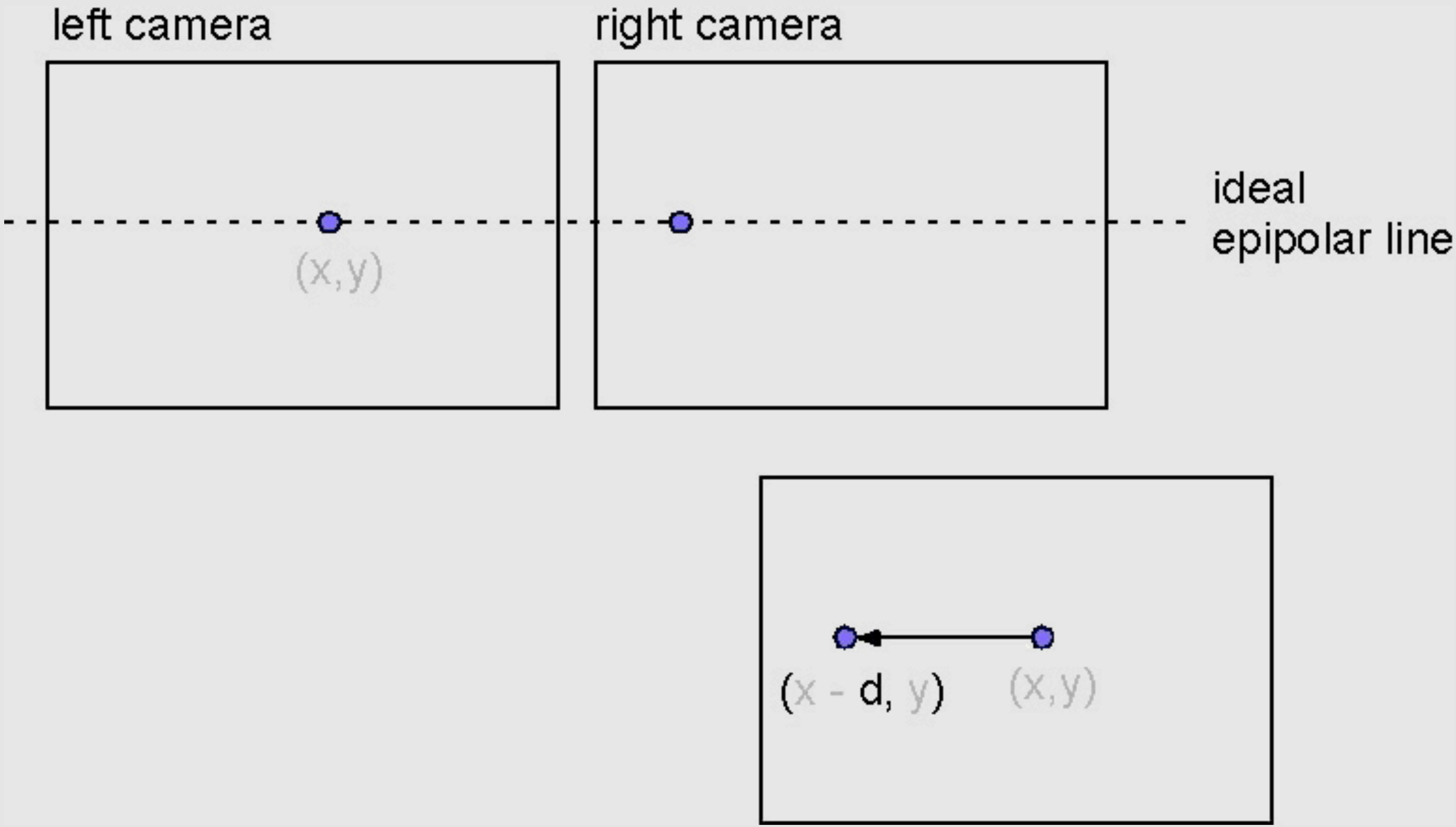
with white = no motion



applied to the synthetic stereo sequence (with ground truth) in Set 2 on

# Stereo Analysis

with two eyes




disparity *d* at *(x,y)* in left image

# Disparities are mapped into depth and shown in gray scale or color.

# A 2007 night-vision stereo sequence (on the German Autobahn)


## Dynamic programming stereo with temporal propagation

[Liu and Klette, ICONIP 2008]

# Belief propagation stereo with Sobel preprocessing

Original left input sequence

BP on original input sequences

Sobel of left input sequence

BP on Sobel input sequences

See [Guan and Klette, Robot Vision 2008]

# Specification of (finally) used BP algorithm

| Number | Max-disparity | Iterations | Image size | Running time | Truncation of discontinuity cost | Truncation of data cost |
|---|---|---|---|---|---|---|
| 1 | 30 *pixel* | 7 | $640 \times 360$ *pixel* | 2.9 *s* | 11 | 30 |
| 2 | 35 *pixel* | 7 | $640 \times 360$ *pixel* | 3.1 *s* | 11 | 25 |
| 3 | 40 *pixel* | 5 | $640 \times 360$ *pixel* | 2.9 *s* | 23 | 20 |
| 4 | 30 *pixel* | 7 | $640 \times 360$ *pixel* | 2.9 *s* | 20 | 60 |
| 5 | 30 *pixel* | 5 | $640 \times 360$ *pixel* | 2.7 *s* | 11 | 30 |
| 6 | 35 *pixel* | 6 | $640 \times 360$ *pixel* | 3.1 *s* | 10 | 30 |
| 7 | 40 *pixel* | 5 | $640 \times 360$ *pixel* | 2.9 *s* | 11 | 30 |

Running time: (for one pair of images)

Truncation of discontinuity cost: (penalty for intensity differences)

Truncation of data cost: (allows to handle occlusions)

Sobel preprocessing
max-product
4-adjacency
quadratic cost function
red-black speed-up method
coarse to fine for more reliable matching (5 to 7 layers; reduces #iterations)

(no initialization with disparities at time t-1, for t>0; future work)

# A modified SGM = Semi-Global Matching

(original SGM was proposed by Hirschmüller, CVPR 2005)

Each image pair is pre-processed by a 3x3 Sobel edge detector.

