The Quest for Robustness AND Accuracy – Mission Impossible?

U. Franke, Daimler AG Research and Development, Böblingen

Stereo Vision is a well understood Problem

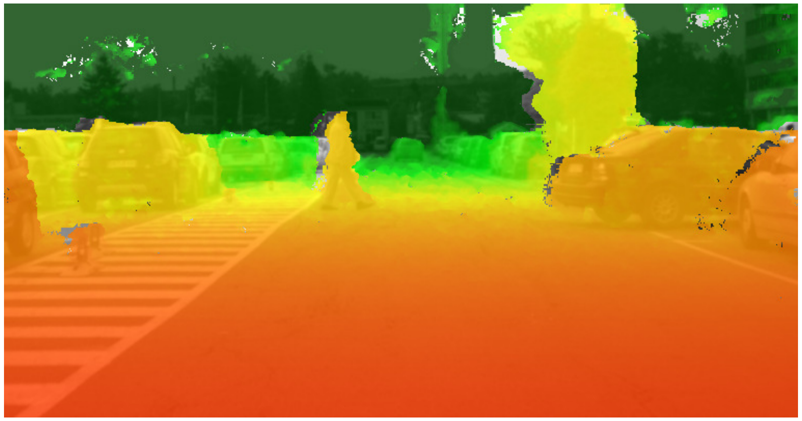
Mercedes-Benz Intelligent Drive

The new Mercedes-Benz S-Class and E-Class offer:

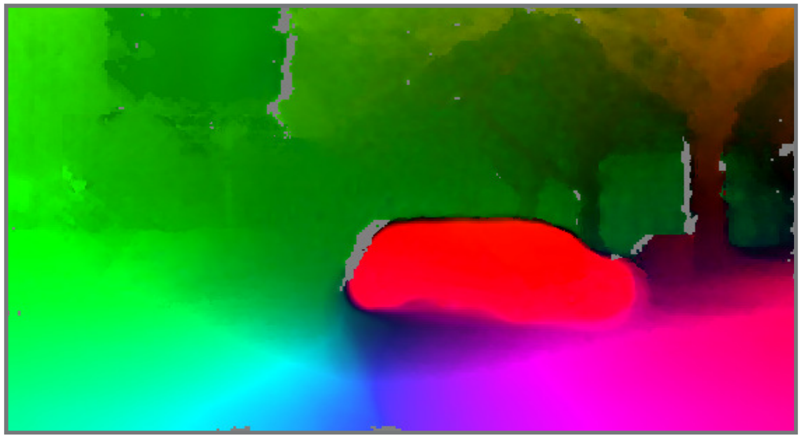
- Pedestrian collision avoidance up to 50km/h by **autonomous braking**, activated up to 72km/h.
- Active braking assistant reacting to crossing traffic.
- Lateral and longitudinal control up to 200km/h even under adverse weather conditions (hands-on recognition).
- Low speed **autonomous driving** in traffic jams.
- Magic Body Control (active body control utilizing road profile).

Overview

- Stereo Vision
  - Robustness
  - Precision
  - Calibration

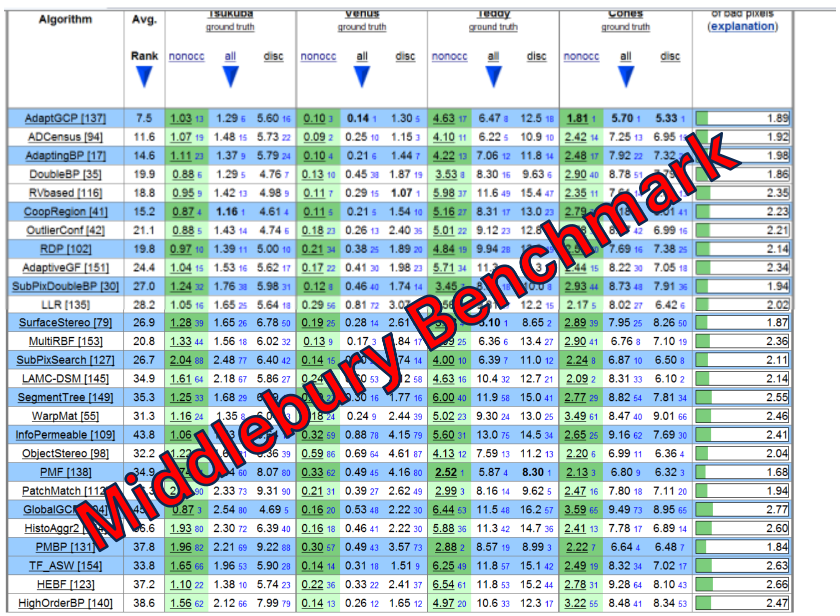


- Optical Flow
  - Illumination Changes
  - Large Displacements
  - Light Artifacts

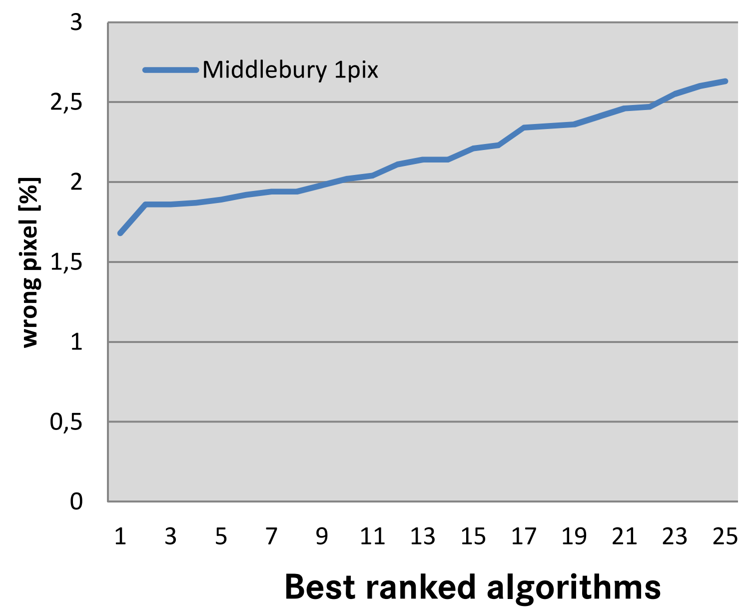


Stereo Vision seems to be a Solved Problem

The 25 top ranked algorithms differ only slightly in the achieved quality.



