Uncalibrated Photometric Stereo Meets Depth Super-Resolution

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

ICCV Workshop Submission

July 18, 2017

Outline

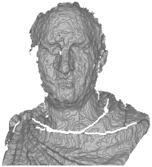
Outline

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

Lambert's Law

Intensity I

$=$

Albedo ρ

$\times$

Shading S

Lambert's Law

$=$

$\times$

$\cdot$

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

Lambert's Law

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

Lambert's Law

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

Super-Resolution Model

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

- 1280×960 RGB image
- 640×480 depth image z_0

$\Longrightarrow$ 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:

$$z_0^i = Kz + \varepsilon_z^i, \quad \forall i \in \{1,\dots,n\},$$

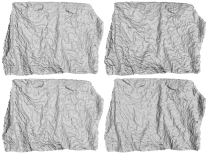
where $\varepsilon_I^i \sim \mathcal{N}(0,\sigma_I^2)$ and $\varepsilon_z^i \sim \mathcal{N}(0,\sigma_z^2)$

Outline

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 SR RGBs w/ different lighting



n LR depth maps of depth sensor

Variational framework

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

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

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

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

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

Variational framework

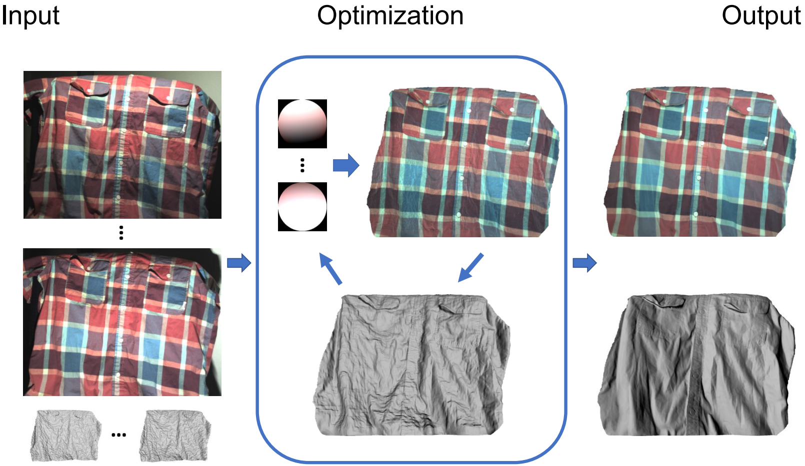
Plugging this together with the standard super-resolution minimization approach we end up minimizing

$$\min_{\substack{z\colon \Omega_{HR}\to\mathbb{R}\\ \rho\colon \Omega_{HR}\to\mathbb{R}\\ \{\mathbf{s}^i\in\mathbb{R}^4\}_i}} \sum_{i=1}^{n} \left\{ \lambda \left\| \mathbf{A}^i(z,\rho,\mathbf{s}^i) \begin{bmatrix} \nabla z \\ z \end{bmatrix} - b^i(\rho,\mathbf{s}^i) \right\|^2_{\mathcal{L}^2(\Omega_{HR})} + \left\| Kz - z_0^i \right\|^2_{\mathcal{L}^2(\Omega_{LR})} \right\},$$

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

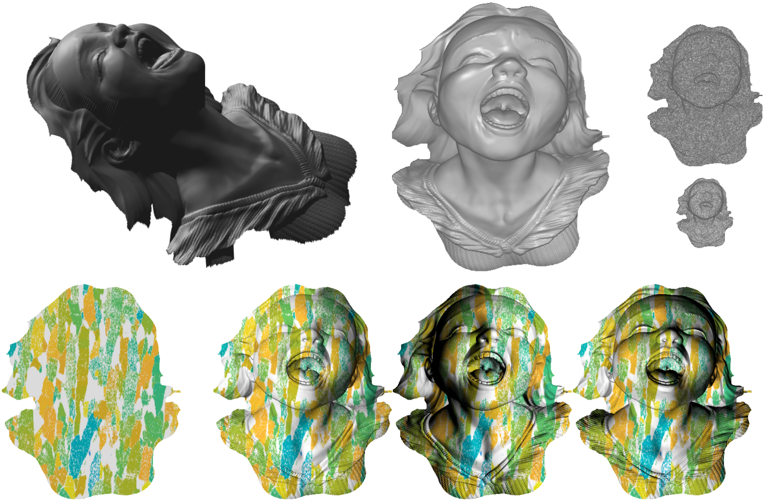
Minimization is done in an alternating scheme.

