High Quality Shape from a Single RGB-D Image under Uncalibrated Natural Illumination

Yudeog Han Joon-Young Lee In So Kweon
Robotics and Computer Vision Lab., KAIST

Problem Definition

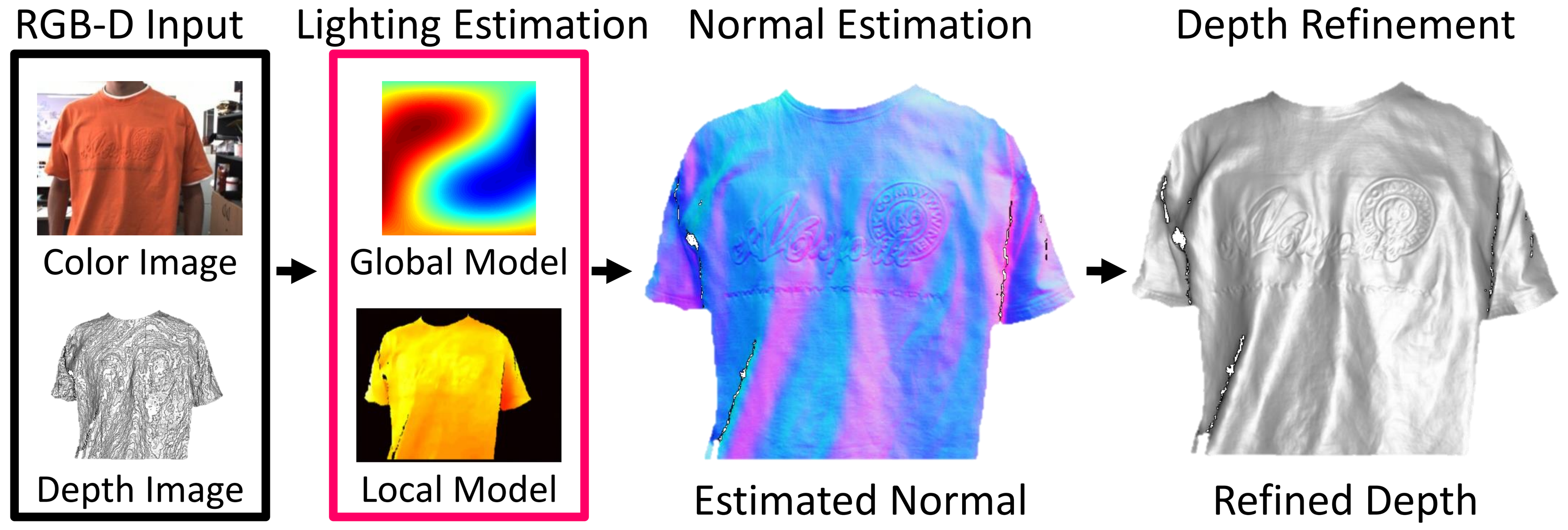
Consumer depth camera

A single RGB-D image
Natural illumination

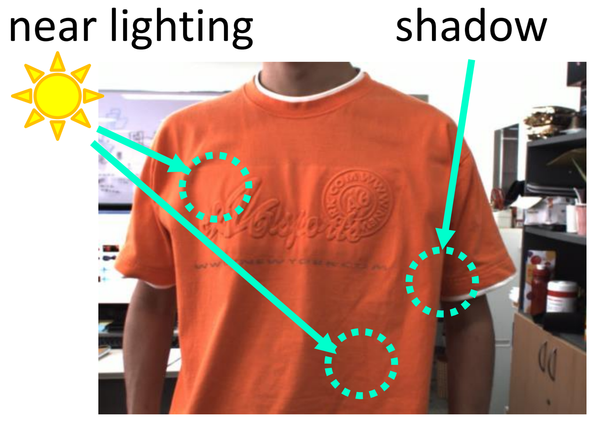
Lighting estimation
High-quality shape

Shape Estimation in Natural Illumination

❖Overall framework



❖Image formation



Conventional quadratic lighting function[§]

$$s(\mathbf{n}) = \mathbf{n}^\intercal A\mathbf{n} + \mathbf{b}^\intercal \mathbf{n} + c$$

