High Quality Shape from a RGB-D Camera using Photometric Stereo

Songyou Peng

Supervised by:
Dr. Yvain Quéau   Prof. Daniel Cremers

Technical University of Munich
Department of Informatics
Computer Vision Group

June 13, 2017

Outline

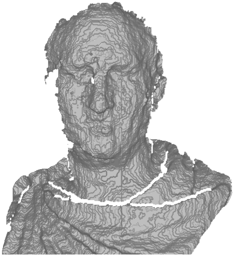
Outline

Problem Statement

Example: RGB-D data from Kinect [Han et al., 2013]

Input RGB image

Depth image



3D shape from the depth

- Missing areas
- Noisy and quantization effect
- No fine details

Goal: improve the quality of depth

Depth refinement

Input RGB image

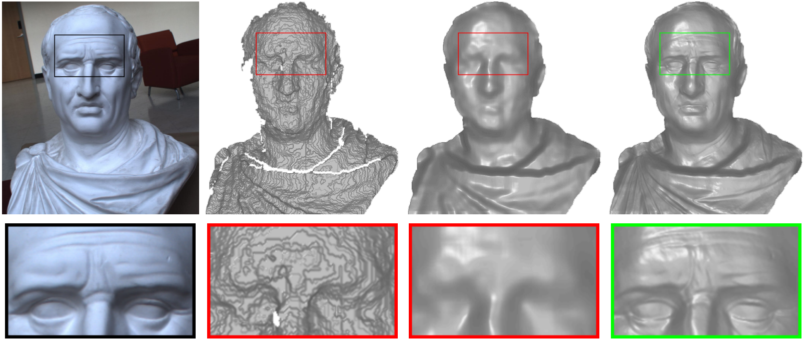
Input depth

Pre-processing

Refined depth

Goal: improve the quality of depth

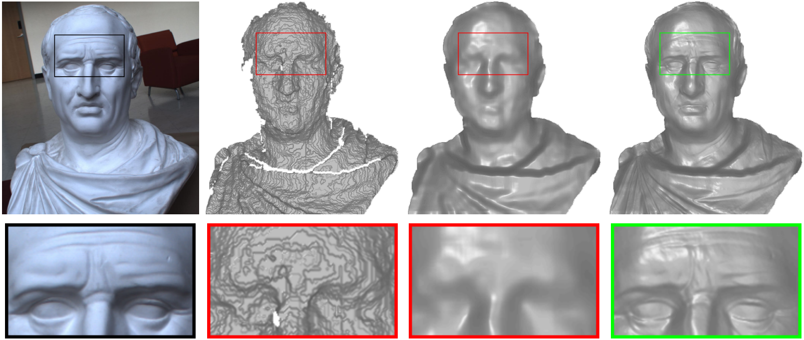
Depth refinement



Input RGB image | Input depth | Pre-processing | Refined depth

Goal: improve the quality of depth

Depth refinement



Input RGB image | Input depth | Pre-processing | Refined depth

Fine details in input RGB $\Rightarrow$ Relation between RGB image & depth?

Outline

Reflectance Model[1]

Intensity I = Albedo ρ × Shading S

[1] Image Courtesy: MIT Intrinsic Images Dataset [Grosse et al., 2009]

Reflectance Model[1]

Intensity I = Albedo ρ × Light $\mathbf{l}$ · Surface normal $\mathbf{n}$

[1] Image Courtesy: MIT Intrinsic Images Dataset [Grosse et al., 2009]

Reflectance Model[1]

 $=$  $\times$  $\cdot$ 

Intensity I — Albedo ρ — Light $\mathbf{l}$ — Surface normal $\mathbf{n}$

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

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

φ: ambient light parameter

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

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

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

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

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

[1] Image Courtesy: MIT Intrinsic Images Dataset [Grosse et al., 2009]

Reflectance Model[1]

Intensity I = Albedo ρ × Light $\mathbf{l}$ · Surface normal $\mathbf{n}$

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

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

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

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

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

φ: ambient light parameter

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

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

[1] Image Courtesy: MIT Intrinsic Images Dataset [Grosse et al., 2009]

Related Work

State-of-the-art depth refinement methods:

- Assume uniform albedo
  - [Yu et al., 2013]
  - [Han et al., 2013]
  - [Haque et al., 2014]

- Estimate the albedo
  - [Wu et al., 2014] $\Rightarrow$ directly I/S
  - [Chatterjee et al., 2015] $\Rightarrow$ using extra IR sources
  - [Kim et al., 2015] $\Rightarrow$ Laplacian on chromaticity
  - **[Or-El et al., 2015]** $\Rightarrow$ Laplacian on the intensity and depth

RGBD-Fusion [Or-El et al., 2015]

$$\min_{\mathbf{s},\rho,\boldsymbol{z}} \|\boldsymbol{I} - \rho\, \tilde{\mathbf{n}}(\boldsymbol{z})\mathbf{s}\|_2^2 + \lambda_\rho \|\Delta_\omega \rho\|_2^2 + \lambda_{\boldsymbol{z}} \|\boldsymbol{z} - \boldsymbol{z}_0\|_2^2 + \lambda_l \|\Delta \boldsymbol{z}\|_2^2$$

Input

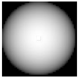
RGBD-Fusion [Or-El et al., 2015]

$$\boxed{\min_{\mathbf{s},\rho,z} \|I - \rho\,\tilde{\mathbf{n}}(z)\mathbf{s}\|_2^2 + \lambda_\rho \|\Delta_\omega \rho\|_2^2 + \lambda_z \|z - z_0\|_2^2 + \lambda_l \|\Delta z\|_2^2} \quad \rho = 1, z = z_0$$

Input

RGBD-Fusion

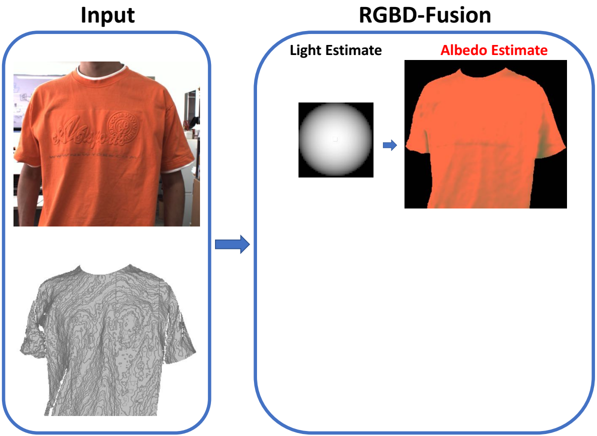
Light Estimate



RGBD-Fusion [Or-El et al., 2015]

$$\min_{\mathbf{s},\rho,z} \|I - \rho\, \tilde{\mathbf{n}}(z)\mathbf{s}\|_2^2 + \lambda_\rho \|\Delta_\omega \rho\|_2^2 + \lambda_z \|z - z_0\|_2^2 + \lambda_l \|\Delta z\|_2^2$$

