ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Dense 3D scene flow estimation for locally rigid scenes

Konrad Schindler

Photogrammetry and Remote Sensing, ETH Zürich

3D Scene Flow

- “Joint stereo and optical flow”
- Given ≥2 video frames from ≥2 different viewpoints

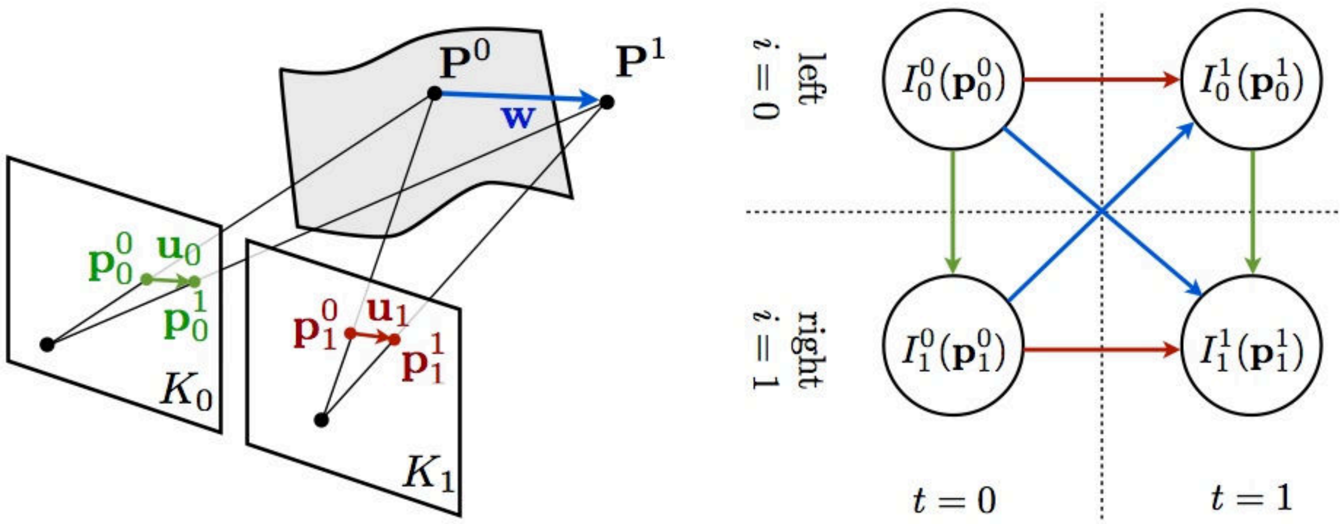
3D Scene Flow

- “Joint stereo and optical flow”
- Given ≥2 video frames from ≥2 different viewpoints
- Estimate dense 3D shape and 3D motion field

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

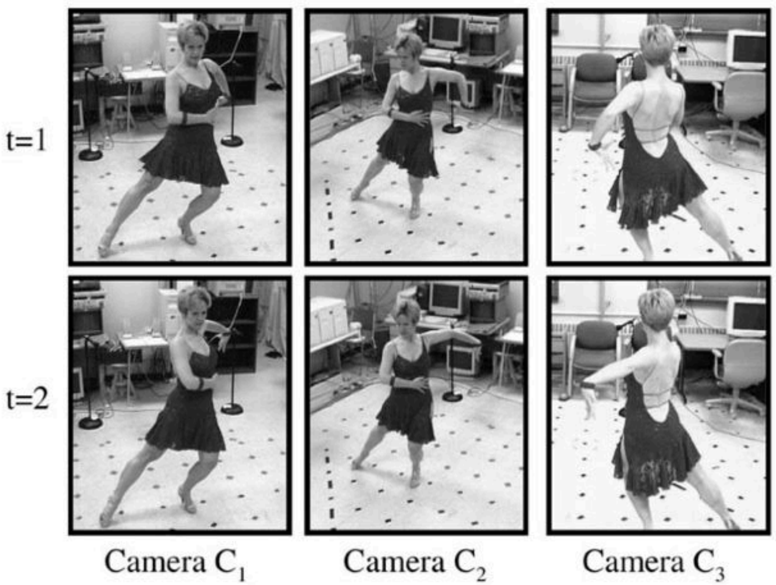
3D Scene Flow

- 4 unknowns per point: depth + 3D motion vector
- stereo/flow constraints between any two images
- (for simplicity, will refer to the 2-view case in the talk)



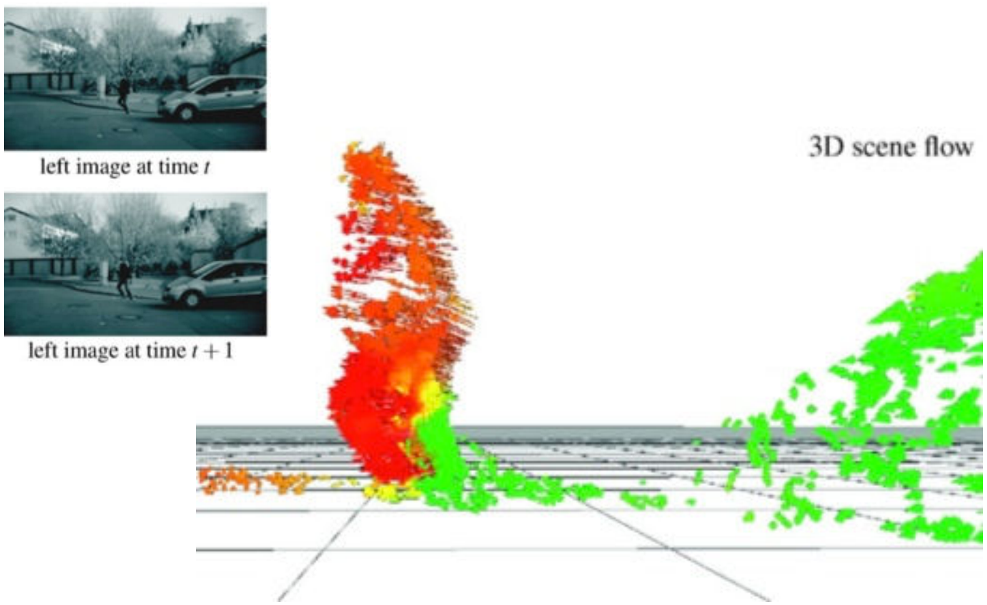
Applications

- Entertainment, 3D TV
- Motion capture
- Driver assistance, autonomous robots



[Vedula, Baker, Rander,

Collins, Kanade 1999]



