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

Songyou Peng Yvain Quéau Daniel Cremers

Computer Vision Group, Technical University of Munich

Problem Statement

Depth images from RGB-D cameras:

- Noisy
- No fine details
- Missing areas

Goal: Improve the quality of depths

Contributions

- Propose a novel RGB ratio model to resolve the nonlinearity and achieve similar accuracy to the previous methods.
- Introduce a robust multi-light method which outperforms the state-of-the-art approaches without any regularization term.
- First depth super-resolution method based on photometric stereo.

Reflectance Model

Lambertian Reflectance Model

Intensity I [1] = Albedo ρ × Shading S

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

$\mathbf{l}$: light direction, $\mathbf{n}$: surface normal.

1st-order Spherical Harmonics (SH)

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

φ: ambient light parameter. SH model accounts for 87.5% real-world illumination.

Surface Normal

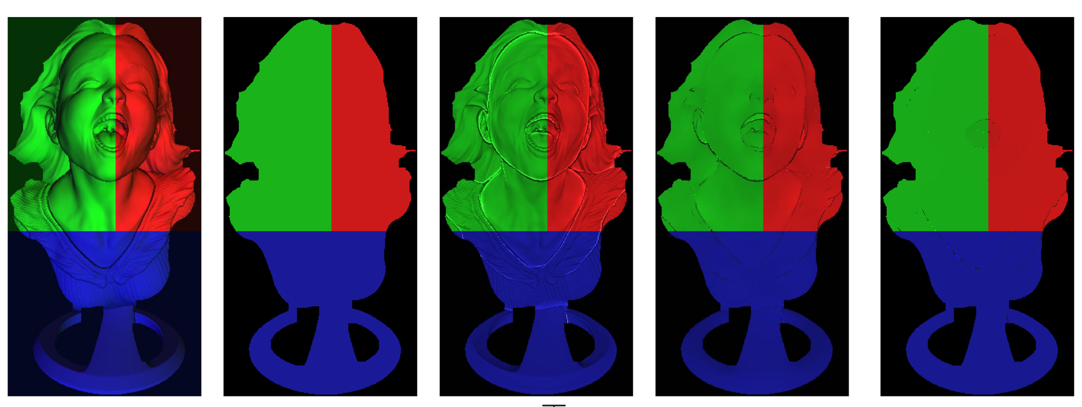
$$\mathbf{n} = \frac{1}{\sqrt{|\nabla z|^2 + 1}} \begin{pmatrix} \nabla z \\ -1 \end{pmatrix}$$

Our goal is to acquire z from the SH model.

Overall Energy

$$E(\mathbf{s}, \rho, z) = E_{\text{data}}(z) + E_{\text{shading}}(\mathbf{s}, \rho, z) + E_{\text{regu}}$$

- $E_{\text{data}}(z) = \lambda_z \|z - z_0\|_2^2$ depth data term
- $E_{\text{shading}}(\mathbf{s}, \rho, z)$ varies from methods
- E_{regu} regularization imposed on ρ or z



I ρ $I/(\mathbf{s}^\top \tilde{\mathbf{n}})$ Fusion [2] Ratio model

References

[1] R. Grosse *et al.* Ground Truth Dataset and Baseline Evaluations for Intrinsic Image Algorithms. In *ICCV '09*

[2] R. Or-El *et al.* RGBD-Fusion: Real Time High Precision Depth Recovery. In *CVPR '15*

[3] Y. Han *et al.* High Quality Shape from a Single RGB-D Image under Uncalibrated Natural Illumination. In *ICCV '13*

RGBD-Fusion Like Method

Modification of RGBD-Fusion method [2].

$$E_{\text{shading}} = \|I - \rho\, \mathbf{s}^\top \tilde{\mathbf{n}}\|_2^2$$
$$E_{\text{regu}} = \lambda_\rho \|\textstyle\sum_{k \in \mathcal{N}} \omega_k (\rho - \rho_k)\|_2^2 + \lambda_l \|\Delta z\|_2^2$$

Input depth [3] Pre-processing Refined depth

RGB Ratio Model

Red and Green channel:

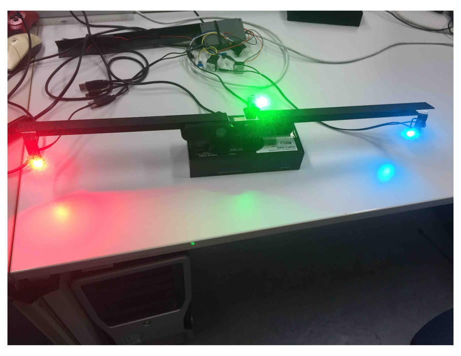
$$\frac{I_R - \rho_R \varphi_R}{I_G - \rho_G \varphi_G} = \frac{\rho_R \mathbf{l}_R^\top \mathbf{n}}{\rho_G \mathbf{l}_G^\top \mathbf{n}}$$

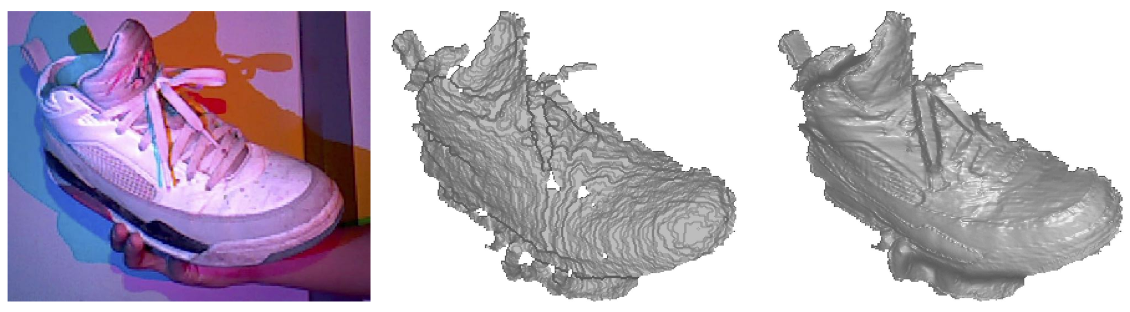
Nonlinearity has been resolved.

$$\rho_G (I_R - \rho_R \varphi_R) \mathbf{l}_G^\top \mathbf{n} - \rho_R (I_G - \rho_G \varphi_G) \mathbf{l}_R^\top \mathbf{n} = 0$$
$$\rho_B (I_G - \rho_G \varphi_G) \mathbf{l}_B^\top \mathbf{n} - \rho_G (I_B - \rho_B \varphi_B) \mathbf{l}_G^\top \mathbf{n} = 0$$
$$\rho_R (I_B - \rho_B \varphi_B) \mathbf{l}_R^\top \mathbf{n} - \rho_B (I_R - \rho_R \varphi_R) \mathbf{l}_B^\top \mathbf{n} = 0$$
$$\Rightarrow \mathcal{R}(\rho, z) = 0 \text{ (RGB ratio model)}$$

$$E_{\text{shading}} = \|\mathcal{R}(\rho, z)\|_2^2$$
$$E_{\text{regu}} = \lambda_\rho \|\omega \nabla \rho\|_2^2$$

* We add active lights to emphasize the difference among 3 channels of color images.





Multi-Light Method

Assuming n illumination conditions

$$E_{\text{shading}} = \textstyle\sum_i^n \|I - \rho\, \mathbf{s}_i^\top \tilde{\mathbf{n}}\|_2^2$$
$$E_{\text{regu}} = 0$$

Depth Super-resolution

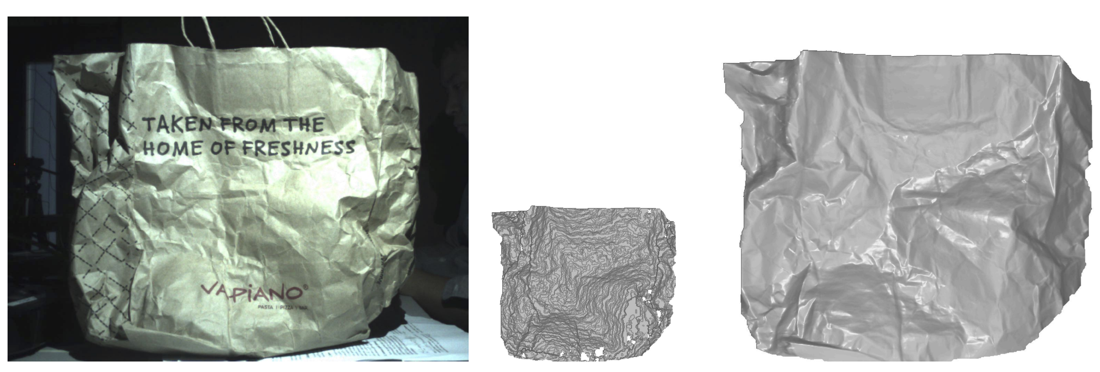
ASUS Xtion Pro Live can provide:

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

Goal: 1280×960 high-quality depth image

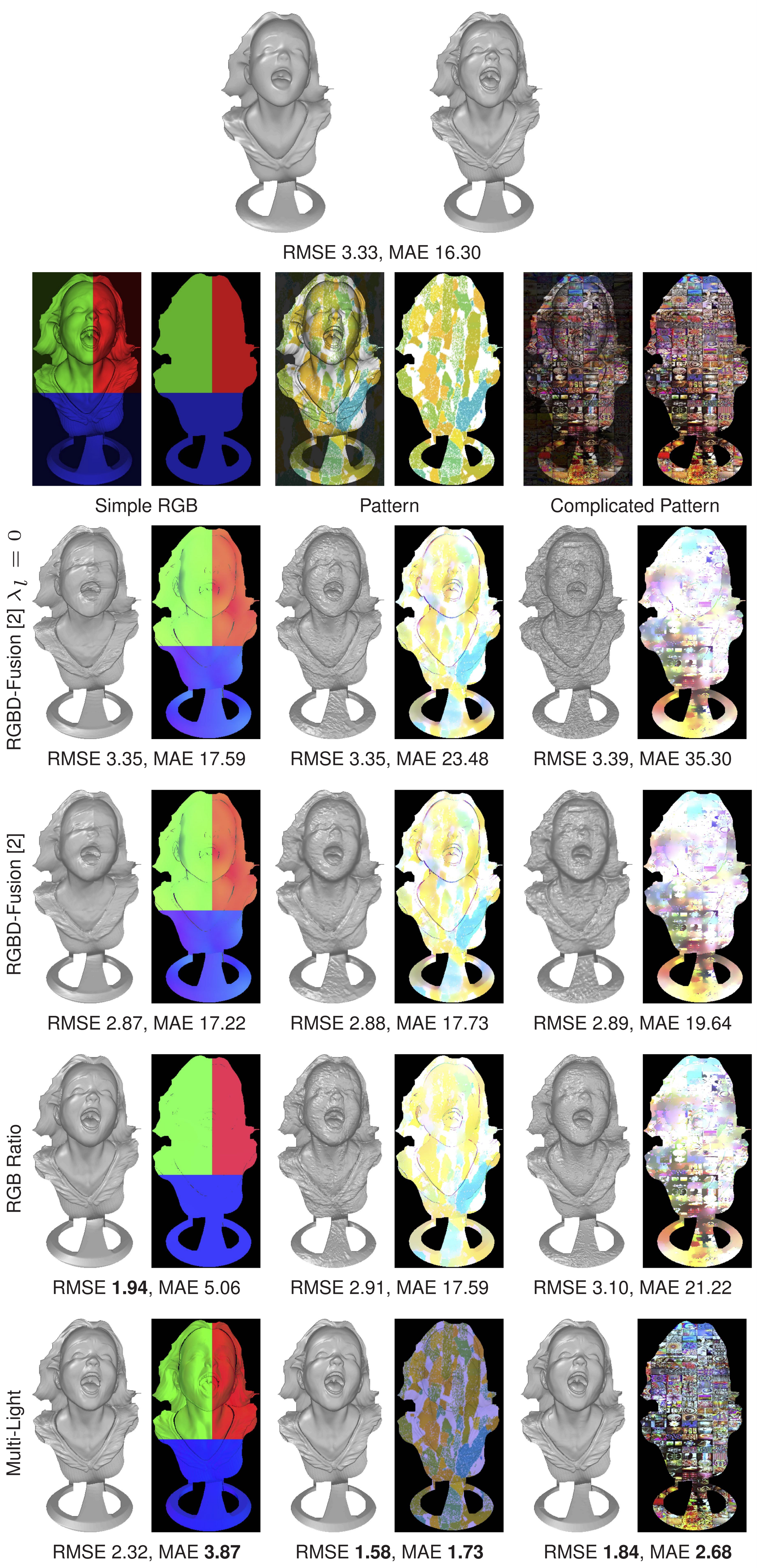
$$E_{\text{data}}(z) = \lambda_z \|Kz - z_0\|_2^2$$

K: downsampling operator.