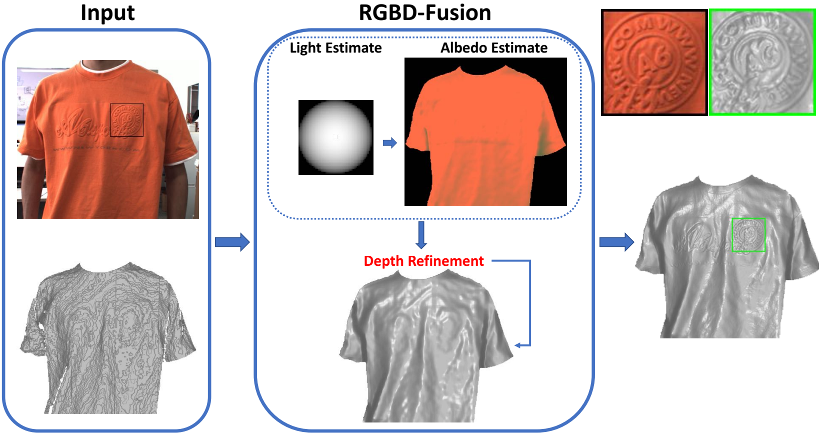
Freeze $\mathbf{s}$, $z = z_0$



RGBD-Fusion [Or-El et al., 2015]

$$\min_{\mathbf{s},\rho,z} \|I - \rho\,\tilde{\mathbf{n}}(z)\mathbf{s}\|_2^2 + \lambda_\rho\|\Delta_\omega \rho\|_2^2 + \lambda_z\|z - z_0\|_2^2 + \lambda_l\|\Delta z\|_2^2$$

Freeze ρ and s



RGBD-Fusion [Or-El et al., 2015]

Problems:

- Use only the intensity image instead of RGB image
- Inaccurate light and albedo estimation
- Surface normal $\mathbf{n}$ is nonlinear w.r.t. depth z, "Fixed Point" scheme is applied in each iteration t:

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

Not convergent!

Outline

1 Introduction

2 Background

3 Methodology
- Proposed I: RGB Ratio Model
- Proposed II: Robust Multi-Light Method

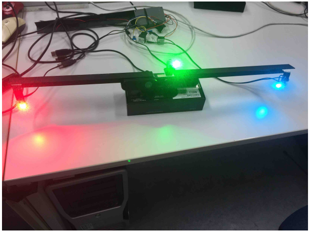
4 Evaluations and Results

5 Conclusions and Future Work

Outline

RGB Ratio Model: Solve the problems faced before

- Alternating optimization on light $\mathbf{s}$, albedo ρ and depth z
- Add red, green and blue LEDs to allow each channel of the RGB image modelled by SH model [Hernández et al., 2011]



(a) LED lights setup

(b) An image from the LED setup

- Resolve the nonlinearity
- Surface normal $\mathbf{n}$ modelled by perspective projection

Build the Ratio Model

Surface normal (perspective):

$$\mathbf{n} = \frac{1}{d}\begin{pmatrix} f\ \nabla \hat{z} \\ -1 - \tilde{x}\hat{z}_x - \tilde{y}\hat{z}_y \end{pmatrix}$$

$\hat{z} = \log z$

f : focal length

(x_0, y_0): principle point

$\tilde{x} = (x - x_0), \tilde{y} = (y - y_0)$

$d = \sqrt{(f\ \nabla \hat{z})^2 + (-1 - \tilde{x}\hat{z}_x - \tilde{y}\hat{z}_y)^2}$

Every channel is treated as an intensity image:

$$I_R = \rho_R(\mathbf{l}_R^\top \mathbf{n} + \varphi_R)$$
$$I_G = \rho_G(\mathbf{l}_G^\top \mathbf{n} + \varphi_G)$$
$$I_B = \rho_B(\mathbf{l}_B^\top \mathbf{n} + \varphi_B)$$

Take red and green channel:

$$\frac{I_R - \rho_R\varphi_R}{I_G - \rho_G\varphi_G} = \frac{\rho_R \mathbf{l}_R^\top \mathbf{n}}{\rho_G \mathbf{l}_G^\top \mathbf{n}}$$

Build the Ratio Model

Surface normal (perspective):

$$\mathbf{n} = \frac{1}{d} \begin{pmatrix} f \, \nabla \hat{z} \\ -1 - \tilde{x}\hat{z}_x - \tilde{y}\hat{z}_y \end{pmatrix}$$

$\hat{z} = \log z$

f : focal length

(x_0, y_0): principle point

$\tilde{x} = (x - x_0), \tilde{y} = (y - y_0)$

$d = \sqrt{(f \, \nabla \hat{z})^2 + (-1 - \tilde{x}\hat{z}_x - \tilde{y}\hat{z}_y)^2}$

Every channel is treated as an intensity image:

$$I_R = \rho_R(\mathbf{l}_R^\top \mathbf{n} + \varphi_R)$$
$$I_G = \rho_G(\mathbf{l}_G^\top \mathbf{n} + \varphi_G)$$
$$I_B = \rho_B(\mathbf{l}_B^\top \mathbf{n} + \varphi_B)$$

Take red and green channel:

$$\frac{I_R - \rho_R\varphi_R}{I_G - \rho_G\varphi_G} = \frac{\rho_R \mathbf{l}_R^\top \mathbf{n}}{\rho_G \mathbf{l}_G^\top \mathbf{n}}$$

Nonlinearity is resolved!
d is cancelled out

Proposed I: RGB Ratio Model

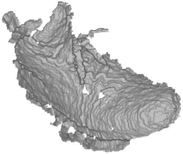
$$\min_{\mathbf{s},\mathcal{P},\boldsymbol{z}} \|\mathcal{R}(\mathcal{P},\boldsymbol{z})\|_2^2 + \lambda_{\boldsymbol{z}}||\boldsymbol{z} - \boldsymbol{z}_0||^2 + \lambda_\rho\|\omega\nabla\mathcal{P}\|^2 + \sum_{c}\|\rho_c\,\tilde{\mathbf{n}}\mathbf{s}_c - I_c\|_2^2,\; c \in \{R,G,B\}$$

Input

Light Estimation

$$\min_{\mathbf{s},\mathcal{P},z} \|\mathcal{R}(\mathcal{P},z)\|_2^2 + \lambda_z\|z - z_0\|^2 + \lambda_\rho\|\omega\nabla\mathcal{P}\|^2 + \sum_c \|\rho_c\, \tilde{\mathbf{n}}\mathbf{s}_c - I_c\|_2^2,\ c \in \{R, G, B\}$$

Input



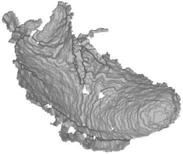
RGB Ratio Model

Light Estimate

Albedo Estimation

$$\min_{\mathbf{s},\mathcal{P},\mathbf{z}} \|\mathcal{R}(\mathcal{P},\mathbf{z})\|_2^2 + \lambda_z\|z - z_0\|^2 + \lambda_\rho\|\omega\nabla\mathcal{P}\|^2 + \sum_c \|\rho_c\,\tilde{\mathbf{n}}\mathbf{s}_c - I_c\|_2^2,\ c \in \{R, G, B\}$$

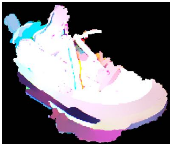
Input



RGB Ratio Model

Light Estimate

Albedo Estimate



Albedo Estimation

$$\boxed{\min_{\mathbf{s},\mathcal{P},\mathbf{z}} \|\mathcal{R}(\mathcal{P},\mathbf{z})\|_2^2 + \lambda_z\|\mathbf{z}-\mathbf{z}_0\|^2 + \lambda_\rho\|\omega\nabla\mathcal{P}\|^2 + \sum_c \|\rho_c\,\tilde{\mathbf{n}}\mathbf{s}_c - I_c\|_2^2,\ c\in\{R,G,B\}}$$