Our lighting model

$$I = \alpha s(\mathbf{n}) = \alpha(\mathbf{n}^\intercal A\mathbf{n} + \mathbf{b}^\intercal \mathbf{n} + c)$$

I: observed intensity
α: (per-pixel) local lighting parameter
$A, \mathbf{b}, c$: (per-channel) global lighting parameter
$\mathbf{n}$: surface normal vector

❖Lighting estimation

Global lighting parameter estimation

the low-dimensional characteristic of a diffuse reflectance model

$$\begin{bmatrix} \mathrm{vec}(\mathbf{n}_1\mathbf{n}_1^T)^T & \mathbf{n}_1^T & 1 \\ \vdots & & \\ \mathrm{vec}(\mathbf{n}_k\mathbf{n}_k^T)^T & \mathbf{n}_k^T & 1 \end{bmatrix} \begin{bmatrix} \mathrm{vec}(A) \\ \mathbf{b} \\ c \end{bmatrix} = \begin{bmatrix} I_1 \\ \vdots \\ I_k \end{bmatrix}$$

Local lighting parameter estimation

- The residual error in the global lighting model
  - local lighting variations
  - initial normal deviations
- Illumination in small neighborhood is smoothly varying

$$\operatorname*{argmin}_{\alpha} \lambda_1^l E_1^l(\alpha) + \lambda_2^l E_2^l(\alpha) + \lambda_3^l E_3^l(\alpha)$$

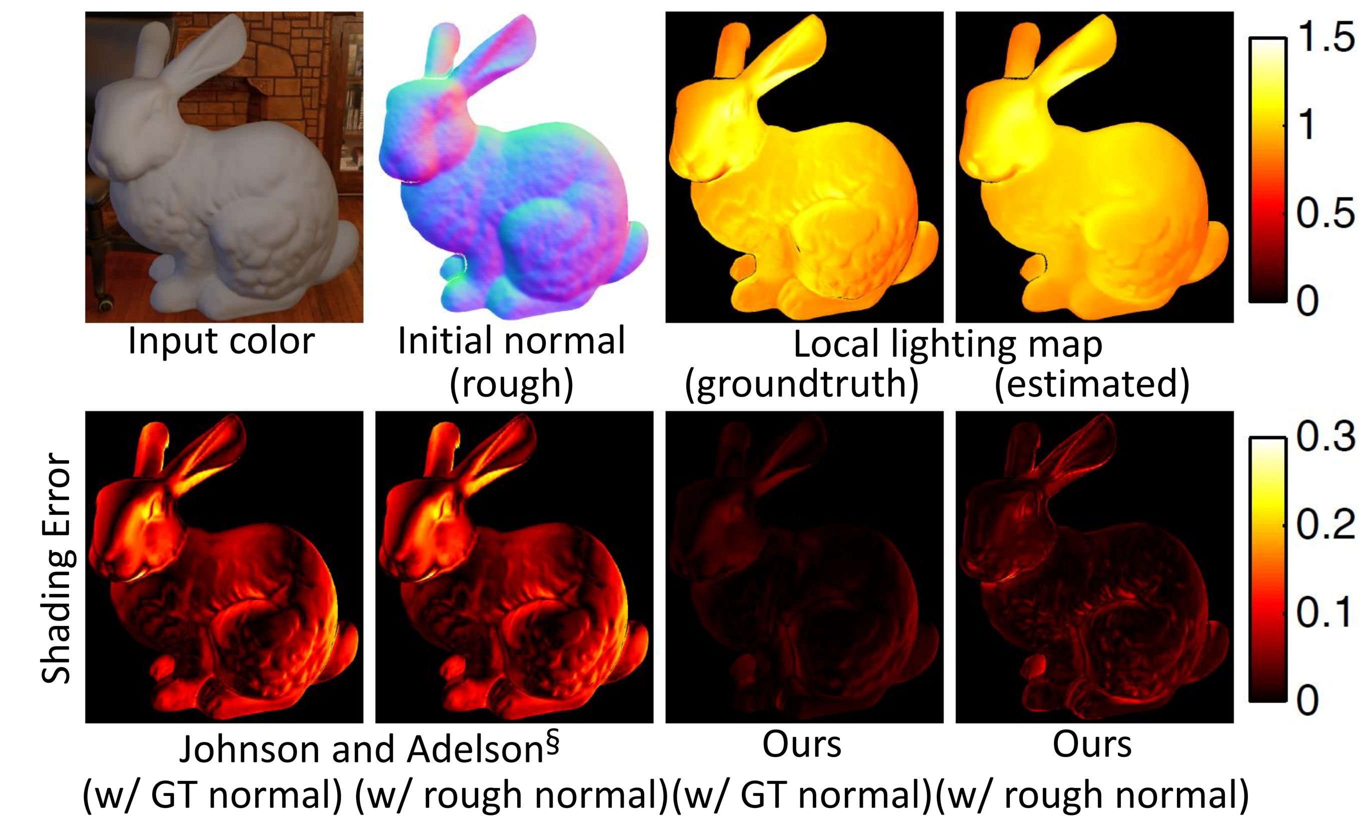
$$E_1^l(\alpha) = \sum_p \|I_p - \alpha_p s(\mathbf{n}_p)\|^2$$

