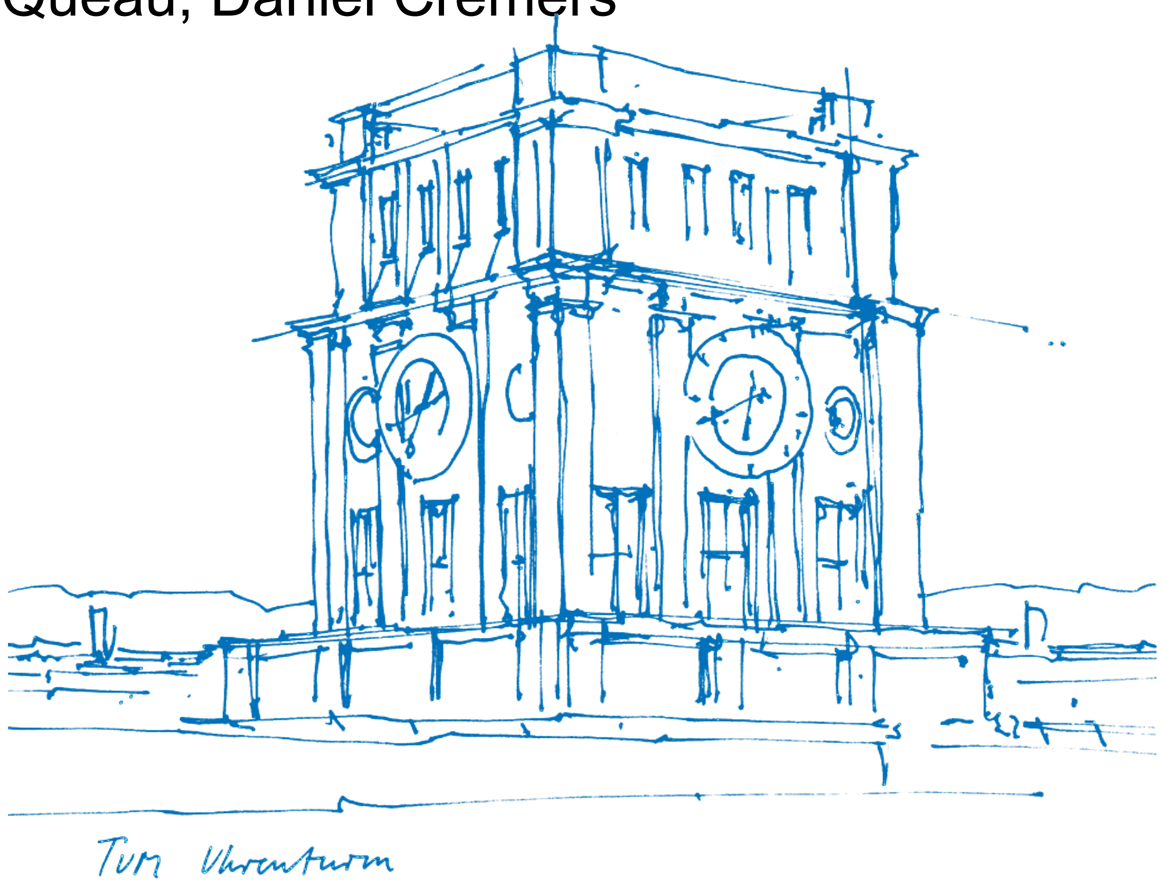
A Seminar Report on
"**Depth Super-Resolution** Meets **Uncalibrated Photometric Stereo**"

Master Seminar "The Evolution of Motion Estimation and Real-time 3D Reconstruction"

Student: Yixuan Liu
Supervisor: Björn Häfner

Paper authors: Songyou Peng, Björn Haefner, Yvain Quéau, Daniel Cremers
Conference: ICCV 2017

Computer Vision Group
Technical University of Munich
Garching, 15. April 2020



Outline

- **Introduction**
- Method description
- Results and evaluation
- Personal comments
- Summary

Introduction

Low-cost RGB-D sensors:

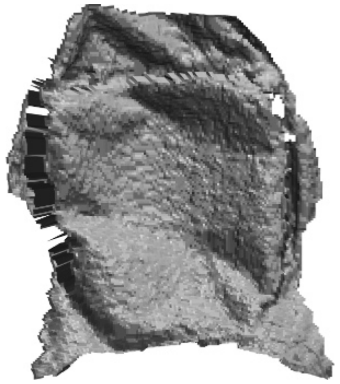
Issue: ***High*** RGB resolution and ***Low*** Depth resolution.

Solutions:

=> Downsample RGB-channel

=> Upsample D-channel

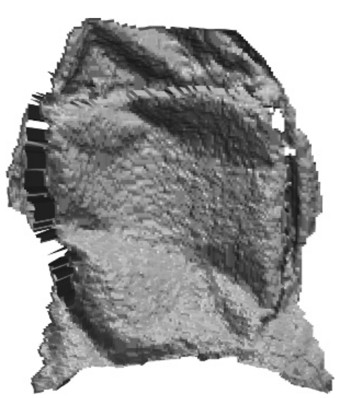
Problem 1: Low-resolutional depth channel

	Low-res.	High-res.
RGB		
Depth		

Given: High-res. RGB channels,
low-res. Depth channel

Source: https://vision.in.tum.de/data/datasets/photometricdepthsr

Problem 1: Low-resolutional depth channel

	Low-res.	High-res.
RGB		
Depth		

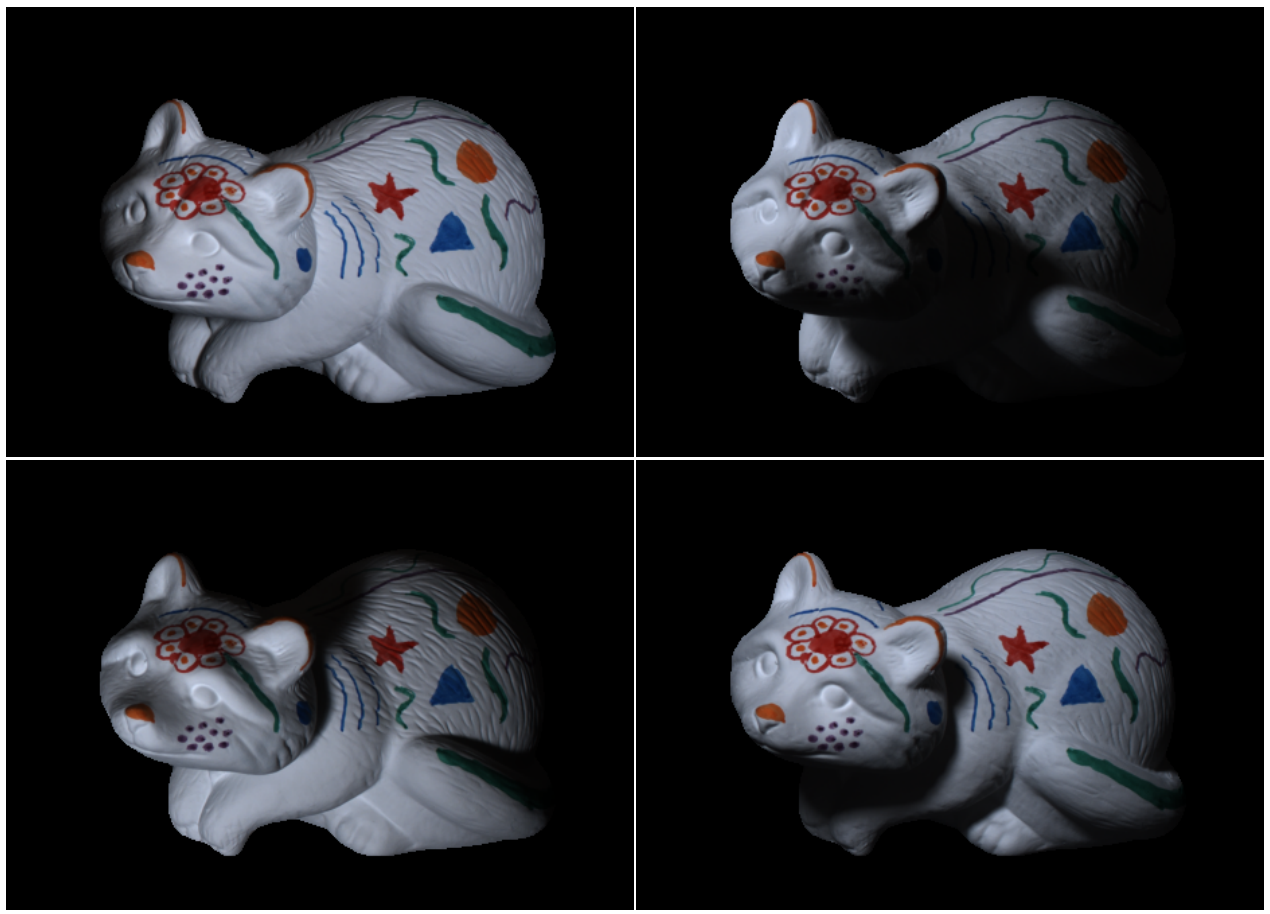
Given: High-res. RGB channels,
low-res. Depth channel