Algorithm	Avg.	Tsukuba ground truth			Venus ground truth			Teddy ground truth			Cones ground truth			of bad pixels (explanation)
	Rank	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	
AdaptGCP [137]	7.5	1.03 13	1.29 6	5.60 16	0.10 3	0.14 1	1.30 5	4.63 17	6.47 8	12.5 18	1.81 1	5.70 1	5.33 1	1.89
ADCensus [94]	11.6	1.07 19	1.48 15	5.73 22	0.09 2	0.25 10	1.15 3	4.10 11	6.22 5	10.9 10	2.42 14	7.25 13	6.95 [illegible]	1.92
AdaptingBP [17]	14.6	1.11 23	1.37 9	5.79 24	0.10 4	0.21 6	1.44 7	4.22 13	7.06 12	11.8 14	2.48 17	7.92 22	7.32 [illegible]	1.98
DoubleBP [35]	19.9	0.88 6	1.29 5	4.76 7	0.13 10	0.45 38	1.87 19	3.53 8	8.30 16	9.63 6	2.90 40	8.78 51	[illegible]	1.86
RVbased [116]	18.8	0.95 9	1.42 13	4.98 9	0.11 7	0.29 15	1.07 1	5.98 37	11.6 49	15.4 47	2.35 11	[illegible]	[illegible]	2.35
CoopRegion [41]	15.2	0.87 4	1.16 1	4.61 4	0.11 5	0.21 5	1.54 10	5.16 27	8.31 17	13.0 23	2.79 [illegible]	[illegible]	[illegible] 41	2.23
OutlierConf [42]	21.1	0.88 5	1.43 14	4.74 6	0.18 23	0.26 13	2.40 35	5.01 22	9.12 23	12.8 [illegible]	[illegible]	[illegible] 42	6.99 16	2.21
RDP [102]	19.8	0.97 10	1.39 11	5.00 10	0.21 34	0.38 25	1.89 20	4.84 19	9.94 28	[illegible]	[illegible]	7.69 16	7.38 25	2.14
AdaptiveGF [151]	24.4	1.04 15	1.53 16	5.62 17	0.17 22	0.41 30	1.98 23	5.71 34	[illegible]	[illegible]	[illegible] 15	8.22 30	7.05 18	2.34
SubPixDoubleBP [30]	27.0	1.24 32	1.76 38	5.98 31	0.12 8	0.46 40	1.74 14	3.45 [illegible]	[illegible] 18	[illegible] 8	2.93 44	8.73 48	7.91 36	1.94
LLR [135]	28.2	1.05 16	1.65 25	5.64 18	0.29 56	0.81 72	[illegible]	[illegible]	[illegible]	12.2 15	2.17 5	8.02 27	6.42 6	2.02
SurfaceStereo [79]	26.9	1.28 39	1.65 26	6.78 50	0.19 25	0.28 14	2.61 [illegible]	[illegible]	5.10 1	8.65 2	2.89 39	7.95 25	8.26 50	1.87
MultiRBF [153]	20.8	1.33 44	1.56 18	6.02 32	0.13 9	0.17 3	[illegible] 17	[illegible] 25	6.36 6	13.4 27	2.90 41	6.76 8	7.10 19	2.36
SubPixSearch [127]	26.7	2.04 88	2.48 77	6.40 42	0.14 15	[illegible]	[illegible] 14	4.00 10	6.39 7	11.0 12	2.24 8	6.87 10	6.50 8	2.11
LAMC-DSM [145]	34.9	1.61 64	2.18 67	5.86 27	0.24 [illegible]	[illegible] 53	[illegible] 58	4.63 16	10.4 32	12.7 21	2.09 2	8.31 33	6.10 2	2.14
SegmentTree [149]	35.3	1.25 33	1.68 29	[illegible]	[illegible]	[illegible] 16	1.77 16	6.00 40	11.9 58	15.0 41	2.77 29	8.82 54	7.81 34	2.55
WarpMat [55]	31.3	1.16 24	1.35 8	[illegible]	[illegible] 24	0.24 9	2.44 39	5.02 23	9.30 24	13.0 25	3.49 61	8.47 40	9.01 66	2.46
InfoPermeable [109]	43.8	1.06 [illegible]	[illegible]	[illegible]	0.32 59	0.88 78	4.15 79	5.60 31	13.0 75	14.5 34	2.65 25	9.16 62	7.69 30	2.41
ObjectStereo [98]	32.2	[illegible]	[illegible]	[illegible] 39	0.59 86	0.69 64	4.61 87	4.13 12	7.59 13	11.2 13	2.20 6	6.99 11	6.36 4	2.04
PMF [138]	34.9	[illegible]	[illegible] 60	8.07 80	0.33 62	0.49 45	4.16 80	2.52 1	5.87 4	8.30 1	2.13 3	6.80 9	6.32 3	1.68
PatchMatch [11[illegible]	[illegible]	[illegible] 90	2.33 73	9.31 90	0.21 31	0.39 27	2.62 49	2.99 3	8.16 14	9.62 5	2.47 16	7.80 18	7.11 20	1.94
GlobalGC[illegible]	[illegible]	0.87 3	2.54 80	4.69 5	0.16 20	0.53 48	2.22 30	6.44 53	11.5 48	16.2 57	3.59 65	9.49 73	8.95 65	2.77
HistoAggr2 [illegible]	[illegible]	1.93 80	2.30 72	6.39 40	0.16 18	0.46 41	2.22 30	5.88 36	11.3 42	14.7 36	2.41 13	7.78 17	6.89 14	2.60
PMBP [131]	37.8	1.96 82	2.21 69	9.22 88	0.30 57	0.49 43	3.57 73	2.88 2	8.57 19	8.99 3	2.22 7	6.64 4	6.48 7	1.84
TF_ASW [154]	33.8	1.65 66	1.96 53	5.90 28	0.14 14	0.31 18	1.51 9	6.25 49	11.8 57	15.1 42	2.49 19	8.32 34	7.02 17	2.63
HEBF [123]	37.2	1.10 22	1.38 10	5.74 23	0.22 36	0.33 22	2.41 37	6.54 61	11.8 53	15.2 44	2.78 31	9.28 64	8.10 43	2.66
HighOrderBP [140]	38.6	1.56 62	2.12 66	7.99 79	0.14 13	0.26 12	1.65 12	4.97 20	10.6 33	12.3 17	3.22 55	8.48 41	8.34 53	2.47

