Resolving Stereo Ambiguities using Object Knowledge

Andreas Geiger
Fatma Güney

MPI for Intelligent Systems
Tübingen

Stereo Matching

Task:

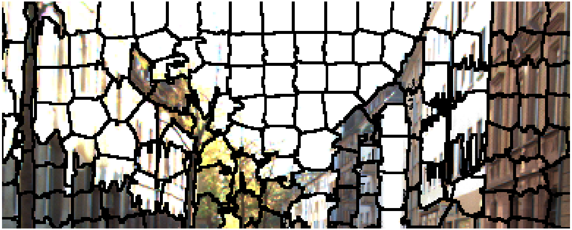
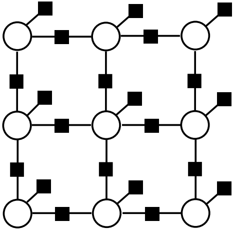
- Given a rectified stereo image pair
- Estimate the disparity (=displacement) at each pixel

Stereo Matching

Task:

- Given a rectified stereo image pair
- Estimate the disparity (=displacement) at each pixel

Stereo Matching

Energy Minimization Problem:

$$E = \sum_{i} E_{data}(d_i) + \sum_{i \sim j} E_{smooth}(d_i, d_j)$$

- d_i: disparity of pixel i or plane parameters of superpixel i
- $i \sim j$: neighbors

Stereo Matching

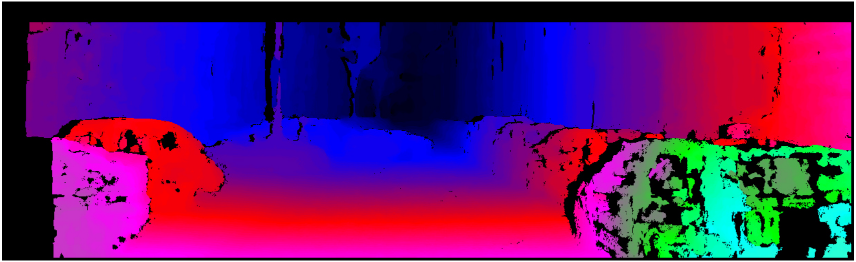
Question:

- Is the stereo problem solved?
- How can we address textureless or specular surfaces?

Key idea:

- Introduce top-down object knowledge into the process!

Stereo Matching



Question:

- Is the stereo problem solved?
- How can we address textureless or specular surfaces?

Key idea:

- Introduce top-down object knowledge into the process!

Stereo Matching

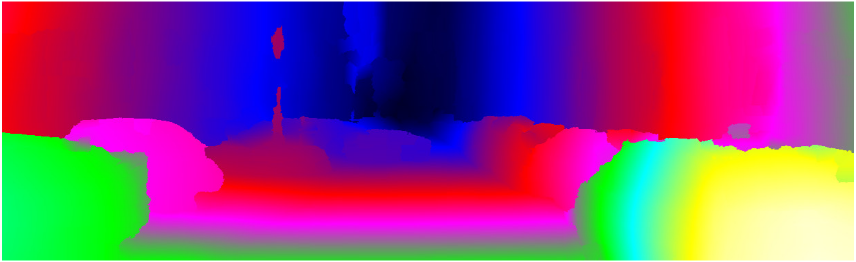
Question:

- Is the stereo problem solved?
- How can we address textureless or specular surfaces?

Key idea:

- Introduce top-down object knowledge into the process!

Stereo Matching



Question:

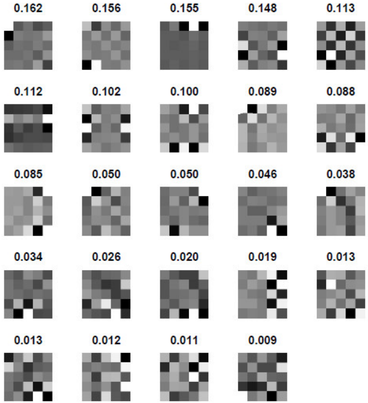
- Is the stereo problem solved?
- How can we address textureless or specular surfaces?

Key idea:

- Introduce top-down object knowledge into the process!