Workflow

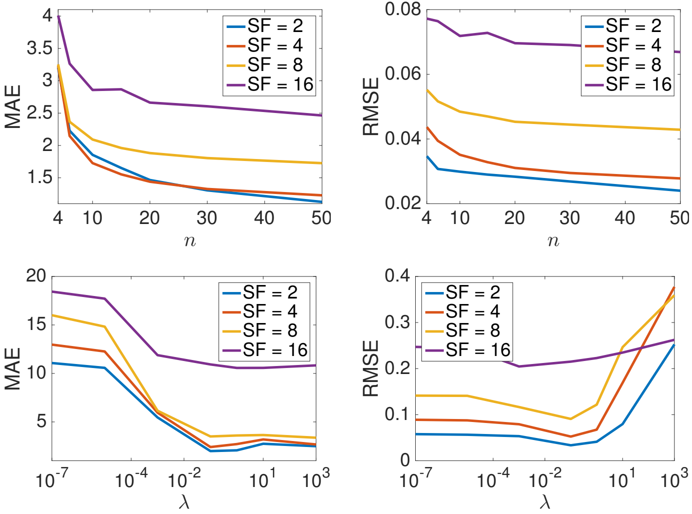


Outline

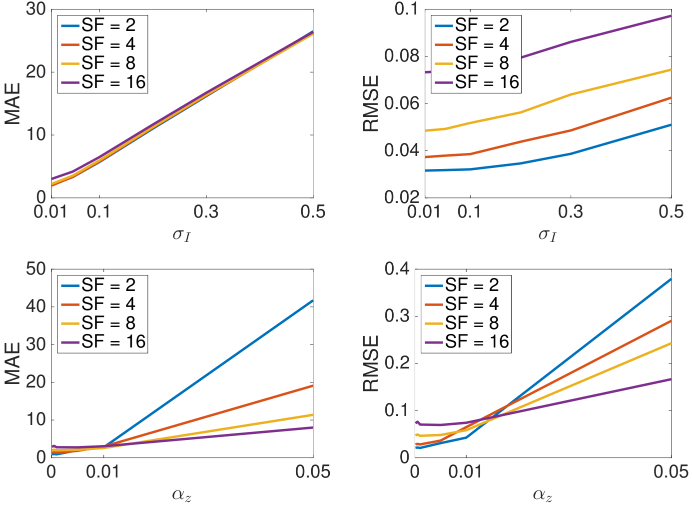
Synthetic Evaluation



error vs. *n* & error vs. λ



Error vs. noise



Comparison with other methods

(a) ID-SR: RMSE $= 0.1237$, MAE $= 38.9402$

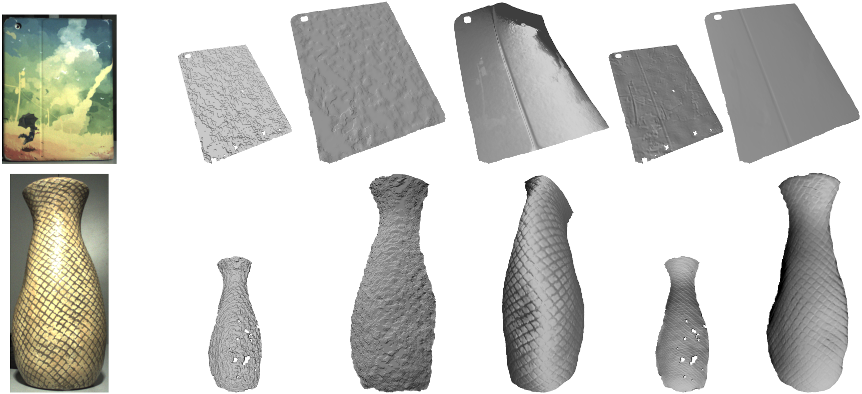
(b) Papadhimitri & Favaro [22]: RMSE $= 0.9199$, MAE $= 41.8041$

(c) Or-El et al [21]: RMSE $= 0.1655$, MAE $= 38.9316$

(d) ours: RMSE $= \mathbf{0.03139}$, MAE $= \mathbf{1.4528}$

Real-world data

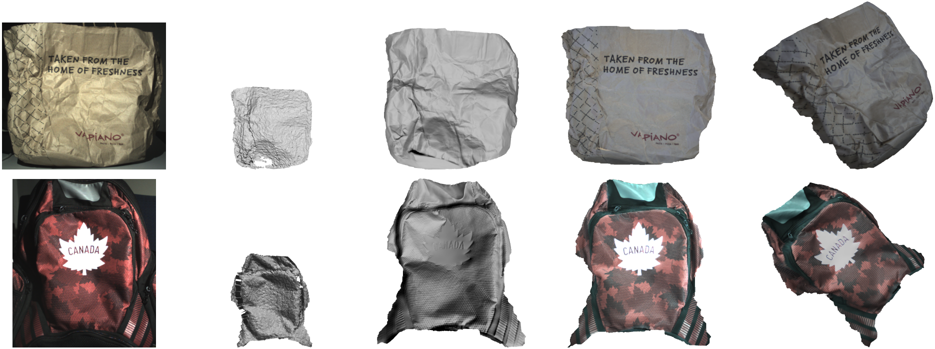
Upscaling by factor of $\xi = 2$ using $n = 20$



(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