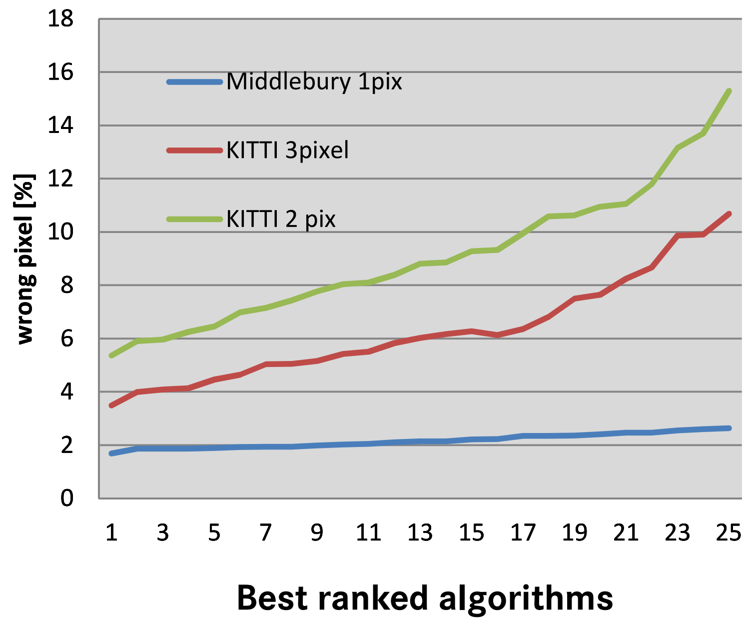


DAIMLER

Stereo Vision seems to be a Solved Problem

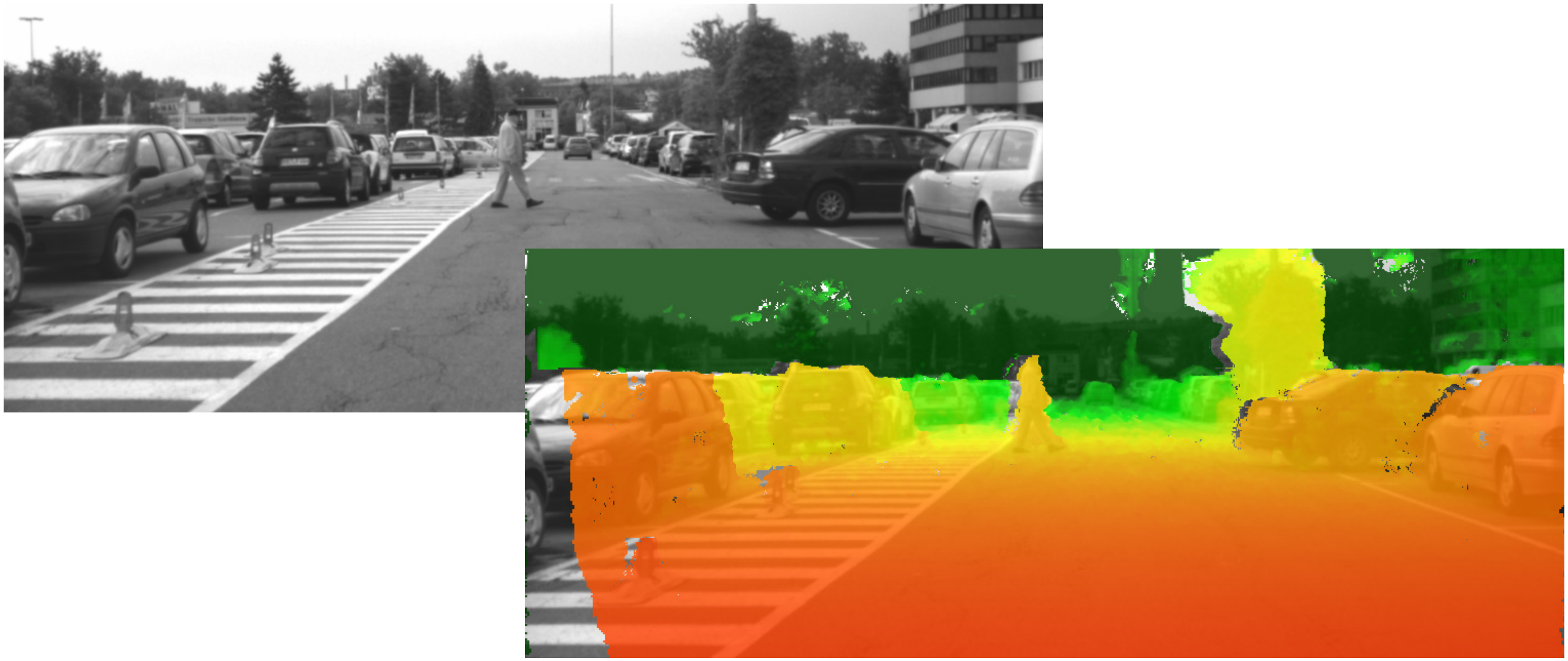
KITTI Benchmark

Rank	Method	Setting	Code	Out-Noc	Out-All	Avg-Noc	Avg-All	Density	Runtime	Environment	Compare
1	PCBP-SS			**3.49 %**	**4.79 %**	**0.8 px**	1.0 px	100.00 %	5 min	4 cores @ 2.5 Ghz (Matlab + C/C++)	
K. Yamaguchi, D. McAllester and R. Urtasun: Robust Monocular Epipolar Flow Estimation. CVPR 2013.											
2	StereoSLIC			3.99 %	5.17 %	0.9 px	1.0 px	99.89 %	2.3 s	1 core @ 3.0 Ghz (C/C++)	
K. Yamaguchi, D. McAllester and R. Urtasun: Robust Monocular Epipolar Flow Estimation. CVPR 2013.											
3	PR-Sf+E			4.09 %	4.95 %	0.9 px	1.0 px	100.00 %	200 s	4 cores @ 3.0 Ghz (Matlab + C/C++)	
Anonymous submission											
4	PCBP			4.13 %	5.45 %	0.9 px	1.2 px	100.00 %	5 min	4 cores @ 2.5 Ghz (Matlab + C/C++)	
K. Yamaguchi, T. Hazan, D. McAllester and R. Urtasun: Continuous Markov Random Fields for Robust Stereo Estimation. ECCV 2012.											
5	PR-Sceneflow			4.46 %	5.32 %	1.0 px	1.1 px	100.00 %	150 sec	4 core @ 3.0 Ghz (Matlab - C/C++)	
Anonymous submission											
6	DDS			4.64 %	5.51 %	1.0 px	1.1 px	100.00 %	30 s	1 core @ 2.5 Ghz ([illegible] + C/C++)	
Anonymous submission											
7	wSGM			5.03 %	6.24 %	1.3 px	1.6 px	97.03 %	6s	[illegible] (C/C++)	
T. Robert Spangenberg and R. Rojas: Weighted Semi-Global Matching and Center-Symmetric Census Transform for Robust D[illegible] 2013.											
8	ATGV			5.05 %	6.91 %	1.0 px	1.6 px	100.00 %	[illegible]	[illegible] @ 3.0 Ghz (Matlab + C/C++)	
R. Ranftl, T. Pock and H. Bischof: Minimizing TGV-based Variational Models with Non-Convex Data terms. [illegible]											
9	iSGM			5.16 %	7.19 %	1.2 px	2.1 px	94.[illegible] %	[illegible]	2 cores @ 2.5 Ghz (C/C++)	
S. Hermann and R. Klette: Iterative Semi-Global Matching for Robust Driver Assistance Syste[illegible]											
10	OCV-SGBM2			5.42 %	6.54 %	[illegible]	[illegible]	[illegible]0.00 %	2 s	1 core @ 2.5 Ghz (C/C++)	
Anonymous submission											
11	AABM			5.5[illegible]	[illegible]60 %	[illegible]	1.3 px	100.00 %	0.43 s	1 core @ 3.0 Ghz (C/C++)	
N. Einecke and J. Eggert: Appearance-Aligned [illegible]. IV 2013.											
12	SGM			5.83 [illegible]	7.08 %	1.2 px	1.3 px	85.80 %	3.7 s	1 core @ 3.0 Ghz (C/C++)	
H. Hirschmueller: Stereo Proces[illegible] Matching and Mutual Information. PAMI 2008.											
13	TGV2ADC			6.02 %	6.94 %	1.1 px	1.2 px	99.99 %	8 s	GPU @ 2.5 Ghz (C/C++)	
Anonymous submission											
14	CD			6.17 %	7.49 %	1.2 px	1.4 px	100.00 %	5 s	1 core @ 2.5 Ghz (C/C++)	
Anonymous submission											
15	SNCC			6.27 %	7.33 %	1.4 px	1.5 px	100.00 %	0.27 s	1 core @ 3.0 Ghz (C/C++)	
N. Einecke and J. Eggert: A Two-Stage Correlation Method for Stereoscopic Depth Estimation. DICTA 2010.											
16	ITGV			6.31 %	7.40 %	1.3 px	1.5 px	100.00 %	7 s	1 core @ 3.0 Ghz (Matlab + C/C++)	
R. Ranftl, S. Gehrig, T. Pock and H. Bischof: Pushing the Limits of Stereo Using Variational Stereo Estimation. IV 2012.											
17	RWR			6.36 %	7.51 %	1.2 px	1.4 px	100.00 %	1 min	1 core @ 2.5 Ghz (C/C++)	
Anonymous submission											
18	LDE			6.81 %	8.92 %	1.8 px	2.5 px	100.00 %	14 s	2 cores @ 2.5 Ghz (C/C++)	
Anonymous submission											
19	BSSM			7.50 %	8.89 %	1.4 px	1.6 px	94.87 %	20.7 s	1 core @ 3.5 Ghz (C/C++)	
Anonymous submission											
20	OCV-SGBM		code	7.64 %	9.13 %	1.8 px	2.0 px	86.50 %	1.1 s	1 core @ 2.5 Ghz (C/C++)	





Stereo Vision seems to be a Solved Problem, even in Practice



Challenge: on-line calibration to guarantee perfect results for the whole lifetime of the car.

Color encoded distance:
close ..... far

Stereo Vision seems to be a Solved Problem, even in Practice

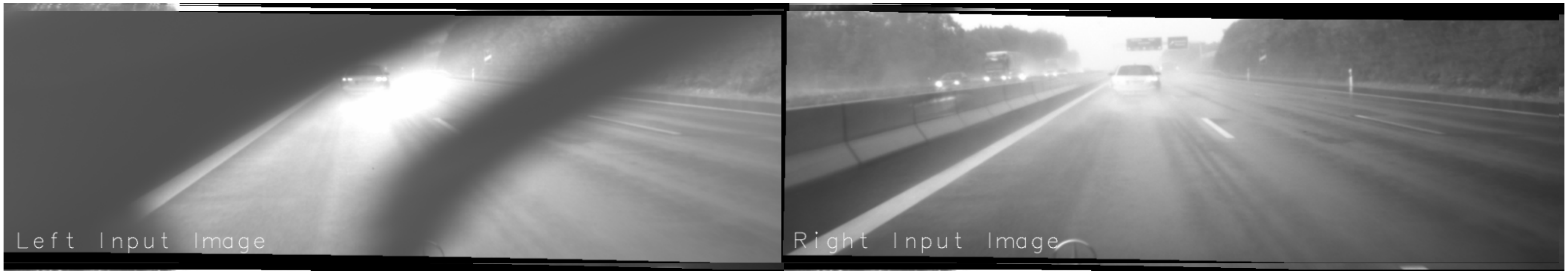
S.Gehrig, F.Eberli, T.Meyer, "**A Real-time Low-Power Stereo Vision Engine Using Semi-Global Matching**", ICVS 2009

Is Stereo a Solved Problem?

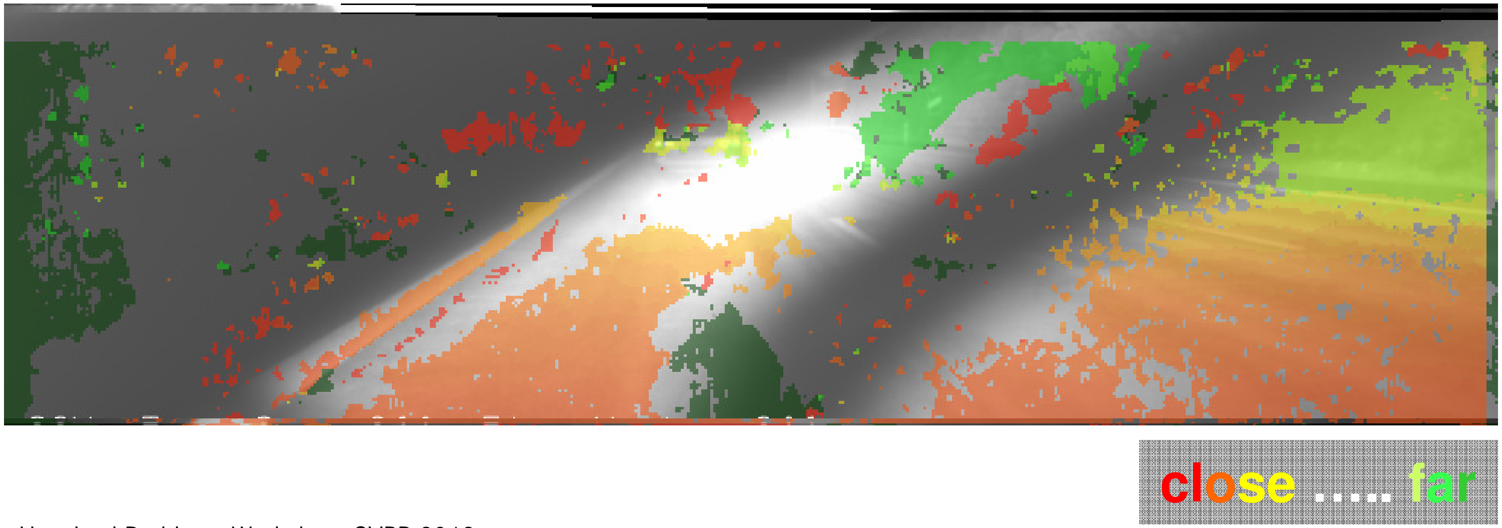
If we want Vision to become an all-day, all-night sensor for safety systems including autonomous driving, we have to cope with such adverse conditions!

Robustness is a Hard Challenge

5 subsequent image pairs taken under heavy rain.



Problems: blockage by wiper, heavy blur, different sharpness, brightness constancy constraint violation, low contrast, spray...