Ratio Model:

$$\begin{aligned}\frac{I_R-\rho_R\varphi_R}{I_G-\rho_G\varphi_G} &= \frac{\rho_R\mathbf{l}_R^\top\mathbf{n}}{\rho_G\mathbf{l}_G^\top\mathbf{n}}\\ \frac{I_G-\rho_G\varphi_G}{I_B-\rho_B\varphi_B} &= \frac{\rho_G\mathbf{l}_G^\top\mathbf{n}}{\rho_B\mathbf{l}_B^\top\mathbf{n}} \Rightarrow\\ \frac{I_B-\rho_G\varphi_B}{I_R-\rho_R\varphi_R} &= \frac{\rho_B\mathbf{l}_B^\top\mathbf{n}}{\rho_R\mathbf{l}_R^\top\mathbf{n}}\end{aligned}$$

$$\begin{aligned}I_G\mathbf{l}_R^\top\mathbf{n}\rho_R - I_R\mathbf{l}_G^\top\mathbf{n}\rho_G &= \rho_R\rho_G(\varphi_G\mathbf{l}_R^\top\mathbf{n} - \varphi_R\mathbf{l}_G^\top\mathbf{n})\\ I_B\mathbf{l}_G^\top\mathbf{n}\rho_G - I_G\mathbf{l}_B^\top\mathbf{n}\rho_B &= \rho_G\rho_B(\varphi_B\mathbf{l}_G^\top\mathbf{n} - \varphi_G\mathbf{l}_B^\top\mathbf{n})\\ I_R\mathbf{l}_B^\top\mathbf{n}\rho_B - I_B\mathbf{l}_R^\top\mathbf{n}\rho_R &= \rho_B\rho_R(\varphi_R\mathbf{l}_B^\top\mathbf{n} - \varphi_B\mathbf{l}_R^\top\mathbf{n})\end{aligned}$$

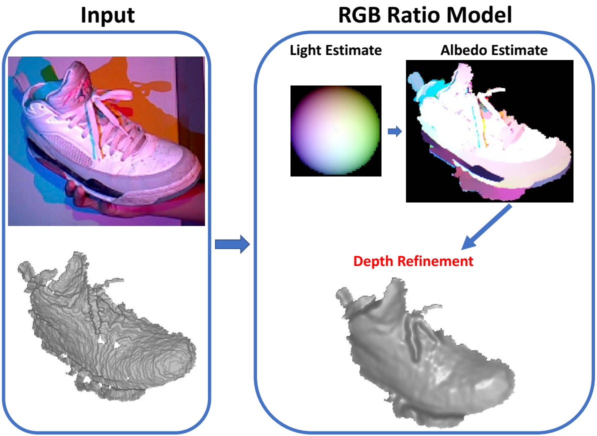
$$\Downarrow$$

$$\underbrace{\begin{pmatrix} I_G\mathbf{l}_R^\top\mathbf{n} & -I_R\mathbf{l}_G^\top\mathbf{n} & 0\\ 0 & I_B\mathbf{l}_G^\top\mathbf{n} & -I_G\mathbf{l}_B^\top\mathbf{n}\\ -I_B\mathbf{l}_R^\top\mathbf{n} & 0 & I_R\mathbf{l}_B^\top\mathbf{n}\end{pmatrix}}_{\mathbf{A}_\rho}\underbrace{\begin{pmatrix}\rho_R\\ \rho_G\\ \rho_B\end{pmatrix}}_{\mathcal{P}} = \underbrace{\begin{pmatrix}\rho_R\rho_G(\varphi_G\mathbf{l}_R^\top\mathbf{n} - \varphi_R\mathbf{l}_G^\top\mathbf{n})\\ \rho_G\rho_B(\varphi_B\mathbf{l}_G^\top\mathbf{n} - \varphi_G\mathbf{l}_B^\top\mathbf{n})\\ \rho_B\rho_R(\varphi_R\mathbf{l}_B^\top\mathbf{n} - \varphi_B\mathbf{l}_R^\top\mathbf{n})\end{pmatrix}}_{\mathbf{b}_\rho}$$

$$\Rightarrow \mathcal{R}(\mathcal{P}) = \mathbf{A}_\rho\mathcal{P} - \mathbf{b}_\rho$$

Depth Refinement

$$\min_{\mathbf{s},\mathcal{P},\mathbf{z}} \|\mathcal{R}(\mathcal{P},\mathbf{z})\|_2^2+\lambda_{\mathbf{z}}\|\mathbf{z}-\mathbf{z}_0\|^2+\lambda_\rho\|\omega\nabla\mathcal{P}\|^2+\sum_c\|\rho_c\,\tilde{\mathbf{n}}\mathbf{s}_c-I_c\|_2^2,\ c\in\{R,G,B\}$$



Depth Refinement

$$\min_{\mathbf{s},\mathcal{P},z} \|\mathcal{R}(\mathcal{P}, z)\|_2^2 + \lambda_z \|z - z_0\|^2 + \lambda_\rho \|\omega \nabla \mathcal{P}\|^2 + \sum_c \|\rho_c \, \tilde{\mathbf{n}} \mathbf{s}_c - I_c\|_2^2, \; c \in \{R, G, B\}$$

Ratio Model:

$$\frac{I_R - \rho_R \varphi_R}{I_G - \rho_G \varphi_G} = \frac{\rho_R \mathbf{l}_R^\top \mathbf{n}}{\rho_G \mathbf{l}_G^\top \mathbf{n}}$$

$$\frac{I_G - \rho_G \varphi_G}{I_B - \rho_B \varphi_B} = \frac{\rho_G \mathbf{l}_G^\top \mathbf{n}}{\rho_B \mathbf{l}_B^\top \mathbf{n}} \Rightarrow$$

$$\frac{I_B - \rho_G \varphi_B}{I_R - \rho_R \varphi_R} = \frac{\rho_B \mathbf{l}_B^\top \mathbf{n}}{\rho_R \mathbf{l}_R^\top \mathbf{n}}$$

$$\rho_G (I_R - \rho_R \varphi_R) \mathbf{l}_G^\top \mathbf{n} - \rho_R (I_G - \rho_G \varphi_G) \mathbf{l}_R^\top \mathbf{n} = 0$$

$$\rho_B (I_G - \rho_G \varphi_G) \mathbf{l}_B^\top \mathbf{n} - \rho_G (I_B - \rho_B \varphi_B) \mathbf{l}_G^\top \mathbf{n} = 0$$

$$\rho_R (I_B - \rho_B \varphi_B) \mathbf{l}_R^\top \mathbf{n} - \rho_B (I_R - \rho_R \varphi_R) \mathbf{l}_B^\top \mathbf{n} = 0$$

$$\Downarrow$$

$$\Psi z = 0$$

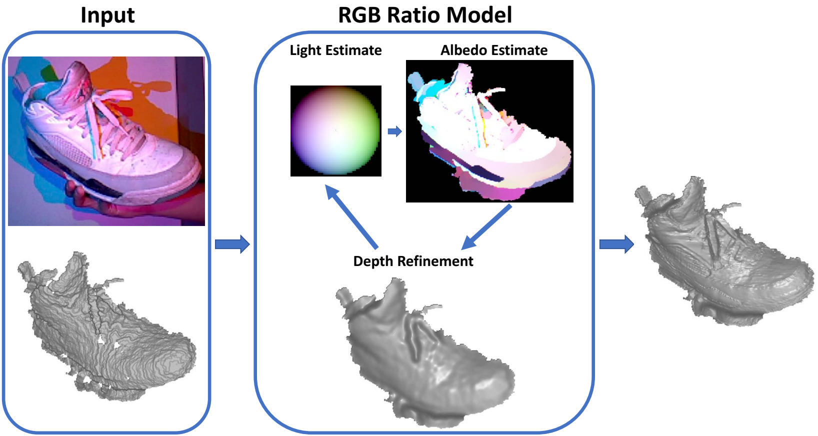
$$\Downarrow$$

$$\mathcal{R}(z) = \Psi z$$

Linear equation

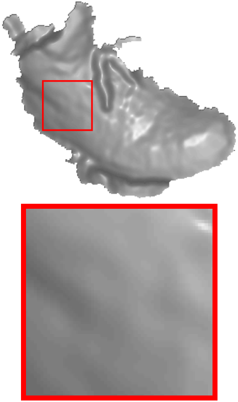
Proposed I: RGB Ratio Model

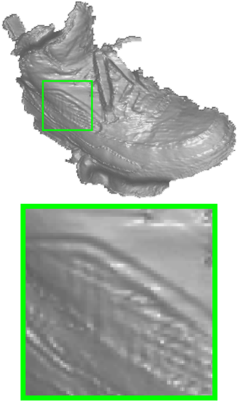
$$\min_{\mathbf{s},\mathcal{P},\mathbf{z}} \|\mathcal{R}(\mathcal{P},\mathbf{z})\|_2^2 + \lambda_{\mathbf{z}}||\mathbf{z}-\mathbf{z}_0||^2 + \lambda_\rho\|\omega\nabla\mathcal{P}\|^2 + \sum_c \|\rho_c\,\tilde{\mathbf{n}}\mathbf{s}_c - I_c\|_2^2,\ c \in \{R,G,B\}$$