$$E_2^l(\alpha) = \sum_p \sum_{q \in \Omega_p} \|\omega^l_{(p,q)} (\alpha_p - \alpha_q)\|^2$$

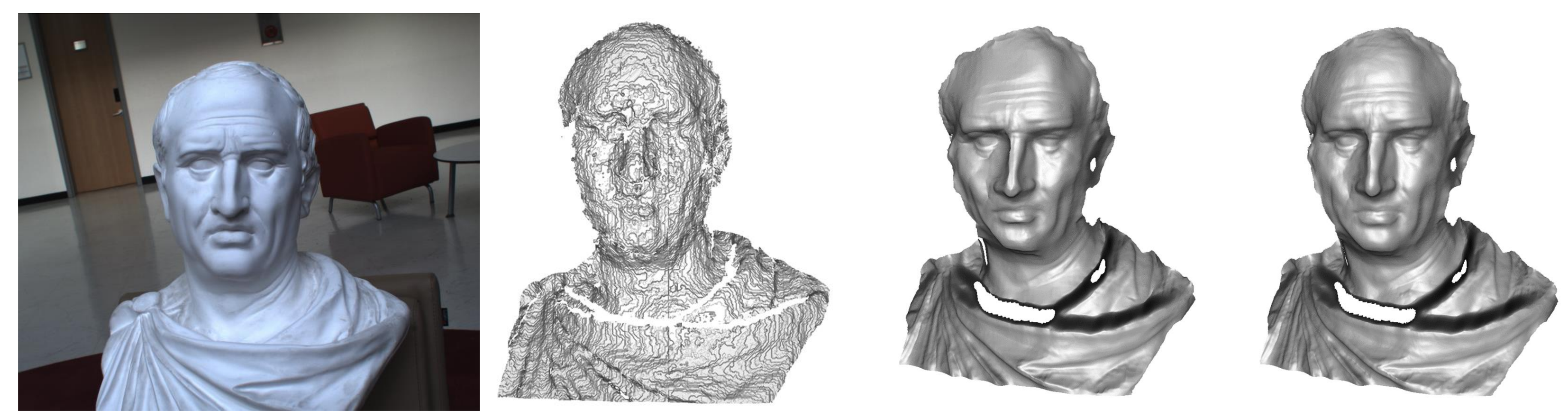
$$\omega^l_{(p,q)} = \begin{cases} 0 & \text{if } \|I_p - I_q\|^2 > \tau^l \\ \exp(-\frac{\|I_p - I_q\|^2}{2\sigma_l^2}) & \text{otherwise} \end{cases}$$

$$E_3^l(\alpha) = \sum_p \|\nabla^2 \alpha_p\|^2$$

➔ The entire lighting estimation is efficiently solved by least squares.