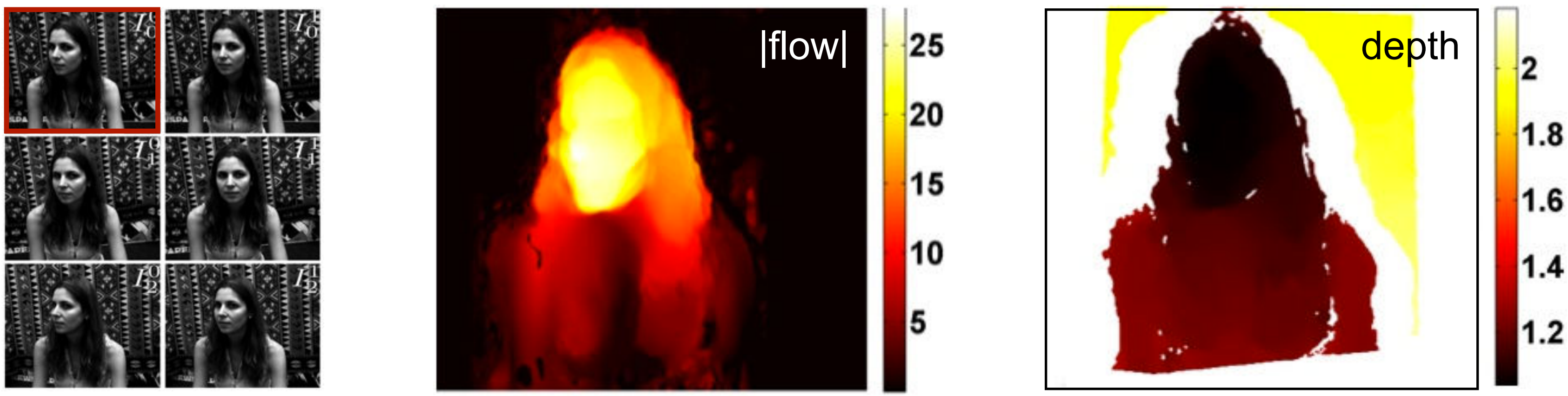
[Wedel, Rabe, Vaudrey,

Brox, Franke, Cremers 2008]

Estimation

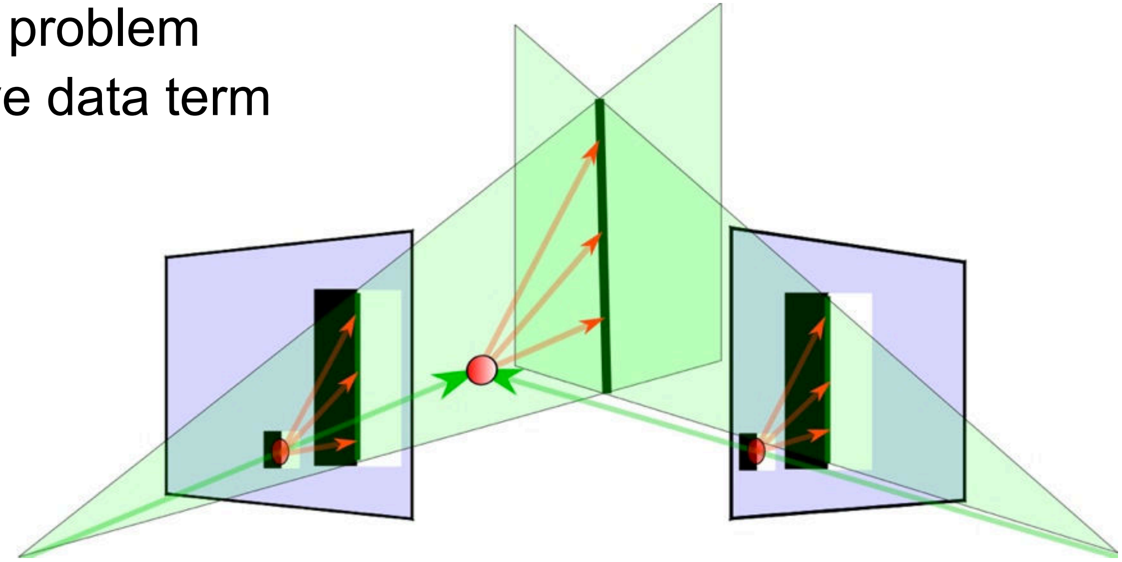
- First 2D flow, then stereo [Vedula et al. 1999]
- First stereo, then 2D flow [Pons, Keriven, Faugeras 2007] [Wedel et al. 2008]
- Everything jointly [Huguet, Devernay 2007] [Basha, Moses, Kiryati 2010]

- Usually parametrized w.r.t. a reference image



Scene Flow is ill-posed

- Available constraint (data term): the flow field should be compatible with all observed images
  - e.g. brightness constancy, cross-correlation, Census, ...
- Like for stereo and optical flow, this is insufficient to determine the flow field
  - 3D aperture problem
  - uninformative data term



Standard regularization

- surface / motion field is piecewise smooth

$$\hat{\mathbf{u}} = \operatorname{argmin}\ E(\mathbf{u}) \quad , \quad E(\mathbf{u}) = E_{data}(\mathbf{u}) + \lambda \cdot E_{smooth}(\mathbf{u})$$

- robust penalty to allow for discontinuities, e.g. total variation (TV-L1)

$$E_{data} = \int \rho(\nabla \mathbf{u}) d\mathbf{x} \quad , \quad \rho(\nabla \mathbf{u}) = |\nabla \mathbf{u}|$$

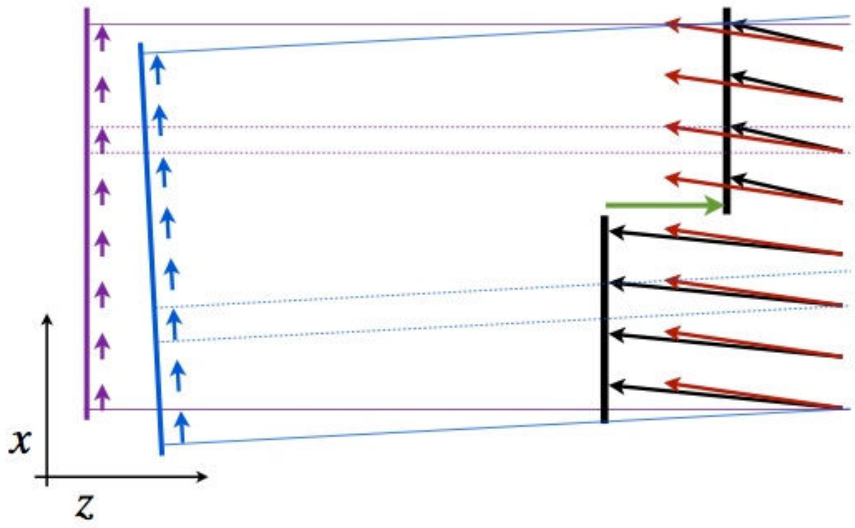
- widely used in stereo and optical flow estimation

TV-L1 applied to 3D scene flow

- Total variation of depth and motion field

$$E_S^{\mathrm{TV}}(d, \mathbf{w}) = \int_\Omega \rho(\nabla d) + \rho(\nabla w_x) + \rho(\nabla w_y) + \rho(\nabla w_z) \mathrm{d}\mathbf{x}$$

- Does not work well for narrow baselines
- Isotropic smoothing is biased against depth discontinuities



Rigidity

- Observation: many scenes of practical importance consist of relatively few rigid, independently moving parts
- Develop a regularizer that encourages rigidity rather than smoothness

Attempt 1 - local rigidity

- For each scene point, penalize the deviation from rigidity
  - look at a small patch around the point
  - find the "non-rigid motion residual", i.e. the difference between the observed motion and its projection onto the rigid motion subspace
  - smoothness term is a robust function of this residual

$$E_S^R(\mathbf{w}) = \int_\Omega \psi(v^R(\mathbf{x};\mathbf{w}))\, \mathrm{d}\mathbf{x}$$

$$v^R(\mathbf{x};\mathbf{w}) = \int_{\mathcal{C}(\mathbf{x})} \|\mathbf{r}(\mathbf{y};\mathbf{w}|_{\mathcal{C}(\mathbf{x})}) - \mathbf{w}(\mathbf{y})\|_2^2 \eta(\mathbf{x},\mathbf{y})\, \mathrm{d}\mathbf{y}$$