Proposed I: RGB Ratio Model


RGB image


After pre-processing


Refined depth

Proposed I: RGB Ratio Model

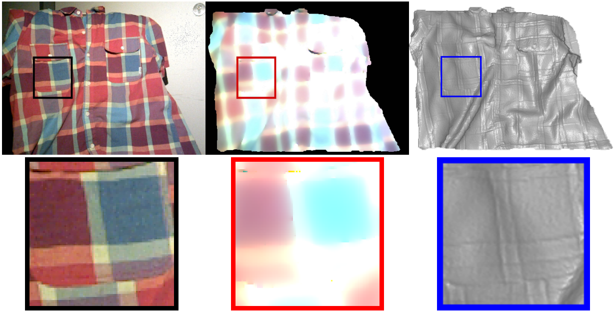
Limitations:

- The red, green and blue LED lights have to be set up far away from each other, otherwise the ratio model may fail
- Three LEDs are likely to bring extra specularity
- **Works well only on the simple albedo cases, like RGBD-Fusion**

Outline

Why propose Robust Multi-Light Method?

- RGBD-Fusion and RGB ratio model only work well on the objects with simple albedos (regularization is not realistic)



RGBD-Fusion on an object with complicated albedo

- Parameter tuning for the regularization is tedious and time-consuming

Proposed II: Robust Multi-Light Method

n various illuminations with the fixed view ($n \geqslant 3$)

Proposed II: Robust Multi-Light Method

n various illuminations with the fixed view ($n \geqslant 3$)

$$\min_{\mathbf{s},\rho,z} \sum_{i}^{n} \sum_{c} \|\rho_c \, \tilde{\mathbf{n}}(z)\mathbf{s}_{i,c} - I_{i,c}\|_2^2 + \lambda_z \|z - z_0\|_2^2, \quad c \in \{R, G, B\}$$

Proposed II: Robust Multi-Light Method

n various illuminations with the fixed view ($n \geqslant 3$)

$$\min_{\mathbf{s},\rho,z} \sum_{i}^{n} \sum_{c} \|\rho_c \cdot \tilde{\mathbf{n}}(z)\mathbf{s}_{i,c} - I_{i,c}\|_2^2 + \lambda_z \|z - z_0\|_2^2, \quad c \in \{R, G, B\}$$

RGBD-Fusion Method:

$$\min_{\mathbf{s},\rho,z} \|\rho\, \tilde{\mathbf{n}}(z)\mathbf{s} - I\|_2^2 + \lambda_\rho \|\Delta_\omega \rho\|_2^2 + \lambda_z \|z - z_0\|_2^2 + \lambda_l \|\Delta z\|_2^2$$

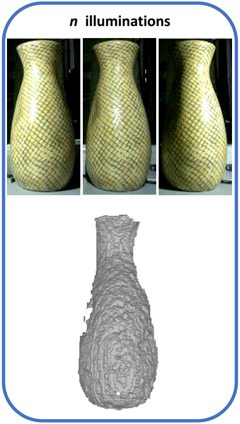
Proposed RGB Ratio Model:

$$\min_{\mathbf{s},\mathcal{P},z} \|\mathcal{R}(\mathcal{P}, z)\|_2^2 + \lambda_\rho \|\omega \nabla \mathcal{P}\|^2 + \lambda_z \|z - z_0\|^2 + \sum_{c} \|\rho_c\, \tilde{\mathbf{n}}\mathbf{s}_c - I_c\|_2^2, \quad c \in \{R, G, B\}$$

Proposed II: Robust Multi-Light Method

$$\min_{\mathbf{s},\rho,z} \sum_{i}^{n} \sum_{c} \| \rho_c \, \tilde{\mathbf{n}}(z)\mathbf{s}_{i,c} - I_{i,c} \|_2^2 + \lambda_z \| z - z_0 \|_2^2, \quad c \in \{R, G, B\}$$

Input

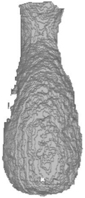


Light Estimation

$$\min_{\mathbf{s},\rho,z} \sum_i^n \sum_c \|\rho_c\, \tilde{\mathbf{n}}(z)\mathbf{s}_{i,c} - I_{i,c}\|_2^2 + \lambda_z \|z - z_0\|_2^2, \quad c \in \{R, G, B\}$$

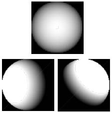
Input

n **illuminations**



Robust Multi-Light Method

Light Estimate

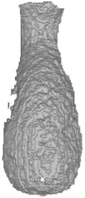


Albedo Estimation

$$\min_{\mathbf{s},\rho,z} \sum_{i}^{n} \sum_{c} \|\rho_c \, \tilde{\mathbf{n}}(z)\mathbf{s}_{i,c} - I_{i,c}\|_2^2 + \lambda_z \|z - z_0\|_2^2, \quad c \in \{R, G, B\}$$

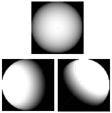
Input

***n* illuminations**



Robust Multi-Light Method

Light Estimate



Albedo Estimate

Albedo Estimation

$$\boxed{\min_{\mathbf{s},\rho,z} \sum_{i}^{n} \sum_{c} \|\rho_c\, \tilde{\mathbf{n}}(z)\mathbf{s}_{i,c} - I_{i,c}\|_2^2 + \lambda_z \|z - z_0\|_2^2, \quad c \in \{R, G, B\}}$$

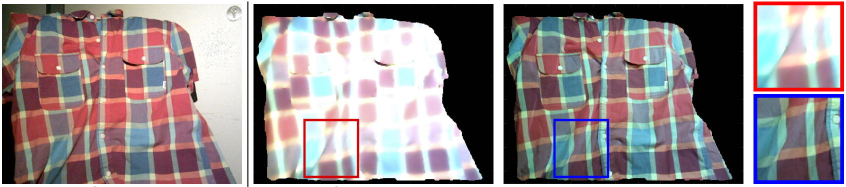
$$\min_{\rho_c} \sum_{i}^{n} \|\rho_c\, \tilde{\mathbf{n}}(z)\mathbf{s}_{i,c} - I_{i,c}\|_2^2 \Leftrightarrow \min_{\rho_c} \|\mathbf{A}_{\rho_c}\rho_c - \mathbf{I}_c\|_2^2$$

where

$$\mathbf{A}_{\rho_c} = \begin{pmatrix} \mathrm{diag}(\tilde{\mathbf{n}} \cdot \mathbf{s}_{1,c}) \\ \vdots \\ \mathrm{diag}(\tilde{\mathbf{n}} \cdot \mathbf{s}_{n,c}) \end{pmatrix} \quad \mathbf{I}_c = \begin{pmatrix} I_{1,c} \\ \vdots \\ I_{n,c} \end{pmatrix}$$