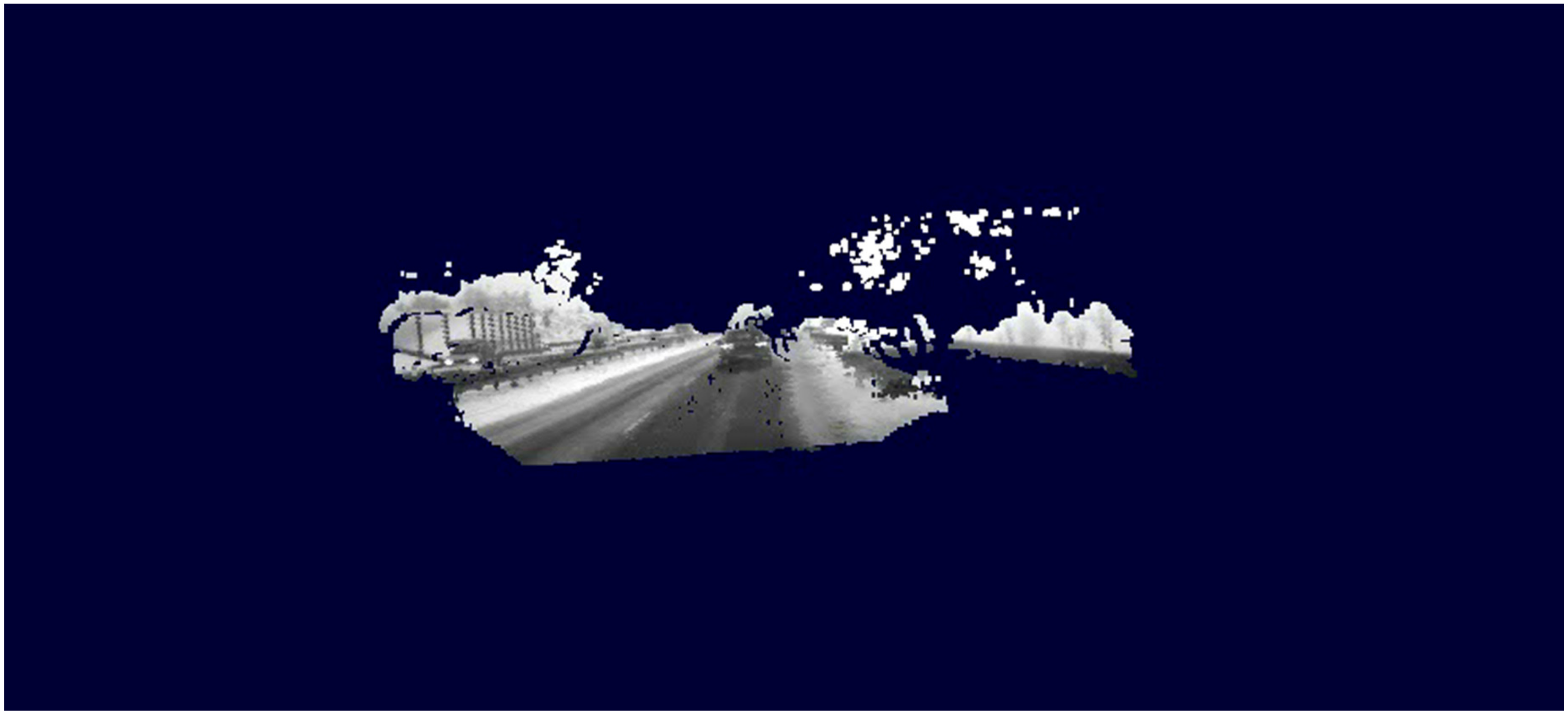


Robustness is a Hard Challenge

Image sequence recorded during heavy snowfall.

Robustness is a Hard Challenge

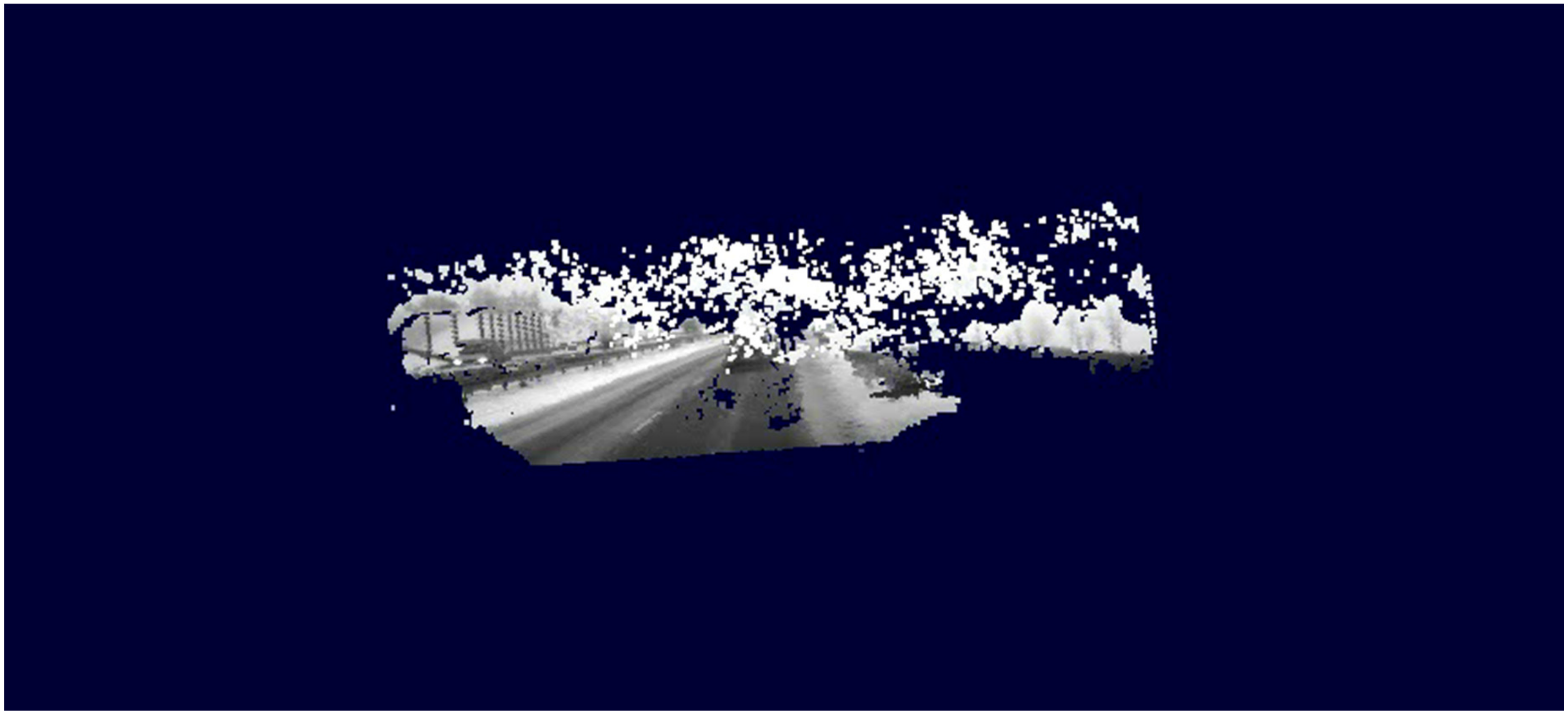
Image sequence recorded during heavy snowfall.



3D-View of an ideal algorithm.

Robustness is a Hard Challenge

Image sequence recorded during heavy snowfall.

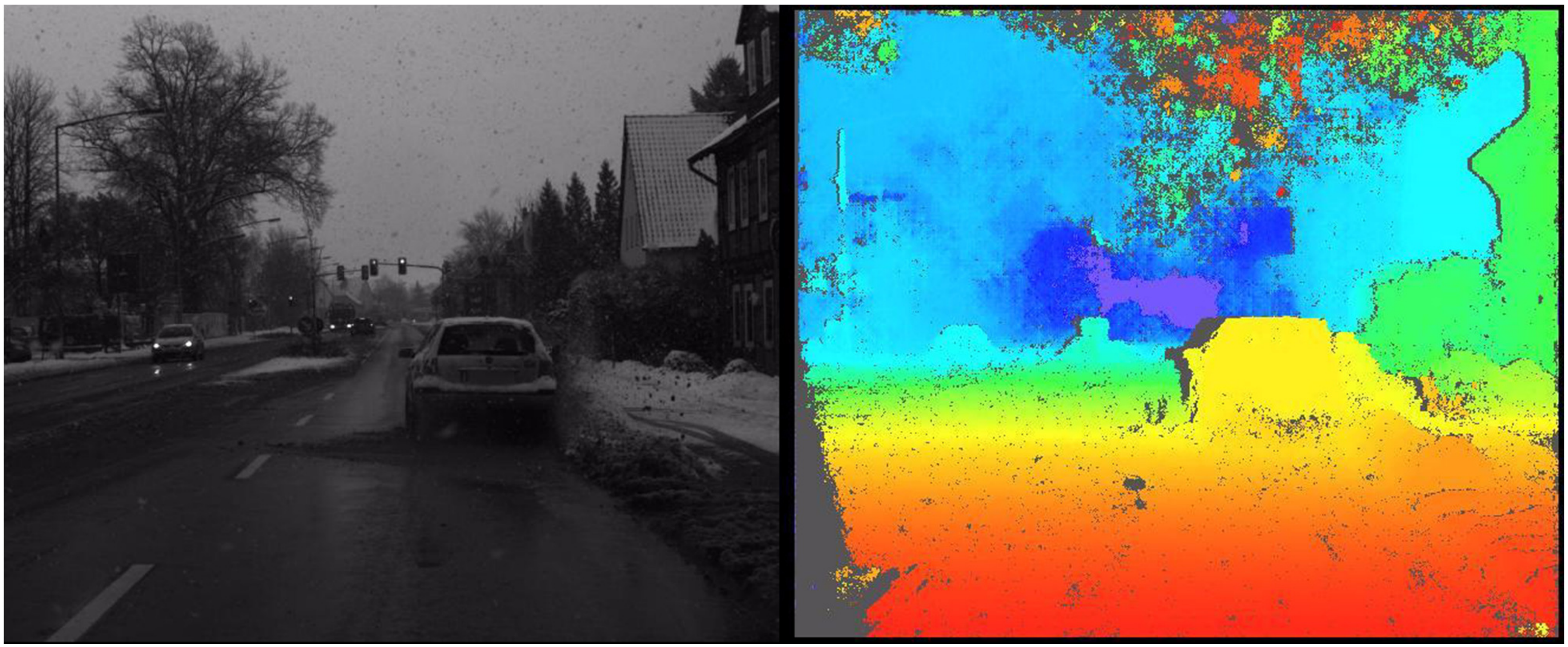


3D-View of the standard estimation scheme.

