

Super-Resolution Techniques for 3D-Reconstruction

Björn Häfner

Technische Universität München
Department of Informatics
Computer Vision Group

13. December 2017

Outline

- 1 Introduction
- 2 Background
- 3 Depth Super-Resolution Meets Uncalibrated Photometric Stereo
 - Methodology
 - Experiments
- 4 Single-Shot Variational Depth Super-Resolution from Shading
 - Methodology
 - Experiments

Outline

- 1 Introduction
- 2 Background
- 3 Depth Super-Resolution Meets Uncalibrated Photometric Stereo
 - Methodology
 - Experiments
- 4 Single-Shot Variational Depth Super-Resolution from Shading
 - Methodology
 - Experiments



Problem Statement

Example: RGB-D data from Kinect



Input RGB image



Depth image



3D shape

- Missing areas
- Noisy and quantization effect
- No fine details
- Lower resolution as RGB image



Goal: Improve depth + increase resolution to RGB image



Input RGB image



Input depth



Refined depth



Goal: Improve depth + increase resolution to RGB image



Input RGB image



Input depth



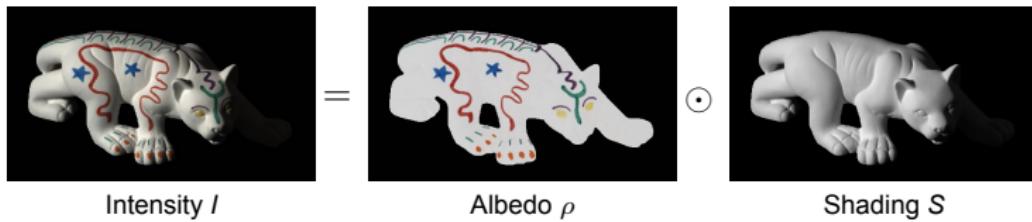
Refined depth

- Filled missing areas
- Removed noisy and quantization effects
- Recovered fine details
- Same resolution as RGB image

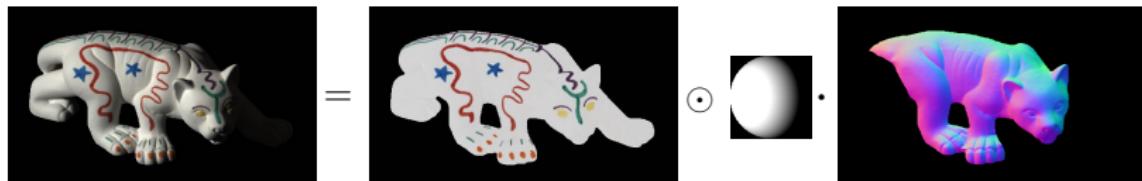
Outline

- 1 Introduction
- 2 Background
- 3 Depth Super-Resolution Meets Uncalibrated Photometric Stereo
 - Methodology
 - Experiments
- 4 Single-Shot Variational Depth Super-Resolution from Shading
 - Methodology
 - Experiments

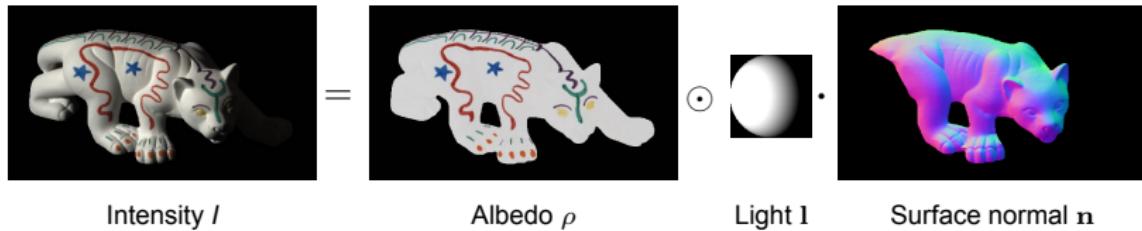
Lambert's Law



Lambert's Law

$$\text{Intensity } I = \text{Albedo } \rho \odot \text{Light 1} \cdot \text{Surface normal } \mathbf{n}$$