Edge images are smoothed using the kernel

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{vmatrix}$$

Resulting image pairs are processed with an SGM Algorithm (parameters $c_1 = 20$, $c_2 = 125$, and 8-path optimization using a new cost function (1x5 window) as published in [Hermann, Klette, and Destefanis, PSIVT 2009]; as common for SGM, outliers are filtered by a 5x5 median.

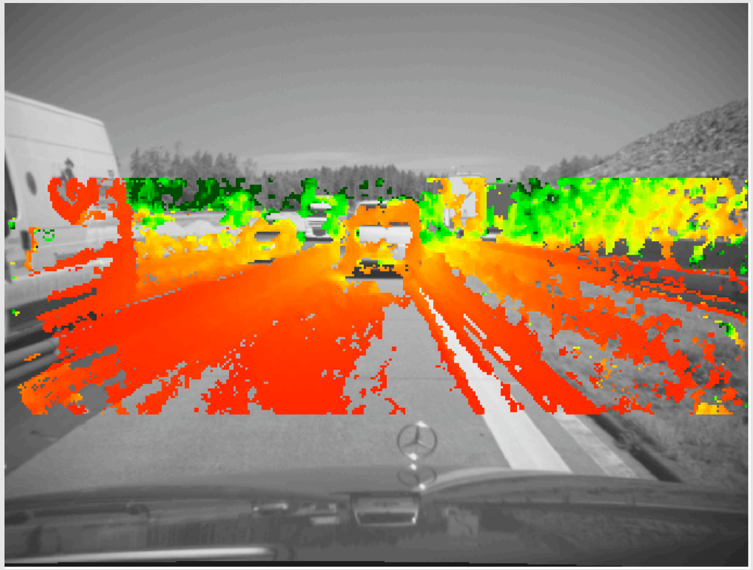
Mismatches are interpolated by a naive scheme (see the same paper) - here is potential for further improvement.

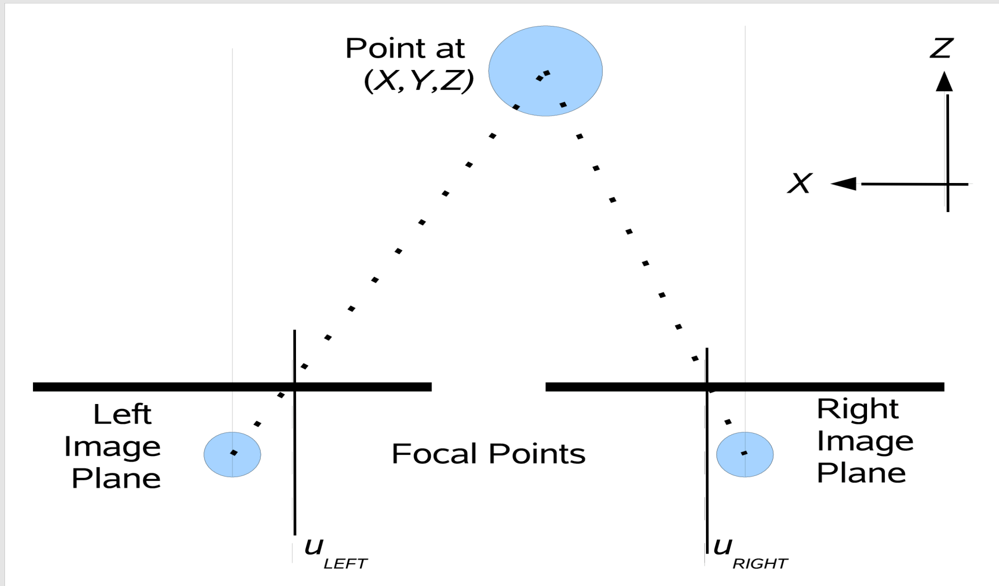
# A difficult night-vision stereo sequence (Daimler AG, Germany) "dancing lights"