State-of-the-Art in Robust Disparity Estimation

iSGM[1] won the “Robust Vision Challenge” at ECCV 2012

(data by Daniel Kondermann, HCI).



1)S. Hermann and R. Klette: „**Iterative Semi-Global Matching for Robust Driver Assistance Systems**“ ACCV, 2012

DAIMLER

Why do we need High Accuracy?

Distance [m]	400	200	100	50	25
Disparity [px]	1	2	4	8	16

Most times we are measuring small disparities. Subpixel precision is a must.

DAIMLER

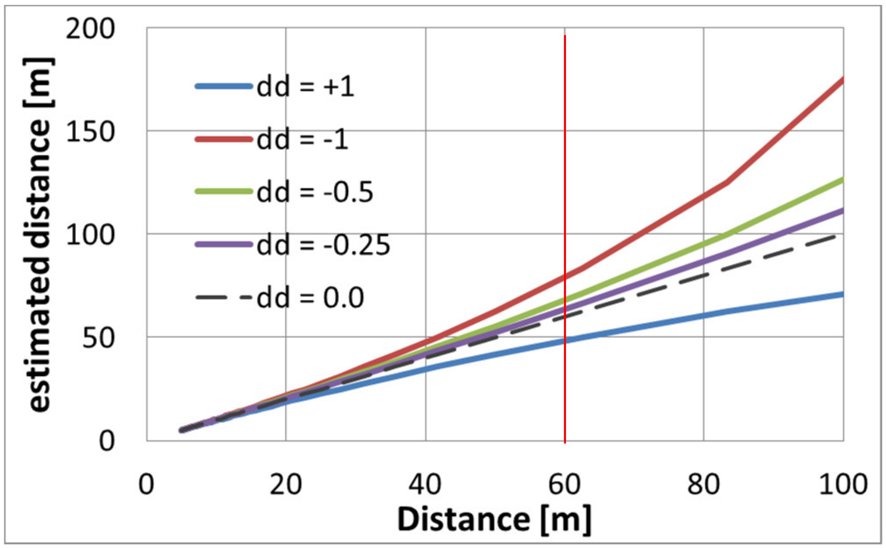
Effects of small Disparity Errors

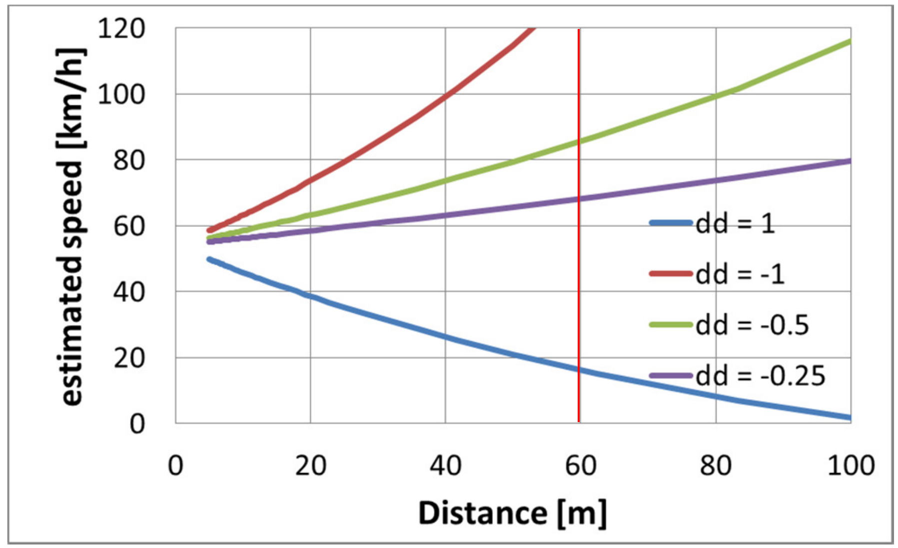
$$Z = \frac{f_x B}{d}, \quad f_x B \approx 250$$

$$\hat{Z} = \frac{Z}{(1 + \Delta d / d)}$$

$$\hat{v}_{rel} = \frac{v_{rel}}{(1 + \Delta d / d)^2}$$

$$\Leftrightarrow \hat{v}_f = v_{ego}(1 - 2\hat{v}_{rel} / v_{rel})$$





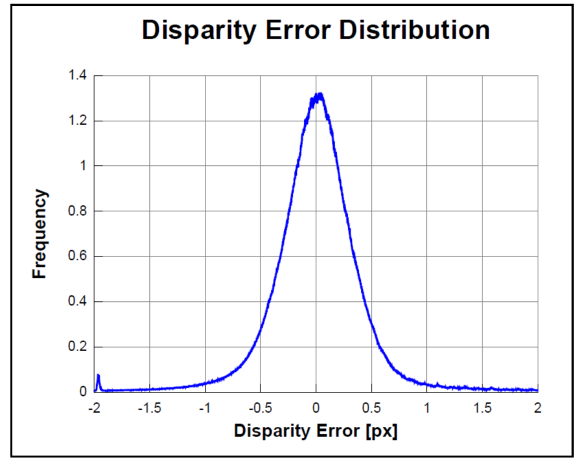
Two cars approaching with 50km/h each

How Precise is Disparity Estimation?

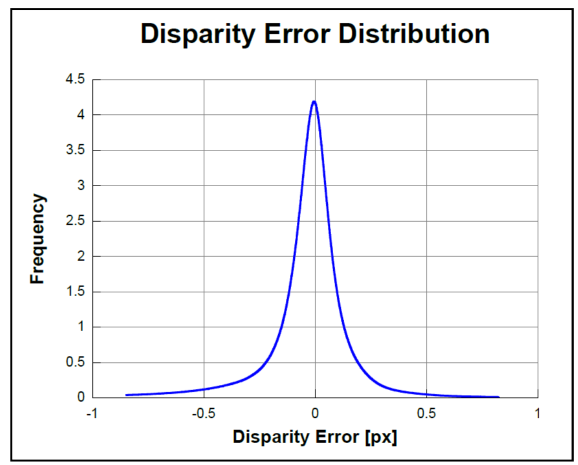
Accuracy investigations using rendered scenes

- Robust SGM about 0.3px.
- Optimal TV-L^1 about 0.1px, but lacking robustness and problems with large disparities.