Lambert's Law



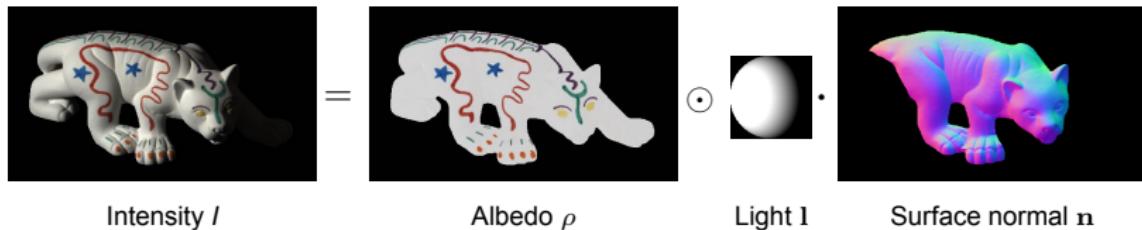
$$I = \rho \mathbf{l}^t \mathbf{n}$$

$$\mathbf{l} = \begin{pmatrix} l^x \\ l^y \\ l^z \end{pmatrix}$$

$$\mathbf{n} = \frac{1}{\sqrt{|\nabla z|^2 + 1}} \begin{pmatrix} \nabla z \\ -1 \end{pmatrix}$$

\mathbf{n} is **nonlinear** w.r.t. z

Lambert's Law



$$I = \rho \mathbf{l}^t \mathbf{n}$$

$$I = \rho (\mathbf{l}^t \mathbf{n} + \varphi) = \rho \tilde{\mathbf{n}} \mathbf{s}$$

$$\mathbf{l} = \begin{pmatrix} l^x \\ l^y \\ l^z \end{pmatrix}$$

$$\mathbf{n} = \frac{1}{\sqrt{|\nabla z|^2 + 1}} \begin{pmatrix} \nabla z \\ -1 \end{pmatrix}$$

\mathbf{n} is **nonlinear** w.r.t. z

φ : ambient light parameter

$$\mathbf{s} = \begin{pmatrix} 1 \\ \varphi \end{pmatrix} \quad \tilde{\mathbf{n}} = \begin{pmatrix} \mathbf{n} \\ 1 \end{pmatrix}^t$$

Spherical Harmonics (SH) model accounts for **87.5%** real-world lights

Super-Resolution Model

ASUS Xtion Pro Live (same as Kinect v1) provides resolution:

- 1280×960 RGB image
 - 640×480 depth image z_0
- ⇒ Scale factor $\xi = 2$

$$z_0 = Kz$$

$K \in \mathbb{R}^{m \times \xi^2 m}$ is a linear operator downsampling z to the same size of z_0 .
 m number of pixels; in our case $m = 640 \cdot 480$

Ugly world

Above setups are true in a perfect world scenario. Real-world data introduces noise, i.e. add noise to

Lambert's model:

$$l^i = \rho \mathbf{s}^i \tilde{\mathbf{n}} + \varepsilon_l^i, \quad \forall i \in \{1, \dots, n\},$$

and super-resolution model:

$$\mathbf{z}_0^i = K\mathbf{z} + \varepsilon_{\mathbf{z}}^i, \quad \forall i \in \{1, \dots, n\},$$

where $\varepsilon_l^i \sim \mathcal{N}(0, \sigma_l^2)$ and $\varepsilon_{\mathbf{z}}^i \sim \mathcal{N}(0, \sigma_{\mathbf{z}}^2)$

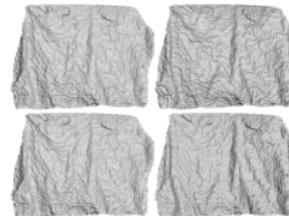
Outline

- 1 Introduction
- 2 Background
- 3 Depth Super-Resolution Meets Uncalibrated Photometric Stereo
 - Methodology
 - Experiments
- 4 Single-Shot Variational Depth Super-Resolution from Shading
 - Methodology
 - Experiments