High-order Models in Computer Vision

Image Denoising

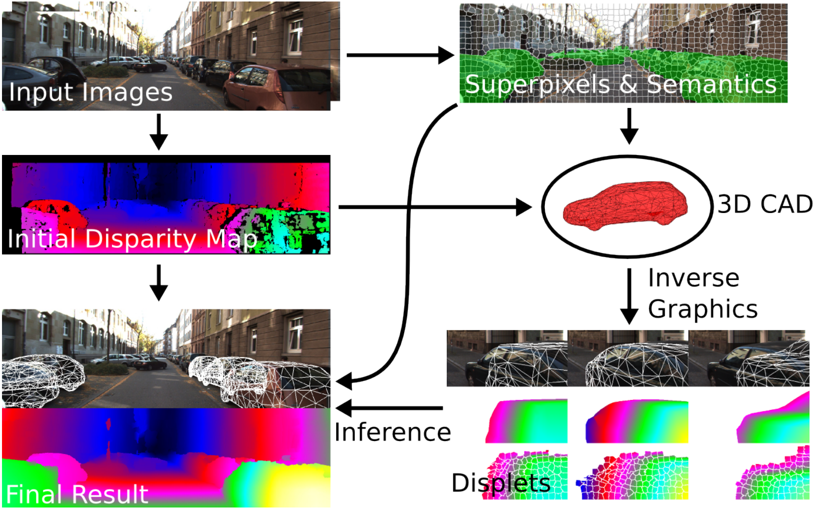


[Roth & Black, 2009]

Image Segmentation

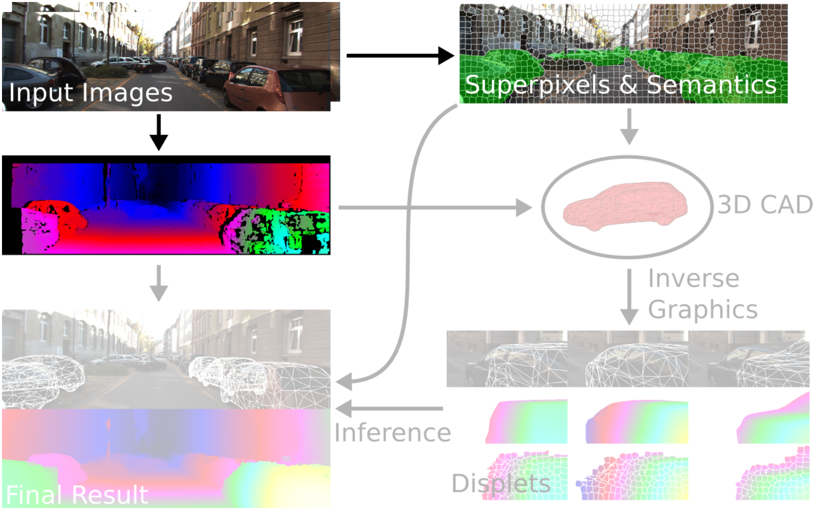
[Kohli, Ladicky & Torr, 2009]

Overview



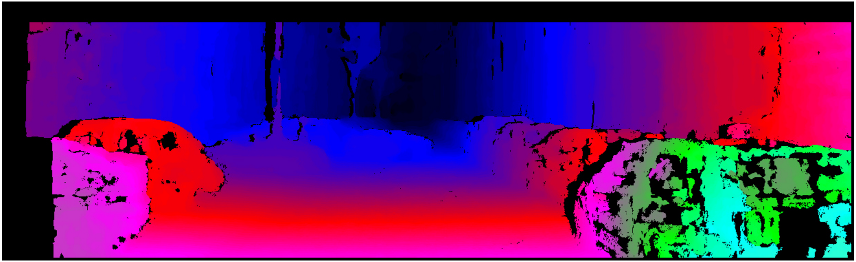
Input Images
Superpixels & Semantics
Initial Disparity Map
3D CAD
Inverse
Graphics
Inference
Displets
Final Result