Goal: **Refined, high-res.** depth maps

=> Depth super-resolution

Source: https://vision.in.tum.de/data/datasets/photometricdepthsr

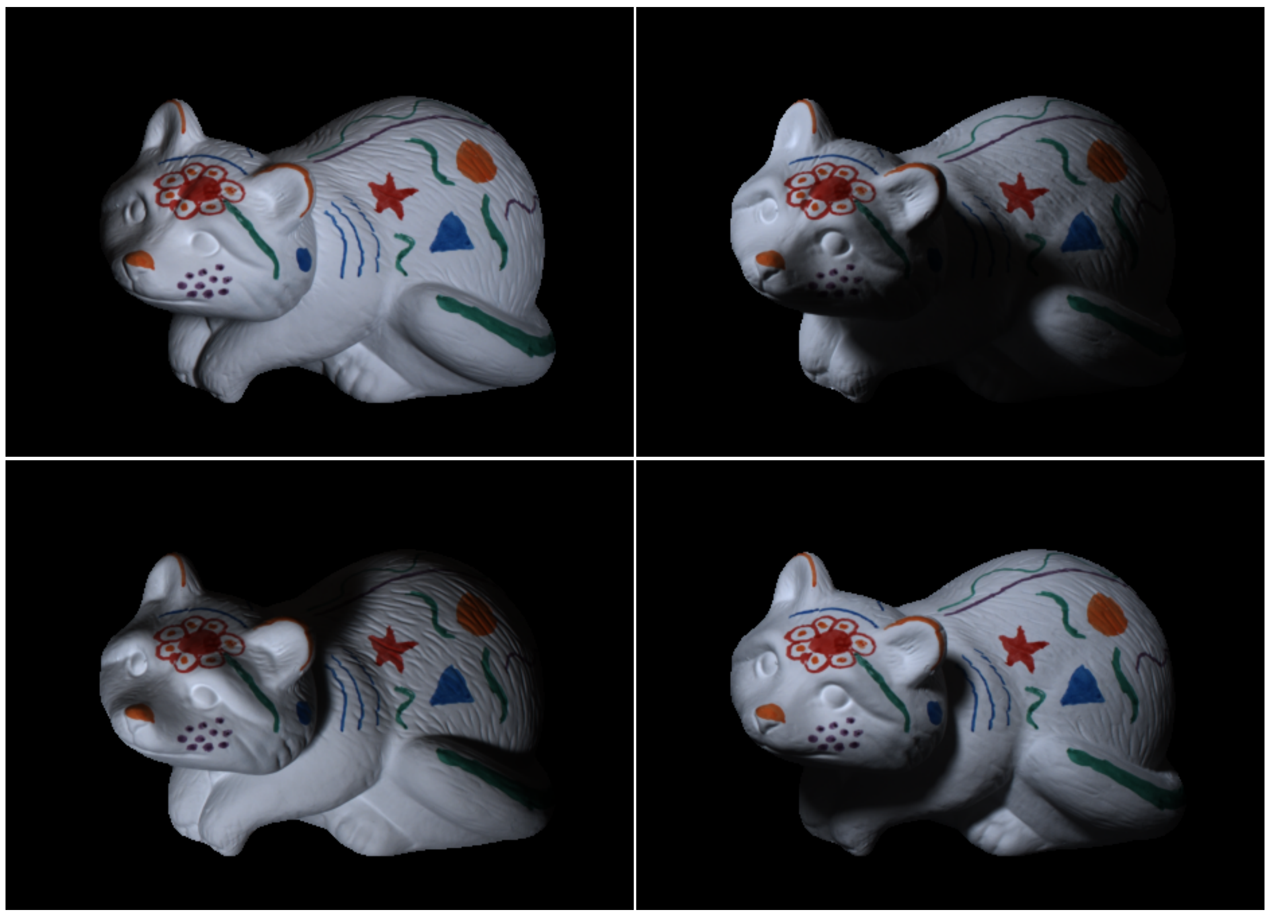
Problem 2: RGB images with various lighting



Source: http://www.cs.toronto.edu/~rgrosse/intrinsic/

Given: multiple differently illuminated RGB images

Problem 2: RGB images with various lighting



Source: http://www.cs.toronto.edu/~rgrosse/intrinsic/

Given: multiple differently illuminated RGB images

Goal: **Geometry** for realistic 3D-reconstruction

=> Uncalibrated photometric stereo

Outline

- Introduction
- **Method description**
- Results and evaluation
- Personal comments
- Summary

Method description

Both problems are simultaneously solved

by combining **depth super-resolution** and **uncalibrated photometric stereo**.

