Depth Super-Resolution Meets Uncalibrated Photometric Stereo

Songyou PENG Bjoern HAEFNER Yvain QUEAU Daniel CREMERS

Computer Vision Group

Technical University of Munich

ICCV 2017 Color and Photometry in Computer Vision Workshop

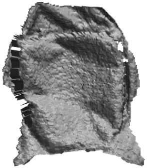
Outline

Outline

Problem Statement

Example: RGB-D data from ASUS Xtion Pro Live

Input RGB image



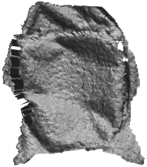
Input depth

+ Good quality
+ High resolution

– Noisy & missing areas
– Low resolution

Goal

Input RGB image



Input depth

Refined depth

Objective:
Use high-resolution photometric clues in the RGB image to turn the low-resolution depth maps into a refined, high resolution one

Contribution

Propose a novel variational model to:

- disambiguate depth super-resolution through high-resolution photometric clues;
- disambiguate uncalibrated photometric stereo through low-resolution depth cues.

Outline

Background

Depth Super-Resolution

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

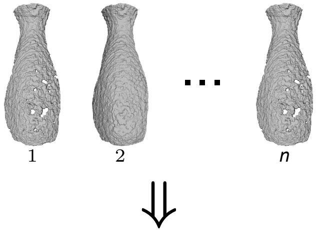
z_0^i: input LR depth maps

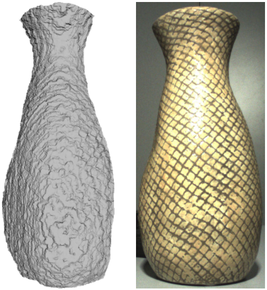
z: output HR depth map

K: down-sampling kernel

ε_z^i: noise $\sim \mathcal{N}(0, {\sigma_z}^2)$

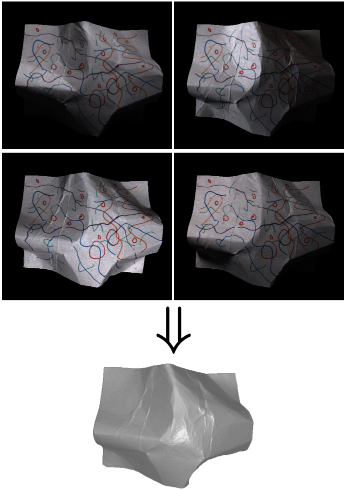
$$\min_z \mathcal{R}_z(z) + \frac{1}{2n} \sum_{i=1}^{n} \|Kz - z_0^i\|_{\ell^2}^2$$





Background

[Grosse et al., *ICCV* 2009]



Photometric Stereo

$$I^i = \rho\, \mathbf{l}^i \cdot \begin{bmatrix} \mathbf{n}(z) \\ 1 \end{bmatrix} + \varepsilon_I^i,\ \forall i \in \{1, \dots, n\}$$

I^i: images under various lightings

$\mathbf{l}^i$: lighting vector $\mathbb{R}^4$

ρ: albedo / reflectance

$\mathbf{n}(z)$: surface normal

$$\min_z \mathcal{R}_I(z) + \frac{1}{2n} \sum_{i=1}^{n} \|\rho\, \mathbf{l}^i \cdot \begin{bmatrix} \mathbf{n}(z) \\ 1 \end{bmatrix} - I^i\|_{\ell^2}^2$$

Background

Depth Super-Resolution

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

z_0^i: input LR depths

z: output HR depth

K: down-sampling kernal

ε_z^i: noise $\sim \mathcal{N}(0, {\sigma_z}^2)$

$$\min_z \mathcal{R}_z(z) + \frac{1}{2n} \sum_{i=1}^{n} \|Kz - z_0^i\|_{\ell^2}^2$$

Photometric Stereo

$$I^i = \rho \, \mathbf{l}^i \cdot \begin{bmatrix} \mathbf{n}(z) \\ 1 \end{bmatrix} + \varepsilon_I^i, \ \forall i \in \{1, \dots, n\}$$