Overview



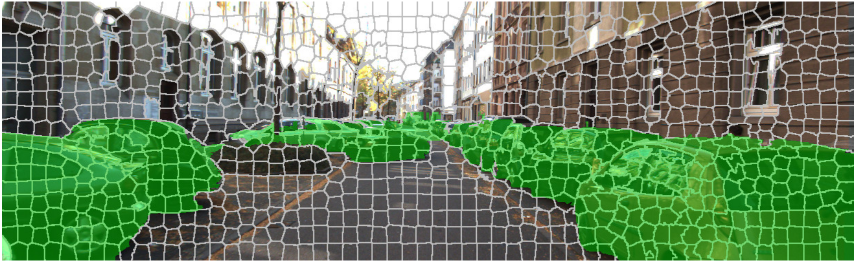
Input Images
Superpixels & Semantics
3D CAD
Inverse
Graphics
Inference
Displets
Final Result

Initial Disparity Map



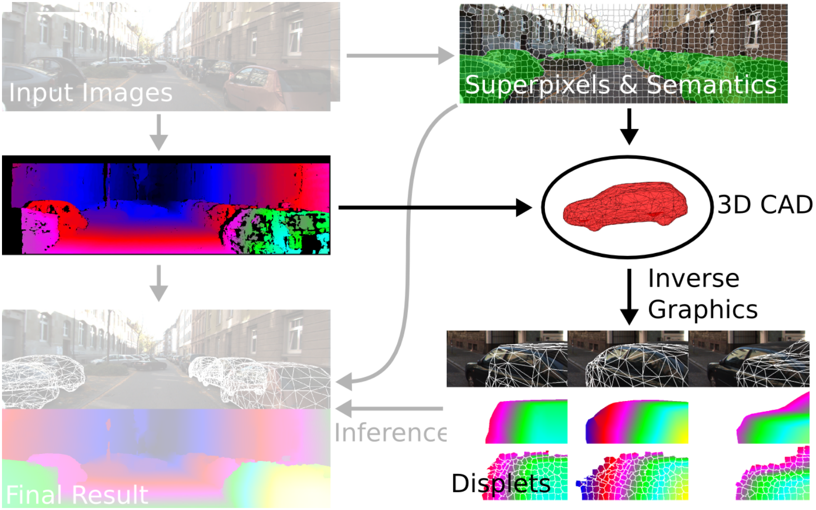
- CNN-based stereo matching cost [Zbontar et al., 2015]
- Semi-global matching [Hirschmüller, 2008]
- Left-right consistency check

Superpixels and Semantic Segmentation

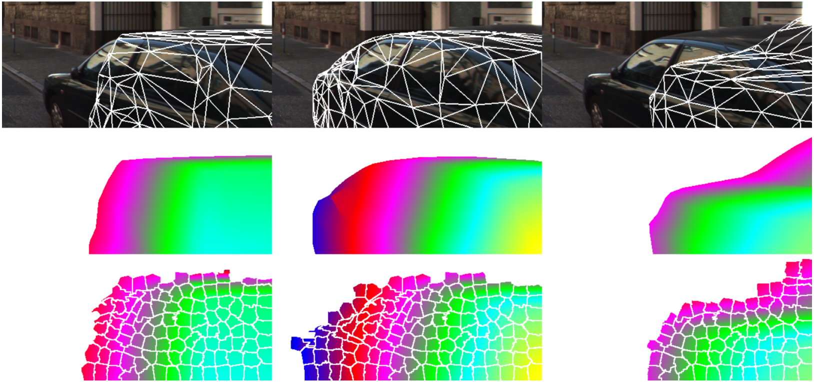


- StereoSLIC superpixels [Yamaguchi et al., 2013]
- Associative Hierarchical Random Fields for vehicle vs. background segmentation [Ladicky et al., 2013]

