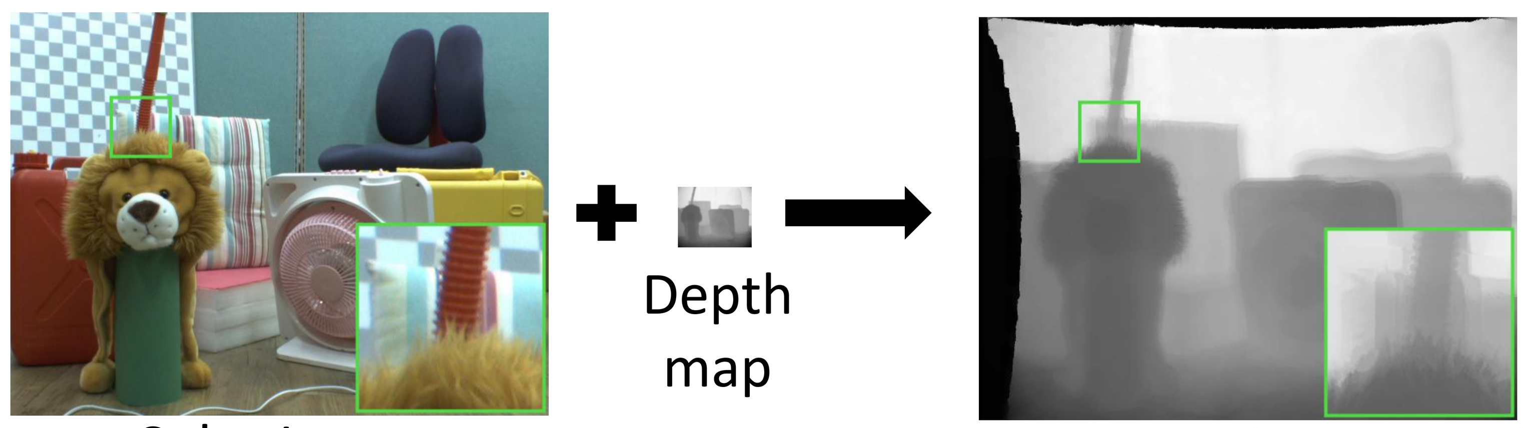
High Quality Depth Map Upsampling for 3D-TOF Cameras

Jaesik Park[1], Hyeongwoo Kim[1*], Yu-Wing Tai[1], Michael S. Brown[2], In So Kweon[1]
Korea Advanced Institute of Science and Technology (KAIST)[1], National University of Singapore (NUS)[2]
*The first and the second authors provided equal contributions to this work.

Problem Definition

Are 'filter based upsampling methods' good enough?



Color Image

Yang *et al*. (CVPR2007)