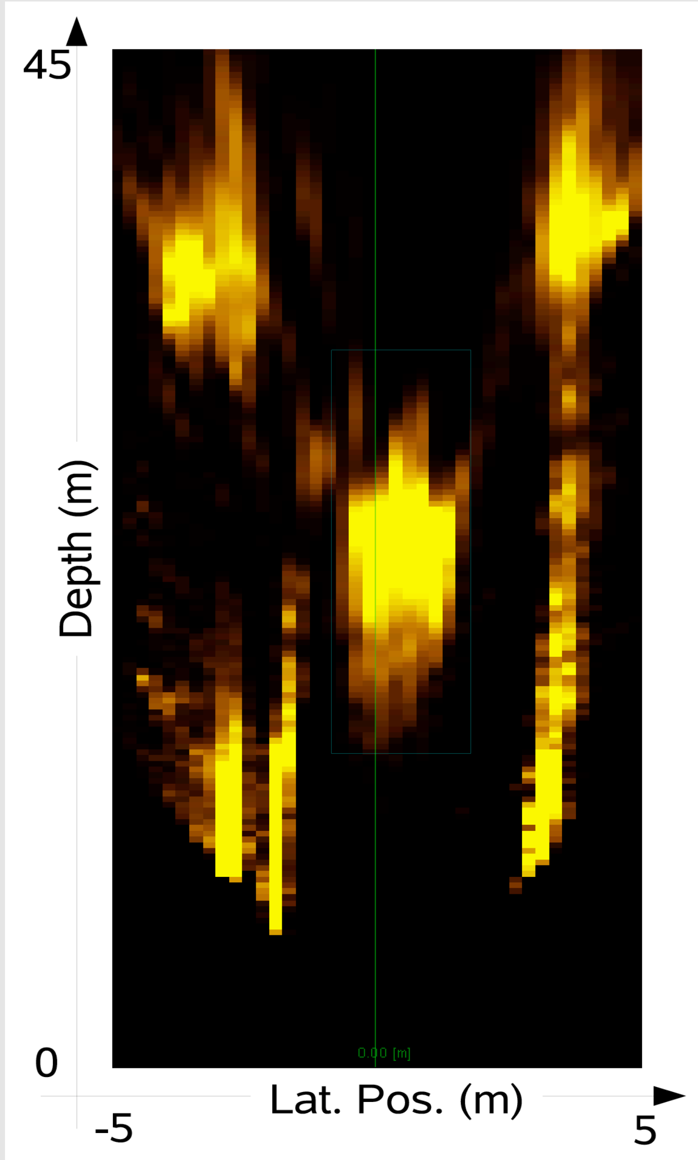

# Modified SGM recovers complex 3D shapes above ground "quite well"

# From Disparity Map to Occupancy Grid

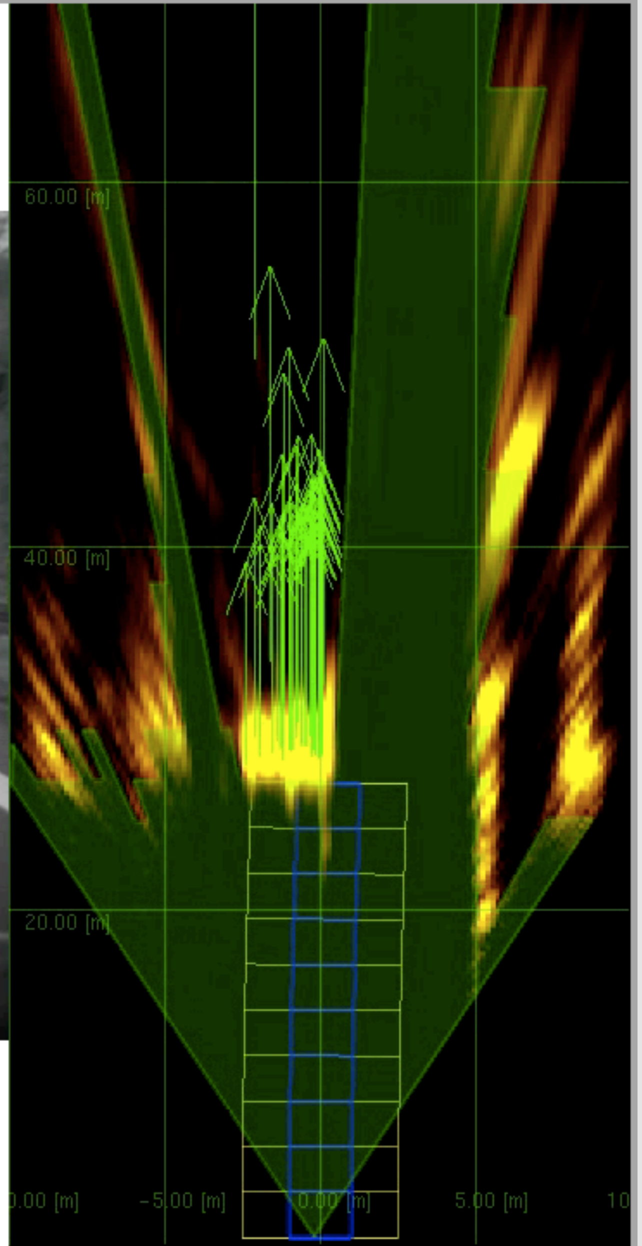
[Vaudrey, Badino, Gehrig, Robot Vision 2008]