Overview

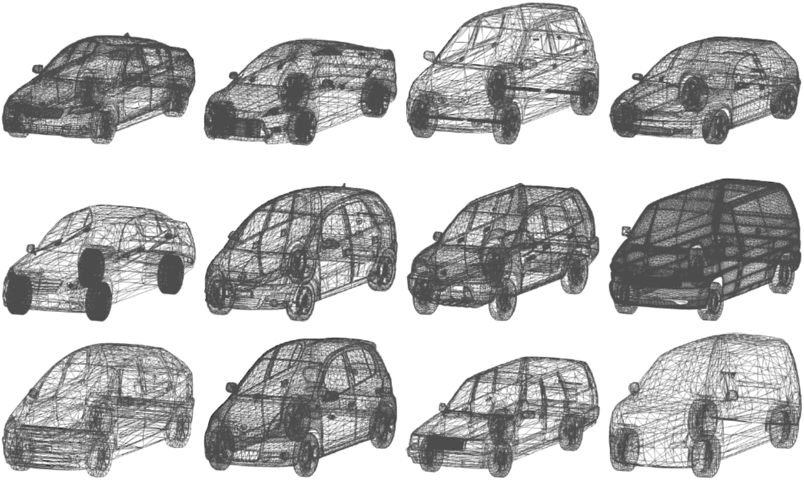

Input Images
Superpixels & Semantics
3D CAD
Inverse
Graphics
Inference
Displets
Final Result

Rapid Inverse Graphics

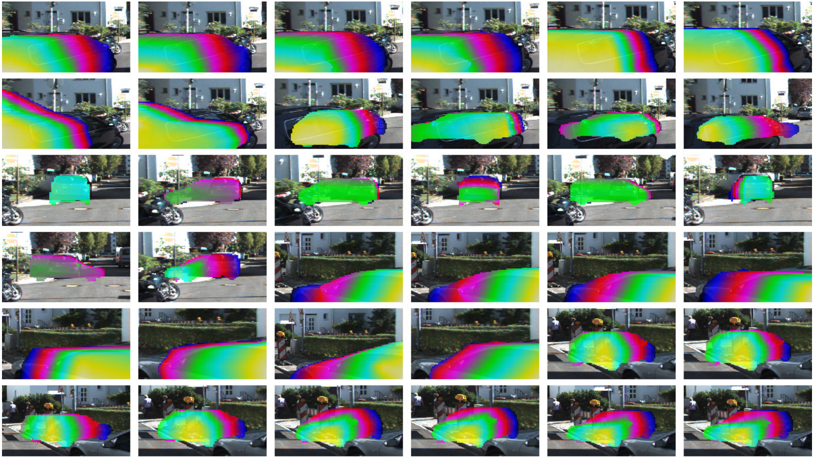


- **Goal:** Generate plausible 3D geometries for a given semantic class
- MCMC simulation based on 3D CAD models ($\sim$8200 fps)

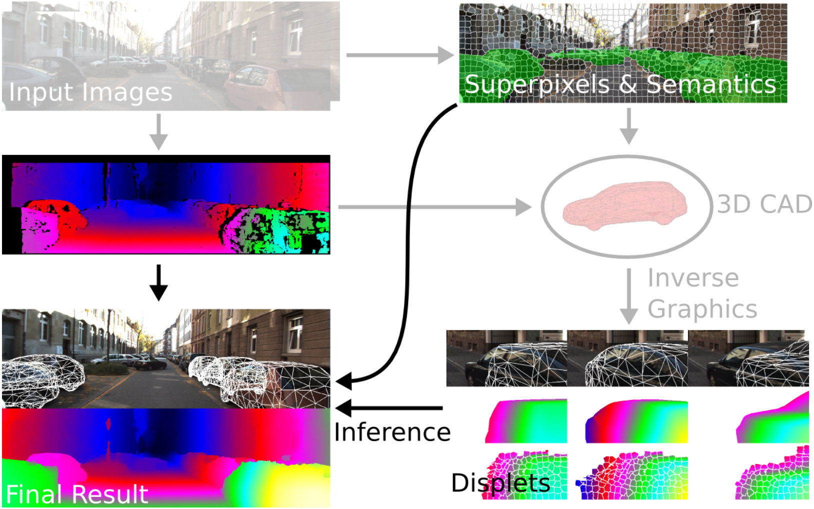
Semi-Convex Hull



Displets



Overview



Input Images
Superpixels & Semantics
3D CAD
Inverse
Graphics
Inference
Displets
Final Result

Inference

Goal: Minimize Gibbs energy ...

$$E = \underbrace{\sum_{i \in \mathcal{S}} \varphi_i^{\mathcal{S}}(\mathbf{n}_i)}_{\text{Data}} + \underbrace{\sum_{i \sim j} \psi_{ij}^{\mathcal{S}}(\mathbf{n}_i, \mathbf{n}_j)}_{\text{Smoothness}} + \underbrace{\sum_{k \in \mathcal{D}} \varphi_k^{\mathcal{D}}(d_k)}_{\text{Displets}} + \underbrace{\sum_{k \in \mathcal{D}} \sum_{i \in \mathcal{S}_k} \psi_{ki}^{\mathcal{D}}(d_k, \mathbf{n}_i)}_{\text{Consistency}}$$