I^i: images under various lightings

$\mathbf{l}^i$: lighting vector $\mathbb{R}^4$

ρ: albedo / reflectance

$\mathbf{n}(z)$: surface normal

$$\min_z \mathcal{R}_I(z) + \frac{1}{2n} \sum_{i=1}^{n} \|\rho \, \mathbf{l}^i \cdot \begin{bmatrix} \mathbf{n}(z) \\ 1 \end{bmatrix} - I^i\|_{\ell^2}^2$$

Background

Depth Super-Resolution

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

z_0^i: input LR depths

z: output HR depth

K: down-sampling kernal

ε_z^i: noise $\sim \mathcal{N}(0, {\sigma_z}^2)$

$$\min_z \mathcal{R}_z(z) + \frac{1}{2n}\sum_{i=1}^{n} \|Kz - z_0^i\|_{\ell^2}^2$$

Photometric Stereo

$$I^i = \rho\, \mathbf{l}^i \cdot \begin{bmatrix} \mathbf{n}(z) \\ 1 \end{bmatrix} + \varepsilon_I^i,\ \forall i \in \{1, \dots, n\}$$

I^i: images under various lightings

$\mathbf{l}^i$: lighting vector $\mathbb{R}^4$

ρ: albedo / reflectance

$\mathbf{n}(z)$: surface normal

$$\min_z \mathcal{R}_I(z) + \frac{1}{2n}\sum_{i=1}^{n} \|\rho\, \mathbf{l}^i \cdot \begin{bmatrix} \mathbf{n}(z) \\ 1 \end{bmatrix} - I^i\|_{\ell^2}^2$$

Proposed Model: $\min_z \frac{1}{2n}\sum_{i=1}^{n} \left\{ \|Kz - z_0^i\|_{\ell^2}^2 + \lambda \|\rho\, \mathbf{l}^i \cdot \begin{bmatrix} \mathbf{n}(z) \\ 1 \end{bmatrix} - I^i\|_{\ell^2}^2 \right\}$

Outline

Methodology

With $(i, \star, \mathbf{p})$ the indices of images, channel and pixel,

$$I_\star^i(\mathbf{p}) = \rho_\star(\mathbf{p})\,\mathbf{l}_\star^i \cdot \begin{bmatrix} \mathbf{n}(\mathbf{p}) \\ 1 \end{bmatrix} + \varepsilon_\star^i(\mathbf{p})$$

Methodology

With $(i, \star, \mathbf{p})$ the indices of images, channel and pixel,

$$\left.\begin{aligned} I_\star^i(\mathbf{p}) &= \rho_\star(\mathbf{p})\, \mathbf{l}_\star^i \cdot \begin{bmatrix} \mathbf{n}(\mathbf{p}) \\ 1 \end{bmatrix} + \varepsilon_\star^i(\mathbf{p}) \\ \mathbf{n}(\mathbf{p}) &= \tfrac{1}{d(z)(\mathbf{p})} \begin{bmatrix} f\, \nabla z(\mathbf{p}) \\ -z(\mathbf{p}) - \nabla z(\mathbf{p}) \cdot \left(\mathbf{p} - \mathbf{p}^0\right) \end{bmatrix} \end{aligned}\right\}$$

f: focal length
$\mathbf{p}^0$: principal point
$d(z)$: normalizer

Methodology

With $(i, \star, \mathbf{p})$ the indices of images, channel and pixel,

$$\left.\begin{aligned} I_\star^i(\mathbf{p}) &= \rho_\star(\mathbf{p})\, \mathbf{l}_\star^i \cdot \begin{bmatrix} \mathbf{n}(\mathbf{p}) \\ 1 \end{bmatrix} + \varepsilon_\star^i(\mathbf{p}) \\ \mathbf{n}(\mathbf{p}) &= \tfrac{1}{d(z)(\mathbf{p})} \begin{bmatrix} f\, \nabla z(\mathbf{p}) \\ -z(\mathbf{p}) - \nabla z(\mathbf{p}) \cdot \left(\mathbf{p} - \mathbf{p}^0\right) \end{bmatrix} \end{aligned}\right\} \mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} = \mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i) + \varepsilon^i$$

f: focal length
$\mathbf{p}^0$: principal point
$d(z)$: normalizer

Proposed Variational Model

Here we have:

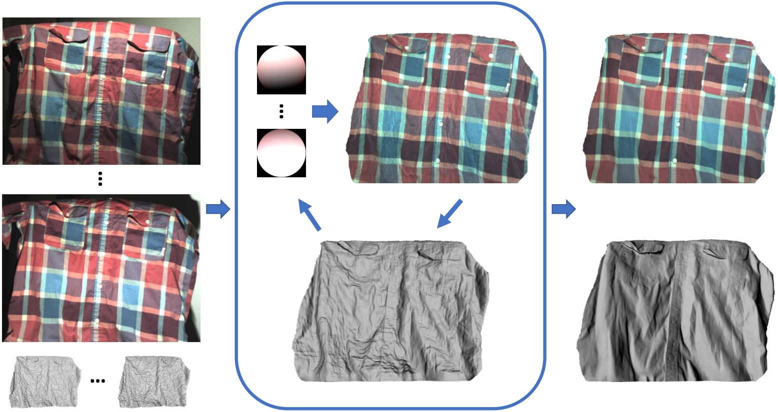
- depth super-resolution cue: $z_0^i = Kz + \varepsilon_z^i,\ \forall i \in \{1, \dots, n\}$
- photometric stereo cue: $\mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} = \mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i) + \boldsymbol{\varepsilon}^i$

The final variational model is acquired from maximum likelihood:

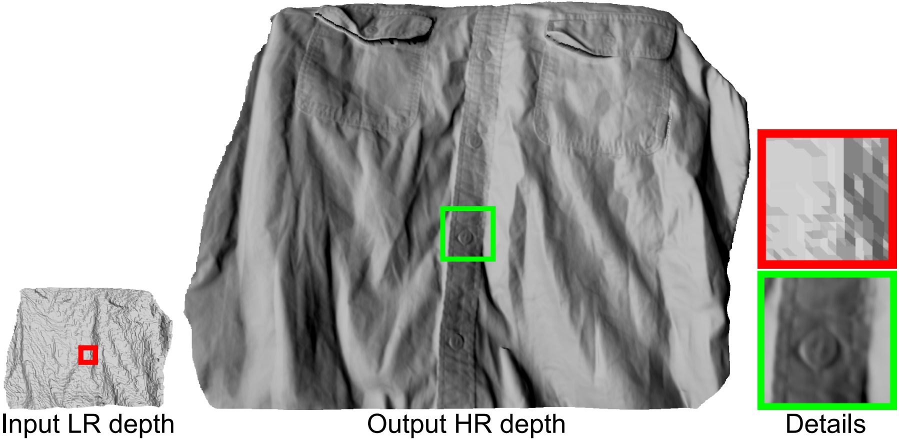
$$\min_{z, \boldsymbol{\rho}, \{\mathbf{l}^i\}_i} \left\{ \sum_{i=1}^{n} \|Kz - z_0^i\|_{\ell^2}^2 + \lambda \sum_{i=1}^{n} \left\| \mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} - \mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i) \right\|_{\ell^2}^2 \right\}$$

Alternating Optimization Workflow

$$\min_{z,\boldsymbol{\rho},\{\mathbf{l}^i\}_i} \left\{ \sum_{i=1}^{n} \|Kz - z_0^i\|_{\ell^2}^2 + \lambda \sum_{i=1}^{n} \left\| \mathbf{A}^i(z,\boldsymbol{\rho},\mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} - \mathbf{b}^i(\boldsymbol{\rho},\mathbf{l}^i) \right\|_{\ell^2}^2 \right\}$$



Alternating Optimization Workflow

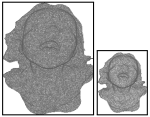


Outline

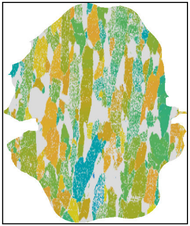
Synthetic Data

3D shape

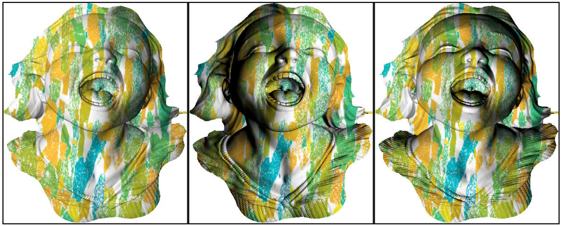
Ground truth HR depth



LR noisy depth



HR albedo map[1]