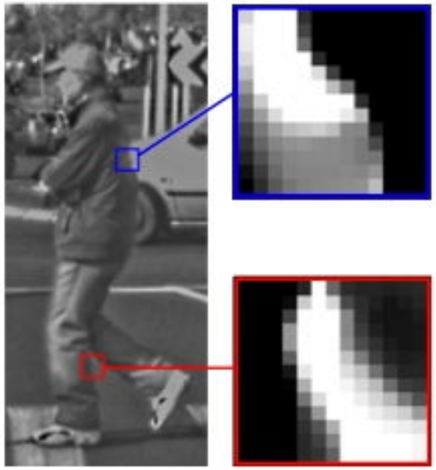
[Vogel, Roth, Schindler ICCV'11]

Local rigidity

- For small motions this can be done quite efficiently
  - linearize rotation, discretize to pixel grid
  - non-rigid motion residual has linear closed form
  - it is not necessary to explicitly compute the rigid motion

$$E_S^R(\mathbf{w}) = \sum_{c \in C} \psi\left(v_{\text{dsc}}^R(c;\mathbf{w})\right) \qquad R \approx I + \alpha[\mathbf{r}]_\times$$

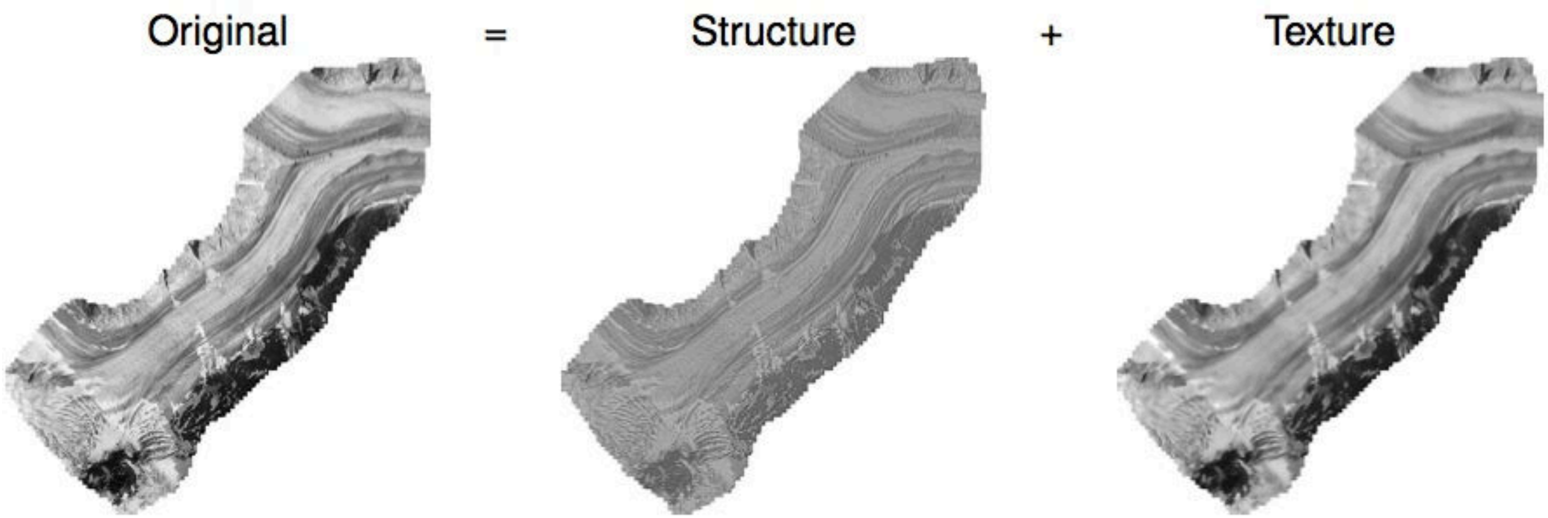
$$v_{\text{dsc}}^R(c;\mathbf{w}) = \left\| A_c \mathbf{w}_{(c)} - \mathbf{w}_{(c)} \right\|_{N_c}^2$$



  - Pixels in the patch are weighted to avoid fitting across depth and motion boundaries

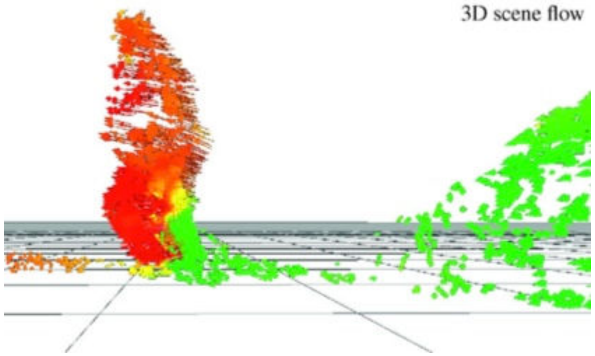
Inference

- variational energy minimization
- usual tricks from optical flow can (and should) be used
  - auxiliary dual variables to decouple data and smoothness terms
  - course-to-fine scheme
  - careful gradient interpolation
  - structure-texture decomposition



Evaluating scene flow

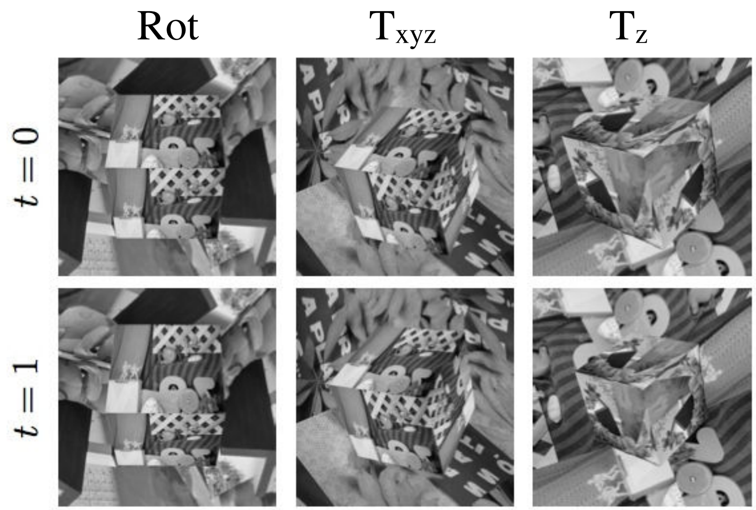
- Ground truth
  - no easy way to obtain dense ground truth
  - no existing benchmark data
  - quantitative results only for synthetic scenes
  - (note, some public ground truth is wrong)



- Error measures
  - angular error of 3D flow component
  - RMS of depth, normalized to scene extent
  - RMS of flow, normalized to scene extent
  - 2D angular of reprojected 2D flow field
  - 2D endpoint error of reprojected flow field

Results

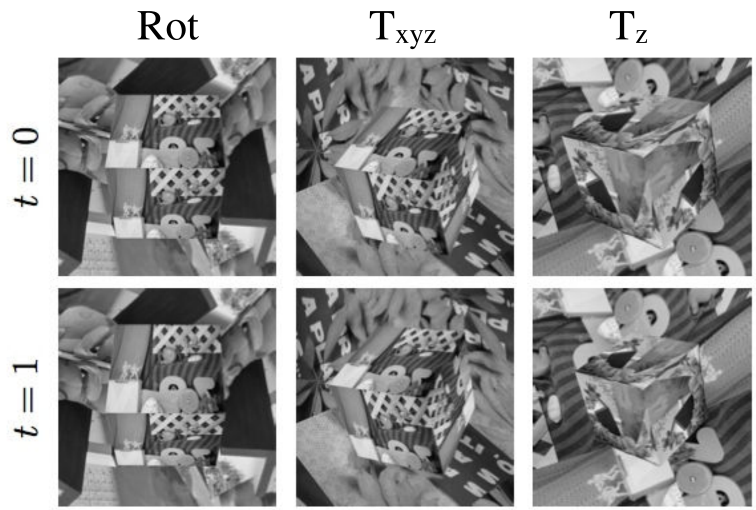
- Synthetic scenes with known ground truth