... with respect to superpixel normals $\mathbf{n}_i \in \mathbb{R}^3$ and displets $d_k \in \{0, 1\}$

Notation:

- $\mathcal{S}$: set of all superpixels $\quad \mathcal{D}$: set of all displets
- $\mathcal{S}_k$: set of superpixels covered by displet k

Inference:

- Max-Product Particle Belief Propagation (MP-PBP)

Inference

Goal: Minimize Gibbs energy ...

$$E = \underbrace{\sum_{i \in \mathcal{S}} \varphi_i^{\mathcal{S}}(\mathbf{n}_i)}_{\text{Data}} + \underbrace{\sum_{i \sim j} \psi_{ij}^{\mathcal{S}}(\mathbf{n}_i, \mathbf{n}_j)}_{\text{Smoothness}} + \underbrace{\sum_{k \in \mathcal{D}} \varphi_k^{\mathcal{D}}(d_k)}_{\text{Displets}} + \underbrace{\sum_{k \in \mathcal{D}} \sum_{i \in \mathcal{S}_k} \psi_{ki}^{\mathcal{D}}(d_k, \mathbf{n}_i)}_{\text{Consistency}}$$

... with respect to superpixel normals $\mathbf{n}_i \in \mathbb{R}^3$ and displets $d_k \in \{0, 1\}$

Data Term:

- Penalizes deviation from initial disparity map (truncated ℓ_1)
- Measures photo-consistency between left and right image

Inference

Goal: Minimize Gibbs energy ...

$$E = \underbrace{\sum_{i \in \mathcal{S}} \varphi_i^{\mathcal{S}}(\mathbf{n}_i)}_{\text{Data}} + \underbrace{\sum_{i \sim j} \psi_{ij}^{\mathcal{S}}(\mathbf{n}_i, \mathbf{n}_j)}_{\text{Smoothness}} + \underbrace{\sum_{k \in \mathcal{D}} \varphi_k^{\mathcal{D}}(d_k)}_{\text{Displets}} + \underbrace{\sum_{k \in \mathcal{D}} \sum_{i \in \mathcal{S}_k} \psi_{ki}^{\mathcal{D}}(d_k, \mathbf{n}_i)}_{\text{Consistency}}$$

... with respect to superpixel normals $\mathbf{n}_i \in \mathbb{R}^3$ and displets $d_k \in \{0, 1\}$

Smoothness Term:

- Penalizes depth discontinuities at superpixel boundaries
- Encourages similar orientations of neighboring superpixel planes

Inference

Goal: Minimize Gibbs energy ...

$$E = \underbrace{\sum_{i\in\mathcal{S}} \varphi_i^{\mathcal{S}}(\mathbf{n}_i)}_{\text{Data}} + \underbrace{\sum_{i\sim j} \psi_{ij}^{\mathcal{S}}(\mathbf{n}_i, \mathbf{n}_j)}_{\text{Smoothness}} + \underbrace{\sum_{k\in\mathcal{D}} \varphi_k^{\mathcal{D}}(d_k)}_{\text{Displets}} + \underbrace{\sum_{k\in\mathcal{D}} \sum_{i\in\mathcal{S}_k} \psi_{ki}^{\mathcal{D}}(d_k, \mathbf{n}_i)}_{\text{Consistency}}$$

... with respect to superpixel normals $\mathbf{n}_i \in \mathbb{R}^3$ and displets $d_k \in \{0, 1\}$

Displet Unary Term:

- Encourages image regions with semantic class label c_k to be explained by a displet of the corresponding class:

$$\varphi_k^{\mathcal{D}}(d_k) = -\theta_3 \, [d_k = 1] \cdot (|[\mathbf{S} = c_k] \cap \mathbf{M}_k| + \kappa_k)$$

- $\mathbf{M}_k$: foreground mask of displet k $\mathbf{S}$: semantic segmentation

Inference

Goal: Minimize Gibbs energy ...