Albedo Estimation

$$\min_{\mathbf{s},\rho,z} \sum_{i}^{n} \sum_{c} \|\rho_c \, \tilde{\mathbf{n}}(z)\mathbf{s}_{i,c} - I_{i,c}\|_2^2 + \lambda_z \|z - z_0\|_2^2, \quad c \in \{R, G, B\}$$



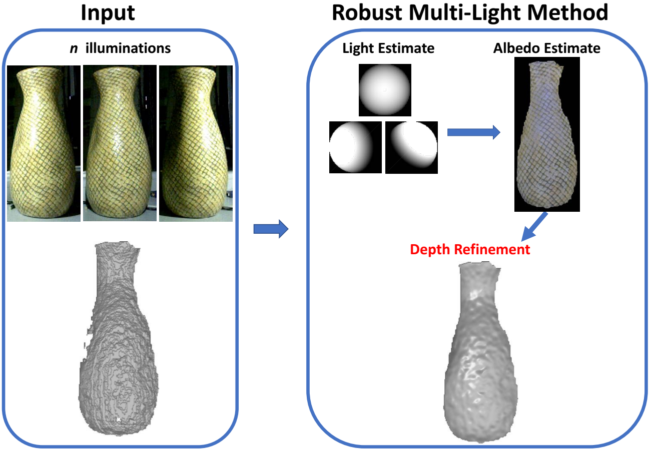
RGB image | RGBD-Fusion | Robust Multi-Light

Comparsion on the albedo estimation between RGBD-Fusion and the proposed method

- Recover most of real albedo details
- Remove the shading on the RGB image

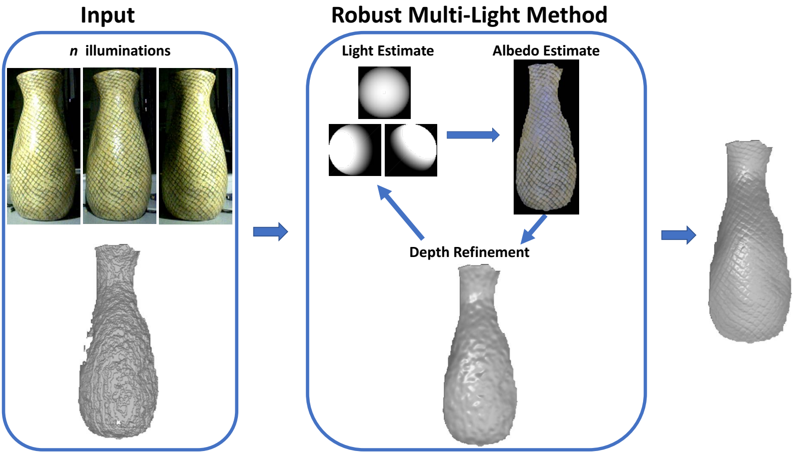
Depth Refinement

$$\min_{\mathbf{s},\rho,z} \sum_{i}^{n} \sum_{c} \|\rho_c\, \tilde{\mathbf{n}}(z)\mathbf{s}_{i,c} - I_{i,c}\|_2^2 + \lambda_z \|z - z_0\|_2^2, \quad c \in \{R, G, B\}$$



Proposed II: Robust Multi-Light Method

$$\boxed{\min_{\mathbf{s},\rho,z} \sum_{i}^{n} \sum_{c} \|\rho_c \, \tilde{\mathbf{n}}(z)\mathbf{s}_{i,c} - I_{i,c}\|_2^2 + \lambda_z \|z - z_0\|_2^2, \quad c \in \{R, G, B\}}$$

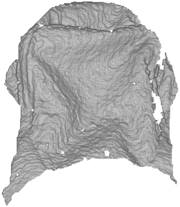


Proposed II: Robust Multi-Light Method

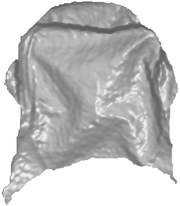
Backpack

Estimated albedo

Refined normal



Input depth



After pre-processing

Refined depth

Proposed II: Robust Multi-Light Method

When Super-Resolution Meets Depth Refinement

ASUS Xtion Pro Live provides maximum:

- 1280×960 RGB image (after cropping)
- 640×480 depth image

Scale factor $\xi = 2$

$$z = KZ$$

$K \in \mathbb{R}^{m \times \xi m}$ is a downsampling operator [Unger et al., 2010]
m: number of pixels for the 640×480 depth z_0

When Super-Resolution Meets Depth Refinement

ASUS Xtion Pro Live provides maximum:

- 1280×960 RGB image (after cropping)
- 640×480 depth image

Scale factor $\xi = 2$

$$z = KZ$$

$K \in \mathbb{R}^{m \times \xi m}$ is a downsampling operator [Unger et al., 2010]
m: number of pixels for the 640×480 depth z_0

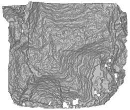
Depth refinement now changes to:

$$\min_{z} \|\mathbf{A}_z z - \mathbf{b}_z\|_2^2 + \lambda_z \|z - z_0\|_2^2 \Rightarrow \min_{Z} \|\mathbf{A}_Z Z - \mathbf{b}_Z\|_2^2 + \lambda_z \|KZ - z_0\|_2^2$$

When Super-resolution meets depth refinement

One input image (10 in total)

Estimated albedo



Input depth

Refined super-resolution depth

Outline

Synthetic Data

"Joy Yells" Dataset:

Input depth z_0 | Ground truth depth z

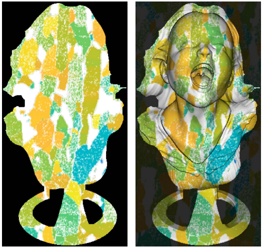
Metric:

- Root mean square error (RMSE)
- Mean angular error (MAE)

Synthetic Data

From left to right: simple to complicated albedo

Simple RGB



Pattern[2]

Complicated Pattern[3]

[2] EBSD map. Image courtesy of `https://mtex-toolbox.github.io/files/doc/EBSDSpatialPlots.html`

[3] 1000 Visual Mashups. Image courtsesy of `https://www.flickr.com/photos/qthomasbower/3470650293`

Parameter setup

Method	Total number	Parameters
RGBD-Fusion [Or-El et al., 2015]	8	$\lambda_\rho = 0.1, \lambda_\beta^1 = 0.1, \lambda_\beta^2 = 0.1, \tau = 0.05, \sigma_c = \sqrt{0.05}, \sigma_d = \sqrt{50}, \lambda_z^1 = 0.004, \lambda_z^2 = 0.0075$
Proposed I: RGB Ratio Model	4	$\lambda_\rho^1 = 10^{15}, \lambda_\rho^2 = 10^{13}, \sigma_c = 100, \lambda_z = 100$
Proposed II: Robust Multi-Light	1	$\lambda_z = 100$

Quantitative evaluation

	Simple RGB		Pattern		Complicated	
Method	RMSE	MAE	RMSE	MAE	RMSE	MAE
Input reference	3.3305	16.3096	3.3305	16.3096	3.3305	16.3096
RGBD-Fusion (ns)	3.3418	18.9115	3.3872	27.0026	3.3411	25.6574
RGBD-Fusion	3.1751	17.2197	3.1890	18.4722	3.1708	18.0850
RGB ratio model	**1.9437**	5.0574	2.9116	17.5238	3.1006	21.2286
Robust multi-light	2.3125	**3.8708**	**1.5794**	**1.7368**	**1.8424**	**2.6815**

ns: no smoothness term $||\Delta z||_2^2$

Observation:

- Smoothness term has a huge impact on the accuracy of RGBD-Fusion

Quantitative evaluation

	Simple RGB		Pattern		Complicated	
Method	RMSE	MAE	RMSE	MAE	RMSE	MAE
Input reference	3.3305	16.3096	3.3305	16.3096	3.3305	16.3096
RGBD-Fusion (ns)	3.3418	18.9115	3.3872	27.0026	3.3411	25.6574
RGBD-Fusion	3.1751	17.2197	3.1890	18.4722	3.1708	18.0850
RGB ratio model	**1.9437**	5.0574	2.9116	17.5238	3.1006	21.2286
Robust multi-light	2.3125	**3.8708**	**1.5794**	**1.7368**	**1.8424**	**2.6815**

ns: no smoothness term $||\Delta z||_2^2$

Observation:

- Smoothness term has a huge impact on the accuracy of RGBD-Fusion
- Proposed RGB ratio model has a better performance than RGBD-Fusion

Quantitative evaluation

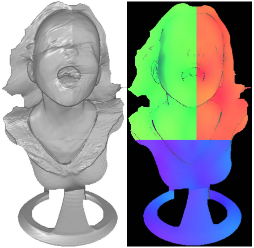
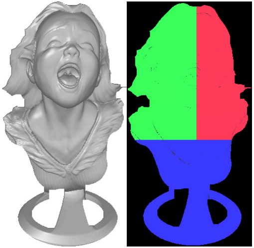
	Simple RGB		Pattern		Complicated	
Method	RMSE	MAE	RMSE	MAE	RMSE	MAE
Input reference	3.3305	16.3096	3.3305	16.3096	3.3305	16.3096
RGBD-Fusion (ns)	3.3418	18.9115	3.3872	27.0026	3.3411	25.6574
RGBD-Fusion	3.1751	17.2197	3.1890	18.4722	3.1708	18.0850
RGB ratio model	**1.9437**	5.0574	2.9116	17.5238	3.1006	21.2286
Robust multi-light	2.3125	**3.8708**	**1.5794**	**1.7368**	**1.8424**	**2.6815**

ns: no smoothness term $||\Delta \boldsymbol{z}||_2^2$

Observation:

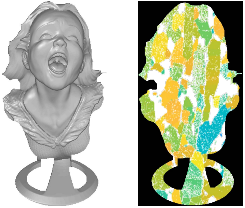
- Smoothness term has a huge impact on the accuracy of RGBD-Fusion
- Proposed RGB ratio model has a better performance than RGBD-Fusion
- Proposed Robust Multi-Light method outperforms other methods in almost all the cases

RGBD-Fusion v.s Proposed RGB Ratio Model

Ground truth	RGBD-Fusion	RGB Ratio Model
		
RMSE = 3.33, MAE = 16.31	RMSE = 2.87, MAE = 17.17	RMSE = **1.94**, MAE = **5.06**

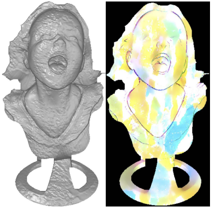
Synthetic Data (Complicated albedo)

Ground Truth



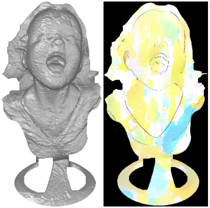
RMSE = 3.33, MAE = 16.31

RGBD-Fusion



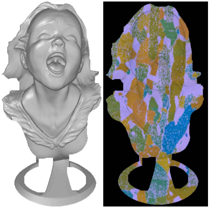
RMSE = 3.35, MAE = 23.48

RGB Ratio Model



RMSE = 2.91, MAE = 17.52

Robust Multi-Light



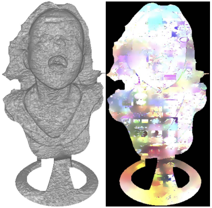
RMSE = **1.58**, MAE = **1.73**

Synthetic Data (Complicated albedo)

Ground Truth

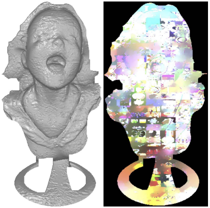
RMSE = 3.33, MAE = 16.31

RGBD-Fusion



RMSE = 3.39, MAE = 35.48

RGB Ratio Model



RMSE = 3.10, MAE = 21.22

Robust Multi-Light

RMSE = **1.85**, MAE = **2.68**

Runtime

For a 540×960 image,

Method	Runtime (s)
RGBD-Fusion	21.64
Proposed I: RGB Ratio	49.33
Proposed II: Multi-Light	52.82

RGBD-Fusion:

- Estimate the light and albedo only once
- Use only intensity value
- Optimization diverges after 1-3 iterations

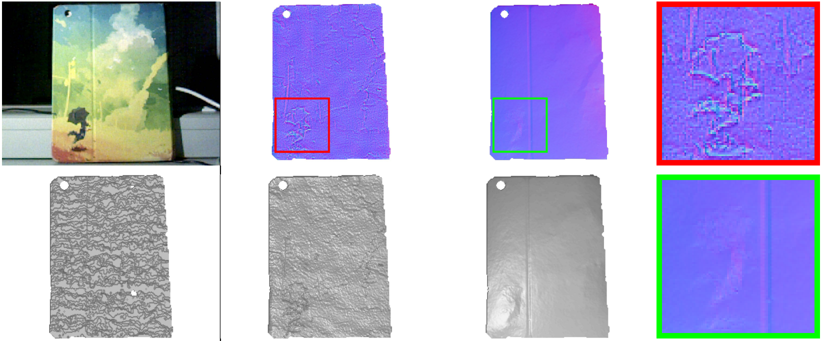
Proposed I:

- Use RGB information and jointly estimate the color albedo

Proposed 2:

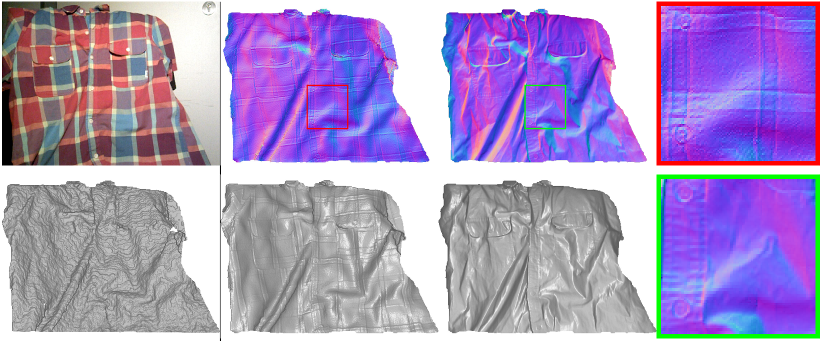
- Use *n* RGB images

Real Data (Complicated albedo)



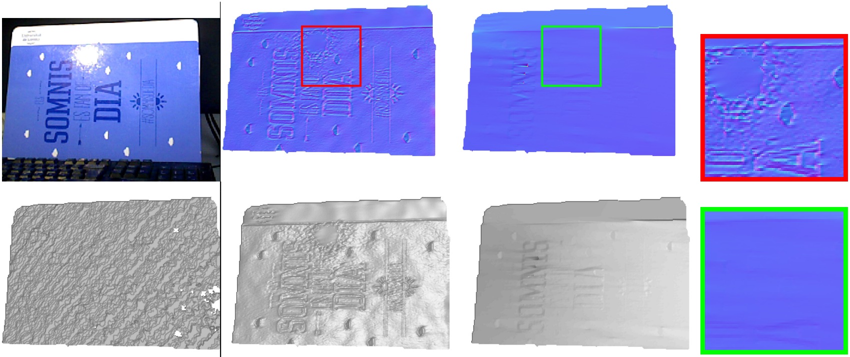
Input RGBD-Fusion **Robust Multi-Light**

