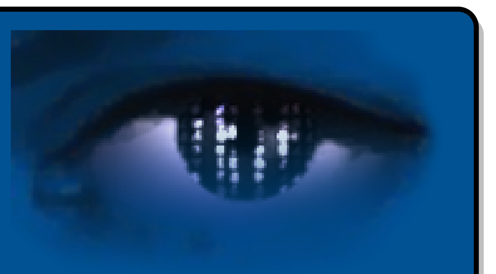
Inferring Super-Resolution Depth from a Moving Light-Source Enhanced RGB-D Sensor: A Variational Approach

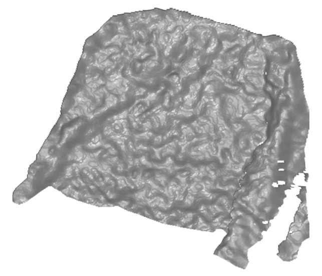
Lu Sang, Bjoern Haefner, Daniel Cremers
Department of Informatics, Technical University of Munich



Problem Statement

- **Issue with depth map from RGB-D sensor:** noise, quantization effects, coarse resolution

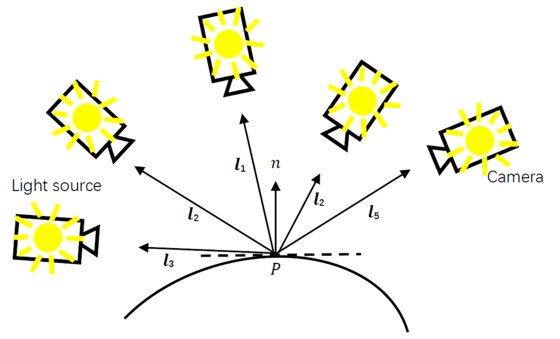
RGB image



input depth

- **Goal:** recover super resolution depth using a light source enhanced moving RGB-D sensor.

Multi-View Photometric Stereo



multi-view PS

data acquisition setup

- An LED light source is attached to a RGB-D sensor.
- Camera and light source move simultaneously.

Background

- Using a first-order spherical harmonic approximation for image formation model [1].

$$\boldsymbol{I} = \boldsymbol{\rho}\,\langle \boldsymbol{l}, \boldsymbol{m}[\boldsymbol{n}[z]] \rangle$$

$$\boldsymbol{m}[\boldsymbol{n}[z]] = \begin{bmatrix} \boldsymbol{n}[z] \\ 1 \end{bmatrix}, \qquad \boldsymbol{n}[z] = \frac{\tilde{\boldsymbol{n}}[z]}{\mathrm{d}\mathcal{A}[z]} = \frac{1}{\sqrt{|f\nabla z|^2 + (-z - \langle \boldsymbol{p}, \nabla z \rangle)^2}} \begin{bmatrix} f\nabla z \\ -z - \langle \boldsymbol{p}, \nabla z \rangle \end{bmatrix}$$

- image $\boldsymbol{I}$
- albedo $\boldsymbol{\rho}$
- lighting $\boldsymbol{l}$
- depth z
- normals $\boldsymbol{n}$
- focal length f
- pixel coords. $\boldsymbol{p}$

- Images from different viewpoints are warped to the reference frame using $\mathbf{R}_i$ and $\mathbf{t}_i$ [3].

$$\tilde{\boldsymbol{I}}_i\,[\mathbf{R}_i, \mathbf{t}_i, z] = \boldsymbol{I} \circ \boldsymbol{w}[\mathbf{R}_i, \mathbf{t}_i, z] = \boldsymbol{I} \circ \pi(\mathbf{T}[\mathbf{R}_i, \mathbf{t}_i]\pi^{-1}(z)), \quad i \in \{1, \dots, n\}$$

- i-th rotation matrix $\mathbf{R}_i$
- i-th translation vector $\mathbf{t}_i$
- warping operator $\boldsymbol{w}$
- projection operator π
- 4×4 transformation matrix $\mathbf{T} = \begin{bmatrix} \mathbf{R}_i & \mathbf{t}_i \\ 0 & 1 \end{bmatrix}$
- re-projection operator π^{-1}

- The low-resolution (LR) depth should match the super-resolution (SR) depth.

$$Dz = z_0$$

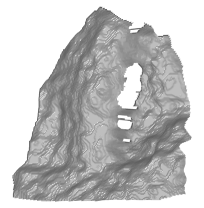
- super-resolution depth z
- linear downsampling operator D
- low-resolution depth z_0

Contribution

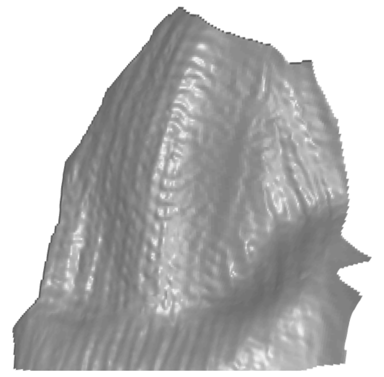
Jointly solve depth super resolution and camera motion estimation

- Simple data acquisition setup.
- No lighting calibration.
- A variational formulation.
- End-to-end frame work.