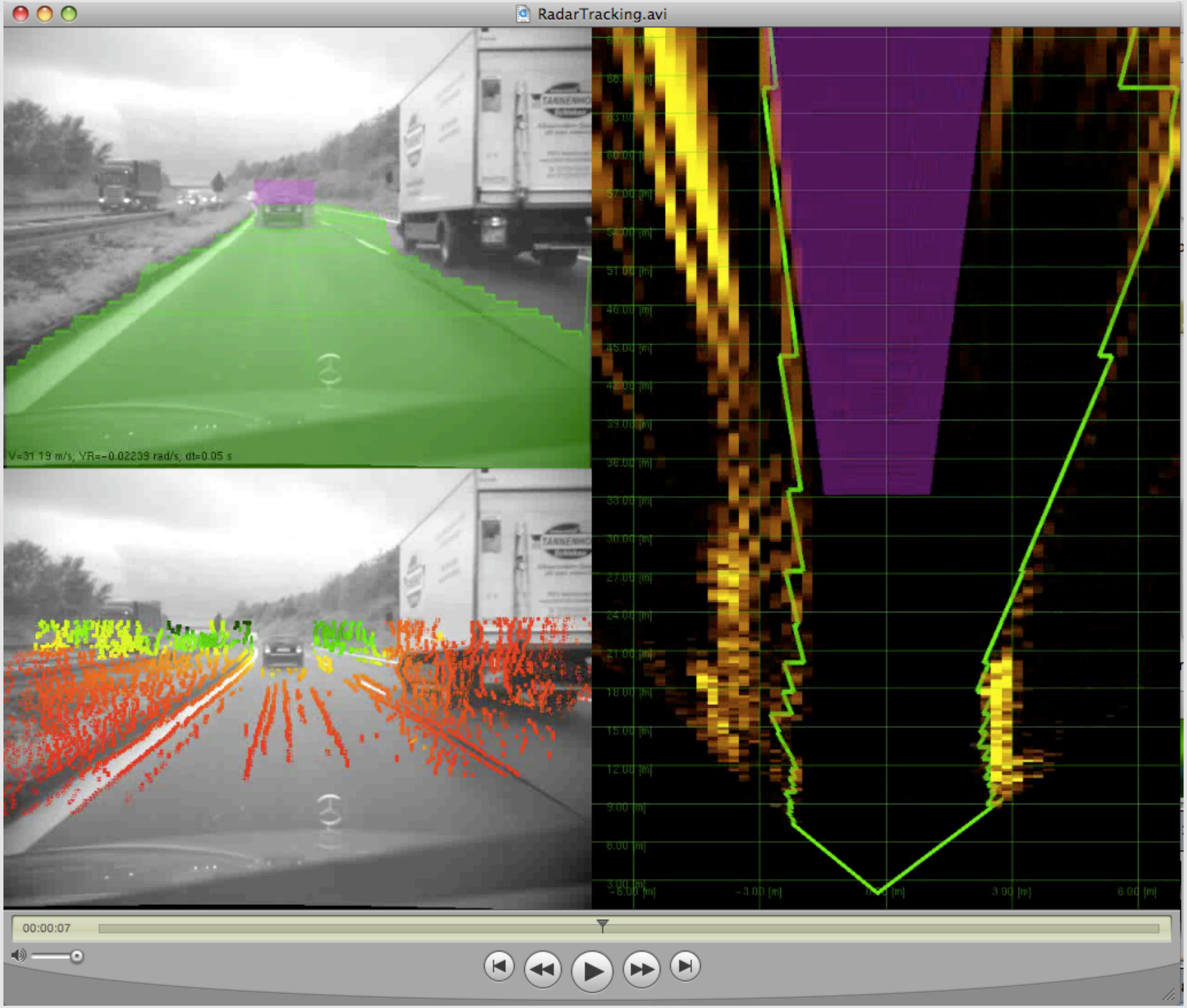



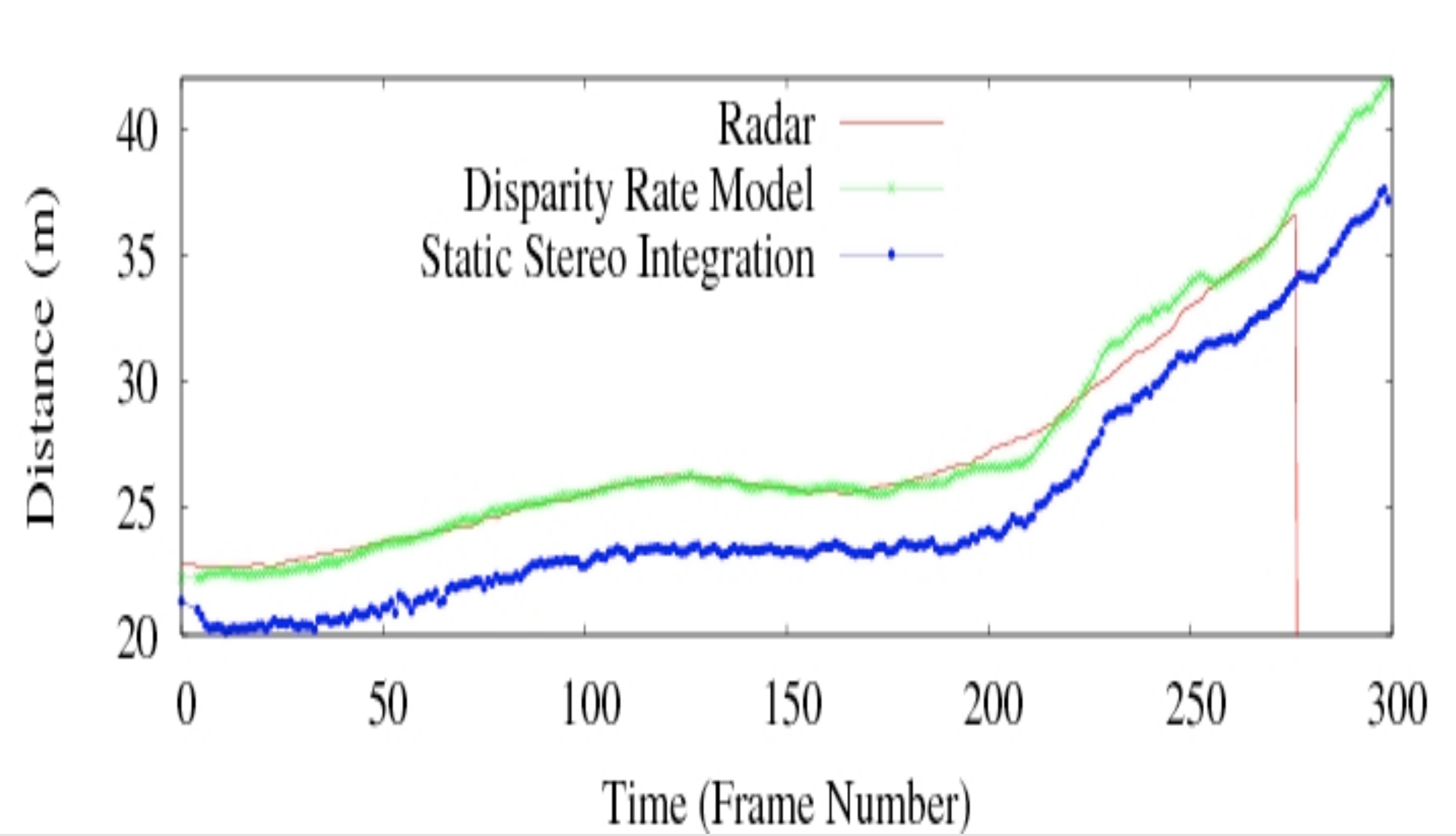
# Occupancy grid improvement by using a Kalman filter for disparity and (!) disparity rate


