

Uncalibrated Photometric Stereo Meets Depth Super-Resolution

S. Peng, B. Häfner, Y. Queau, D. Cremers

ICCV Workshop Submission

July 18, 2017

Outline

1 Introduction

2 Background

3 Methodology

4 Evaluation

Outline

1 Introduction

2 Background

3 Methodology

4 Evaluation

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

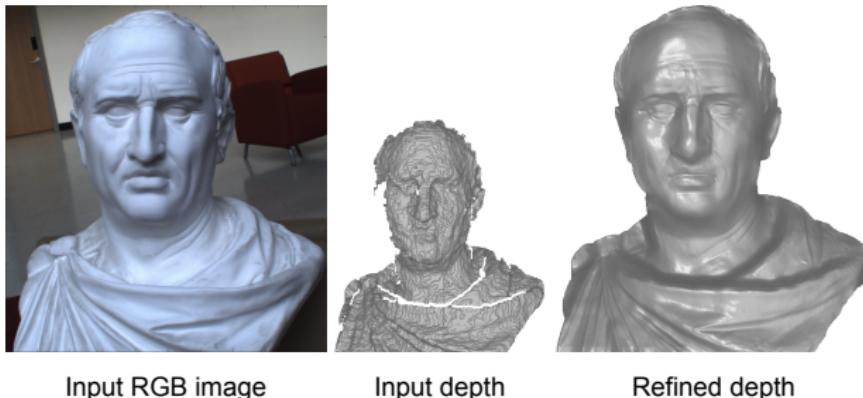


Input depth



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 Methodology

4 Evaluation

Lambert's Law

$$\text{Intensity } I = \text{Albedo } \rho \times \text{Shading } S$$

The diagram illustrates the components of Lambert's Law for a 3D model of a white tiger. On the left, the 'Intensity I ' is shown as a black image with red outlines and blue stars indicating highlights. In the center, the 'Albedo ρ ' is shown as a black image with red outlines and yellow dots indicating shading. On the right, the 'Shading S ' is shown as a grayscale image of the tiger's body. The equation shows the intensity being the product of albedo and shading.

Lambert's Law

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

The diagram illustrates the components of Lambert's Law. On the left, a white tiger model is shown with various colored lines and stars overlaid, labeled "Intensity I ". To its right is an equals sign. Next is the "Albedo ρ " component, which shows the same tiger model with a grayscale texture overlay. To the right of the equals sign is a circular light source labeled "Light 1". To the right of the light source is a vector labeled "Surface normal \mathbf{n} ", represented by a tiger model with a color gradient from purple to yellow.

Lambert's Law

Intensity I Albedo ρ 

Light 1

Surface normal \mathbf{n}

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

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

φ : ambient light parameter

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

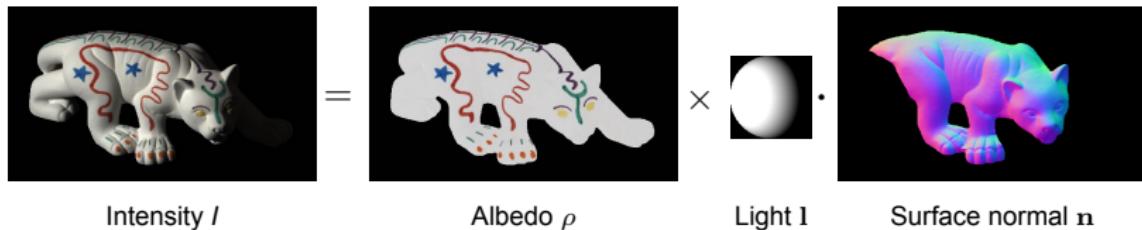
$$\mathbf{n} = \frac{1}{\sqrt{|\nabla z|^2 + 1}} \begin{pmatrix} \nabla z \\ -1 \end{pmatrix}$$

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

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

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

Lambert's Law



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

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

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

n is nonlinear w.r.t. z

φ : ambient light parameter

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

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:

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

and super-resolution model:

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

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

Outline

1 Introduction

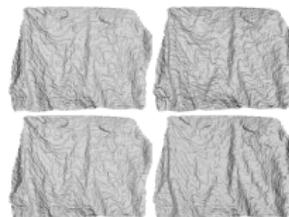
2 Background

3 Methodology

4 Evaluation

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

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

to get

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

$d(\mathbf{z})(\mathbf{p})$ 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