Setup

- Position an RGB-D device, e.g Kinect or ASUS Xtion Pro Live in front of object
- From same viewpoint capture n RGB-D images under different lighting conditions
- Depth images z_0^i must be registered with RGB images I^i ; same viewpoint condition.
- Provide intrinsic parameters



n LR depth maps of depth sensor

n SR RGBs w/ different lighting

Variational framework

Use Lambert's law and normal representation to arrive at a non-linear PDE in \mathbf{z} , ρ , $\{\mathbf{s}^i\}_{i \in \{1, \dots, n\}}$, i.e. plug

$$\mathbf{n}(\mathbf{p}) = \frac{1}{d(\mathbf{z})(\mathbf{p})} \begin{pmatrix} f \nabla \mathbf{z}(\mathbf{p}) \\ -\mathbf{z}(\mathbf{p}) - \nabla \mathbf{z}(\mathbf{p}) \cdot (\mathbf{p} - \mathbf{p}^0) \end{pmatrix}$$

in

$$l^i(\mathbf{p}) = \rho(\mathbf{p}) \mathbf{s}^i \cdot \begin{pmatrix} \mathbf{n}(\mathbf{p}) \\ 1 \end{pmatrix} + \varepsilon_l^i, \quad \forall i \in \{1, \dots, n\}$$

to get

$$\mathbf{A}^i(\mathbf{z}, \rho, \mathbf{s}^i) \begin{pmatrix} \nabla \mathbf{z} \\ \mathbf{z} \end{pmatrix} = \mathbf{b}^i(\rho, \mathbf{s}^i) + \varepsilon_l^i, \quad \forall i \in \{1, \dots, n\}.$$

$d(\mathbf{z})(\mathbf{p})$ is normalizing constant at pixel \mathbf{p} ; f is focal length; \mathbf{p}^0 is principal point; $\mathbf{A}^i(\mathbf{z}, \rho, \mathbf{s}^i) : \Omega_{HR} \mapsto \mathbb{R}^{1 \times 3}$; $\mathbf{b}^i(\rho, \mathbf{s}^i) : \Omega_{HR} \mapsto \mathbb{R}$, with Ω_{HR} being the high resolution image domain.

Variational framework

Plugging this together with the standard super-resolution minimization approach we end up minimizing

$$\min_{\substack{\mathbf{z}: \Omega_{HR} \rightarrow \mathbb{R} \\ \rho: \Omega_{HR} \rightarrow \mathbb{R} \\ \{\mathbf{s}^i \in \mathbb{R}^4\}_i}} \sum_{i=1}^n \left\{ \lambda \left\| \mathbf{A}^i(\mathbf{z}, \rho, \mathbf{s}^i) \begin{bmatrix} \nabla \mathbf{z} \\ \mathbf{z} \end{bmatrix} - \mathbf{b}^i(\rho, \mathbf{s}^i) \right\|_{\mathcal{L}^2(\Omega_{HR})}^2 + \left\| K\mathbf{z} - \mathbf{z}_0^i \right\|_{\mathcal{L}^2(\Omega_{LR})}^2 \right\},$$

where λ is a trade-off parameter and $\|\cdot\|_{\mathcal{L}^2(\Omega)}^2$ is the \mathcal{L}^2 -norm over the image domain Ω .

Minimization is done in an alternating scheme.

Workflow

Input



⋮



⋮



Optimization

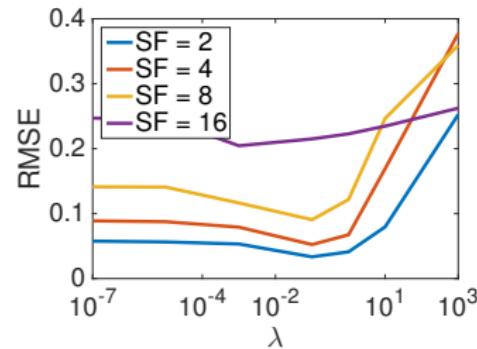
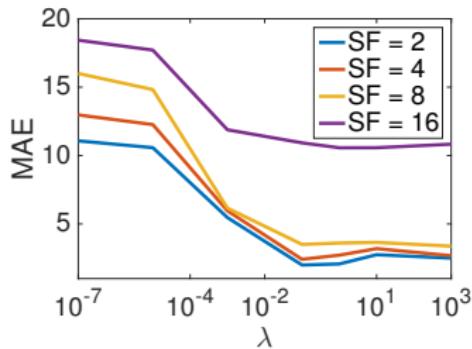
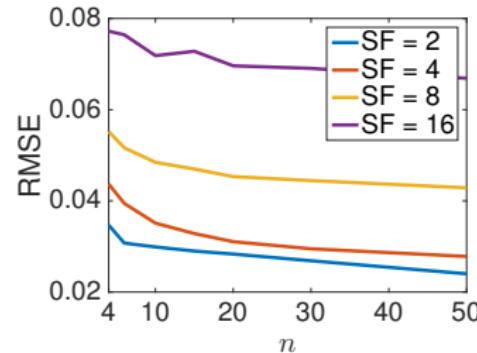
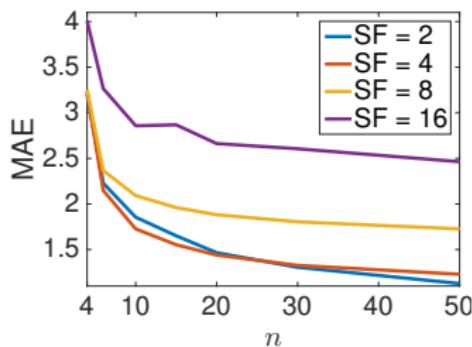