Scene		3D Error AAE$_w$	3D Error NRMS$_w$	3D Error NRMS$_d$	2D Error AAE	2D Error AEP
Rot	Rig	**4.5°**	**7.3%**	11.7 %	**1.6°**	0.36
	TV	8.5°	9.8%	**11.6%**	1.7°	**0.35**
T_{xyz}	Rig	**2.5°**	**11.9%**	11.8 %	**1.5°**	**0.39**
	TV	8.6°	25.6 %	**11.7%**	2.3°	0.42
T_z	Rig	**3.9°**	**14.0%**	**9.9%**	**1.8°**	**0.35**
	TV	7.8°	15.3%	10.7 %	2.4°	0.37

Results

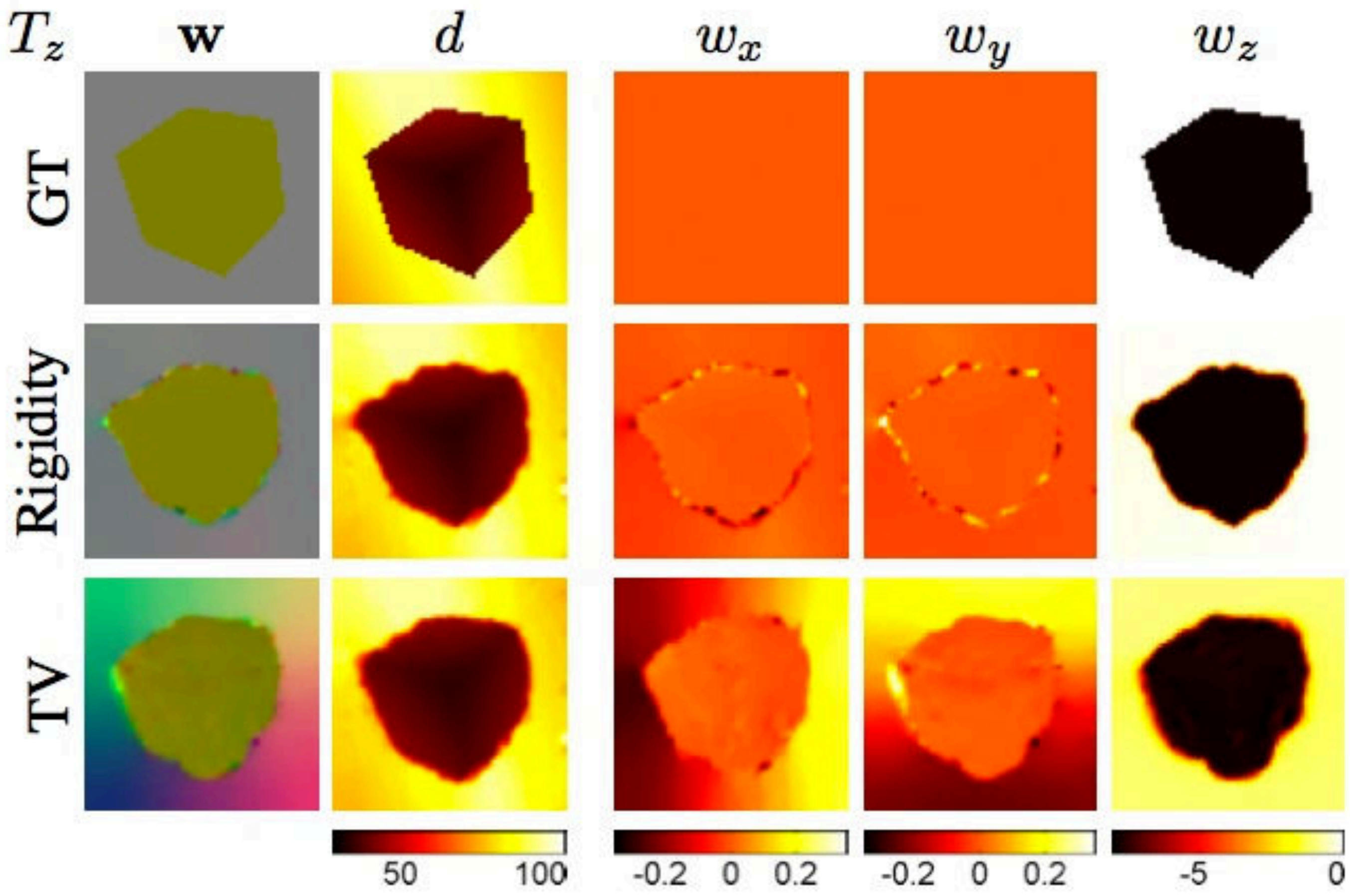
- Synthetic scenes with known ground truth



Scene		AAE_w [°]			$NRMS_w$ [%]		
		—	OC	OC&DC	—	OC	OC&DC
Rot	Rig	**4.5**	**4.1**	**3.3**	**7.3**	**6.7**	**4.5**
	TV	8.5	8.2	7.4	9.8	9.4	7.9
T_{xyz}	Rig	**2.5**	**2.1**	**1.5**	**11.9**	**10.3**	**6.4**
	TV	8.6	8.0	7.7	25.6	24.4	23.6
T_z	Rig	**3.9**	**2.5**	**1.2**	**14.0**	**10.6**	**5.0**
	TV	7.8	6.5	4.9	15.3	13.3	8.2