Outline

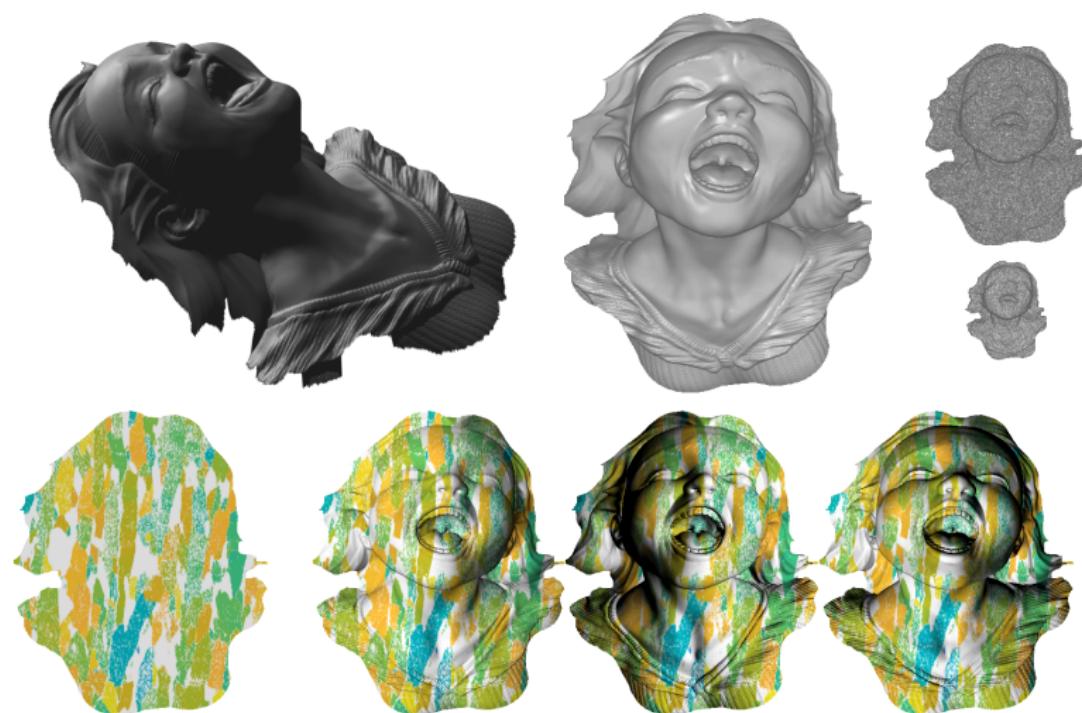
1 Introduction

2 Background

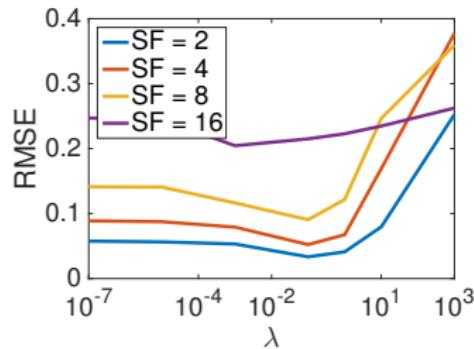
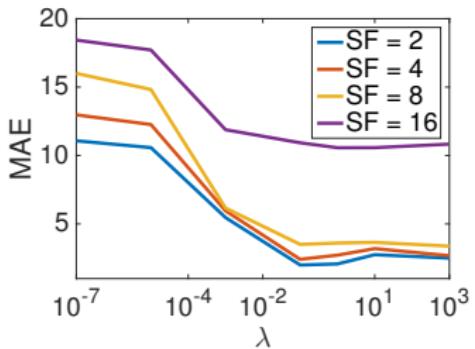
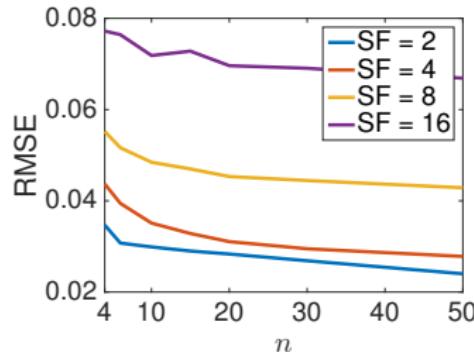
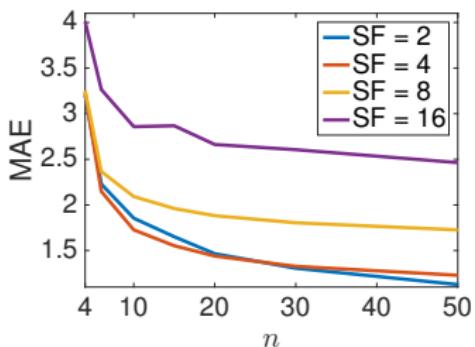
3 Methodology