[§] DM. K. Johnson and E. H. Adelson. Shape estimation in natural illumination. CVPR 2011



Input image | Input depth | Our result (the global model only) | Our result (final)

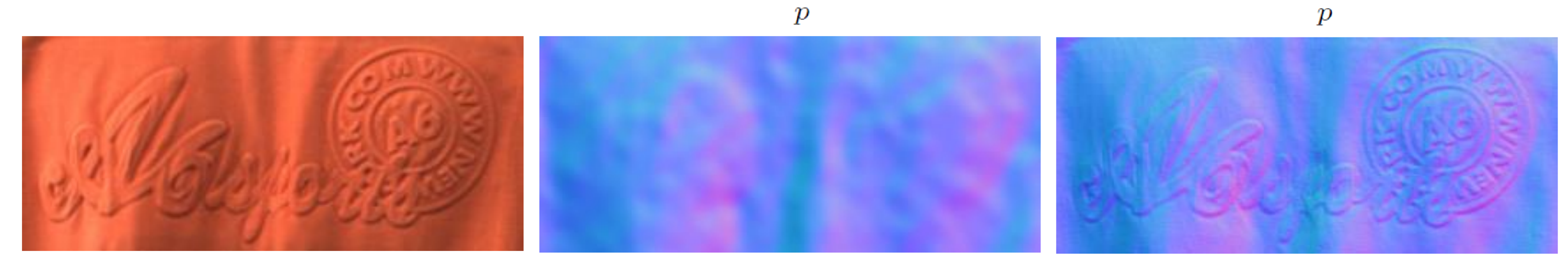
❖Normal estimation

- Shading constraint
- Normal constraint
- Integrability constraint

$$\operatorname*{argmin}_{\mathbf{n}} \lambda_1^n E_1^n(\mathbf{n}) + \lambda_2^n E_2^n(\mathbf{n}) + \lambda_3^n E_3^n(\mathbf{n})$$

$$E_1^n(\mathbf{n}) = \sum_p \|I_p - \tilde{s}(\mathbf{n}_p)\|^2$$

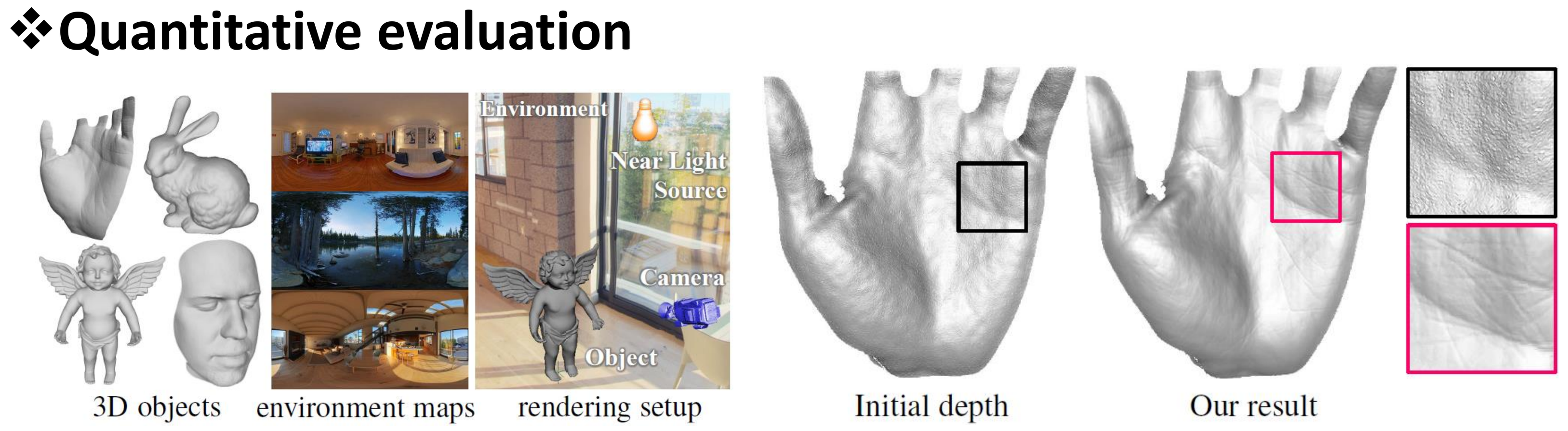
$$E_2^n(\mathbf{n}) = \sum_p \|1 - \mathbf{n}_p \cdot \mathbf{n}_p^0\|^2 \quad E_3^n(\mathbf{n}) = \sum_p \|\nabla \times \mathbf{n}_p\|^2$$