$$E = \underbrace{\sum_{i \in \mathcal{S}} \varphi_i^{\mathcal{S}}(\mathbf{n}_i)}_{\text{Data}} + \underbrace{\sum_{i \sim j} \psi_{ij}^{\mathcal{S}}(\mathbf{n}_i, \mathbf{n}_j)}_{\text{Smoothness}} + \underbrace{\sum_{k \in \mathcal{D}} \varphi_k^{\mathcal{D}}(d_k)}_{\text{Displets}} + \underbrace{\sum_{k \in \mathcal{D}} \sum_{i \in \mathcal{S}_k} \psi_{ki}^{\mathcal{D}}(d_k, \mathbf{n}_i)}_{\text{Consistency}}$$

... with respect to superpixel normals $\mathbf{n}_i \in \mathbb{R}^3$ and displets $d_k \in \{0, 1\}$

Displet Consistency Term:

- Encourages superpixels to agree with the geometry of the displets:

$$\psi_{ki}^{\mathcal{D}}(d_k, \mathbf{n}_i) = \lambda_{ki} \, [d_k = 1] \cdot (1 - \delta(\mathbf{n}_i, \hat{\mathbf{n}}_{k,z_i}))$$

- λ_{ki}: weight depending on distance to boundary:

Results on KITTI Validation Set (Ablation)

	Reflective Regions	All Regions
Input $\hat{\Omega}$ (Interpolated)	19.84 %	3.35 %
Unary Only	17.72 %	3.31 %
Unary + Pair (Boundary)	15.96 %	3.21 %
Unary + Pair (Normal)	17.06 %	3.28 %
Unary + Pair	14.78 %	3.12 %
Unary + Pair + Occ	15.32 %	3.04 %
Unary + Pair + Disp	7.08 %	2.87 %
Unary + Pair + Occ + Disp	7.16 %	2.78 %

- Metric: Percentage of erroneous pixels ($>$ 3px disparity error)

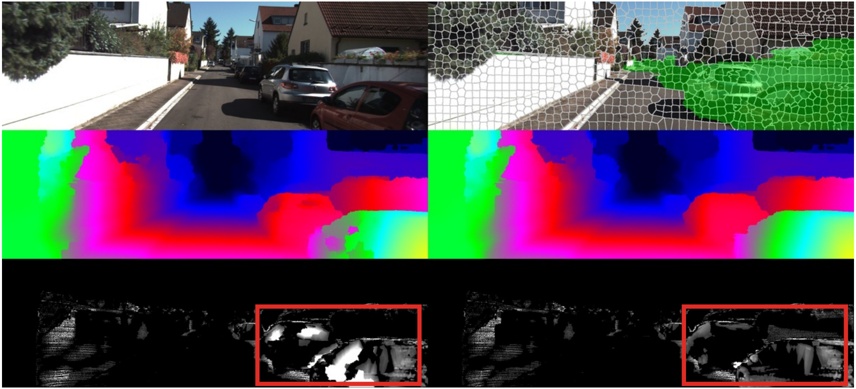
Results on KITTI Test Set (Reflective Regions)

Rank	Method	Out-Noc	Out-All	Avg-Noc	Avg-All
1	Our Method	**8.40 %**	**9.89 %**	**1.9 px**	**2.3 px**
2	VC-SF *	11.58 %	12.29 %	2.7 px	2.8 px
3	PCBP-SS	14.26 %	18.33 %	2.4 px	3.9 px
4	SPS-StFl *	14.74 %	18.00 %	2.9 px	3.6 px
5	CoP	15.30 %	19.15 %	2.7 px	4.1 px
6	SPS-St	16.05 %	19.34 %	3.1 px	3.6 px
7	DDS-SS	16.23 %	19.39 %	2.5 px	3.0 px
8	PCBP	16.28 %	20.22 %	2.8 px	4.4 px
9	PR-Sf+E *	17.85 %	20.82 %	3.3 px	4.0 px
10	StereoSLIC	18.22 %	21.60 %	2.8 px	3.6 px
11	MC-CNN	18.45 %	21.96 %	3.5 px	4.3 px
12	PR-Sceneflow *	19.22 %	22.07 %	3.3 px	4.0 px
⋮	⋮	⋮	⋮	⋮	
62	ALE-Stereo	83.80 %	84.37 %	24.6 px	25.4 px