Scene by C. Rabe, available at www.mi.auckland.ac.nz



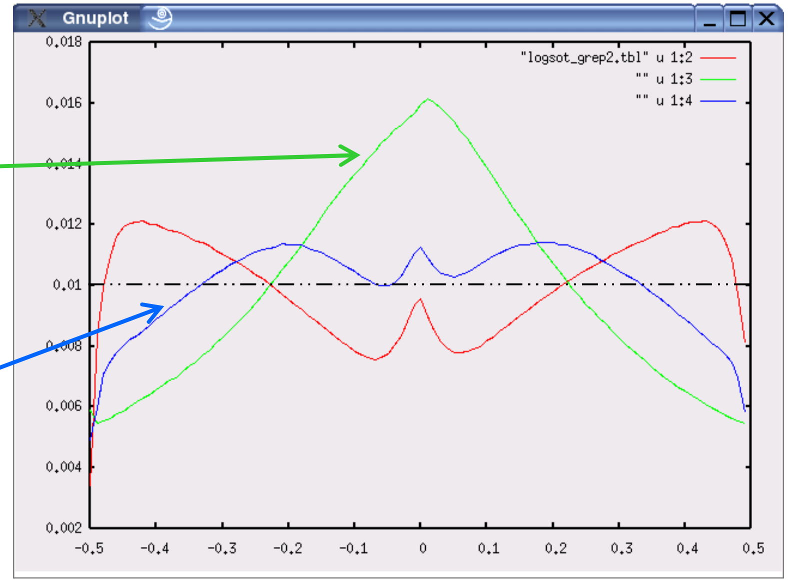
SGM with Census + Interpolation

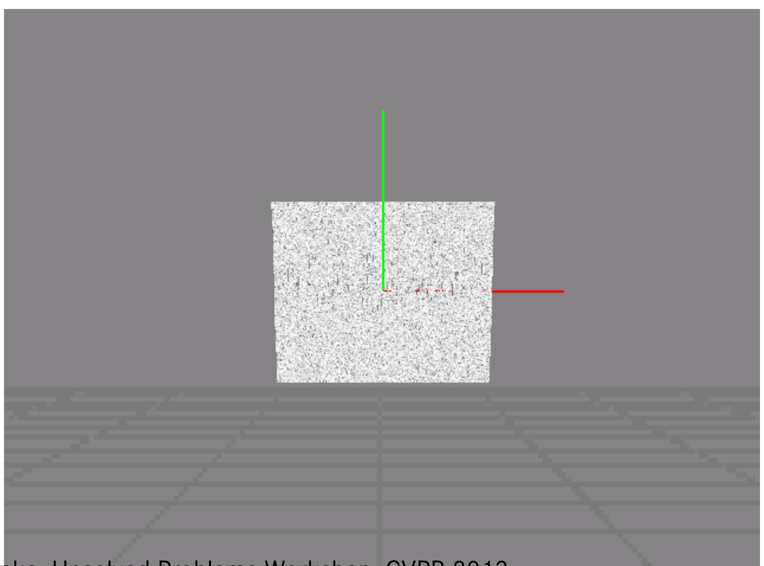


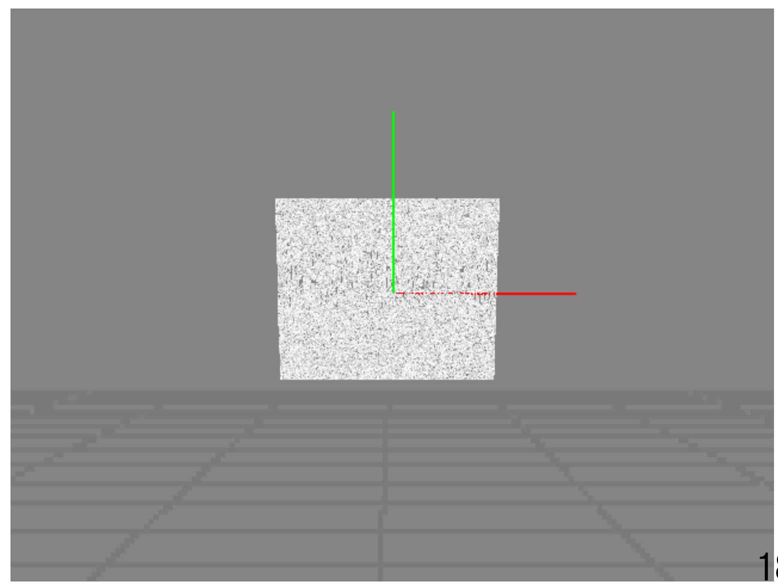
TV-L^1 , abs. diff, no window

The Pixel Locking Effect

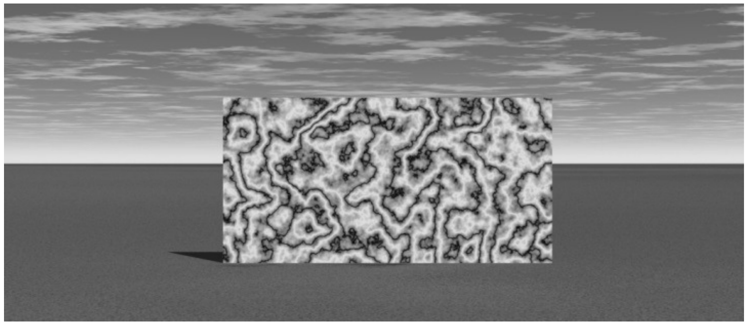
- Standard sub-pixel interpolation suffers from pixel-locking (unequal probability of the fractional disparities).
- Shimizu/Okutomi suggested a second sampling round, reducing the effect.
- Maximum error up to 0.15 px





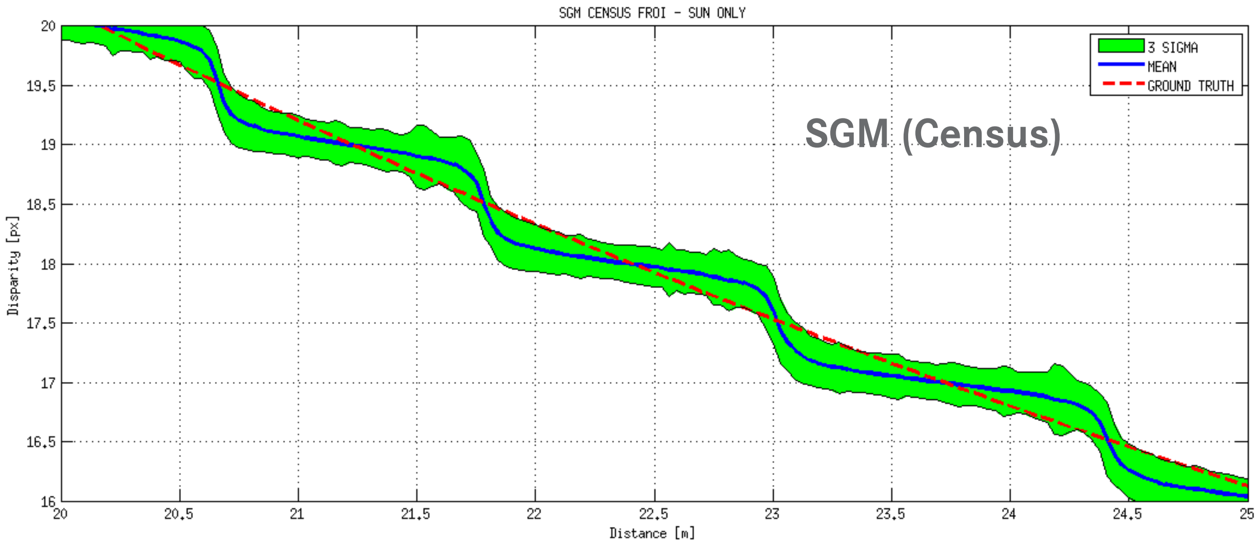


The Pixel Locking Effect



Rendered scene used for the investigation

Investigation revealed that the regularization (smoothness term!) leads to an additional Pixel Locking effect with a maximum error of up to 0.3px.



Effects of De-Calibration

Calibration cannot be considered perfect over the vehicle's life-time. However, even small epipolar errors can result in undesired effects (SGM+Census).

Good calibration

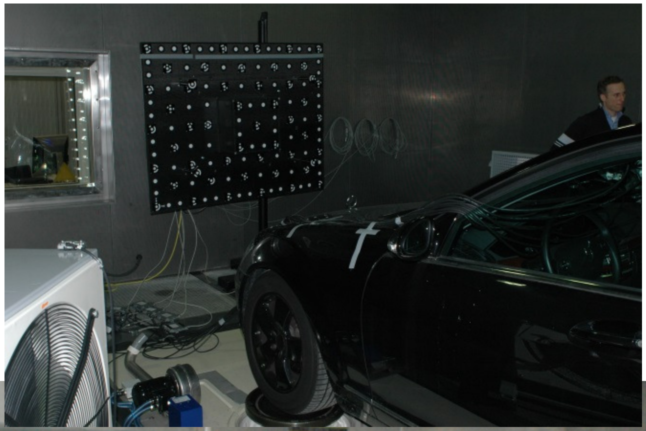
3D reconstruction good calibration

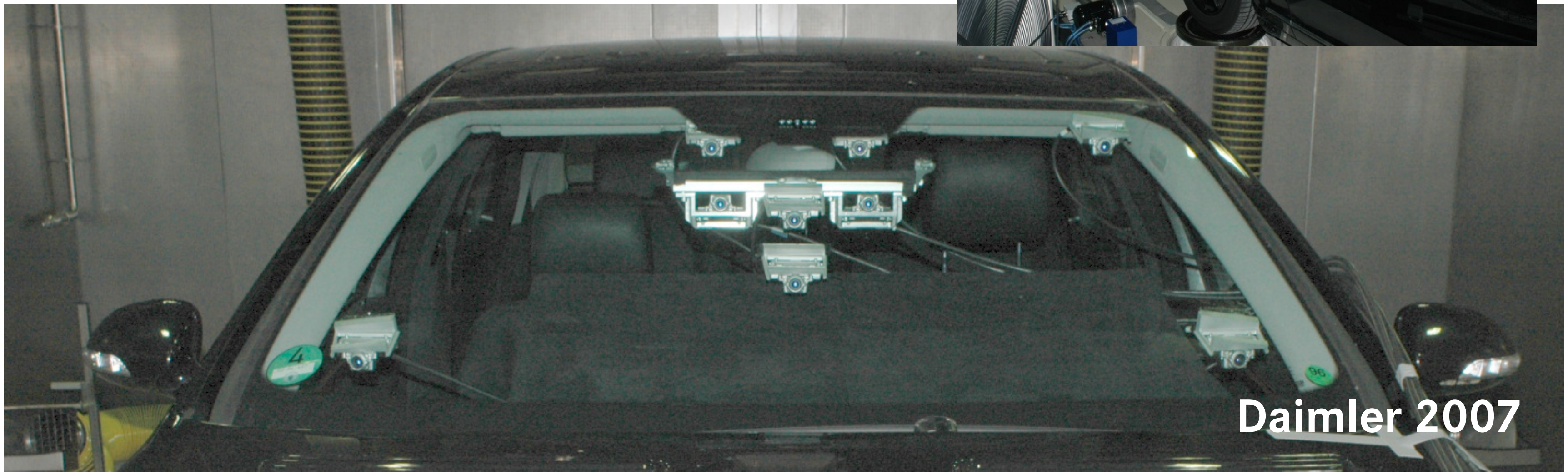
3D reconstruction -0.2px off

=> Epipolar error must be kept < 0.2px (for horizontal structures, e.g. stop lines)