also allows (pixel-wise) good estimates of speed of lead vehicle

# Use of radar tracking (of lead vehicle) for generating `ground truth'






Incorporating disparity rate improved distance (speed) estimates to lead vehicle, compared to static stereo integration

[Vaudrey, Badino, Gehrig, Robot Vision 2008]


More accurate (right) detection of the bicyclist compared to disparity integration only in the middle

[Vaudrey, Badino, Gehrig, Robot Vision 2008]

# Stereo and Motion Analysis Combined

Stereo and motion data may be combined into

3D spatial + 3D motion =

6D combined scene representation

known as scene flow

# Scene Flow

Combines flow and disparity into one frame-work:

dense optical flow (BBPW)
Dense stereo (SGM)

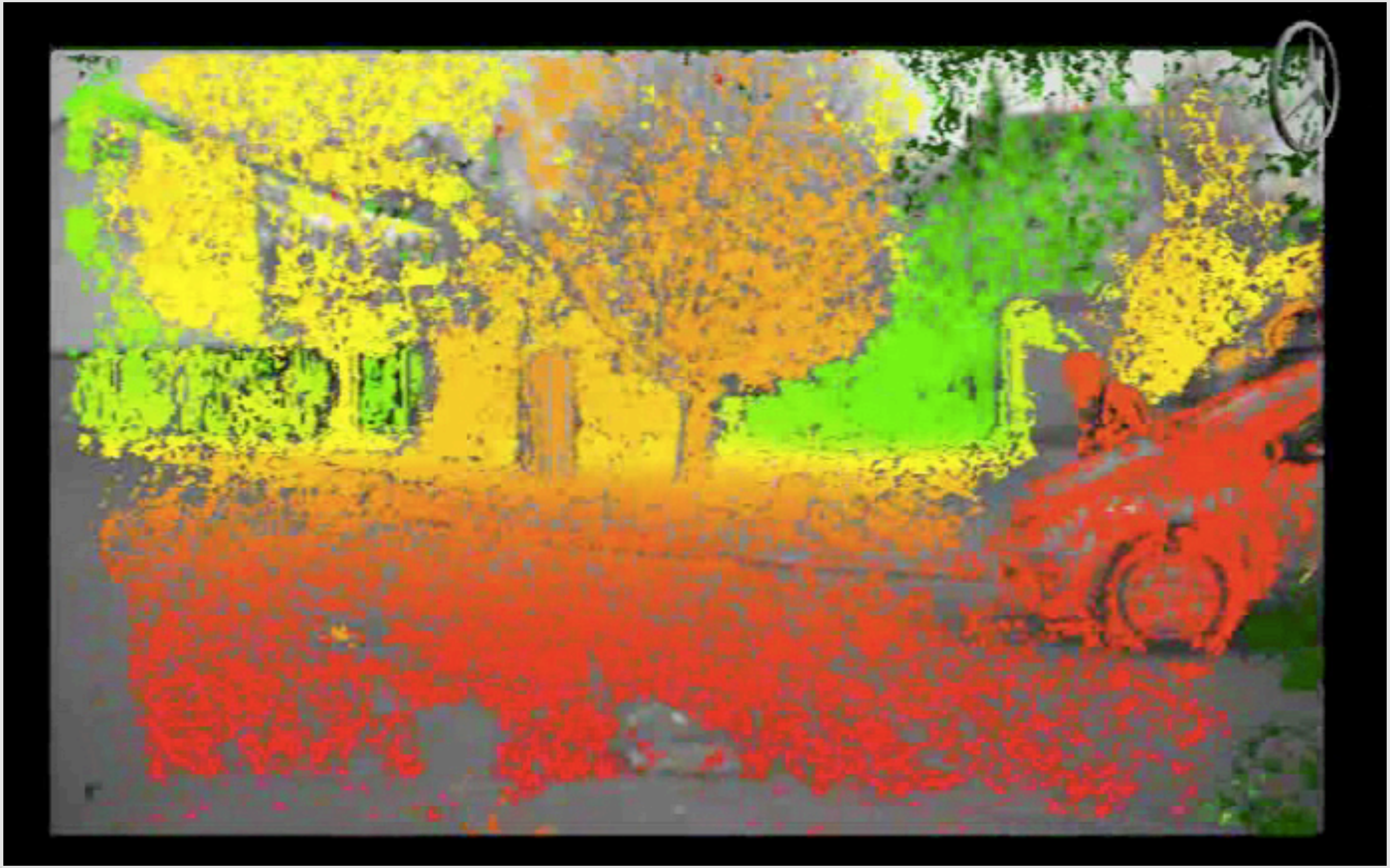
Estimates the flow (*u* and *v*) and the disparity rate (*d'*) at every pixel