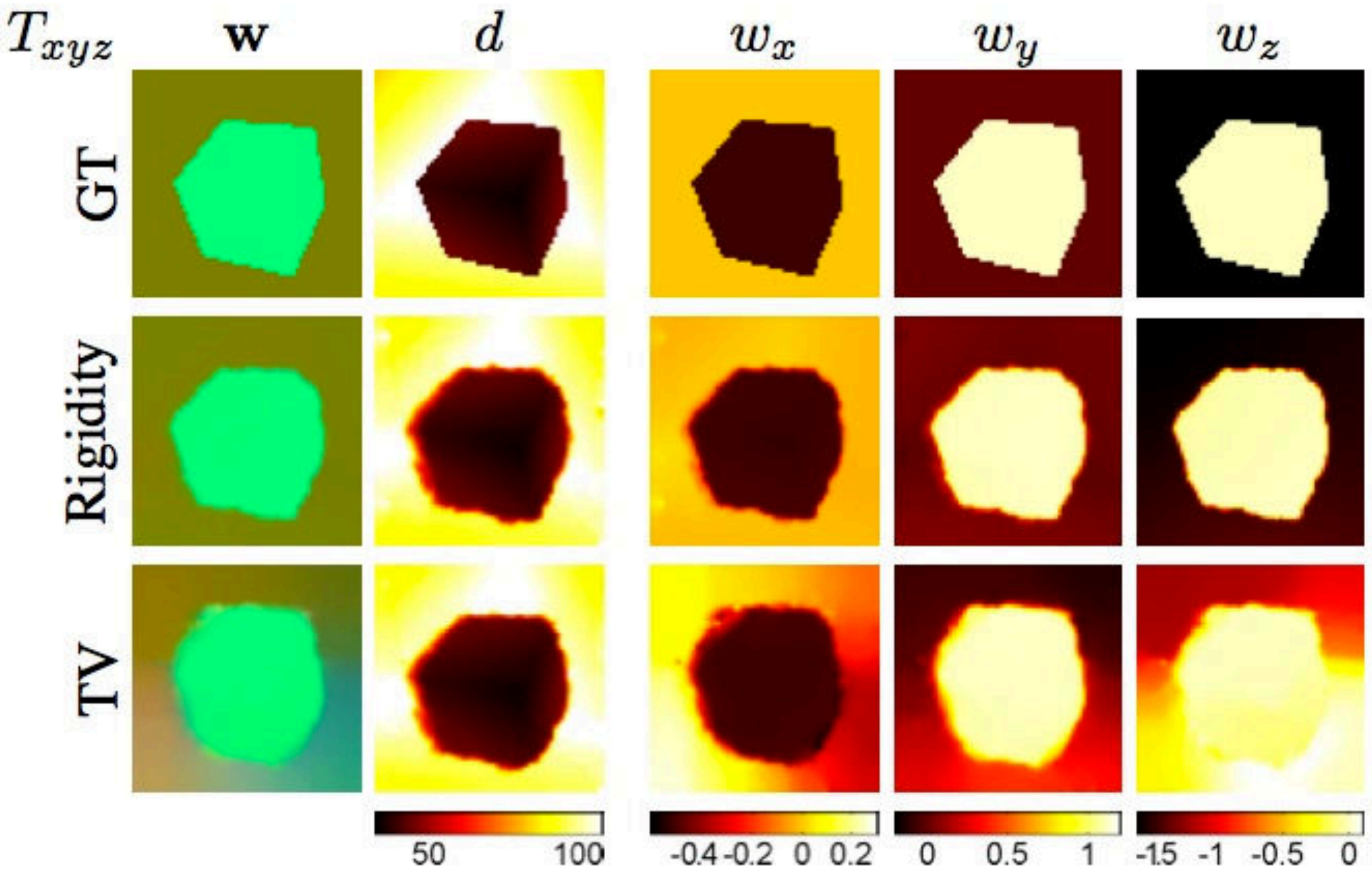
Results

- Qualitative results



Results

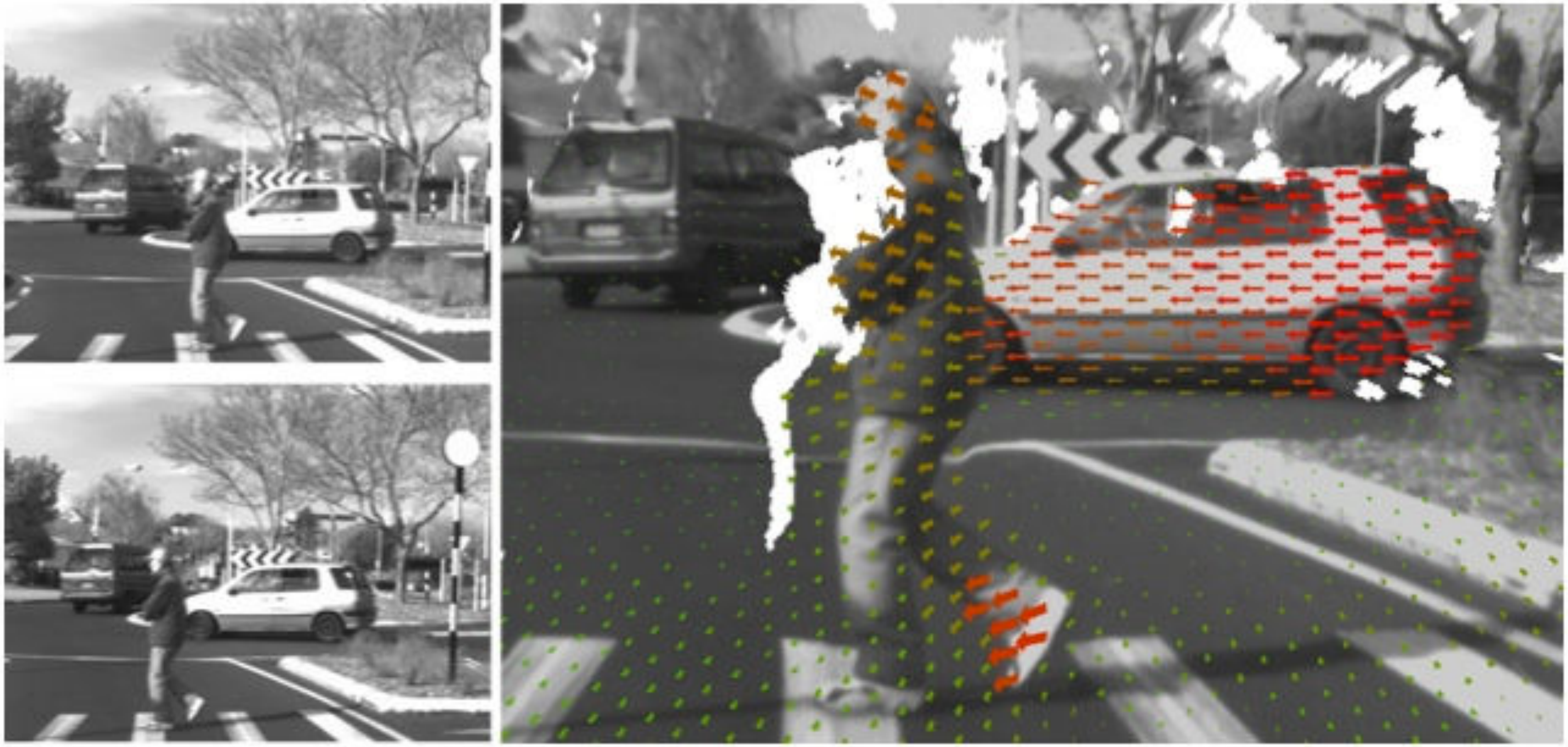
- Qualitative results



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Results

- Stereo camera on car



Results

- Maria (three views)

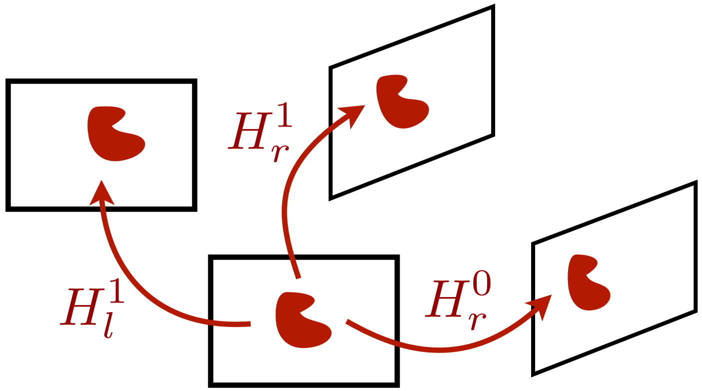
Attempt 2 - piecewise planarity and rigidity

- Explicitly model scene as a collection of planar patches
- Recent trend in both
  - stereo [e.g. Bleyer et al. 2011] and
  - optical flow [e.g. Sun, Sudderth, Black 2010]

[Bleyer, Gelautz, Rother, Rhemann 2009]