Input Image | Input Normal | Estimated Normal

Experiments

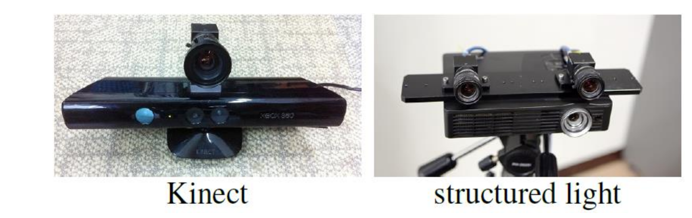
❖Quantitative evaluation



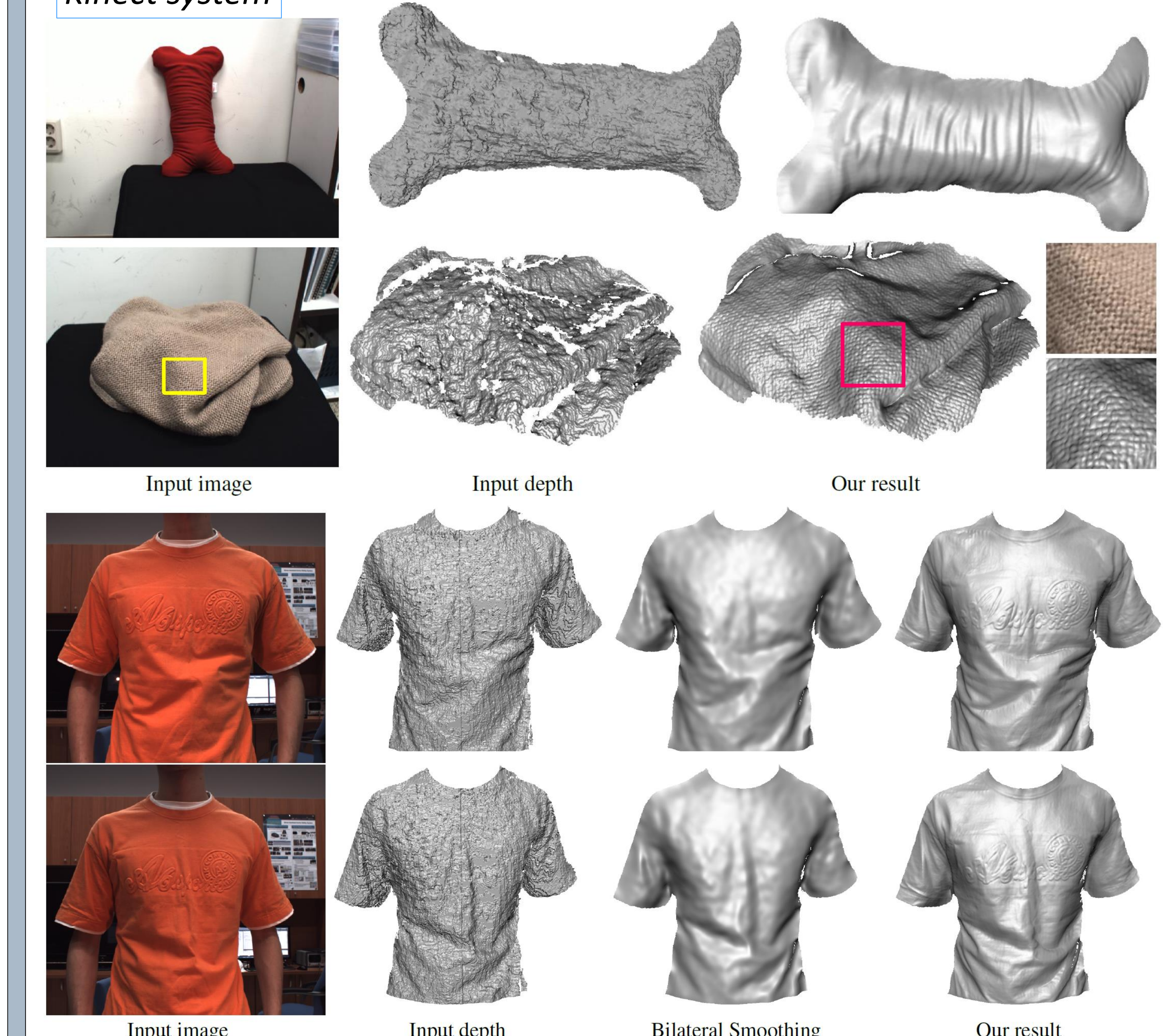
		Average (°)				R10 (%)				A75 (°)			
		initial	GL	GC	Ours	initial	GL	GC	Ours	initial	GL	GC	Ours
ANGEL	Env1	6.360	7.216	7.553	**6.099**	17.54	21.18	25.61	**14.31**	7.950	9.284	10.139	**7.587**
	Env2	6.360	6.765	7.292	**5.577**	17.54	19.62	24.18	**13.11**	7.950	8.893	9.825	**6.777**
	Env3	6.360	7.541	6.923	**6.118**	17.54	19.99	18.79	**15.02**	7.950	9.120	8.620	**7.604**
BUNNY	Env1	4.933	6.633	8.413	**4.857**	7.72	17.65	31.33	**7.20**	6.396	8.699	10.989	**6.286**
	Env2	4.933	6.590	8.872	**4.826**	7.72	18.32	33.73	**7.40**	6.396	8.828	11.476	**6.170**
	Env3	4.933	6.203	6.467	**4.893**	7.72	11.96	16.72	**7.18**	6.396	7.854	8.418	**6.307**
FACE	Env1	4.513	6.080	6.466	**3.648**	4.85	9.84	13.68	**4.16**	5.756	7.714	8.486	**4.525**
	Env2	4.513	5.948	6.610	**3.815**	4.85	8.75	15.47	**3.56**	5.756	7.440	8.702	**4.830**
	Env3	4.513	5.324	5.198	**3.887**	4.85	6.63	6.77	**3.63**	5.756	6.663	6.698	**4.845**
PALM	Env1	4.679	6.288	6.244	**3.463**	5.55	13.30	16.68	**2.50**	6.059	8.319	8.612	**4.562**
	Env2	4.679	6.856	6.590	**3.522**	5.55	17.54	20.58	**3.02**	6.059	8.951	9.238	**4.625**
	Env3	4.679	5.415	5.577	**3.213**	5.55	7.76	9.12	**1.51**	6.059	7.081	7.480	**4.171**