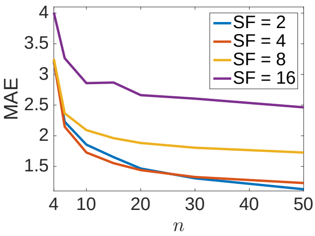


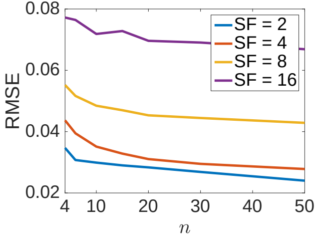
HR photometric stereo images

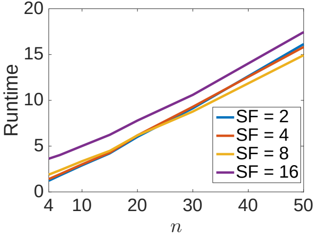
[1] Source: `https://mtex-toolbox.github.io/files/doc/EBSDSpatialPlots.html`

Quantitative Evaluation

Number of images





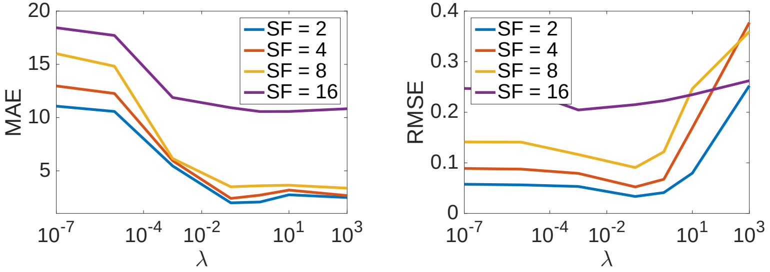


n $\in [10, 30]$ is a good compromise between accuracy and speed

Quantitative Evaluation

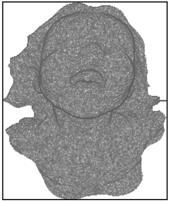
Parameter tuning

$$\min_{z,\rho,\{\mathbf{l}^i\}_i} \left\{ \sum_{i=1}^{n} \|Kz - z_0^i\|_{\ell^2}^2 + \lambda \sum_{i=1}^{n} \left\| \mathbf{A}^i(z,\rho,\mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} - \mathbf{b}^i(\rho,\mathbf{l}^i) \right\|_{\ell^2}^2 \right\}$$



$\lambda \in [10^{-2}, 10^{1}]$ provide satisfactory results

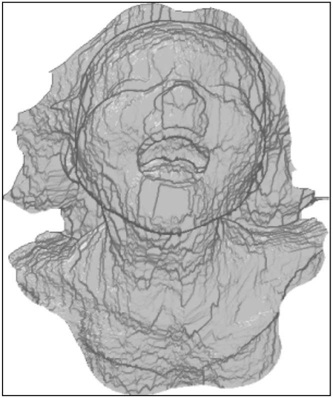
Quantitative Evaluation



RMSE $= 0.0579$

MAE $= 65.7150$

Input depth

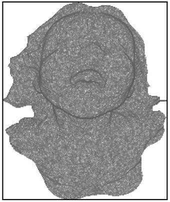


RMSE $= 0.0728$

MAE $= 34.4129$

Depth super-resolution with TV

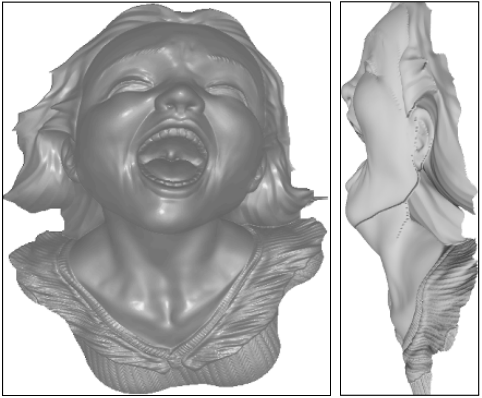
Quantitative Evaluation



$\text{RMSE} = 0.0579$

$\text{MAE} = 65.7150$

Input depth

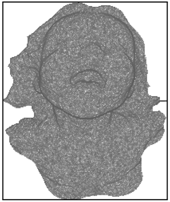


$\text{RMSE} = 0.9199$

$\text{MAE} = 41.8041$

LDR Photometric Stereo
[Papadhimitri and Favaro, *IJCV* 2014]