Piecewise planarity and rigidity

- For 3D scene flow, represent scene as a collection of rigidly moving planes
- Why that?
  - a good approximation for most object surfaces
  - stronger regularization
  - large, well-delineated support regions for estimation
  - simple mapping with homographies
  - (potential for implicit (or even explicit) object segmentation)



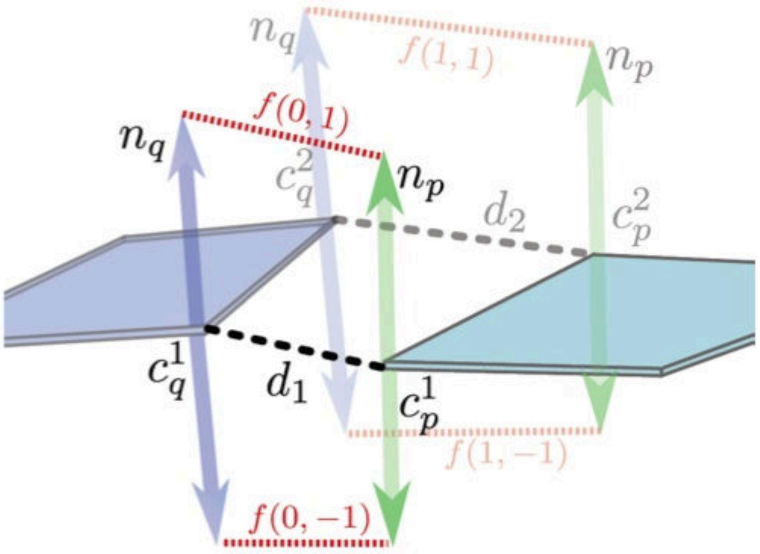
[Vogel, Roth, Schindler ICCV'13]

Energy

- In addition to data fidelity and smoothness, encourage "good" segmentation

$$E(\mathcal{P},\mathcal{S}) = E_D(\mathcal{P},\mathcal{S}) + \lambda E_R(\mathcal{P},\mathcal{S}) + \mu E_S(\mathcal{S})$$

- Data term
  - two optical flow pairs
  - two stereo pairs
- Smoothness at boundary pixels
  - 3D distances between patches small
  - curvature small (computed from distances of auxiliary points)



Segmentation regularization

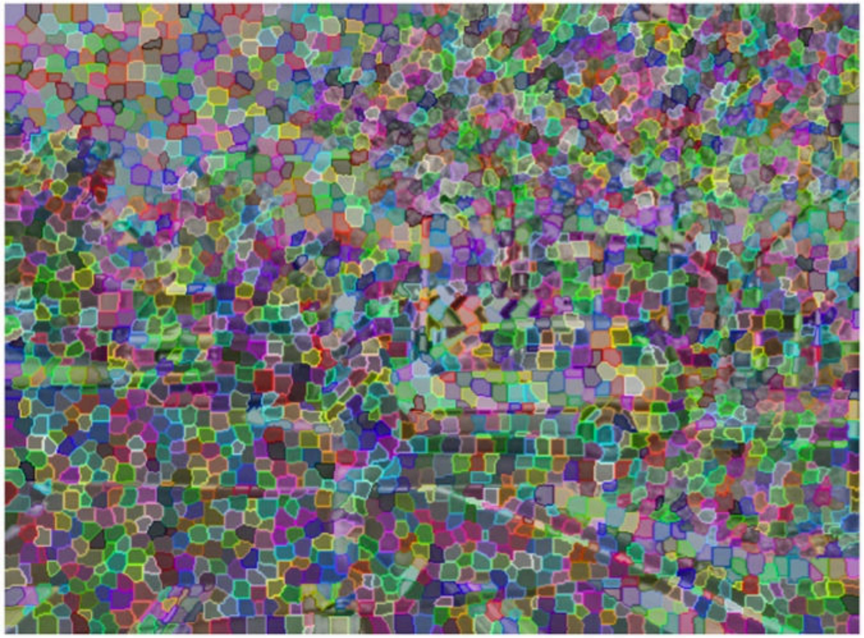
- Segments should be spatially coherent (≠ compact)
- Suitable models exist in energy-based segmentation
  - Potts model to encourage segment boundaries at high gradients
  - allow only assignments to nearby segments

$$E_S(\mathcal{S}) = \sum_{\substack{(\mathbf{p},\mathbf{q})\in\mathcal{N},\\ \mathcal{S}(\mathbf{p})\neq\mathcal{S}(\mathbf{q})}} \exp\left(\frac{-a|I_l^0(\mathbf{p}) - I_l^0(\mathbf{q})|}{\sigma_I(\mathbf{p},\mathbf{q})+\epsilon}\right) + \sum_{\mathbf{p}\in I_l^0} \begin{cases} 0, & \exists \mathbf{e}\in\mathcal{E}(s_i) : ||\mathbf{e}-\mathbf{p}||_\infty < N_S \\ \infty, & \text{else.} \end{cases}$$

[Veksler, Boykov, Mehrani 2010]

Inference

- Initial segmentation based only on intensity
- Compute multiple scene flow proposals for each segment
  - run simpler scene flow, 2D flow + stereo
  - fit rigidly moving planes to segments

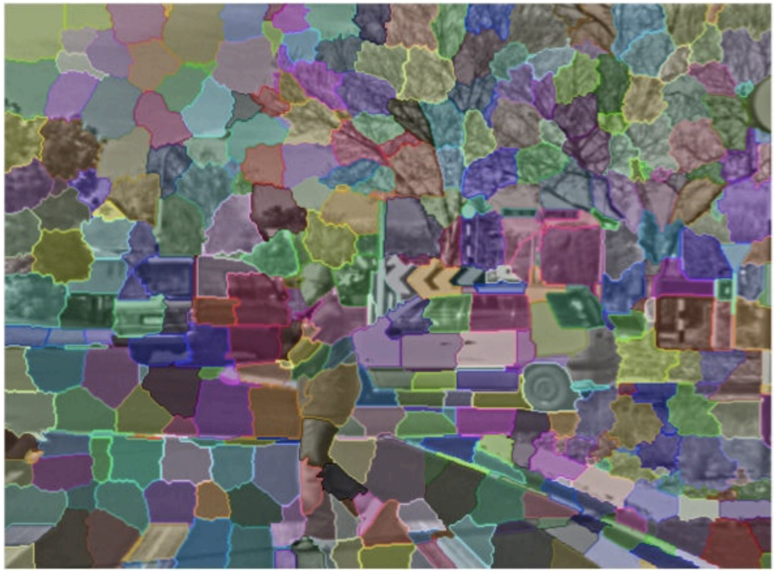


Inference

- Assign each segment to one of the proposals
  - minimize energy function over all pixels of all segments
  - α-expansion with QPBO

Occlusions

- Scene flow is 3D, allows for explicit occlusion modeling
  - occluded pixels do not pay a data penalty
  - instead they are assigned a fixed occlusion penalty
- Integration into inference scheme
  - in each binary expansion step, find cases "*if super-pixel **p** is on plane **X** and super-pixel **q** is on plane **Y**, then **p** occludes **q***"
  - there can be >1 segments on the line of sight to ***q*** → higher-order cliques
  - reduce to binary cliques with auxiliary variables (carefully, to introduce few non-submodular cliques)