Background

1. **Depth Super-resolution**

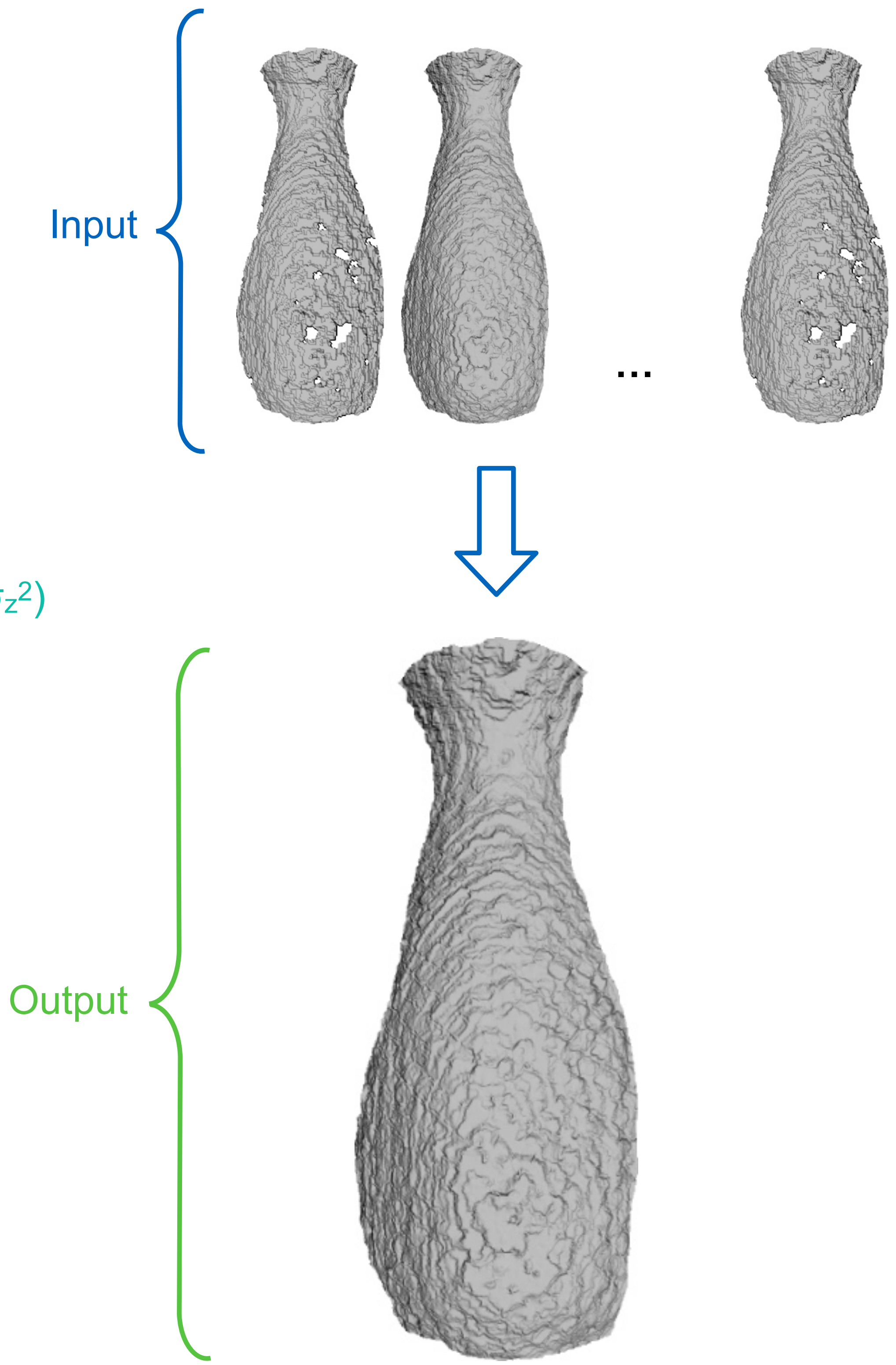
$\forall i \in \{1, \ldots, n\}$: $$z_0^i = Kz + \varepsilon_z^i$$

Background

1. **Depth Super-resolution**

$$\forall i \in \{1, ..., n\} : \quad \boxed{z_0^i} = \boxed{K}\boxed{z} + \boxed{\varepsilon_z^i}$$

Input LR depth maps
Output HR depth map
Down-sampling kernel
Noise ~$N(0, \sigma_z^2)$



Background

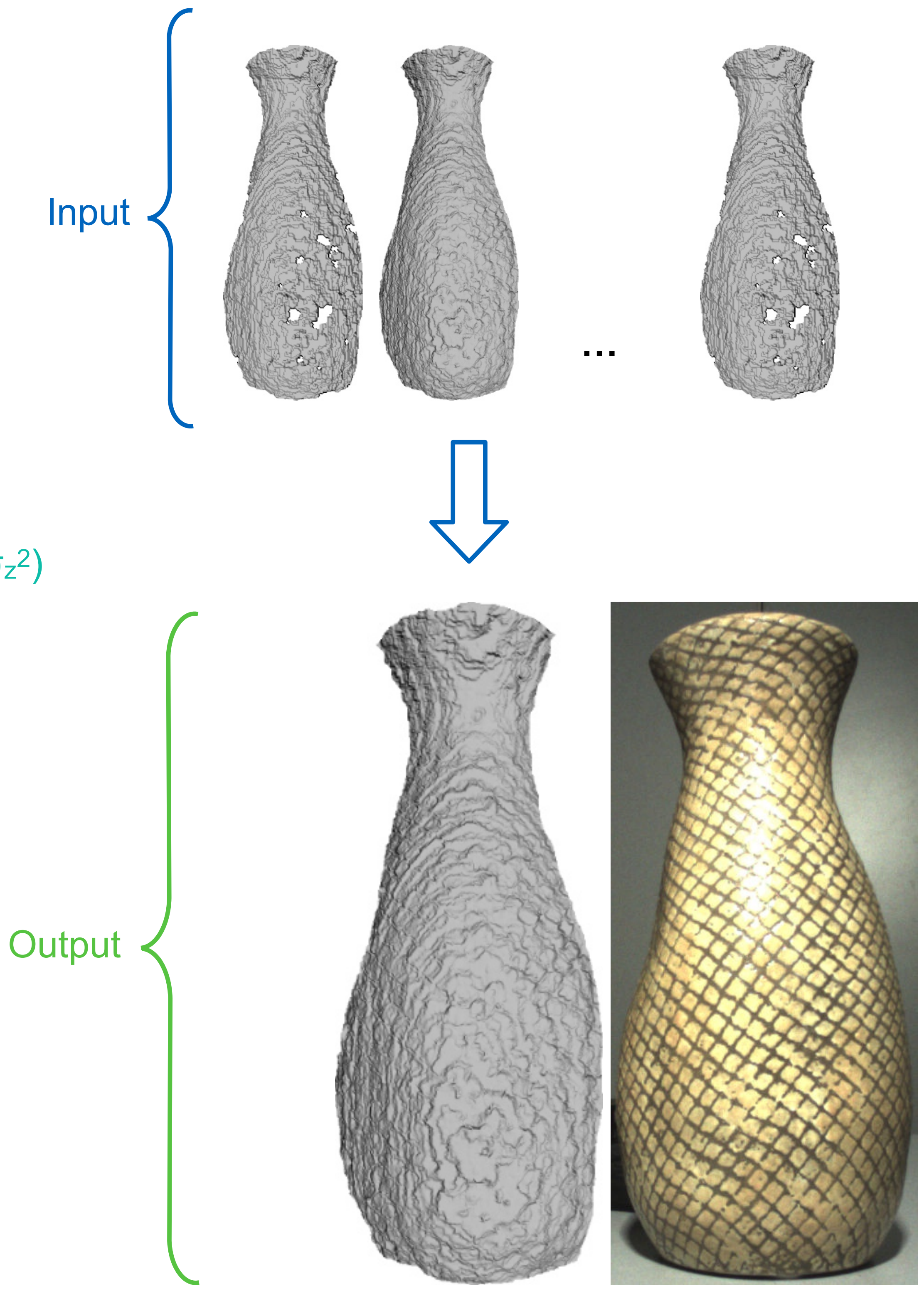
1. **Depth Super-resolution**

$$\forall i \in \{1, \dots, n\}: \quad \boxed{z_0^i} = \boxed{K}\boxed{z} + \boxed{\varepsilon_z^i}$$

Input LR depth maps
Output HR depth map
Down-sampling kernel
Noise $\sim N(0, \sigma_z^2)$

The variational method with regularizer ensures well-posedness:

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



Background

2. **Uncalibrated Photometric Stereo**

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

Albedo
≈ (Diffuse) reflectance
≈ Real color

Background

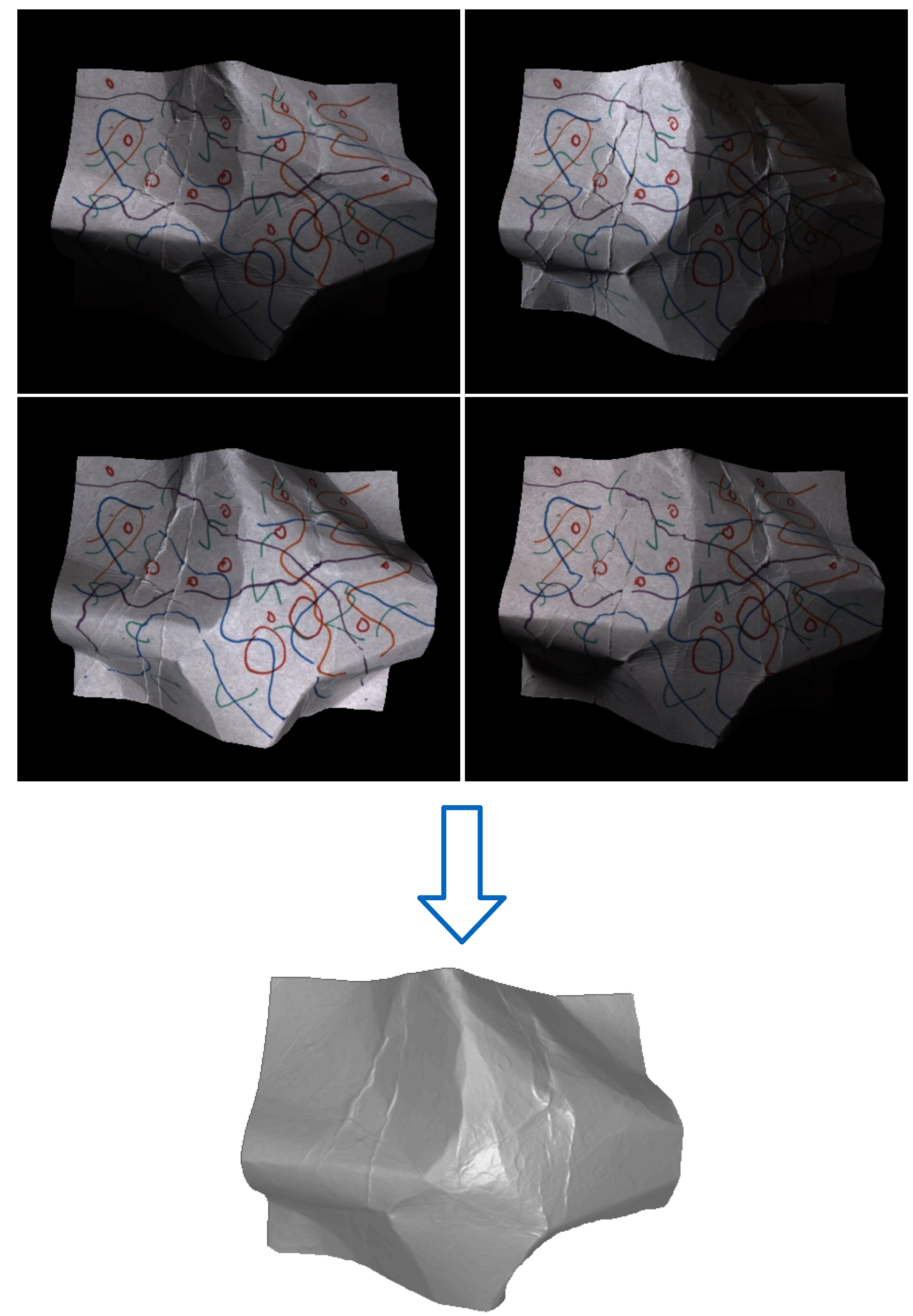
2. **Uncalibrated Photometric Stereo**

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

Images under various lighting — Lighting vector — Noise ~$N(0, \sigma_I^2)$

Albedo — Surface normal

Input



Output

Albedo
≈ (Diffuse) reflectance
≈ Real color

Background

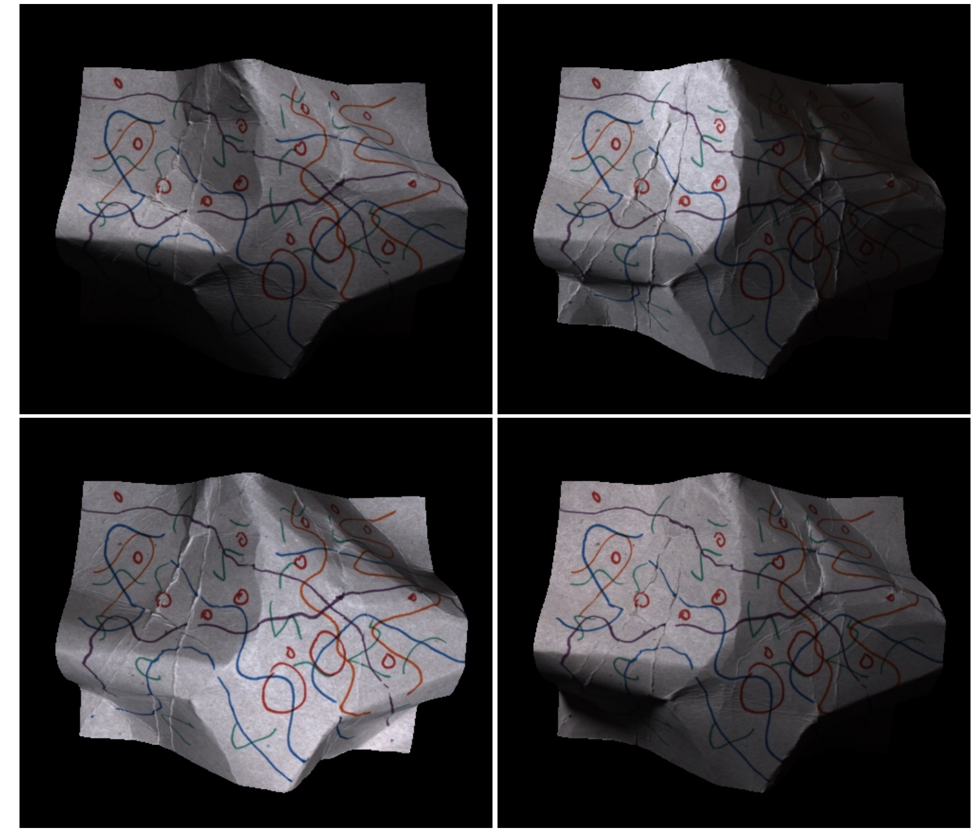
2. **Uncalibrated Photometric Stereo**

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

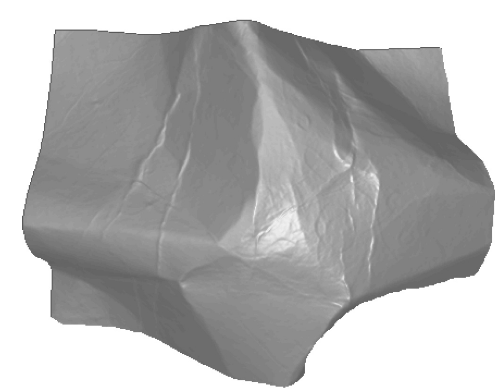
Images under various lighting — Lighting vector — Noise ~N(0, σ_I^2)

Albedo — Surface normal

Input



Output



The variational method with regularizer ensures well-posedness:

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

Methodology

Depth Super-resolution & **Uncalibrated Photometric Stereo**

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

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

Methodology

Depth Super-resolution & **Uncalibrated Photometric Stereo**

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

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

Proposed model:

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

$$\lambda = \frac{{\sigma_z}^2}{{\sigma_I}^2}$$

Methodology

PDE-based **photometric stereo** model

(i, ⋆, p) := (the indices of images, channel, pixel)