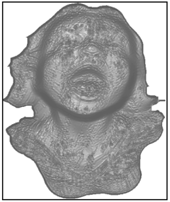
Quantitative Evaluation



RMSE $= 0.0579$

MAE $= 65.7150$

Input depth



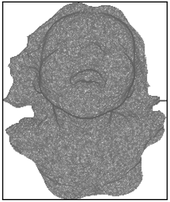
RMSE $= 0.1655$

MAE $= 38.9316$

RGBD-Fusion

[Or-El et al., *CVPR* 2015]

Quantitative Evaluation



RMSE $= 0.0579$

MAE $= 65.7150$

Input depth

RMSE $= \mathbf{0.0314}$

MAE $= \mathbf{1.45280}$

Ours

Quantitative Evaluation

	RMSE	MAE
Input	0.0579	65.7150
Depth SR	0.0728	34.4129
LDR PS	0.0919	41.8041
RGBD-Fusion	0.1655	38.9316
Ours	**0.0314**	**1.45280**

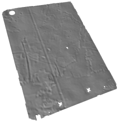
Qualitative Evaluation

Input RGB image

Input depth

Depth SR

LDR-PS

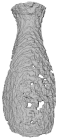


RGBD-Fusion

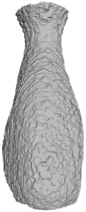
Ours

Qualitative Evaluation

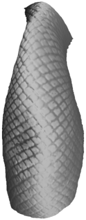
Input RGB image



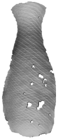
Input depth



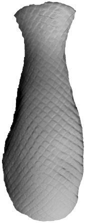
Depth SR



LDR-PS



RGBD-Fusion



Ours

Outline

Conclusions and Future work

- We proposed a novel variational framework for joint depth super-resolution and reflectance/light estimation
- Our method can be used out-of-the-box with common devices
- Theoretical analysis of this approach will be the next step

Data and codes are available on
`https://github.com/pengsongyou/SRmeetsPS`

Depth Super-Resolution Meets Uncalibrated Photometric Stereo

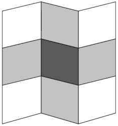
Songyou Peng Bjoern Haefner Yvain Queau Daniel Cremers

Computer Vision Group

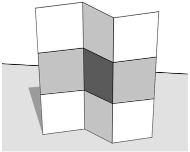
Technical University of Munich

ICCV 2017 Color and Photometry in Computer Vision Workshop

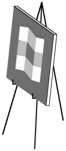
Shape from shading ambiguity



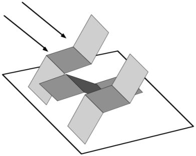
(a) An image



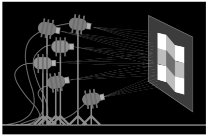
(b) A possible explanation



(c) painter's



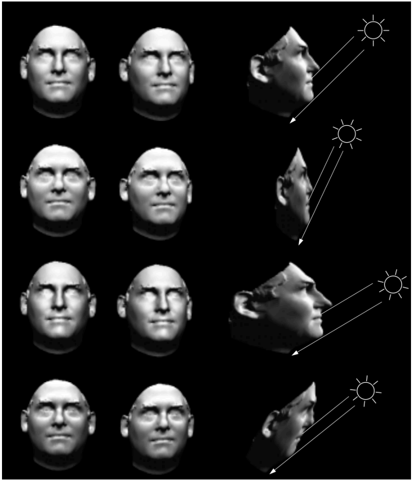
(d) sculptor's



(e) Lighting designer's

[Adelson and Pentland, 1996]

Generalized Bas-Relief (GBR)



[belhumeur et al., IJCV 99]