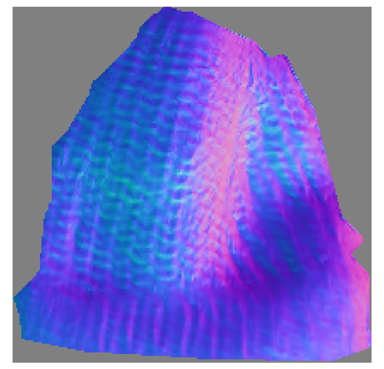
4 out of 20 input images



input depth



refined depth



refined normals

Variational Formulation

$$\min_{\substack{\{\mathbf{R}_i\} \in SO(3),\ \{\mathbf{t}_i\} \in \mathbb{R}^3 \\ z:\Omega_{\mathsf{HR}} \to \mathbb{R},\ \boldsymbol{\rho}:\Omega_{\mathsf{HR}} \to \mathbb{R}^3 \\ \{\boldsymbol{l}_i\} \in \mathbb{R}^4}} \sum_{i=1}^{n} \Phi_{\mathsf{C}} \underbrace{\left(\tilde{\boldsymbol{I}}_i\,[\mathbf{R}_i, \mathbf{t}_i, z] - \boldsymbol{\rho}\,\langle \boldsymbol{l}_i, \boldsymbol{m}[\boldsymbol{n}[z]] \rangle \right)}_{\text{PS term}} + \tau \underbrace{\| Dz - z_0 \|_2^2}_{\text{SR term}}$$

- PS term is evaluated using robust estimator Φ_{C} with $\Phi_{\mathsf{C}}(r) = \frac{\lambda^2}{2}\log(1 + \frac{r^2}{\lambda^2})$.

Numerical Resolution

- Rotation $\mathbf{R}_i$ and translation $\mathbf{t}_i$ are parametrized by the corresponding 6-dim twist coords $\boldsymbol{\xi}_i$, i.e. $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\xi}_i)$, $\mathbf{t}_i = \mathbf{t}_i(\boldsymbol{\xi}_i)$
- Separate normals into linear part by $\tilde{\boldsymbol{n}}[z]$ and nonlinear part $\mathrm{d}\mathcal{A}[z] = \|\tilde{\boldsymbol{n}}[z]\|$ with $\boldsymbol{n}[z] = \frac{\tilde{\boldsymbol{n}}[z]}{\mathrm{d}\mathcal{A}[z]}$
- Solve nonlinear nonconvex problem by alternating reweighted least square.

$$\boldsymbol{\xi}_i^{(k+1)} = \operatorname*{argmin}_{\substack{\boldsymbol{\xi}_i \in \mathbb{R}^6, \\ \forall i \in \{1,\dots,n\}}} \Phi_{\mathsf{C}}^{(k)}\left(\tilde{\boldsymbol{I}}_i \left[\mathbf{R}_i(\boldsymbol{\xi}_i), \mathbf{t}_i(\boldsymbol{\xi}_i), z^{(k)} \right] - \boldsymbol{\rho}^{(k)} \left\langle \boldsymbol{m}\left[\frac{\tilde{\boldsymbol{n}}\left[z^{(k)}\right]}{\mathrm{d}\mathcal{A}\left[z^{(k)}\right]} \right], \boldsymbol{l}_i^{(k)} \right\rangle \right),$$