Real Data (Complicated albedo)



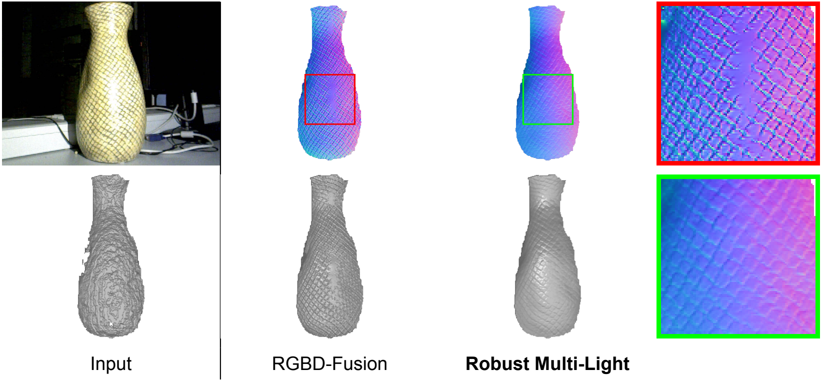
Input | RGBD-Fusion | **Robust Multi-Light**

Real Data (Specular objects)



Input RGBD-Fusion **Robust Multi-Light**

Real Data (Specular objects)



Input | RGBD-Fusion | **Robust Multi-Light**

Outline

Conclusions

- Develop a RGB Ratio Model which overcomes the nonlinearity problem
- Propose the Robust Multi-Light method which outperforms the state-of-the-arts quantitatively and qualitatively
- Combine image super-resolution with depth refinement

Future works

- Prepare for the submission of ICCV 2017 workshop
- Integrate the Robust Multi-Light method with the 3D modelling

Bibliography I

[Chatterjee et al., 2015] Chatterjee, A., Madhav Govindu, V., et al. (2015).
Photometric refinement of depth maps for multi-albedo objects.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 933–941.

[Grosse et al., 2009] Grosse, R., Johnson, M. K., Adelson, E. H., and Freeman, W. T. (2009).
Ground truth dataset and baseline evaluations for intrinsic image algorithms.
In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2335–2342. IEEE.

[Han et al., 2013] Han, Y., Lee, J.-Y., and So Kweon, I. (2013).
High quality shape from a single rgb-d image under uncalibrated natural illumination.
In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1617–1624.

[Haque et al., 2014] Haque, M., Chatterjee, A., Madhav Govindu, V., et al. (2014).
High quality photometric reconstruction using a depth camera.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2275–2282.

[Heide et al., 2014] Heide, F., Steinberger, M., Tsai, Y.-T., Rouf, M., Pajak, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K., et al. (2014).
FlexISP: A flexible camera image processing framework.
ACM Transactions on Graphics (TOG), 33(6):231.

[Hernández et al., 2011] Hernández, C., Vogiatzis, G., and Cipolla, R. (2011).
Overcoming shadows in 3-source photometric stereo.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(2):419–426.

[Khoshelham and Elberink, 2012] Khoshelham, K. and Elberink, S. O. (2012).
Accuracy and resolution of kinect depth data for indoor mapping applications.
Sensors, 12(2):1437–1454.

[Kim et al., 2015] Kim, K., Torii, A., and Okutomi, M. (2015).
Joint estimation of depth, reflectance and illumination for depth refinement.
In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 1–9.

Bibliography II

[Meinhardt et al., 2017] Meinhardt, T., Möller, M., Hazirbas, C., and Cremers, D. (2017).
Learning proximal operators: Using denoising networks for regularizing inverse imaging problems.
arXiv preprint arXiv:1704.03488.

[Or-El et al., 2015] Or-El, R., Rosman, G., Wetzler, A., Kimmel, R., and Bruckstein, A. M. (2015).
Rgbd-fusion: Real-time high precision depth recovery.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5407–5416.

[Unger et al., 2010] Unger, M., Pock, T., Werlberger, M., and Bischof, H. (2010).
A convex approach for variational super-resolution.
In *Joint Pattern Recognition Symposium*, pages 313–322. Springer.

[Wu et al., 2014] Wu, C., Zollhöfer, M., Nießner, M., Stamminger, M., Izadi, S., and Theobalt, C. (2014).
Real-time shading-based refinement for consumer depth cameras.
ACM Transactions on Graphics (TOG), 33(6):200.

[Yu et al., 2013] Yu, L.-F., Yeung, S.-K., Tai, Y.-W., and Lin, S. (2013).
Shading-based shape refinement of rgb-d images.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1415–1422.

High Quality Shape from a RGB-D Camera using Photometric Stereo

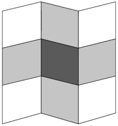
Songyou Peng

Supervised by:
Dr. Yvain Quéau   Prof. Daniel Cremers

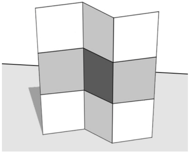
Technical University of Munich
Department of Informatics
Computer Vision Group

June 13, 2017

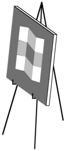
Shape from shading ambiguity



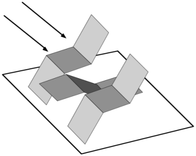
(a) An image



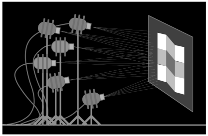
(b) A possible explanation



(c) painter's



(d) sculptor's



(e) Lighting designer's

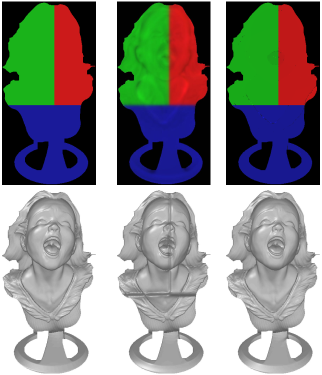
Albedo Estimation

$$\min_{\mathbf{s},\mathcal{P},\mathbf{z}} \|\mathcal{R}(\mathcal{P},\mathbf{z})\|_2^2 + \lambda_z\|\mathbf{z}-\mathbf{z}_0\|^2 + \lambda_\rho\|\boldsymbol{\omega}\nabla\mathcal{P}\|^2 + \sum_c \|\rho_c\,\tilde{\mathbf{n}}\mathbf{s}_c - I_c\|_2^2,\ c\in\{R,G,B\}$$

$$\omega = \mathrm{diag}\left(\begin{pmatrix}\omega_R\\ \omega_G\\ \omega_B\end{pmatrix}\right)$$

$$\omega_c = \exp\left(-\frac{\sigma_c\|\nabla I_c\|^2}{\max\|\nabla I_c\|^2}\right)$$

depends on the gradient of a certain channel



Ground truth $\quad \omega = 1 \quad \omega$ varies

Synthetic data generation

"Joy Yells" Dataset:

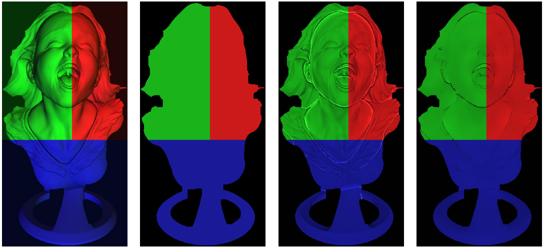
(a) Input depth

(b) Ground truth depth

An image can be generated using SH model, $\mathbf{n}$ is given from the ground truth depth.