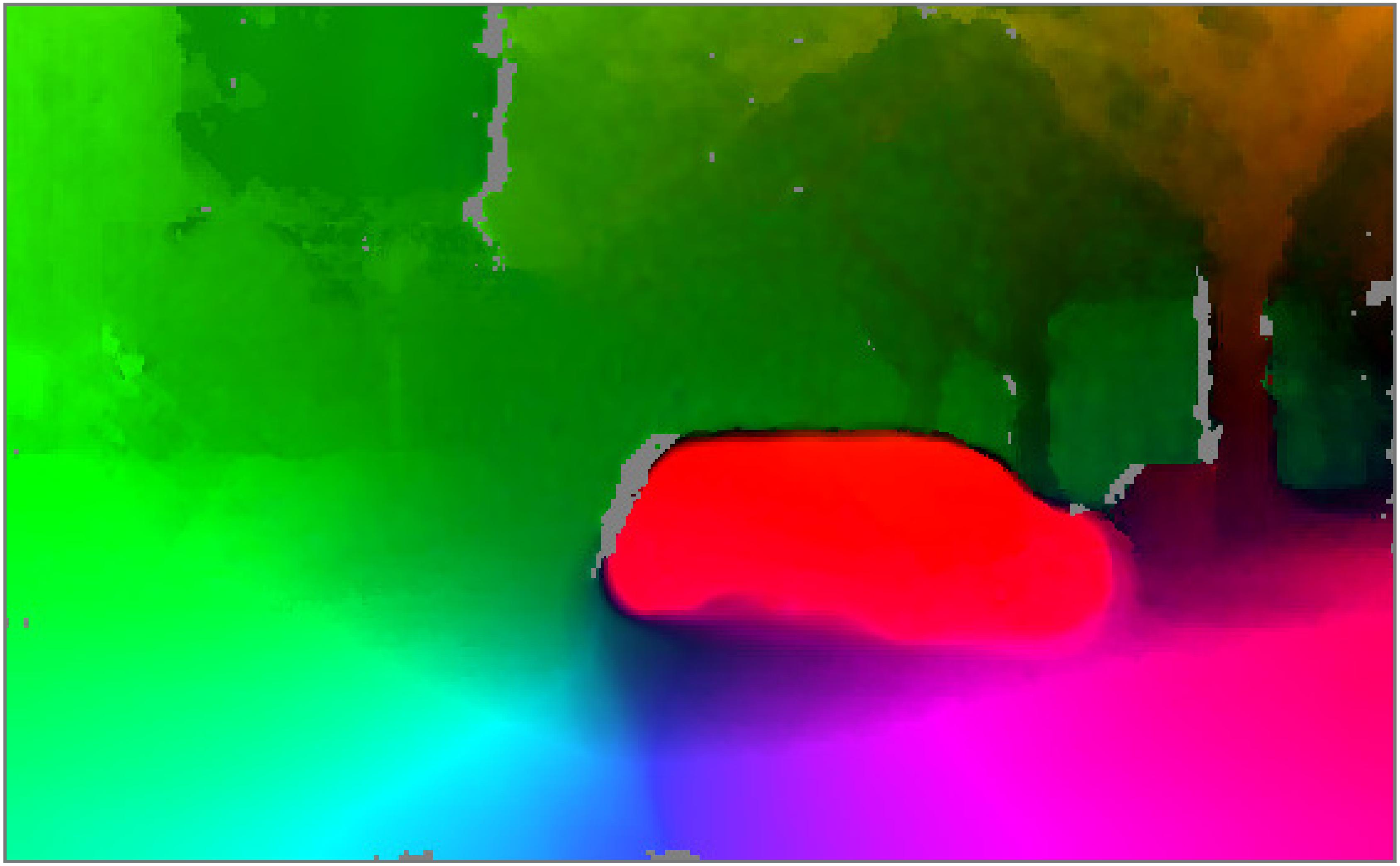
The Challenge of on-line Calibration

- Thermal stress tests (-40°C … +50°/80°C) revealed squint angle drifts equivalent to several pixel disparity error.
- Vision only all-weather on-line calibration in dynamic scenes not better than 0.2px !?
- Better results possible with radar reference.





Problems in Optical Flow

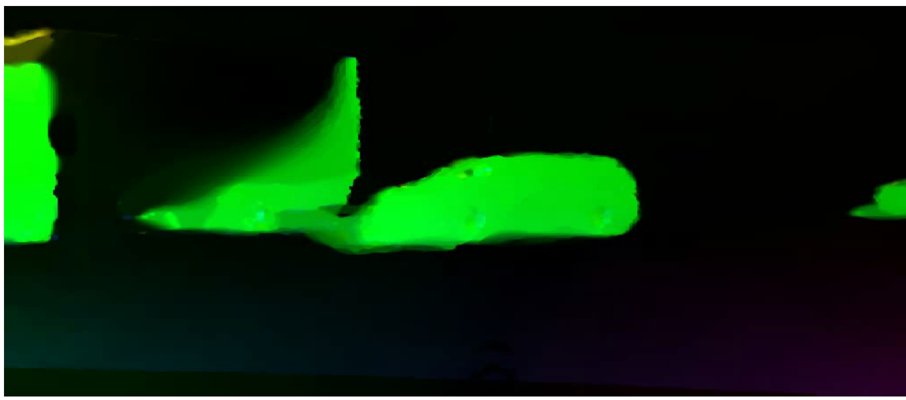


Illumination Changes and Large Displacements

Data from the GCPR webpage with ground truth available

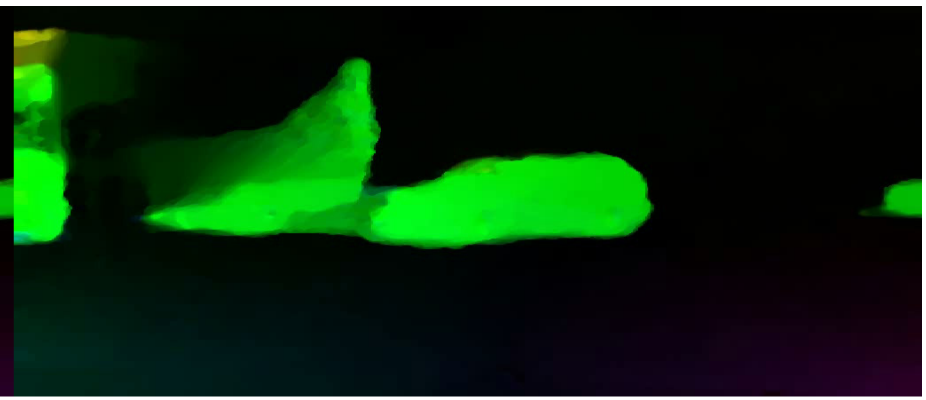
Illumination Changes

Illumination changes between consecutive images can be caused by shadows or changing exposure times, leading to wrong results.



Traffic scene with barely visible illumination changes

TV-L1 classic



TV-L1 with Census

TV-L1 classic

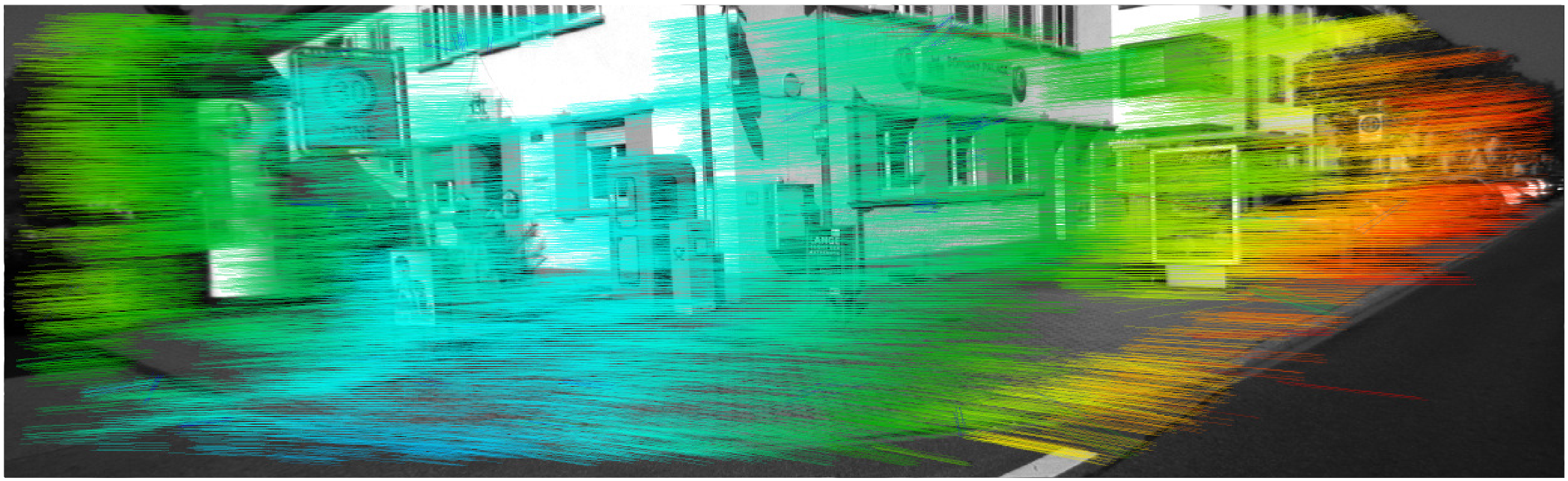
Using Census as matching criterion overcomes this problem.