R*X*: the percentage of pixels that have an angular normal error above *X* degrees
A*X*: the angular normal error at the *X*th percentile after sorting the errors from low to high

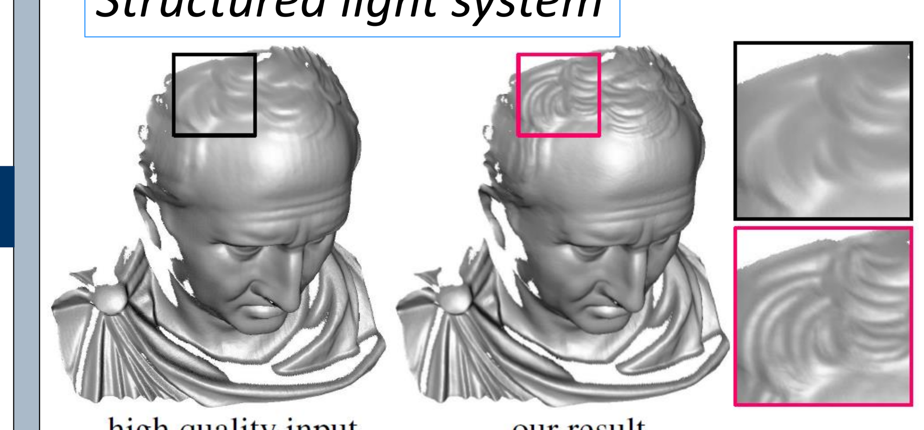
❖Real-world experiments



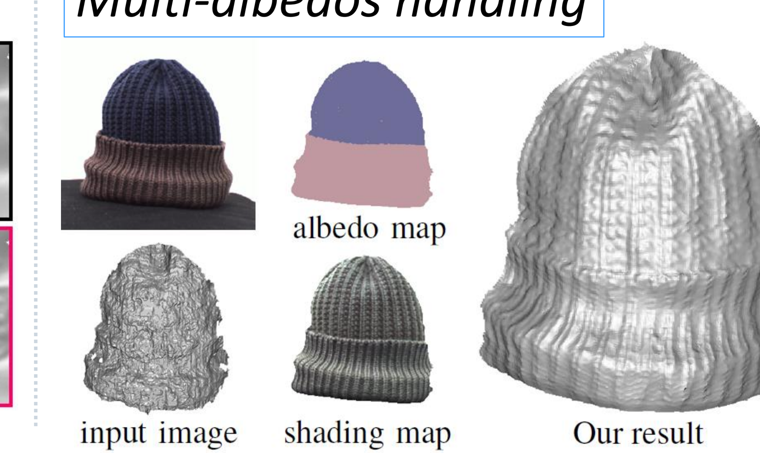
Kinect system



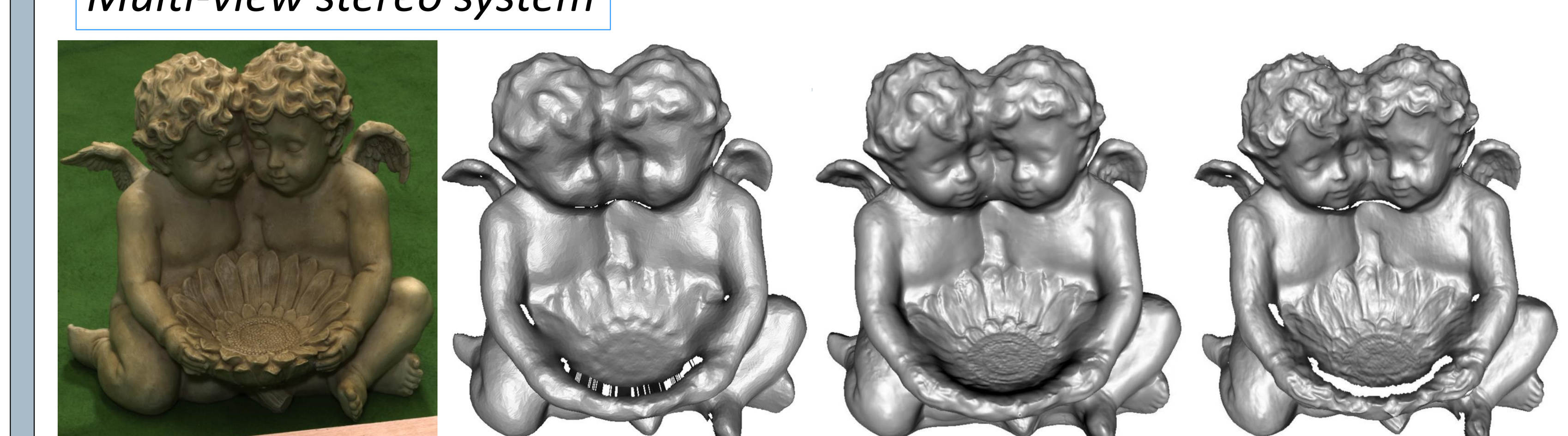
Structured light system



Multi-albedos handling



Multi-view stereo system



Input image | Initial depth | Wu *et al.*[†] | Our result

[†] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt. High-quality shape from multi-view stereo and shading under general illumination. CVPR 2011.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No.2010-0028680)1