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

Images under various lighting Lighting vector Noise ~$N(0, \sigma_z^2)$

Albedo Surface normal

Methodology

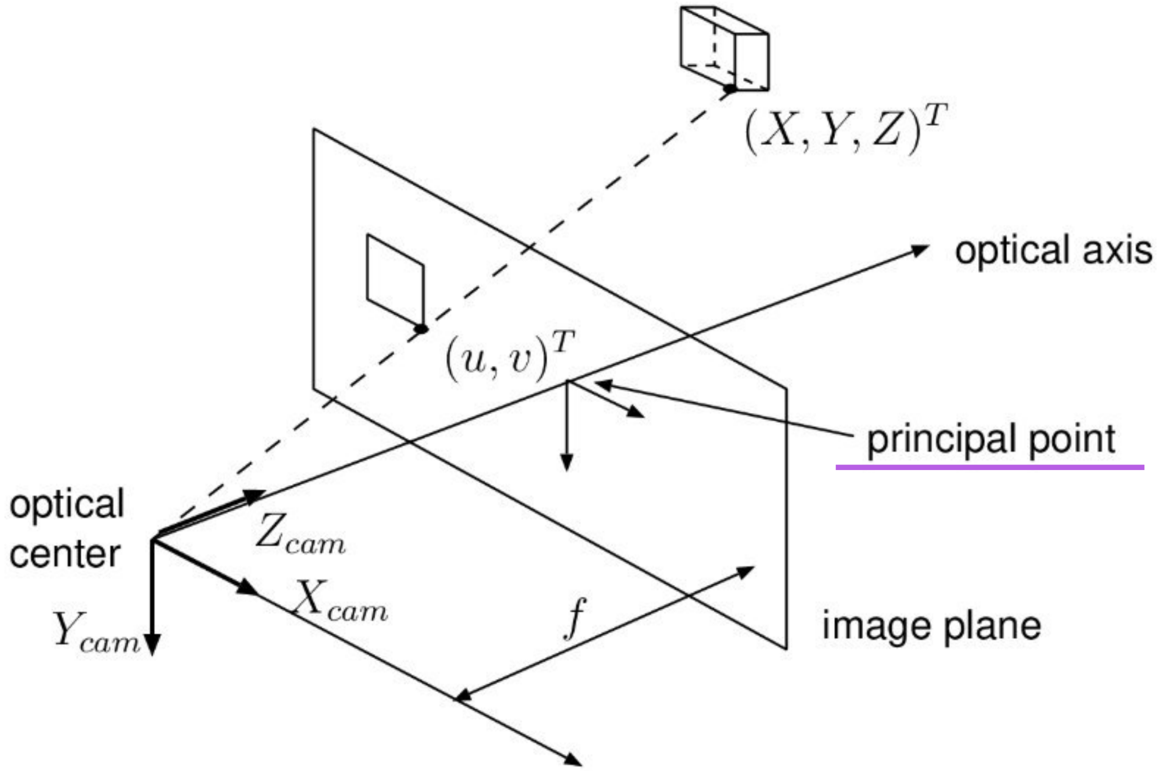
PDE-based **photometric stereo** model

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

Geometry: *Depth map* → *Normal map*

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

Surface normal
Normalizer
Focal length
Depth map
Principal point



Methodology

PDE-based **photometric stereo** model

$$I^i_\star(\mathbf{p}) = \rho_\star(\mathbf{p})\, \mathbf{l}^i_\star \cdot \begin{bmatrix} \boxed{\mathbf{n}(\mathbf{p})} \\ 1 \end{bmatrix} + \varepsilon^i_\star(\mathbf{p})$$

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

A x = b + noise

$$\mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} = \mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i) + \boldsymbol{\varepsilon}^i$$

Methodology

PDE-based **photometric stereo** model

$$I^i_\star(\mathbf{p}) = \rho_\star(\mathbf{p})\, \mathbf{l}^i_\star \cdot \begin{bmatrix} \boxed{\mathbf{n}(\mathbf{p})} \\ 1 \end{bmatrix} + \varepsilon^i_\star(\mathbf{p})$$

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

A x = b + noise

$$\mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} = \mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i) + \boldsymbol{\varepsilon}^i$$

$$\mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)(\mathbf{p}) = \frac{1}{d(z)(\mathbf{p})} \left(f \begin{bmatrix} l^i_{R,1} & l^i_{G,1} & l^i_{B,1} \\ l^i_{R,2} & l^i_{G,2} & l^i_{B,2} \\ 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} \mathbf{p} - \mathbf{p}^0 \\ 1 \end{bmatrix} \left[l^i_{R,3},\, l^i_{G,3},\, l^i_{B,3} \right] \right) \mathrm{Diag}(\boldsymbol{\rho}(\mathbf{p}))$$

$$\mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i)(\mathbf{p}) = \mathbf{I}^i(\mathbf{p}) - \begin{bmatrix} l^i_{R,4} & & \\ & l^i_{G,4} & \\ & & l^i_{B,4} \end{bmatrix} \boldsymbol{\rho}(\mathbf{p})$$

Methodology

Proposed variational model

depth super-resolution clue $\quad z_0^i = Kz + \varepsilon_z^i$

photometric stereo clue $\quad \mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} = \mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i) + \boldsymbol{\varepsilon}^i$

Methodology

Proposed variational model

depth super-resolution clue $z_0^i = Kz + \varepsilon_z^i$

photometric stereo clue $\mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} = \mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i) + \boldsymbol{\varepsilon}^i$

Final model:

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

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

Proposed variational framework

Input

high-res. RGB images
low-res. depth maps

Output

high-res. RGB image
high-res. depth map
high-res. albedo

By-product

Reflectance,
Lighting

RGB-D images n ≥ 4

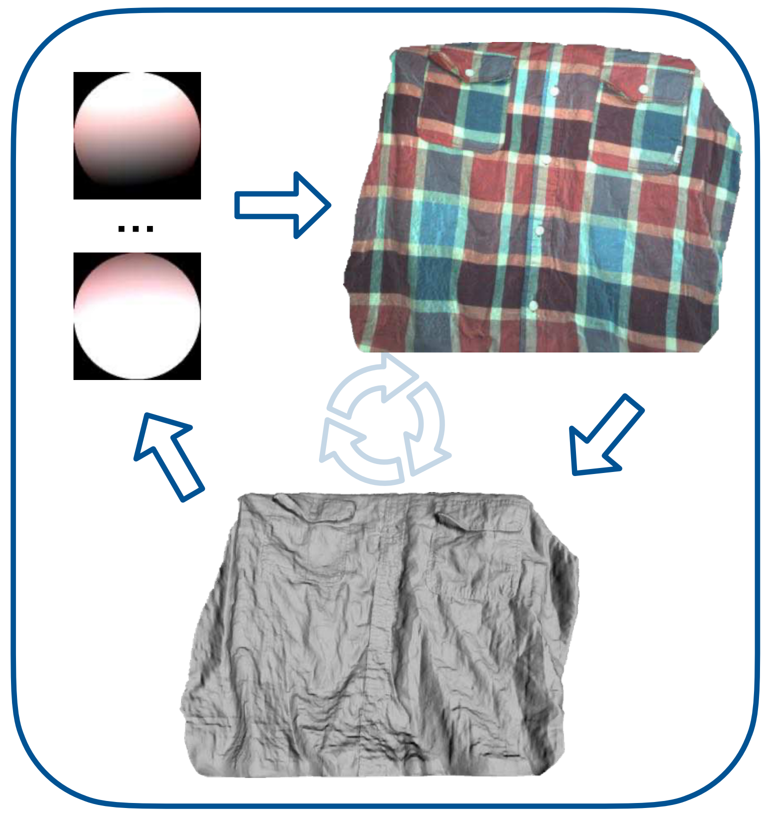
Alternating optimization

Input

high-res. RGB images
low-res. depth maps

RGB-D images n ≥ 4

Iterative optimization



high-res. RGB images
high-res. depth maps
high-res. albedo
lighting

Output

high-res. RGB image
high-res. depth map
high-res. albedo

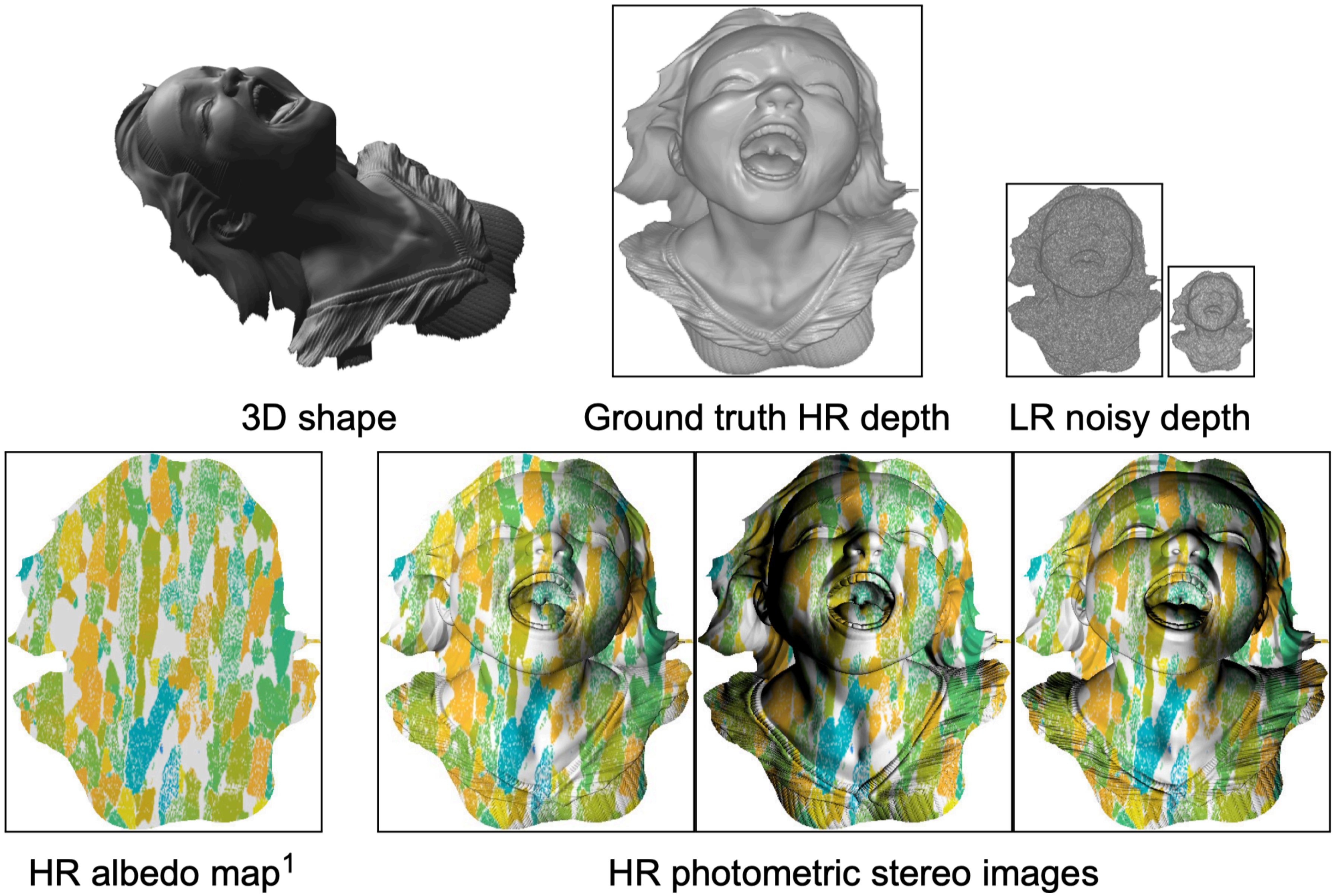
Outline

- Introduction
- Method description
- **Results and evaluation**
- Personal comments
- Summary

Results and evaluation

… of synthetic and real-world datasets

Evaluation on synthetic datasets

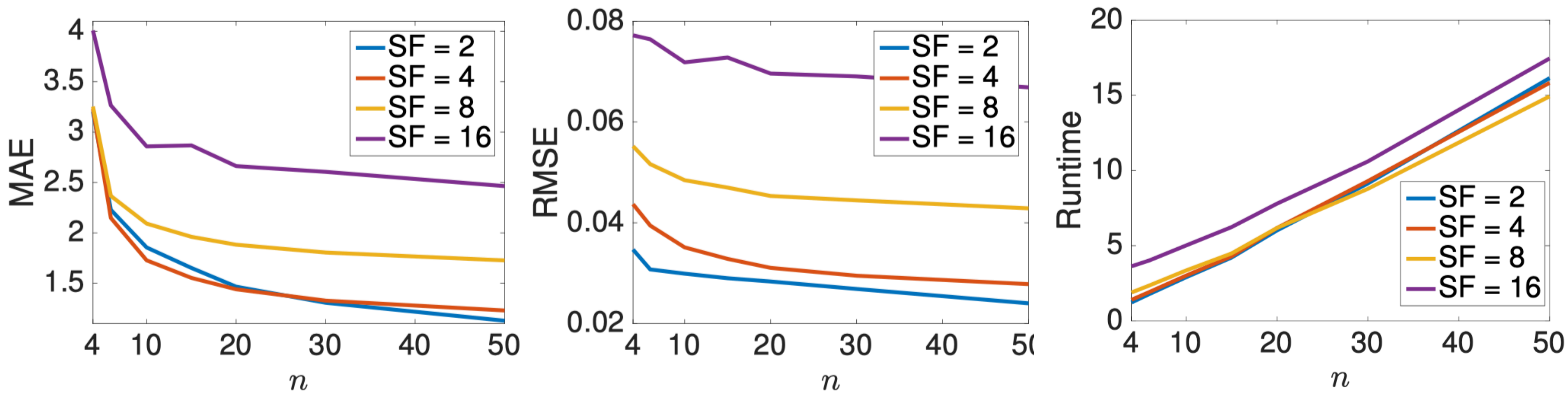


1 source: https://mtex-toolbox.github.io/files/doc/EBSDSpatialPlots.html

RMSE := Root Mean Square Error (on depth)
MAE := Mean Angular Error (on normals)

Evaluation on synthetic dataset

1. Number of images **n**

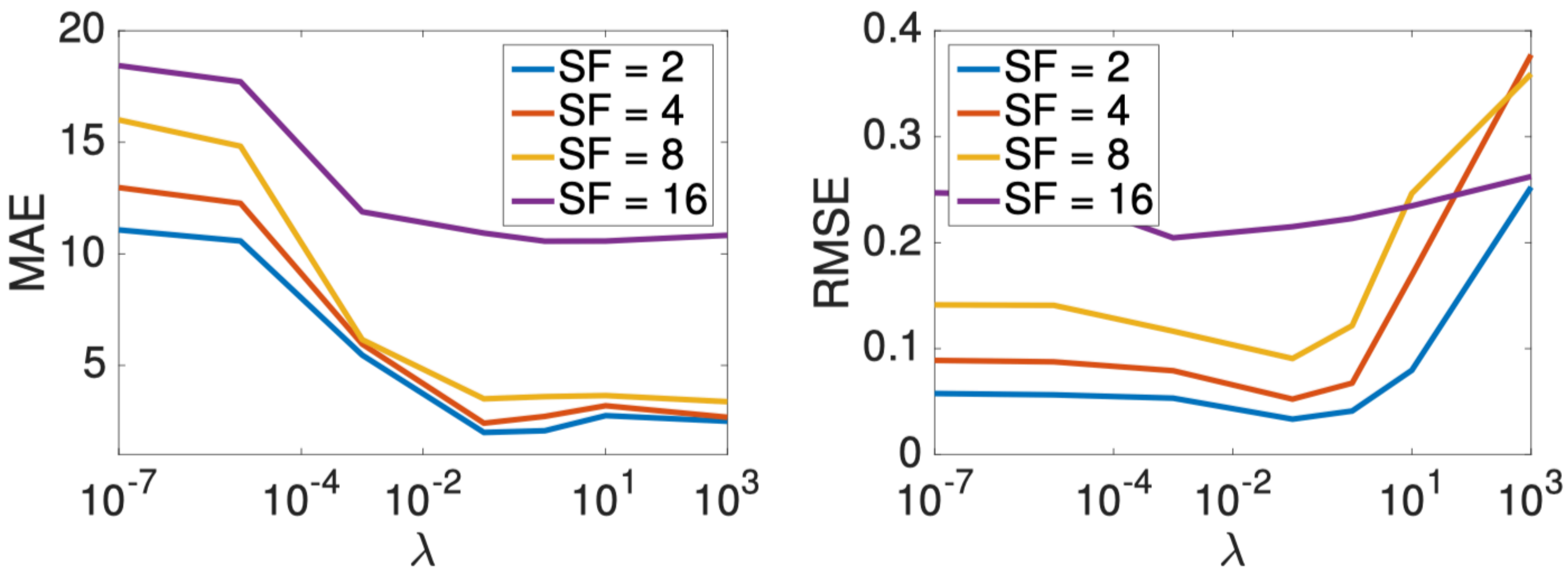


n ∈ **[10, 30]** is a good compromise between accuracy and speed.