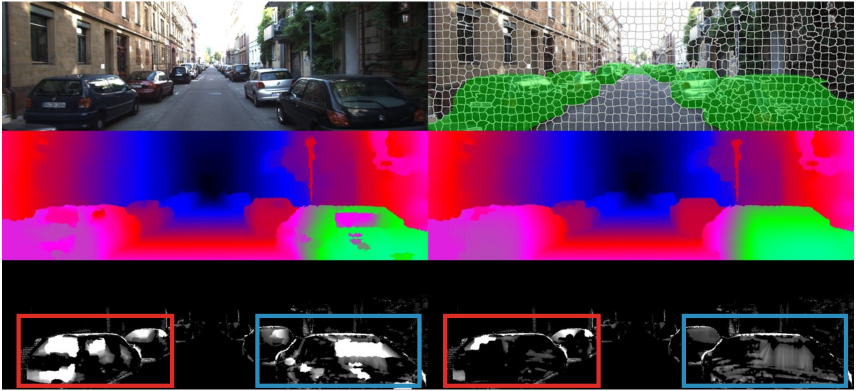
Results on KITTI Test Set (All Regions)

Rank	Method	Out-Noc	Out-All	Avg-Noc	Avg-All
1	Our Method	**2.47 %**	**3.27 %**	**0.7 px**	**0.9 px**
2	MC-CNN	2.61 %	3.84 %	0.8 px	1.0 px
3	SPS-StFl *	2.83 %	3.64 %	0.8 px	**0.9 px**
4	VC-SF *	3.05 %	3.31 %	0.8 px	0.8 px
5	SPS-St	3.39 %	4.41 %	0.9 px	1.0 px
6	PCBP-SS	3.40 %	4.72 %	0.8 px	1.0 px
7	CoP	3.78 %	4.63 %	0.9 px	1.1 px
8	DDS-SS	3.83 %	4.59 %	0.9 px	1.0 px
9	StereoSLIC	3.92 %	5.11 %	0.9 px	1.0 px
10	PR-Sf+E *	4.02 %	4.87 %	0.9 px	1.0 px
11	PCBP	4.04 %	5.37 %	0.9 px	1.1 px
12	PR-Sceneflow *	4.36 %	5.22 %	0.9 px	1.1 px
⋮	⋮	⋮	⋮	⋮	⋮
62	ALE-Stereo	50.48 %	51.19 %	13.0 px	13.5 px