Results

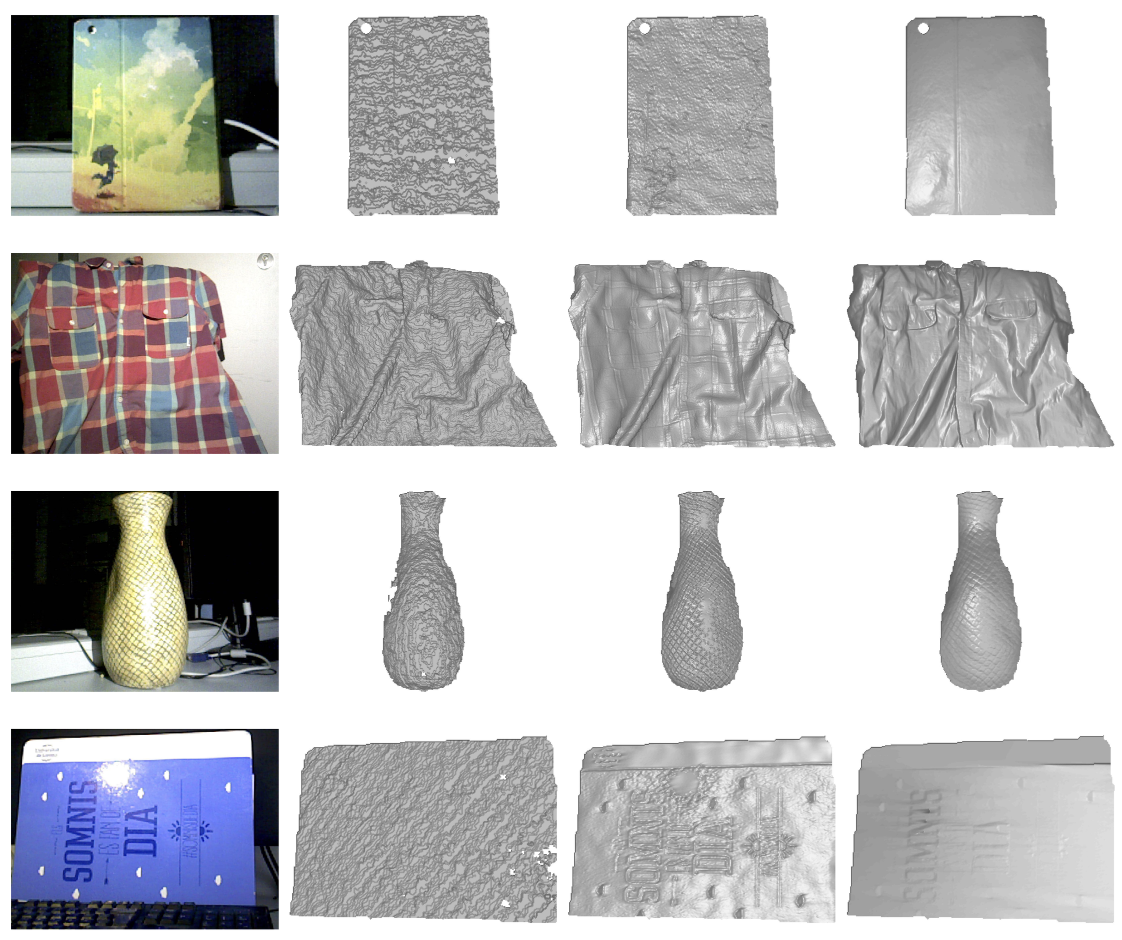
Synthetic Data



RMSE: root mean square error (in mm)

MAE: mean angular error (in degree°)

Real Data



RGB Image Input Depth RGBD-Fusion [2] Multi-Light Method

Acknowledgements

Université de Bourgogne, Universitat de Girona, Heriot Watt University, Erasmus Mundus, Vibot