Compensated using ego-motion estimation

See [Wedel, Rabe, Vaudrey, Brox, Franke, Cremers, ECCV 2008]

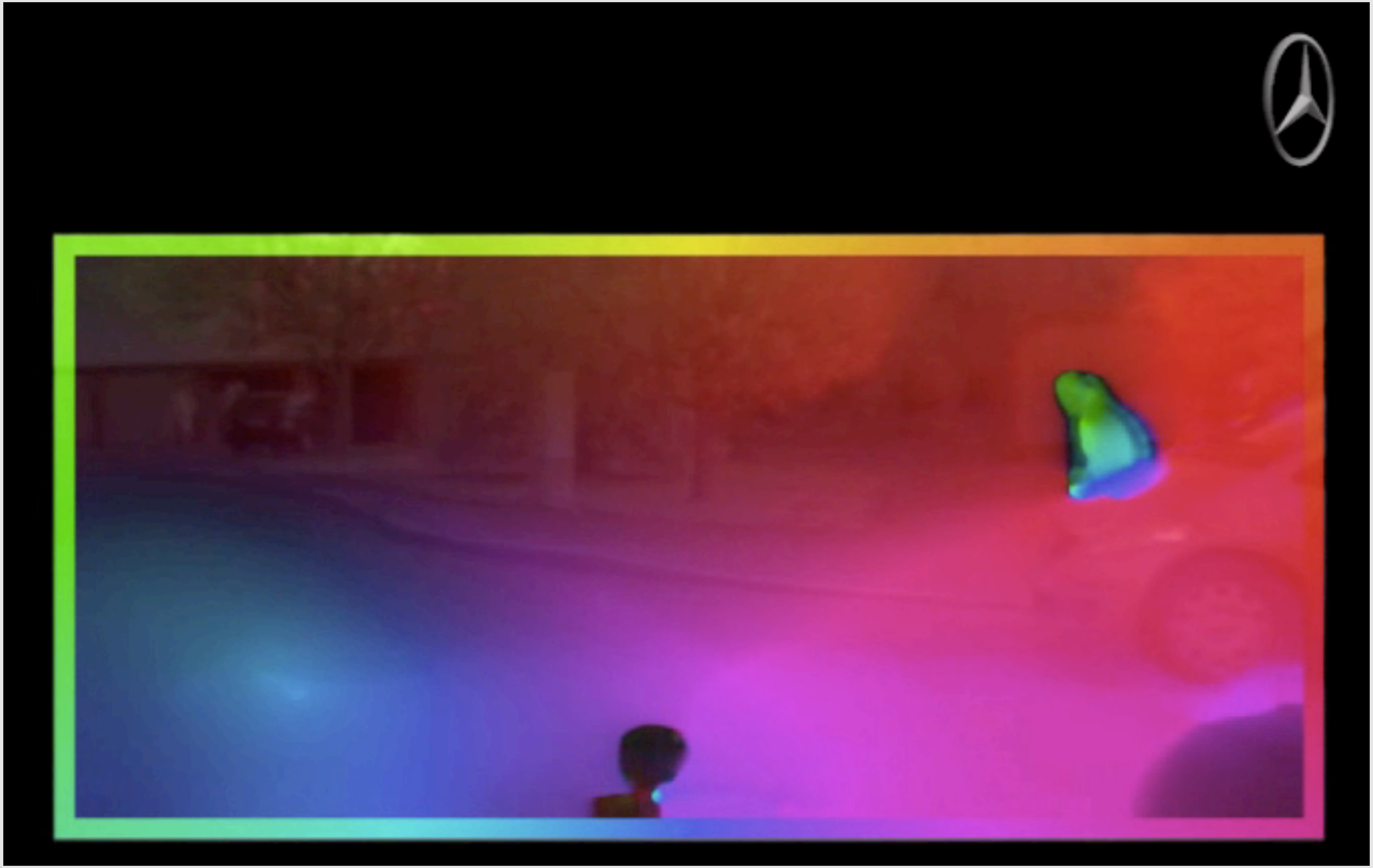