RMSE := Root Mean Square Error (on depth)
MAE := Mean Angular Error (on normals)

Evaluation on synthetic dataset

2. Parameter tuning **λ**

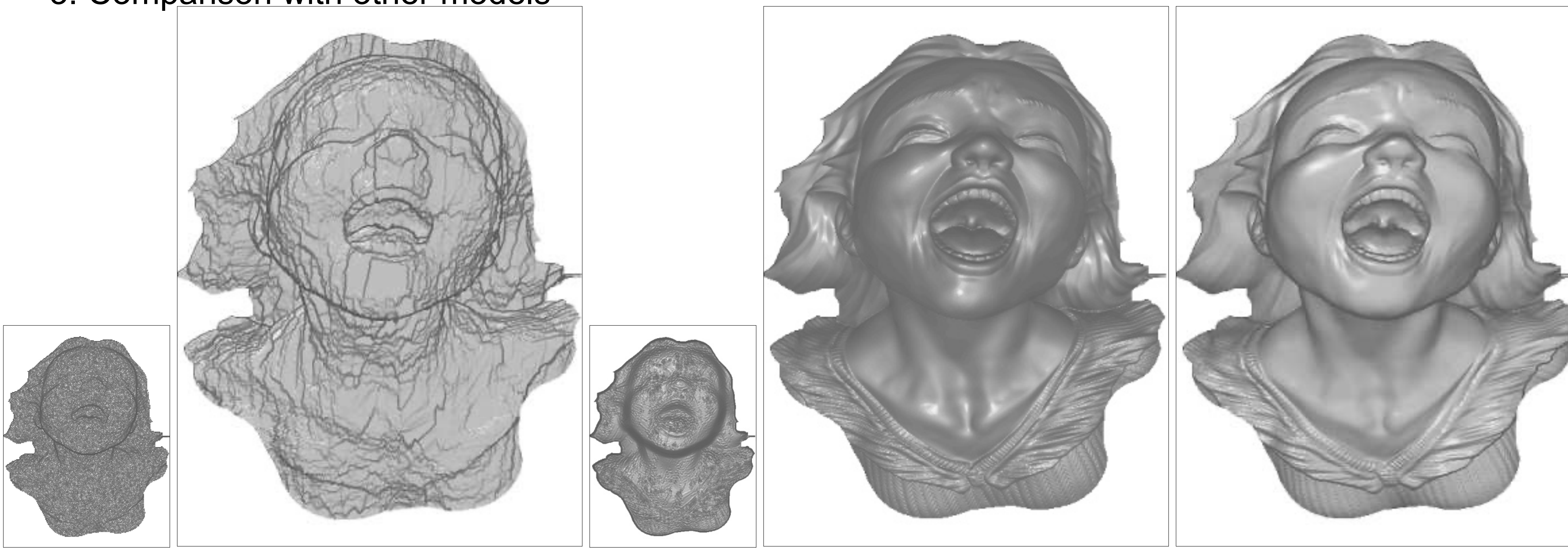


$\lambda \in [10^{-2}, 10^{1}]$ provide satisfactory results.

RMSE := Root Mean Square Error (on depth)
MAE := Mean Angular Error (on normals)

Evaluation on synthetic dataset

3. Comparison with other models

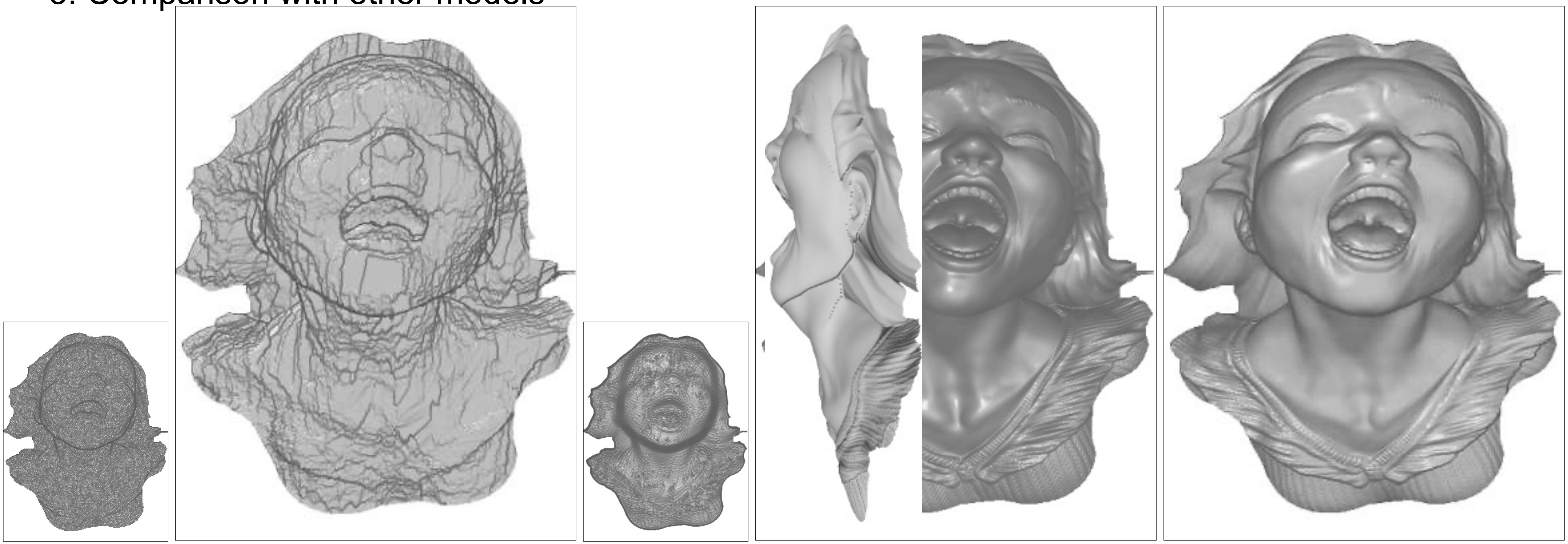


	Input	Depth super-resolution with TV	RGBD-Fusion	LDR Photometric Stereo	**SRmeetsPS (this paper)**
RMSE =	0.0579	0.0728	0.1655	0.9199	**0.03139**
MAE =	65.7150	34.4129	38.9316	41.8041	**1.4528**

RMSE := Root Mean Square Error (on depth)
MAE := Mean Angular Error (on normals)

Evaluation on synthetic dataset

3. Comparison with other models



	Input	Depth super-resolution with TV	RGBD-Fusion	LDR Photometric Stereo	**SRmeetsPS (this paper)**
RMSE =	0.0579	0.0728	0.1655	0.9199	**0.03139**
MAE =	65.7150	34.4129	38.9316	41.8041	**1.4528**