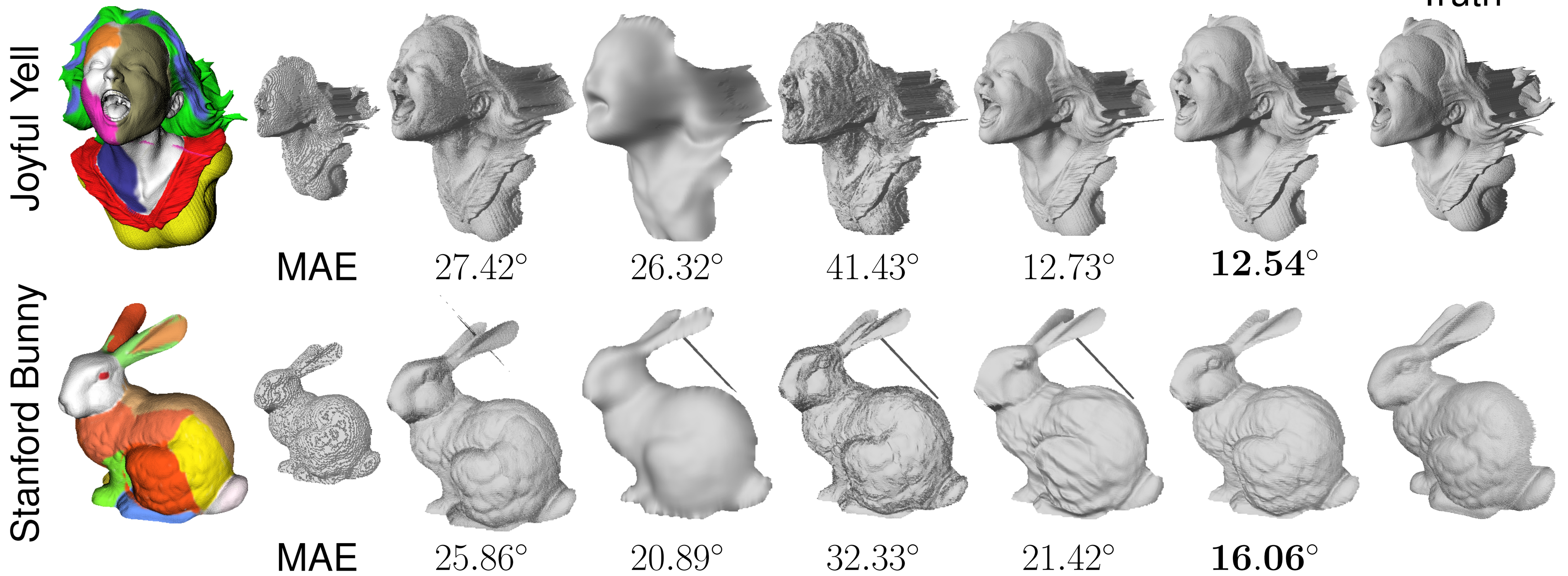
$$z^{(k+1)} = \operatorname*{argmin}_{z:\Omega_{\mathsf{HR}} \to \mathbb{R}} \sum_{i \in \mathcal{I}} \Phi_{\mathsf{C}}^{(k+\frac{1}{4})}\left(\tilde{\boldsymbol{I}}_i \left[\mathbf{R}_i\left(\boldsymbol{\xi}_i^{(k+1)}\right), \mathbf{t}_i\left(\boldsymbol{\xi}_i^{(k+1)}\right), z^{(k)} \right] - \boldsymbol{\rho}^{(k)} \left\langle \boldsymbol{m}\left[\frac{\tilde{\boldsymbol{n}}[z]}{\mathrm{d}\mathcal{A}\left[z^{(k)}\right]} \right], \boldsymbol{l}_i^{(k)} \right\rangle \right) + \tau \| Dz - z_0 \|^2_{\ell_2(\Omega_{\mathsf{LR}})},$$

$$\boldsymbol{\rho}^{(k+1)} = \operatorname*{argmin}_{\boldsymbol{\rho}:\Omega_{\mathsf{HR}} \to \mathbb{R}^3} \sum_{i \in \mathcal{I}} \Phi_{\mathsf{C}}^{(k+\frac{2}{4})}\left(\tilde{\boldsymbol{I}}_i \left[\mathbf{R}_i\left(\boldsymbol{\xi}_i^{(k+1)}\right), \mathbf{t}_i\left(\boldsymbol{\xi}_i^{(k+1)}\right), z^{(k+1)} \right] - \boldsymbol{\rho} \left\langle \boldsymbol{m}\left[\frac{\tilde{\boldsymbol{n}}\left[z^{(k+1)}\right]}{\mathrm{d}\mathcal{A}\left[z^{(k+1)}\right]} \right], \boldsymbol{l}_i^{(k)} \right\rangle \right),$$

$$\boldsymbol{l}_i^{(k+1)} = \operatorname*{argmin}_{\substack{\boldsymbol{l}_i \in \mathbb{R}^4 \\ \forall i \in \{1,\dots,n\}}} \Phi_{\mathsf{C}}^{(k+\frac{3}{4})}\left(\tilde{\boldsymbol{I}}_i \left[\mathbf{R}_i\left(\boldsymbol{\xi}_i^{(k+1)}\right), \mathbf{t}_i\left(\boldsymbol{\xi}_i^{(k+1)}\right), z^{(k+1)} \right] - \boldsymbol{\rho}^{(k+1)} \left\langle \boldsymbol{m}\left[\frac{\tilde{\boldsymbol{n}}\left[z^{(k+1)}\right]}{\mathrm{d}\mathcal{A}\left[z^{(k+1)}\right]} \right], \boldsymbol{l}_i \right\rangle \right).$$