# Input from the left camera

# SGM (real time at 25 Hz) dense stereo result



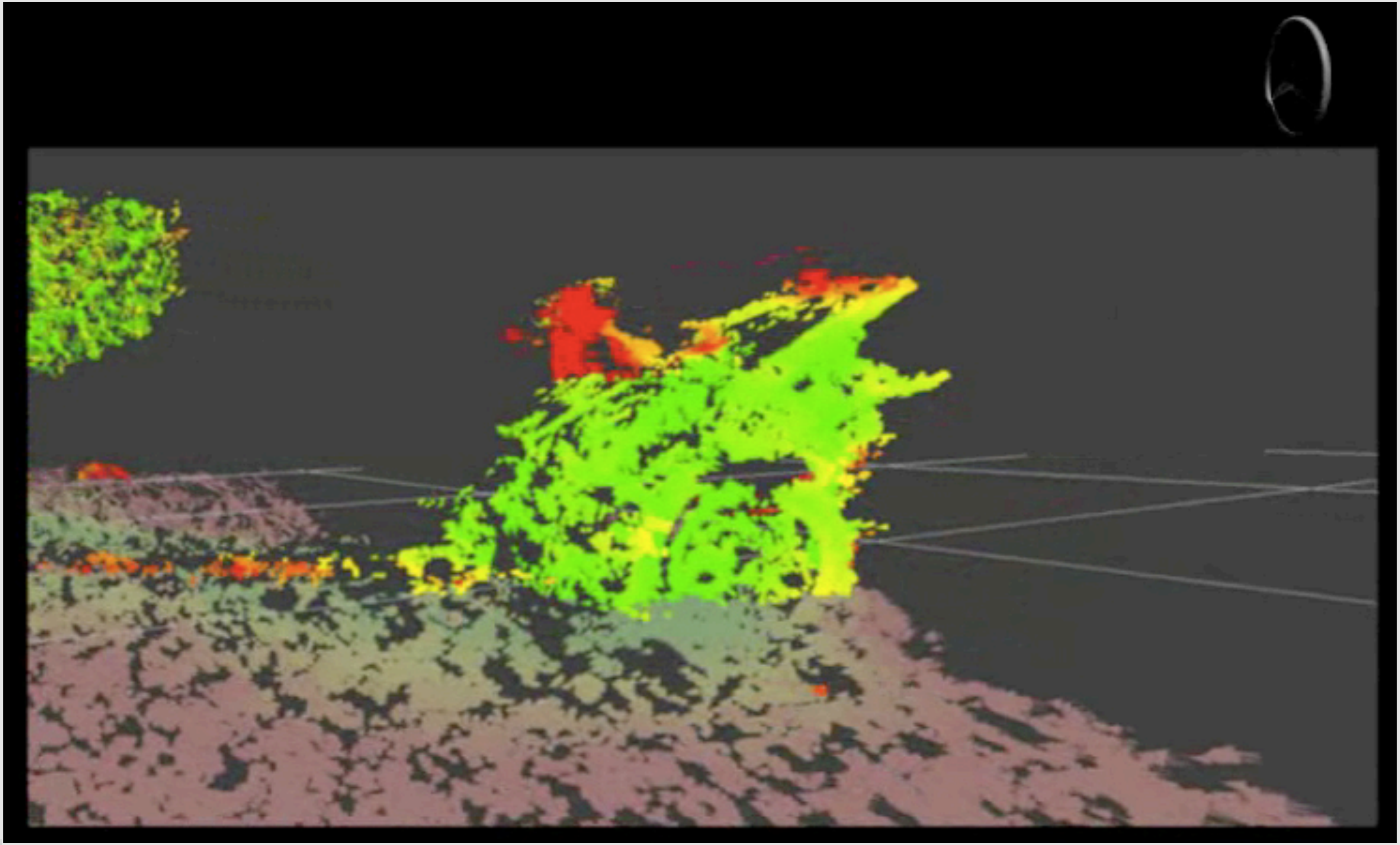
color encodes depth (distance)