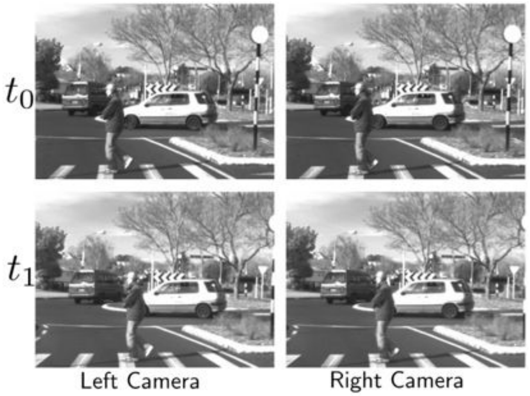


Inference

- Estimate segment-level occlusions
  - include occlusion potentials
  - solve assignment again

Inference

- Keep plane parameters fixed, reassign pixels to segments
  - simpler now, because only few segments within the allowed distance
- (iterate)



Results

- Quantitative results
  - KITTI flow+stereo benchmark
  - errors of reprojected 2D disparity/flow (no other ground truth)

Optical Flow Evaluation

This table ranks general optical flow methods, performing a full 2D search, as com

Rank	Method	Setting	Code	Out-Noc	Out-All	Avg-Noc	Avg-All	Density
1	PR-Sf+E			4.08 %	7.79 %	0.9 px	1.7 px	100.00 %
Anonymous submission								
2	PCBP-Flow			4.08 %	8.70 %	0.9 px	2.2 px	100.00 %
K. Yamaguchi, D. McAllester and R. Urtasun: Robust Monocular Epipolar Flow Estimation. CVPR 2013.								
3	MotionSLIC			4.36 %	10.91 %	1.0 px	2.7 px	100.00 %
K. Yamaguchi, D. McAllester and R. Urtasun: Robust Monocular Epipolar Flow Estimation. CVPR 2013.								
4	PR-Sceneflow			4.48 %	8.98 %	1.3 px	3.3 px	100.00 %
Anonymous submission								
5	TGV2ADCSIFT			6.55 %	15.35 %	1.6 px	4.5 px	100.00 %
6	Data-Flow			7.47 %	14.85 %	1.9 px	5.5 px	100.00 %
C. Vogel, S. Roth and K. Schindler: An Evaluation of Data Costs for Optical Flow. German Conference on Pattern R								
7	DeepMatching			8.04 %	18.60 %	1.6 px	5.7 px	100.00 %
Anonymous submission								
8	TVL1-HOG			8.31 %	19.21 %	2.0 px	6.1 px	100.00 %
Anonymous submission								

Stereo Evaluation

Rank	Method	Setting	Code	Out-Noc	Out-All	Avg-Noc	Avg-All	Density
1	PCBP-SS			3.49 %	4.79 %	0.8 px	1.0 px	100.00 %
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 %
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 %
Anonymous submission								
4	PCBP			4.13 %	5.45 %	0.9 px	1.2 px	100.00 %
K. Yamaguchi, T. Hazan, D. McAllester and R. Urtasun: Continuous Markov Random Fields for Robust Stereo Esti								
5	PR-Sceneflow			4.46 %	5.32 %	1.0 px	1.1 px	100.00 %
Anonymous submission								
6	DDS			4.63 %	5.44 %	1.0 px	1.1 px	100.00 %
Anonymous submission								
7	wSGM			5.03 %	6.24 %	1.3 px	1.6 px	97.03 %
T. Robert Spangenberg and R. Rojas: Weighted Semi-Global Matching and Center-Symmetric Census Transform f								
8	ATGV			5.05 %	6.91 %	1.0 px	1.6 px	100.00 %
R. Ranftl, T. Pock and H. Bischof: Minimizing TGV-based Variational Models with Non-Convex Data terms. ICSSV								

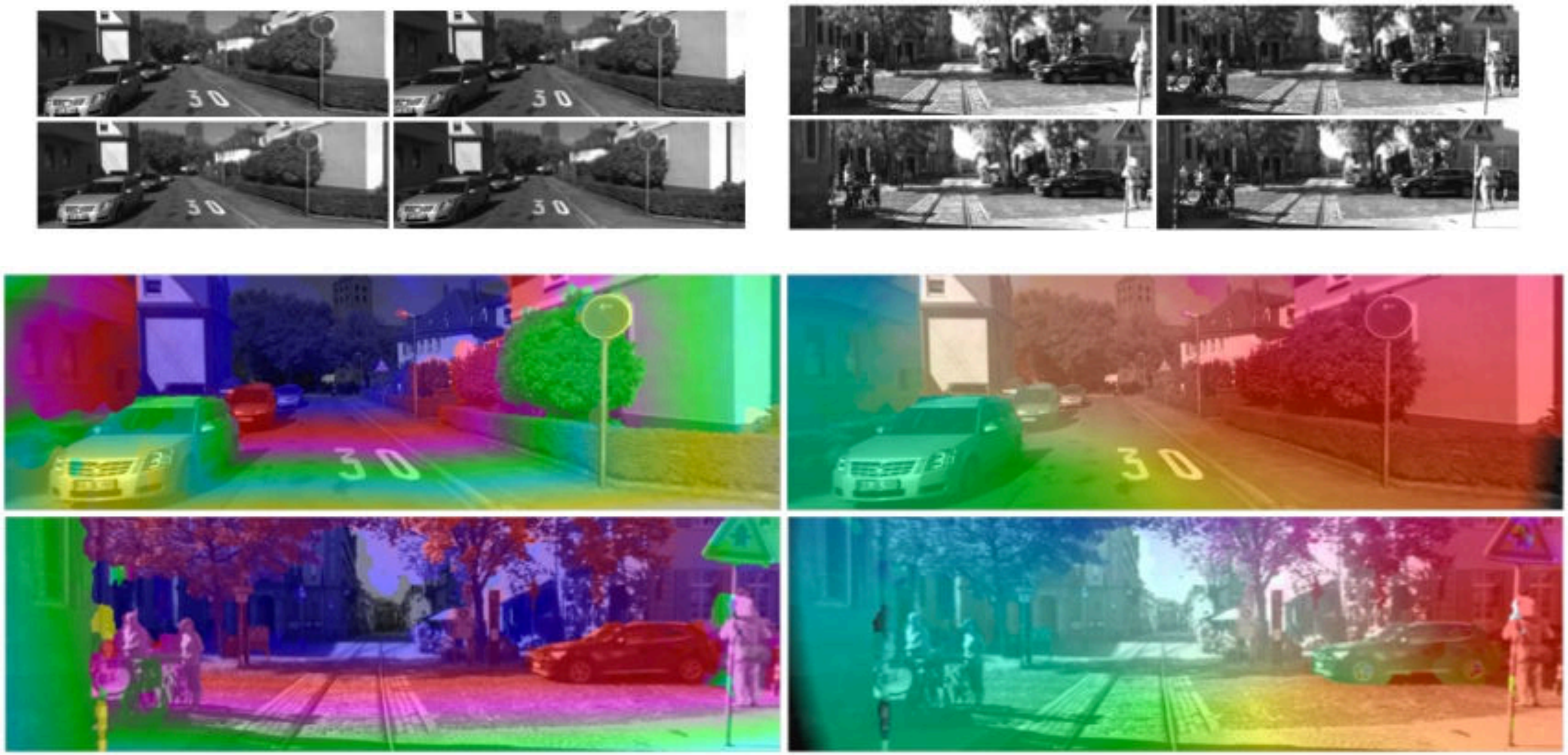
Results

- Comparison
  - other 3D scene flow algorithms
  - intermediate stages and variants of the framework