4 Evaluation

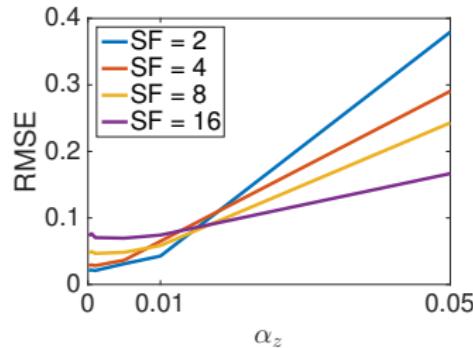
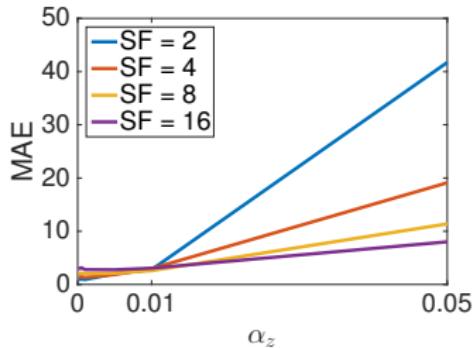
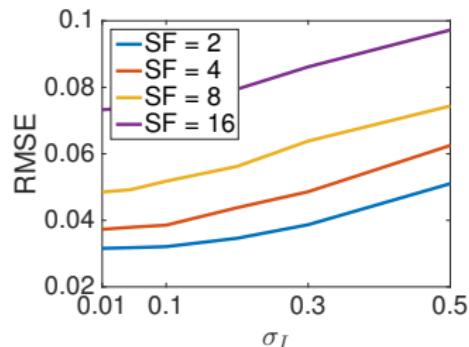
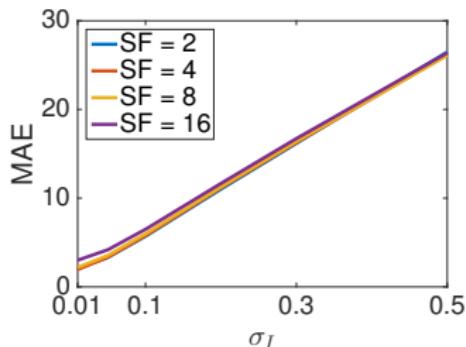
Synthetic Evaluation



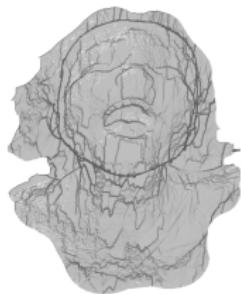
error vs. n & error vs. λ



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]



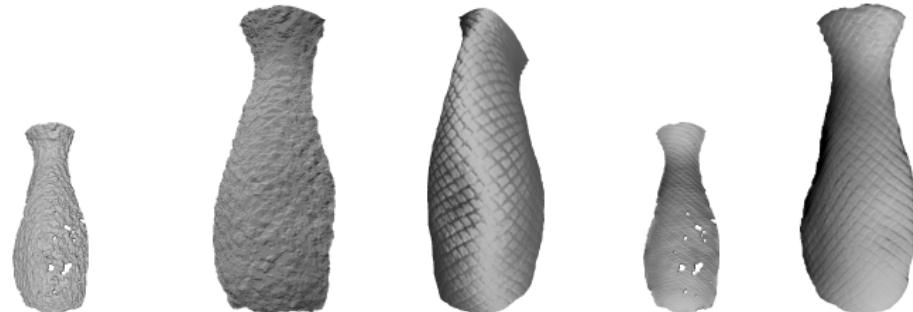
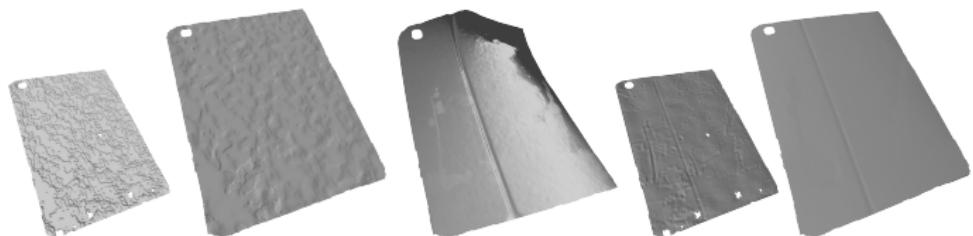
RMSE = **0.03139**

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

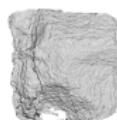
(d) [22]

(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