Output

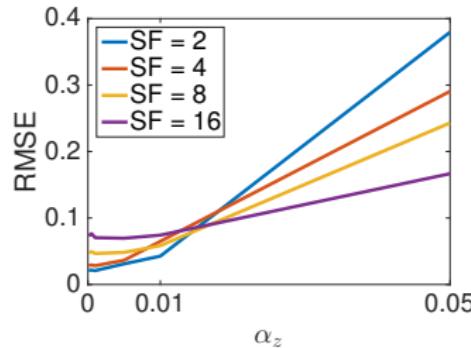
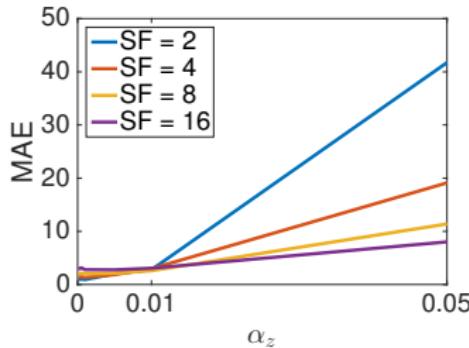
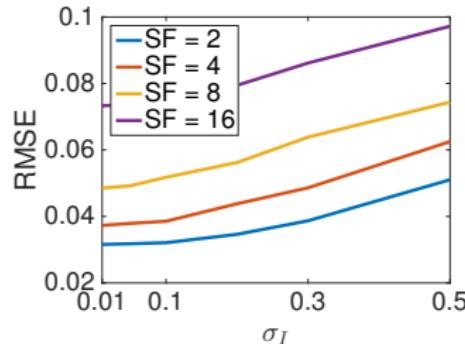
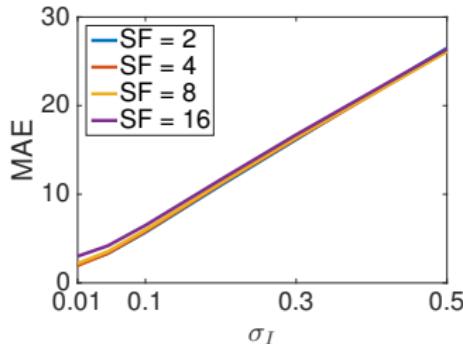


Synthetic Evaluation

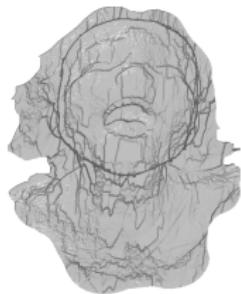


error vs. n & error vs. λ 

Error vs. noise



Comparison with other methods



RMSE = 0.1237

MAE = 38.9402

(a) ID-SR



RMSE = 0.9199

MAE = 41.8041

(b) Papadimitri
& Favaro [22]



RMSE = 0.1655

MAE = 38.9316

(c) Or-El et al [21]