	FLOW (*All*)				FLOW (*Noc*)				STEREO(*All*)				STEREO (*Noc*)			
Error threshold Z	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
LSF [3]	21.6	16.9	14.3	12.7	16.0	12.0	10.0	8.8	17.6	12.0	9.0	7.2	16.4	10.8	8.0	6.3
Rig [24]	16.1	12.1	10.1	8.8	10.6	7.3	5.7	4.8	15.0	10.6	8.3	6.8	13.7	9.5	7.2	5.8
2D [11, 26]	18.9	15.0	12.8	11.3	11.0	7.9	6.5	5.7	13.5	9.9	8.0	6.7	12.3	8.9	7.0	5.8
PRSSeg-3D	13.8	10.1	8.2	7.1	8.4	5.6	4.5	3.9	9.4	6.8	5.4	4.6	8.4	6.0	4.8	4.0
PRSPix-3D	12.8	9.3	7.6	6.6	7.2	4.7	3.7	3.2	8.1	5.8	4.6	3.9	7.1	5.0	4.0	3.3
PRSSeg-2D	12.4	9.0	7.3	6.4	7.4	5.0	3.9	3.4	8.9	6.4	5.1	4.3	7.9	5.6	4.4	3.7
PRSPix-2D	11.8	8.5	6.9	6.0	6.9	4.5	3.5	3.0	8.3	5.9	4.7	3.9	7.3	5.1	4.0	3.3
PRSPix-O-2D	11.2	7.7	5.9	5.1	6.8	4.4	3.3	2.8	8.3	5.9	4.7	4.0	7.4	5.2	4.1	3.4
PRSPix-2D+R	10.9	7.6	6.0	5.1	6.3	4.1	3.1	2.7	7.9	5.7	4.5	3.8	6.9	4.8	3.8	3.2
PRSPix-2D+R+E	**10.0**	**6.7**	**5.0**	**4.1**	**5.8**	**3.6**	**2.7**	**2.2**	**7.4**	**5.3**	**4.2**	**3.5**	**6.4**	**4.5**	**3.6**	**3.0**

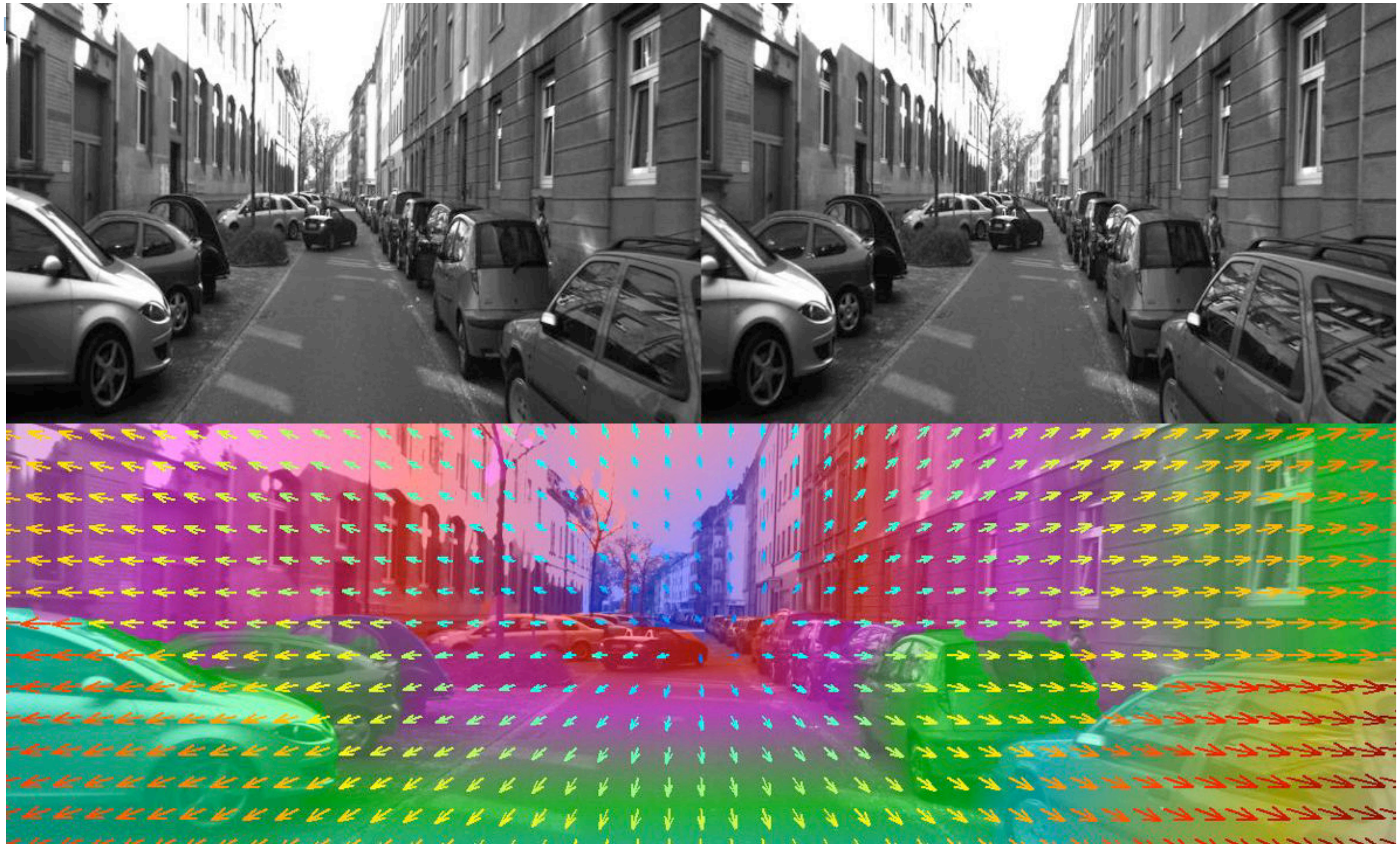
Results

- Qualitative results



Results

- Qualitative results



Summary and Outlook

- Does rigidity help? Yes!
  - we believe it is the best regularizer for scene flow thus far
- Local or piecewise rigidity? Hard to say;
  - obviously depends on the scene - “horses for courses”
  - in practice piecewise so far works much better
- Something obvious missing? Yes!
  - >2 time steps must be beneficial (although notoriously hard to show)
  - rigidity should benefit more than agnostic smoothing
- Is rigidity the final word? No!
  - scene flow is still in its infancy compared to mature computer vision problems (stereo, flow, SfM, categorization...)
  - we as a community can certainly do a lot better; a good area if you want to make an impact ;-)

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Cast listing

Christoph Vogel

Stefan Roth

Konrad Schindler

ECCV 2014 – European Conference on Computer Vision

Zurich, September 5-12, 2014

HOME IMPORTANT DATES LOCAL ARRANGEMENT PEOPLE

Important Dates

- ECCV 2014 Submission Deadline: 7 March 2014 – Friday
- ECCV 2014 Supplementary Materials Deadline: 14 March 2014 – Friday
- ECCV 2014 Announcement of Decisions : 16 June 2014 – Monday

Photogrammetric Computer Vision - PCV 2014

ISPRS Technical Commission III Midterm Symposium

5th - 7th September 2014, Zurich, Switzerland

In Conjunction with the European Conference on Computer Vision

Home | People | Program | Submission | Important Dates | ECCV 2014

Important dates

- submission deadline for ISPRS Annals (full paper peer-review) — 13 April 2014
- submission deadline for ISPRS Archives (abstract review) — 19 June 2014

isprs

information from imagery

www.isprs.org/pcv2014