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

$$\min_{\mathbf{s},\rho,z} \underbrace{\|I - \rho\, \tilde{\mathbf{n}}(\mathbf{z})\mathbf{s}\|_2^2}_{E_S} + \underbrace{\lambda_\rho \|\Delta_\omega \rho\|_2^2}_{E_R} + \lambda_z \|z - z_0\|_2^2 + \lambda_l \|\Delta z\|_2^2$$

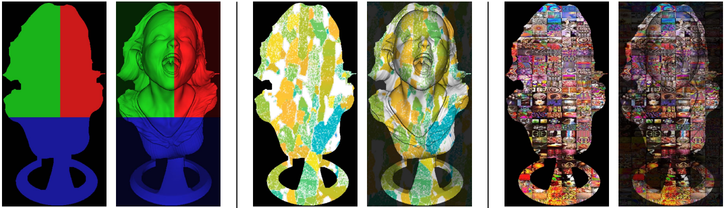


I | True ρ | E_S | $E_S + E_R$

Effectiveness of the anisotropic Laplacian $\|\Delta_\omega \rho\|_2^2$

Synthetic data generation

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



Simple RGB | Pattern[4] | Complicated Pattern[5]

[4] EBSD map. Image courtesy of `https://mtex-toolbox.github.io/files/doc/EBSDSpatialPlots.html`

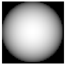
[5] 1000 Visual Mashups. Image courtsesy of `https://www.flickr.com/photos/qthomasbower/3470650293`

Synthetic data generation

$$I = \rho\, \mathbf{l}^\top \mathbf{n} + \varphi = \rho\, \tilde{\mathbf{n}}\mathbf{s}, \quad \mathbf{s} = \begin{bmatrix} \mathbf{l}_1 & \mathbf{l}_2 & \mathbf{l}_3 & \varphi \end{bmatrix}^\top$$

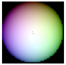
For RGBD-Fusion:

$$\mathbf{s}_R = \begin{bmatrix} 0 & 0 & -1 & 0.2 \end{bmatrix}^\top$$
$$\mathbf{s}_G = \begin{bmatrix} 0 & 0 & -1 & 0.2 \end{bmatrix}^\top$$
$$\mathbf{s}_B = \begin{bmatrix} 0 & 0 & -1 & 0.2 \end{bmatrix}^\top$$



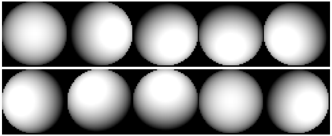
For RGB Ratio Model:

$$\mathbf{s}_R = \begin{bmatrix} 0 & 0 & -1 & 0.2 \end{bmatrix}^\top$$
$$\mathbf{s}_G = \begin{bmatrix} 0.3 & 0.2 & -1 & 0.2 \end{bmatrix}^\top$$
$$\mathbf{s}_B = \begin{bmatrix} -0.2 & 0.3 & -1 & 0.2 \end{bmatrix}^\top$$



For Multi-Light Method:

$$\begin{pmatrix} 0.5 & 0 & -1 & 0.2 \\ 0.3 & 0.4 & -1 & 0.2 \\ 0 & 0.5 & -1 & 0.2 \\ -0.4 & 0.3 & -1 & 0.2 \\ -0.5 & 0 & -1 & 0.2 \\ -0.3 & -0.4 & -1 & 0.2 \\ 0 & -0.5 & -1 & 0.2 \\ 0.4 & -0.3 & -1 & 0.2 \\ 0 & 0 & -1 & 0.2 \\ 0.45 & 0.2 & -1 & 0.2 \end{pmatrix}^\top$$



Metrics

- Root mean square error (RMSE): global quality
- Mean angular error (MAE): precision

$$e_{RMSE} = \sqrt{\frac{\sum_{i}^{m} (z(i) - z_g(i))^2}{m}} \qquad \text{(in } mm\text{)}$$

$$e_{MAE} = \frac{\sum_{i}^{m} \arccos(\mathbf{n}(i) \cdot \mathbf{n}_g(i))}{m} \qquad \text{(in degree}^\circ\text{)}$$

m: number of pixels inside the mask
z: depth
n: surface normal

Depth refinement of Multi-light method

Spherical Harmonics model

$$I(x,y) = \rho(x,y)\cdot \begin{pmatrix} l^1 & l^2 & l^3 \end{pmatrix} \begin{pmatrix} fz_x(x,y) \\ fz_y(x,y) \\ -1-(x-x_0)z_x(x,y)-(y-y_0)z_y(x,y) \end{pmatrix} /d(x,y) + \rho(x,y)\cdot\varphi$$

where $d = \sqrt{(fz_x(x,y))^2 + (fz_y(x,y))^2 + (-1-(x-x_0)z_x(x,y)-(y-y_0)z_y(x,y))^2}$ After rearranging, we have:

$$\frac{l^1 f - l^3(x-x_0)}{d(x,y)} z_x(x,y) + \frac{l^2 f - l^3(y-y_0)}{d(x,y)} z_y(x,y) = I(x,y) + \frac{l^3}{d(x,y)} - \rho(x,y)\varphi$$

we extend this equation to all the pixels in the mask $\mathcal{M}$ and acquire:

$$\frac{l^1 f - l^3 \tilde{x}}{d} \cdot z_x + \frac{l^2 f - l^3 \tilde{y}}{d} \cdot z_y = I + \frac{l^3}{d} - \varphi\cdot\rho$$

Provided we denote the gradient matrices in x and y directions as D_x and D_y,

$$\left(\mathrm{diag}\left(\frac{l^1 f - l^3\tilde{x}}{d}\right)D_x + \mathrm{diag}\left(\frac{l^2 f - l^3\tilde{y}}{d}\right)D_y\right) z = I + \frac{l^3}{d} - \varphi\cdot\rho$$

Depth refinement of Multi-light method

This is the linear equation for one image. Now we define the $\mathbf{A}_{z_c}$ and $\mathbf{b}_{z_c}$ for our system:

$$\mathbf{A}_{z_c} = \begin{pmatrix} \mathrm{diag}(\frac{l^1_{1,c}f - l^3_{1,c}\tilde{x}}{d_1})D_x + \mathrm{diag}(\frac{l^2_{1,c}f - l^3_{1,c}\tilde{y}}{d_1})D_y \\ \vdots \\ \mathrm{diag}(\frac{l^1_{n,c}f - l^3_{n,c}\tilde{x}}{d_n})D_x + \mathrm{diag}(\frac{l^2_{n,c}f - l^3_{n,c}\tilde{y}}{d_n})D_y \end{pmatrix}_{mn \times m}$$

$$\mathbf{b}_{z_c} = \begin{pmatrix} I_{1,c} + \frac{l^3_{1,c}}{d_1} - \varphi_{1,c} \cdot \rho_c \\ \vdots \\ I_{n,c} + \frac{l^3_{n,c}}{d_n} - \varphi_{n,c} \cdot \rho_c \end{pmatrix}_{mn \times 1}$$

Finally, we stack $\mathbf{A}_{z_c}$ and $\mathbf{b}_{z_c}$ for each channel $c \in \{R, G, B\}$:

$$\mathbf{A}_z = \begin{pmatrix} \mathbf{A}_{z_R} \\ \mathbf{A}_{z_G} \\ \mathbf{A}_{z_B} \end{pmatrix}, \quad \mathbf{b}_z = \begin{pmatrix} \mathbf{b}_{z_R} \\ \mathbf{b}_{z_G} \\ \mathbf{b}_{z_B} \end{pmatrix}$$

$$\min_{z} \|\mathbf{A}_z z - \mathbf{b}_z\|_2^2 + \lambda_z \|z - z_0\|_2^2$$

Different Lighting conditions

Runtime for proposed method