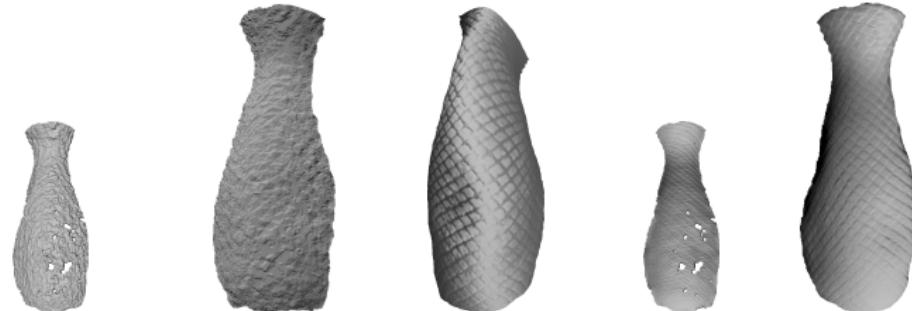
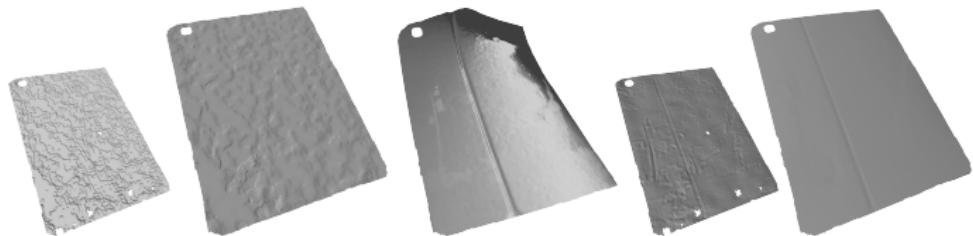
RMSE = **0.03139**

MAE = **1.4528**

(d) ours

Real-world data

Upscaling by factor of $\xi = 2$ using $n = 20$



(a)RGB input

(b)depth input

(c)ID-SR

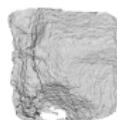
(d) [22]

(e) [21]

(f) ours

Real-world data

Upscaling by factor of $\xi = 4$ using $n = 20$



(a)RGB input

(b)depth input

(c) z

(d) ρ

(e)new pose

Outline

- 1 Introduction
- 2 Background
- 3 Depth Super-Resolution Meets Uncalibrated Photometric Stereo
 - Methodology
 - Experiments
- 4 Single-Shot Variational Depth Super-Resolution from Shading
 - Methodology
 - Experiments

Shape from Shading

Shape from shading aims at inferring shape from a single image

$$\mathbf{z}^* = \operatorname{argmin}_{\mathbf{z}} \|I - \mathcal{R}(\mathbf{z}|\mathbf{l}, \rho)\|_2^2,$$

where \mathcal{R} describes the radiance. Assuming frontal-lighting, uniform Lambertian reflectance, Lipschitz-continuous depth and orthographic projection yields the ambiguous solution

$$|\nabla \mathbf{z}| = \sqrt{\frac{1}{P^2} - 1}.$$

⇒ Under more realistic assumptions one might expect even more ambiguities.

Setup

Assumptions:

- Twice differentiable depth map z .
- Piecewise constant Lambertian reflectance (albedo).

Input:

- One low-resolution depth map z
- and its corresponding RGB image with shading information.



Variational Model

Under perspective projection, we get for the normal

$$\mathbf{n}_z = \frac{1}{\sqrt{|f\nabla z|^2 + (-z - p \cdot \nabla z)^2}} \begin{bmatrix} f\nabla z \\ -z - p \cdot \nabla z \end{bmatrix},$$

where $f > 0$ is focal length and p the pixel w.r.t. the principal point.
We define \mathbf{m} as

$$\mathbf{m}_{a,b} := \left[\frac{fb^t}{dA_{a,b}}, \frac{-a - p \cdot b}{dA_{a,b}}, 1 \right]^t,$$

with

$$dA_{a,b} = \sqrt{|fb|^2 + (-a - p \cdot b)^2}$$

being the area of the corresponding surface element. The piecewise constant albedo assumption can be achieved using the ℓ^0 -norm

$$\|\nabla \rho\|_0 = \sum_{p \in \Omega} \begin{cases} 0, & \text{if } |\nabla \rho(p)|_2 = 0, \\ 1, & \text{otherwise.} \end{cases}$$

Variational Model

$$\min_{\rho, \mathbf{l}, \mathbf{z}} \gamma \|(\mathbf{l} \cdot \mathbf{m}_{\mathbf{z}, \nabla \mathbf{z}}) \rho - \mathbf{l}\|_2^2 + \mu \|K\mathbf{z} - \mathbf{z}_0\|_2^2 + \nu \|\mathbf{d}\mathcal{A}_{\mathbf{z}, \nabla \mathbf{z}}\|_1 + \lambda \|\nabla \rho\|_0$$

solved in an alternating manner using a multi-block ADMM strategy with trade-off parameters γ , μ , ν and λ .