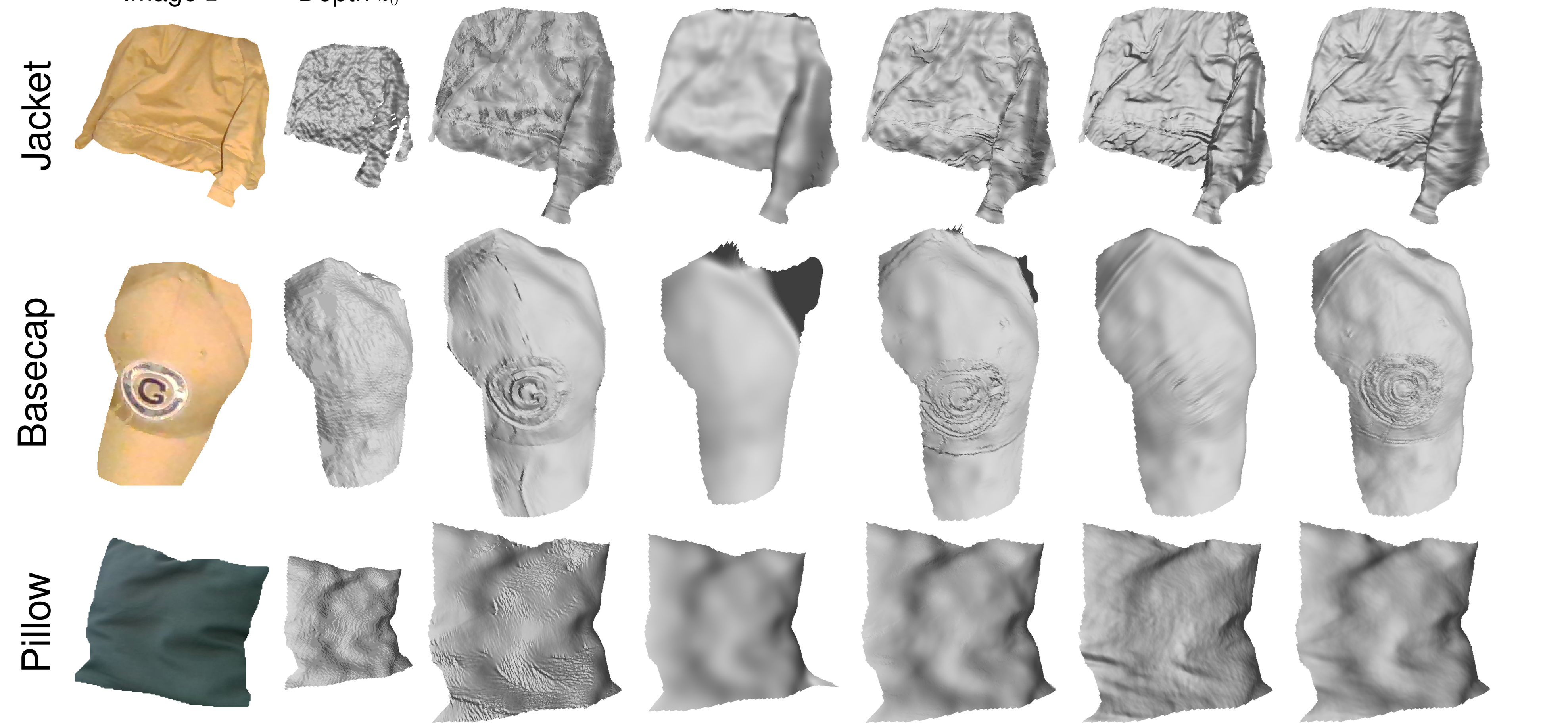
Experimental Results

Synthetic Datasets



	Input Depth z_0	[2]	[4]	[5]	[6]	Ours
Joyful Yell	MAE	27.42°	26.32°	41.43°	12.73°	**12.54°**
Stanford Bunny	MAE	25.86°	20.89°	32.33°	21.42°	**16.06°**

Real-World Datasets



Reference

[1] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, Feb 2003.

[2] B. Haefner, Y. Quéau, T. Möllenhoff, and D. Cremers. Fight ill-posedness with ill-posedness: Single-shot variational depth super-resolution from shading. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[3] C. Kerl, J. Sturm, and D. Cremers. Robust odometry estimation for rgb-d cameras. In *International Conference on Robotics and Automation (ICRA)*, May 2013.

[4] Y. Li, D. Min, M. N. Do, and J. Lu. Fast guided global interpolation for depth and motion. In *ECCV*, 2016.

[5] W. Liu, X. Chen, J. Yang, and Q. Wu. Robust color guided depth map restoration. *IEEE Transactions on Image Processing*, 26(1):315–327, Jan 2017.

[6] S. Peng, B. Haefner, Y. Quéau, and D. Cremers. Depth super-resolution meets uncalibrated photometric stereo. In *International Conference on Computer Vision Workshops (ICCVW)*, 2017.