Color encodes direction
Intensity encodes magnitude

T.Müller,C. Rabe, J.Rannacher, U.Franke: **Illumination-Robust Dense Optical Flow Using Census Signatures,** DAGM 2011

Large Optical Flow

Sparse schemes based on descriptor matching strategies can deliver arbitrarily large displacement vectors in constant time.



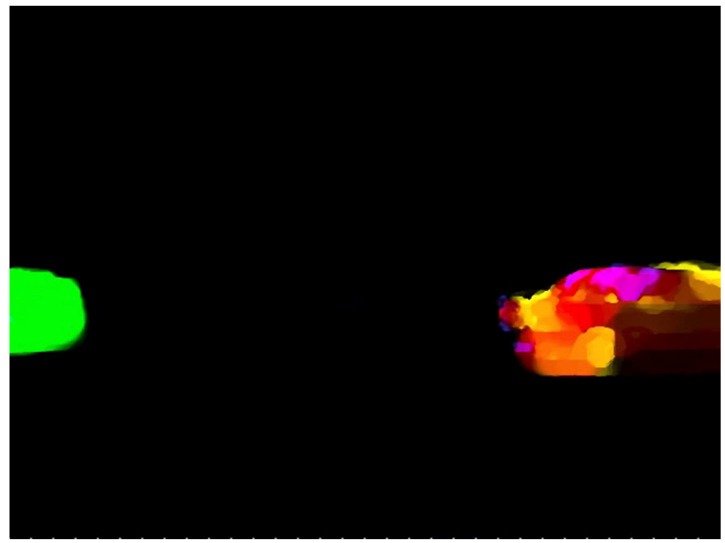
Displacements:
large ..... small
100 .... 0 pixel

F. Stein: "**Efficient Computation of Optical Flow Using the Census Transform**", DAGM 2004

Large Optical Flow

Dense schemes eg. TV-L^1 are based on gradient descent strategies and tend to stack in local minima.

TV-L^1 with support



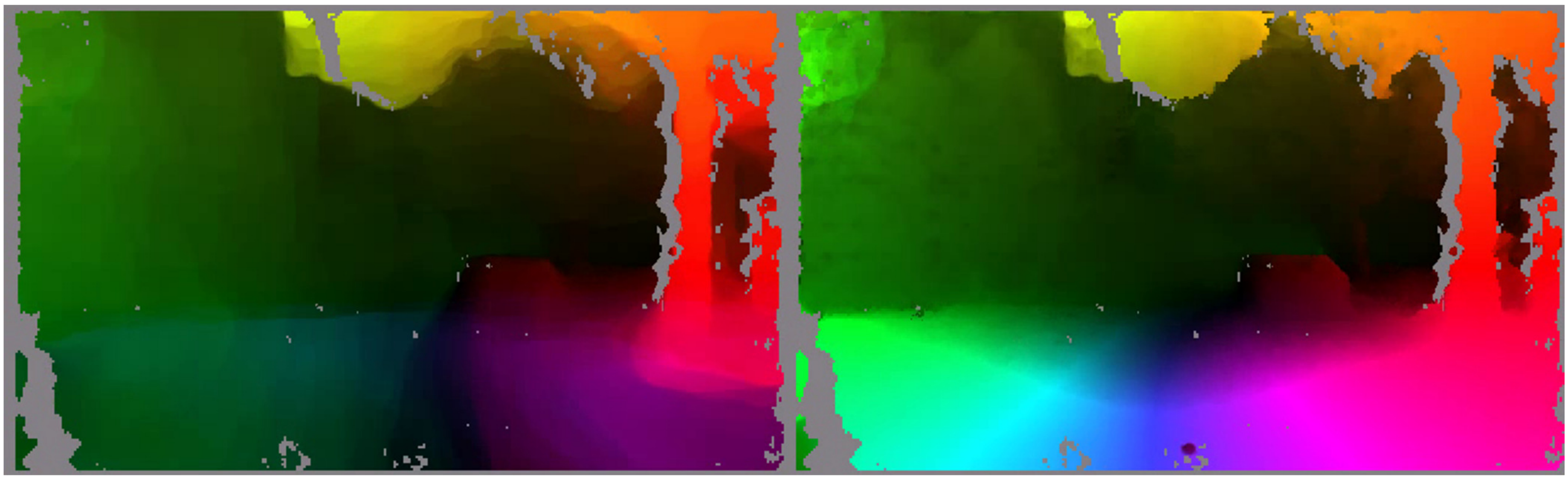
TV-L^1 without support

Additional sparse flow information can guide dense schemes to the correct solution.

T.Müller, J.Rannacher, C.Rabe, U.Franke, R.Mester: **Feature- and Depth-Supported Modified Total Variation Optical Flow for 3D Motion Estimation**, CVPR 2011

Large Optical Flow

Dense schemes eg. TV-L[1] are based on gradient descent strategies and tend to stack in local minima.



TV-L[1] without support   TV-L[1] with support

Alternatively, known depth and ego-motion can constrain the solution.

Sparse flow additionally used here to find the correct solution.

T.Müller, J.Rannacher, C.Rabe, U.Franke, R.Mester: **Feature- and Depth-Supported Modified Total Variation Optical Flow for 3D Motion Estimation**, CVPR 2011

More Problems in Optical Flow & Tracking

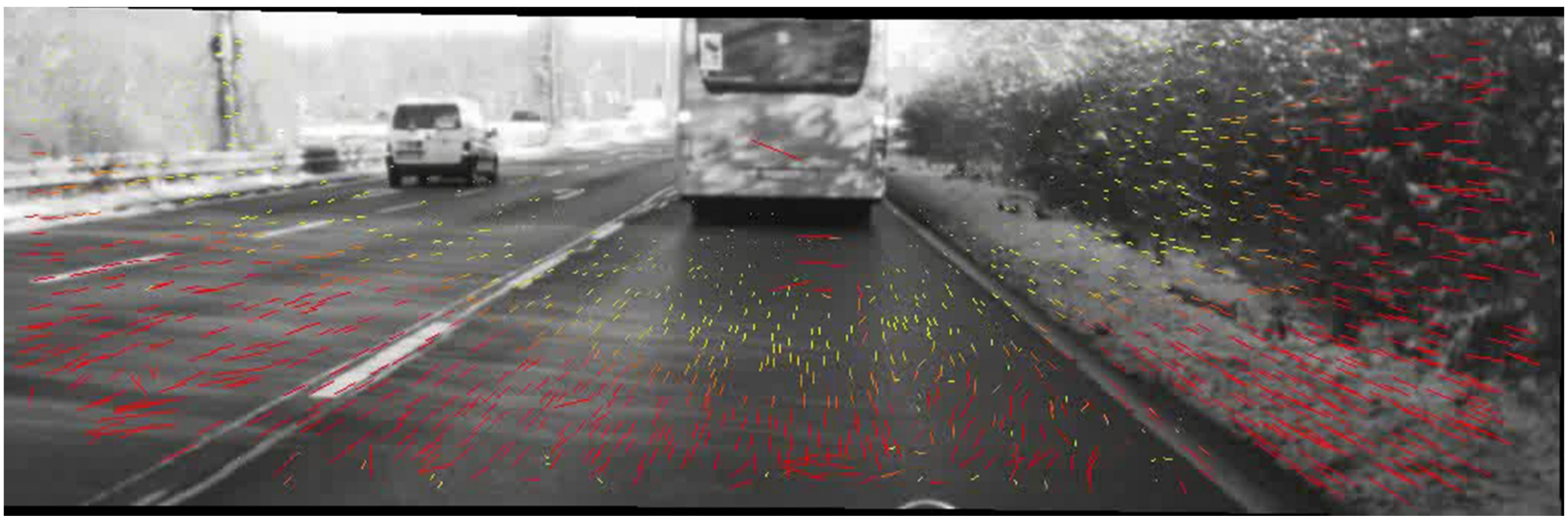
Reflections can be widely eliminated by sun visors.
However, design would be happy if we could live without.

KLT-Tracking, color coded displacement vectors

DAIMLER

More Problems in Optical Flow & Tracking

Moving shadows often occur in traffic scenes.
Tracking of the leading bus fails if optical flow is used.



KLT-Tracking, color coded displacement vectors

DAIMLER

Problems that need to be solved

- Improve the robustness of disparity estimation under adverse weather and illumination conditions.
- Improve the accuracy of disparity estimation.
- Improve vision-only on-line calibration precision.
- Improve the robustness of optical flow, start looking at adverse weather conditions!
- Generate ground truth data for adverse weather image sequences!!