Ambiguity visualization:



Input depth



RGB image



$\gamma = 0$

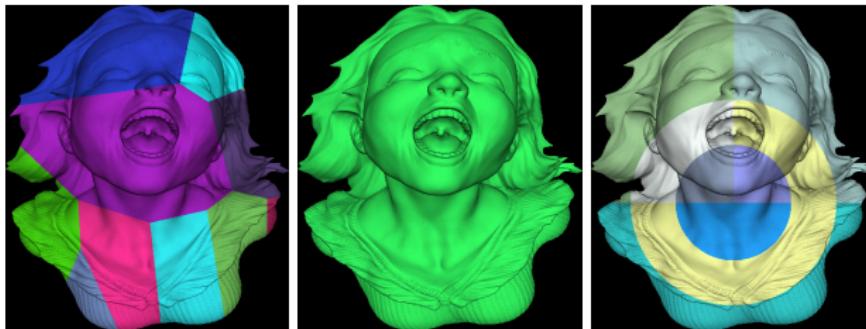


$\mu = 0$



$\gamma > 0$
 $\mu > 0$

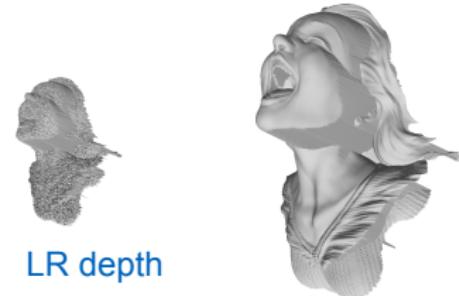
Synthetic Data



Voronoi

Constant

Rectcircle

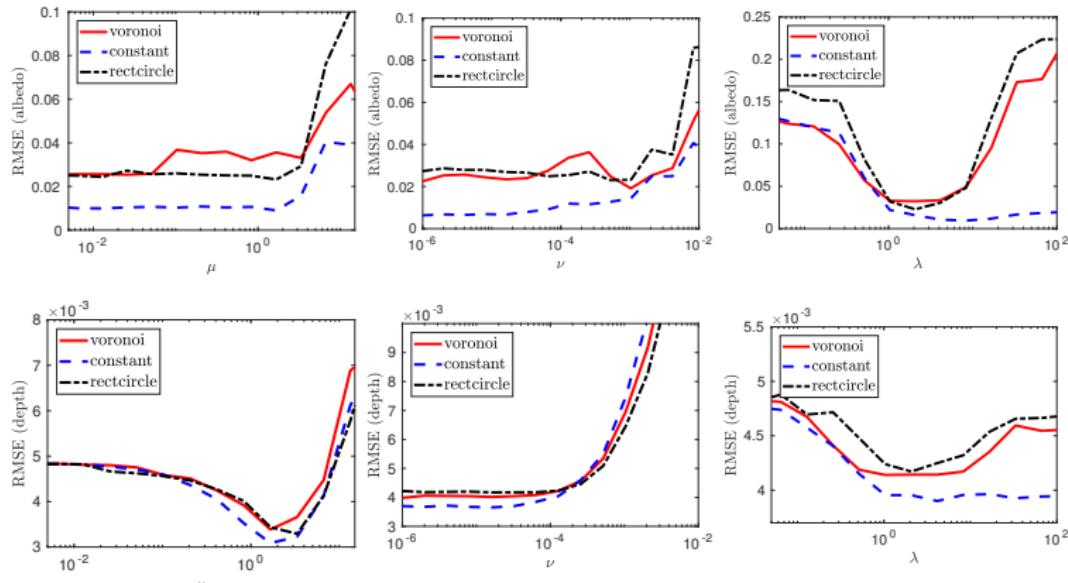


LR depth

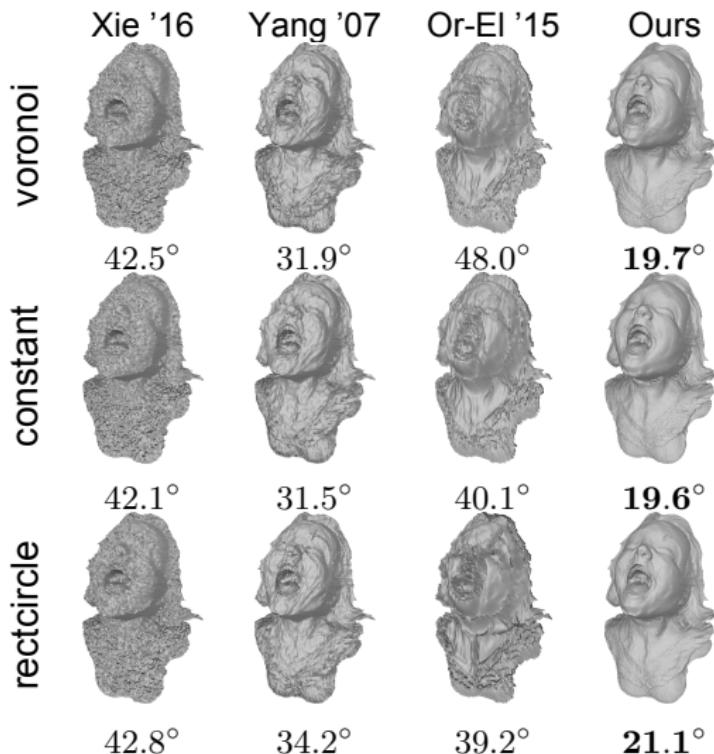
GT depth

Parameter Estimation

$$\min_{\rho, \mathbf{l}, \mathbf{z}} \|(\mathbf{l} \cdot \mathbf{m}_{\mathbf{z}, \nabla \mathbf{z}}) \rho - \mathbf{l}\|_2^2 + \mu \|\mathbf{Kz} - \mathbf{z}_0\|_2^2 + \nu \|\mathbf{dA}_{\mathbf{z}, \nabla \mathbf{z}}\|_1 + \lambda \|\nabla \rho\|_0$$



Quantitative Comparison with other Methods





Qualitative Evaluation



Qualitative Evaluation



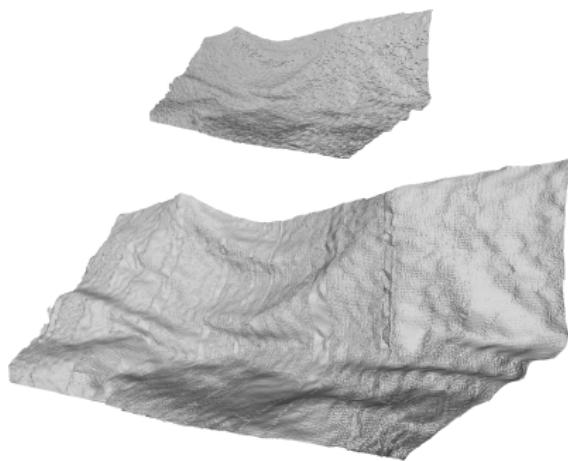


Qualitative Evaluation





Qualitative Evaluation





Qualitative Comparison with other methods

RGB



Input depth



Xie '16



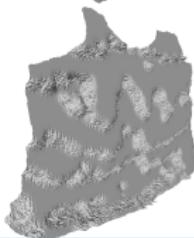
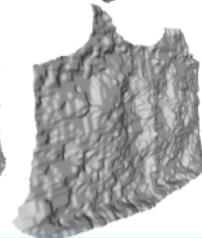
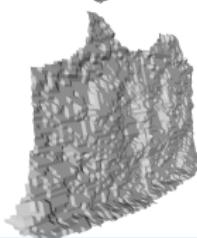
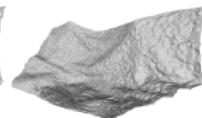
Yang '07