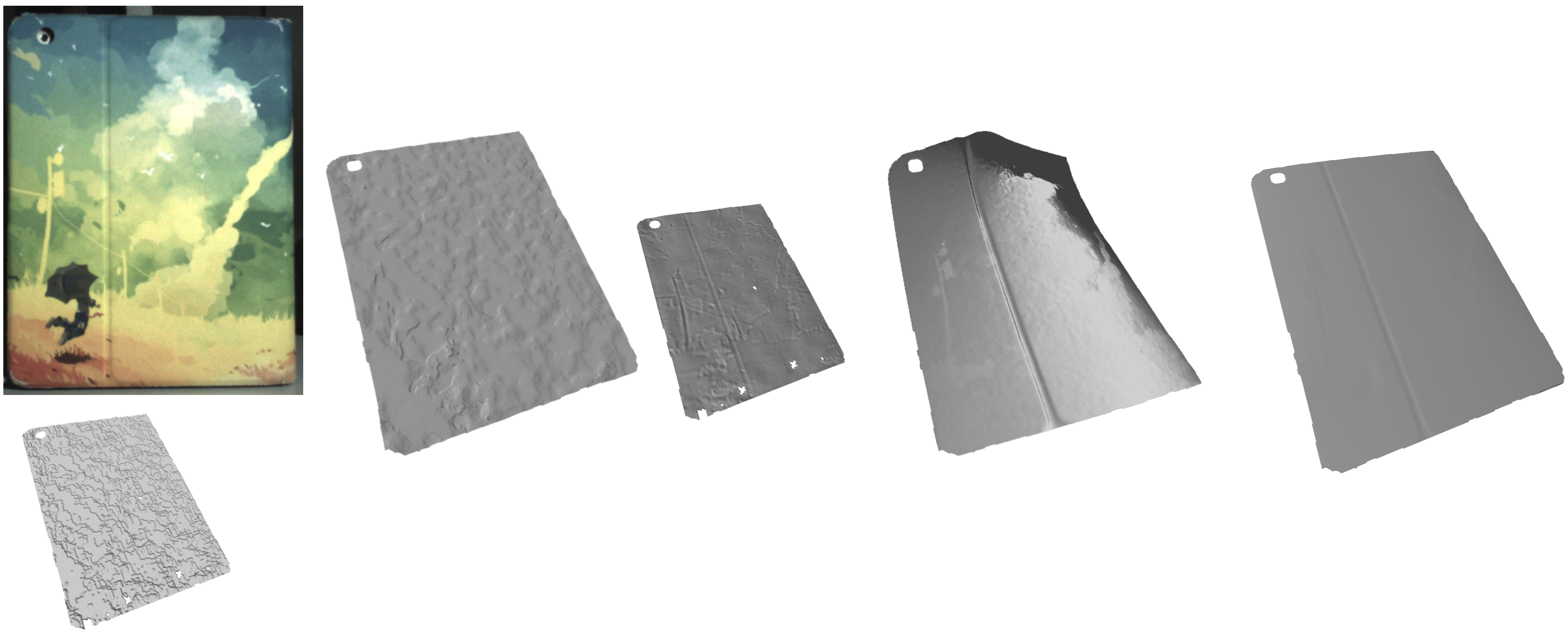
Evaluation on real-world datasets

1. Qualitative results

1 of n input HR RGB | 1 of n input LR depth | Estimated HR depth map | Estimated HR reflectance map | Relighting

Evaluation on real-world datasets

2. Comparison with other models



Input | Depth super-resolution | RGBD-Fusion | LDR Photometric Stereo | **SRmeetsPS (this paper)**

Outline

- Introduction
- Method description
- Results and evaluation
- **Personal comments**
- Summary

Personal comments

The advantage of a **multiple-light** setup and **high-res. RGB-D**

Personal comments

1. Much less restricted environments, e.g. different illumination

2. RGB image & depth map have same resolution
   → e.g. useful for realistic 3D reconstruction in AR

Outline

- Introduction
- Method description
- Results and evaluation
- Personal comments
- **Summary**

Summary

A novel variational framework for **depth super-resolution** in RGB-D sensing with the **photometric stereo** technique

Conclusion

A sequence of
(n >= 4)

Low-res. depth maps
+
High-res. RGB images under **uncalibrated illumination**

Joint

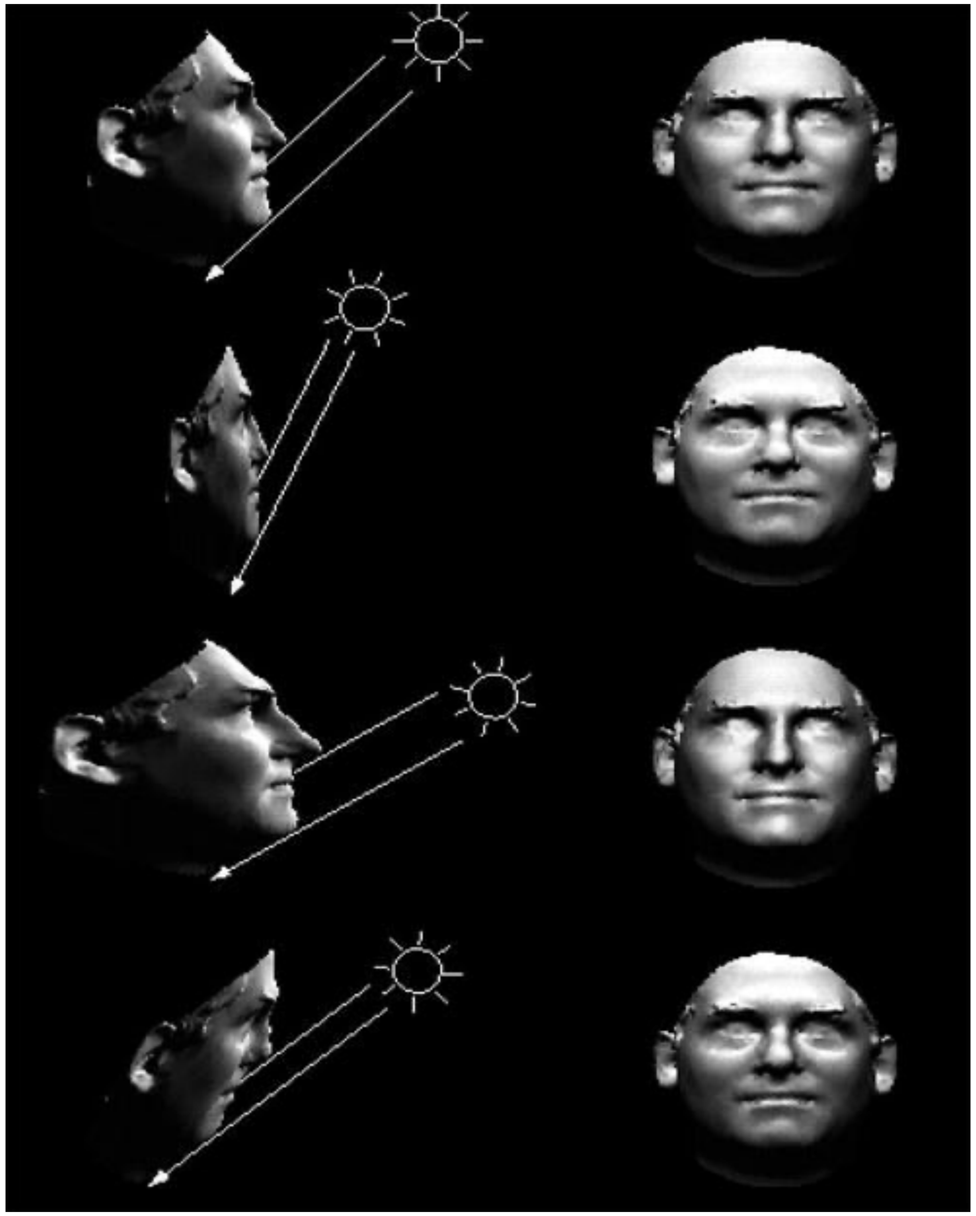
Depth super-resolution

Unambiguous Shape + Reflectance + Lighting Estimation

This method can be used **out-of-the-box** with **common devices**.

Thanks for listening!

Generalized Bas-Relief (GBR) Ambiguity



Source: https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-31439-6_542