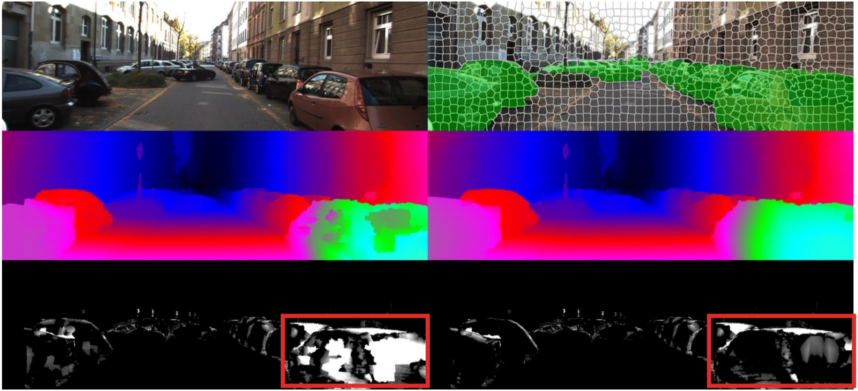
Qualitative Results: Success Cases



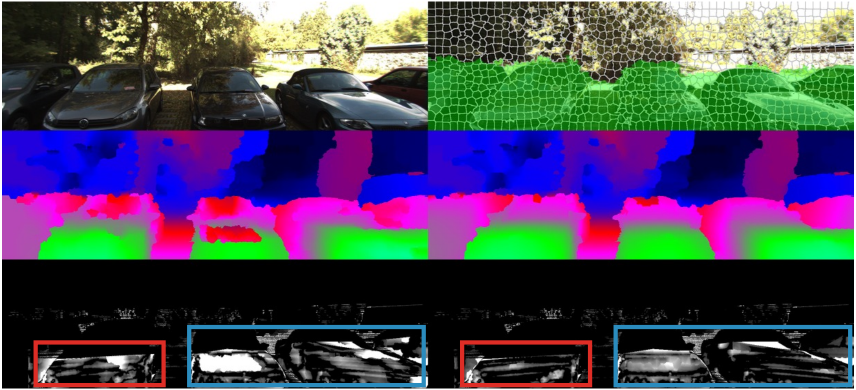
Qualitative Results: Success Cases



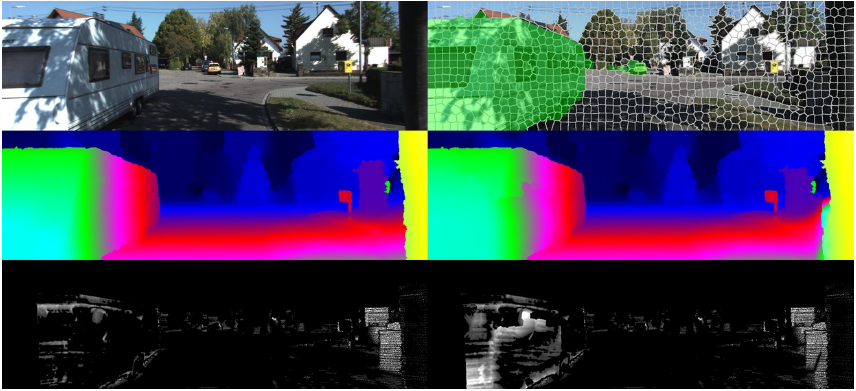
Qualitative Results: Success Cases



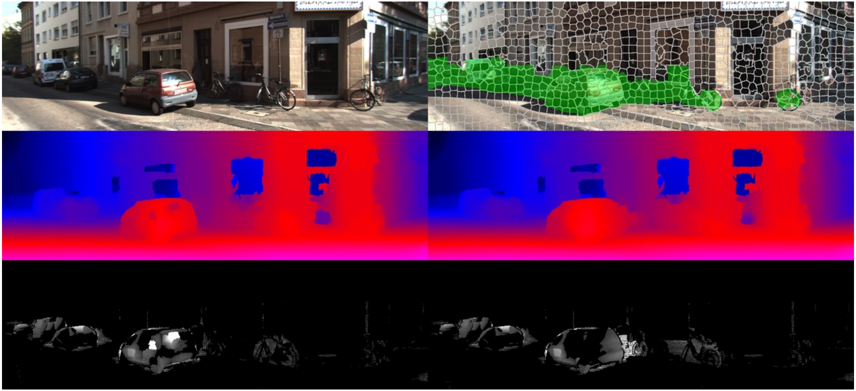
Qualitative Results: Success Cases



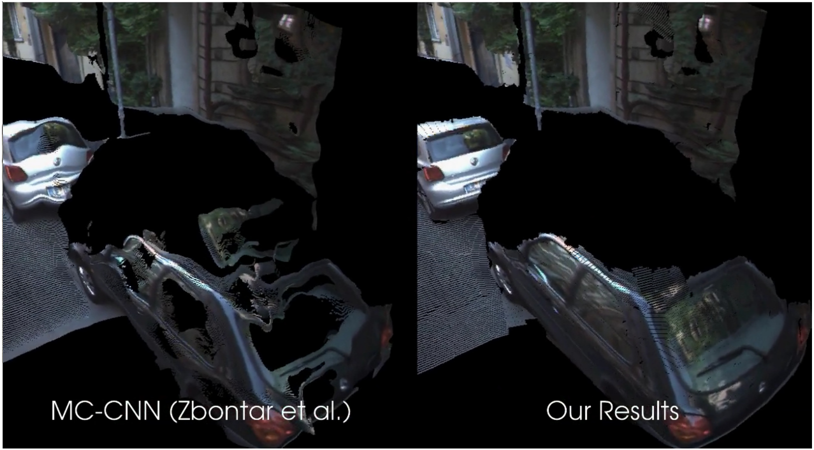
Qualitative Results: Failure Cases



Qualitative Results: Failure Cases



Qualitative Results: Video



Thank you!