Or-El '15

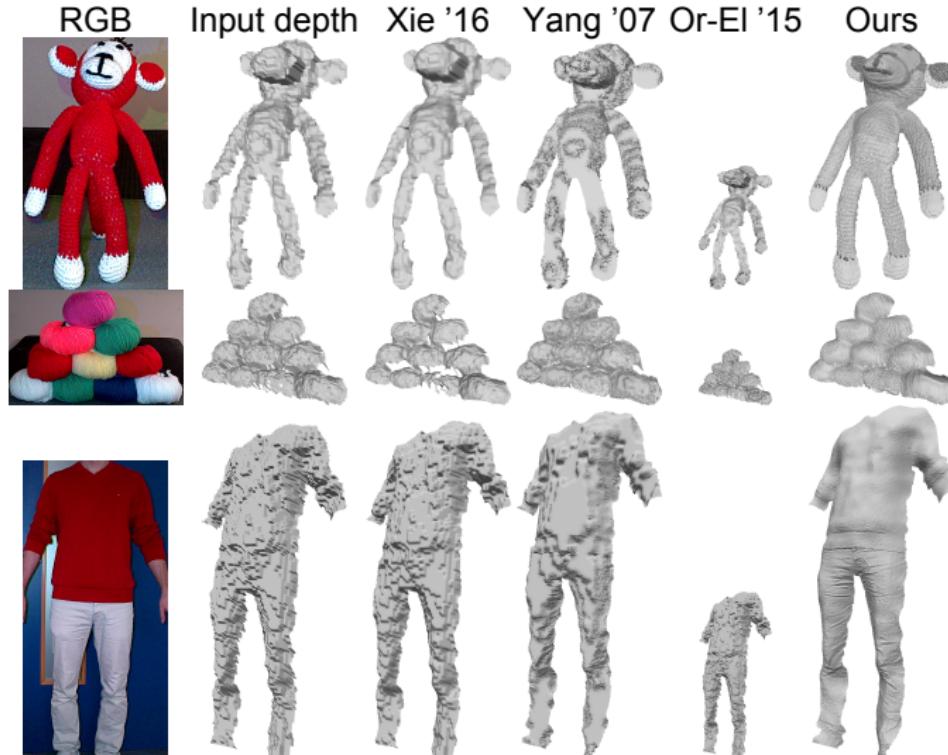


Ours



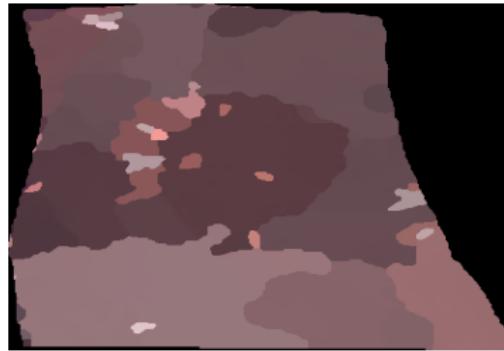
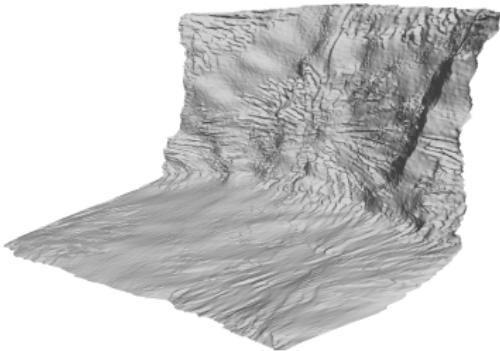
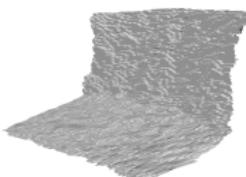


Qualitative Comparison with other methods





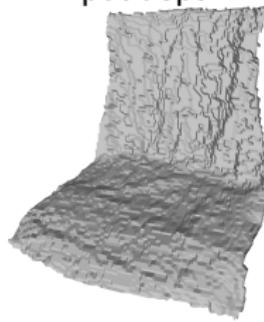
Failure Case



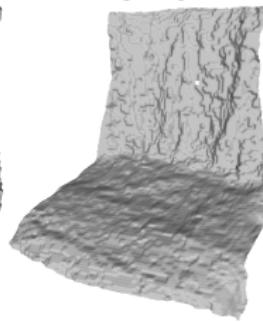


Failure Case with other methods

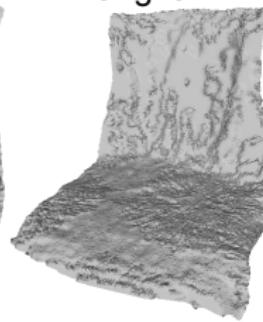
Input depth



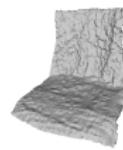
Xie '16



Yang '07



Or-El '15



Ours