# BBPW for the left camera (real time at 25 Hz)



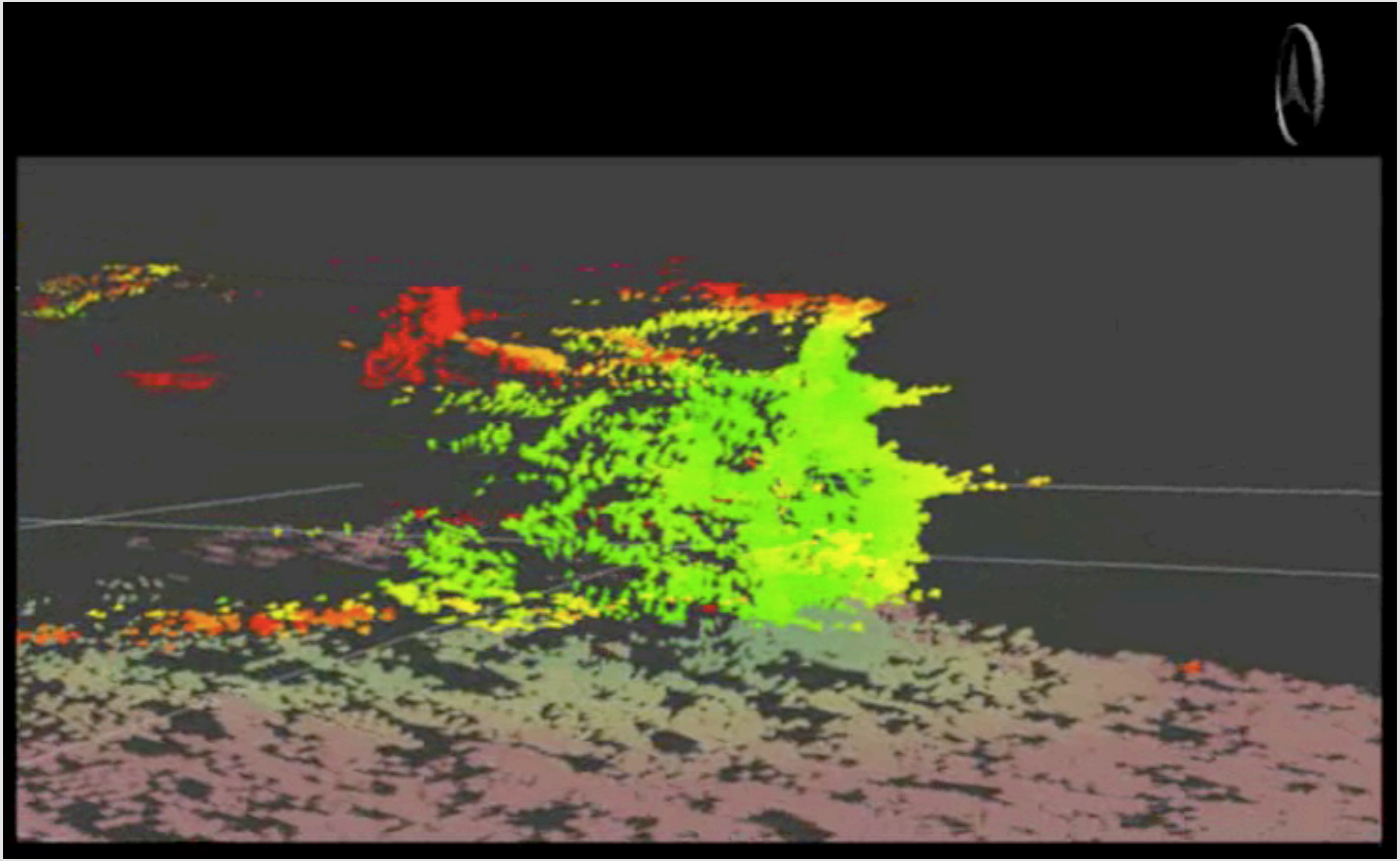
see color on the border for flow key

Segmented depth map also using motion data
(using scene flow = detected motion in 3D)



gray: below 1 meter of height
green - no motion, yellow - slow motion, red - fast motion

# Obviously, 3D data allow various perspectives on the 3D scene



gray: below 1 meter of height
green - no motion, yellow - slow motion, red - fast motion

# Conclusions

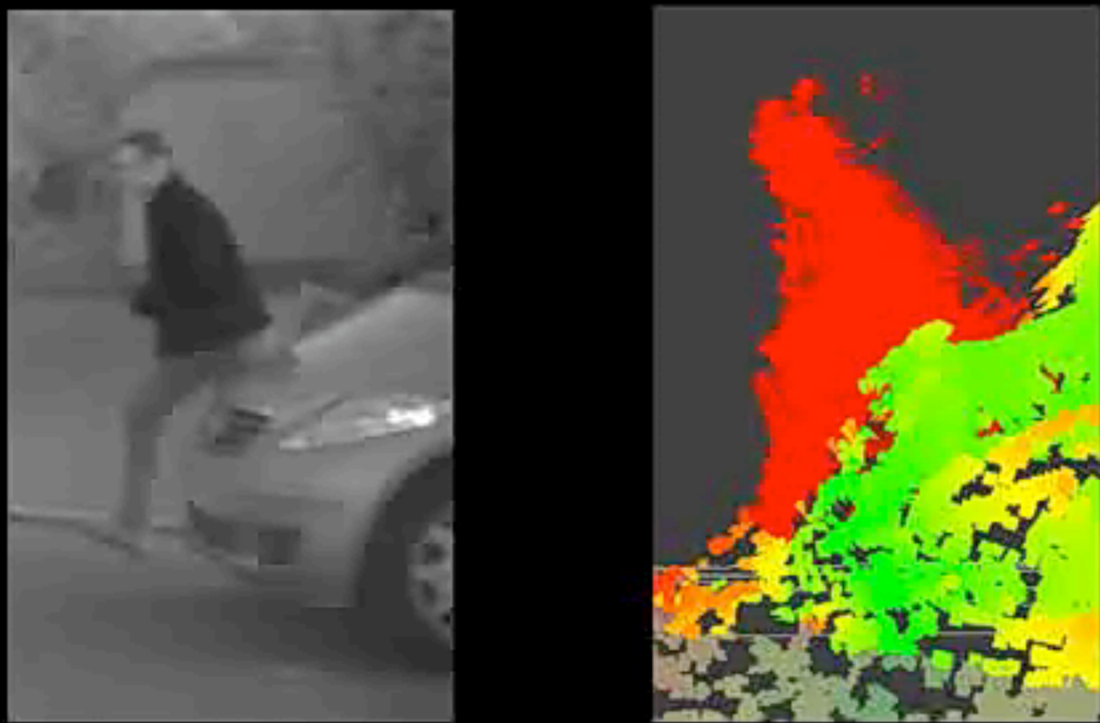
Vision-based driver assistance is imminent for standard cars; currently it moves from top-end products into less expensive models.

There are several mature components (lane departure warning, dis–tance to lead vehicle, blind spot surveillance, parking assistant, …); stereo and motion based solutions are currently still `leading edge' material with ongoing challenges.

The stereo and motion `data collection phase' is likely to turn into a stereo and motion `data analysis phase' (scene flow, segmentation, analysis of moving objects, forward-looking evaluation of current situations,…), followed by a `complex traffic-scene understanding phase', possibly also allowing incremental 3D scene modeling using 3D models on local (stationary) servers, such as current GPS allows the integration of traffic flow updates. A scene studied before should not be forgotten; it should be memorized and used during subsequent visits.

Analysis, recognition and understanding will define new challenges, especially for the neural network community. We are ready for collaboration! Please contact the .*enpeda*.. team if interested. We have the low-level vision data.

# Beyond 6D Vision: Dense Scene Flow



DAIMLER