Limitations on Filtering based methods

- May suffer over-smoothing
- Harder to regularize outlier sample explicitly

Overview of Our Approach

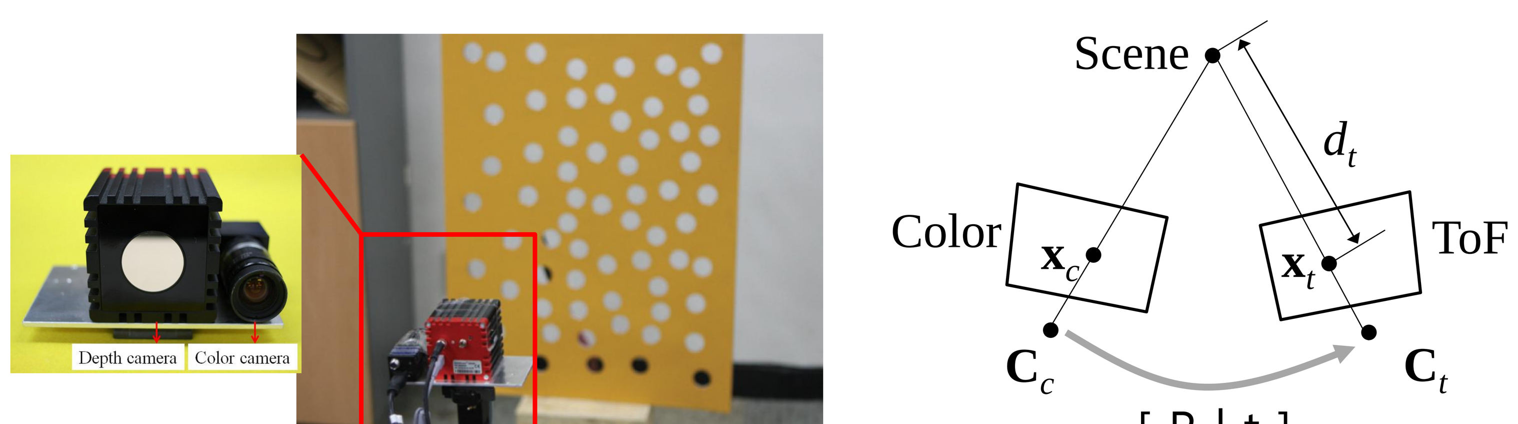
Our Result

- **Accurate alignment of two sensors**
  - Calibration using 'hole' pattern
  - Outlier detection
- **An extended optimization framework**
  - Prevents depth bleeding (guided depth map)
  - Recovers fine structure (non-local means term)
  - Allows simple user markup.

System Setup and Preprocessing

Range Sensor (176x144) + CCD camera (1280x960)

Sensor Calibration using a 'hole' pattern



Projection 3D samples from ToF cameras into high resolution image coordinate

$$s\mathbf{x}_c = \mathrm{K_c}\,[\ \mathrm{R}\ \ |\ \ \mathrm{t}\ ]\,\mathrm{K_t}^{-1}[\mathbf{x}_t\ \ d_t\ \ 1]^T$$

We use Zhang's method to get extrinsic parameters.

Outlier Rejection

Removing blurred depth boundaries using MRF regularization

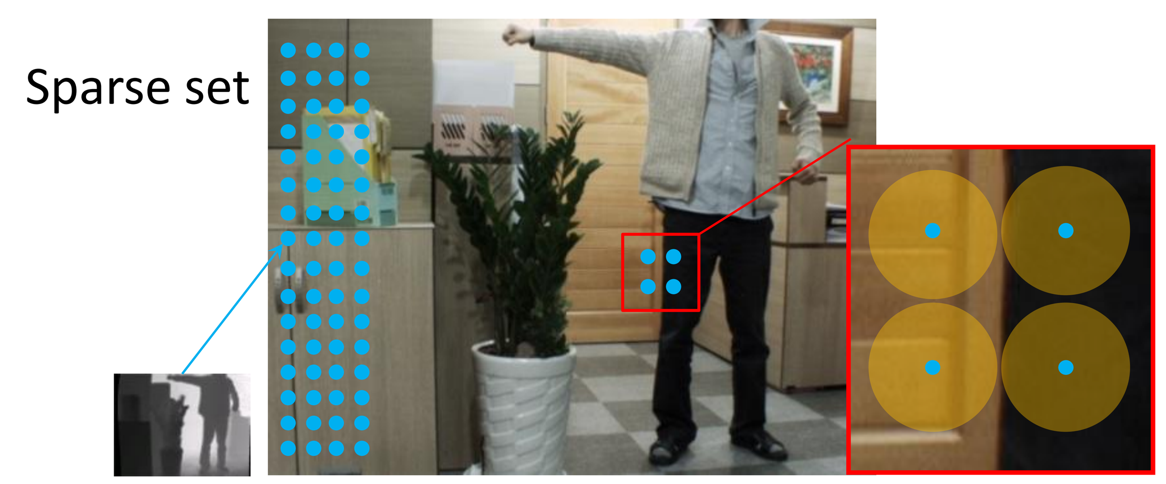
$$E(\mathrm{l}) = \sum_p \left(\mathrm{O}_p(\mathrm{l}) + \lambda_{pq} \sum_{q \in N(p)} \mathrm{O}_{pq}(\mathrm{l}) \right)$$

l is a map of binary label

$\mathbf{O}_p(\mathbf{l})$: Local variance

$\mathbf{O}_{pq}(\mathbf{l})$: Hamming distance

Optimization Framework



$$\boxed{E(\mathbf{D}) = E_{\mathrm{d}}(\mathbf{D}) + \lambda_{\mathrm{s}} E_{\mathrm{s}}(\mathbf{D}) + \lambda_{\mathrm{N}} E_{\mathrm{NLM}}(\mathbf{D})}$$

$E_{\mathrm{d}}(\mathbf{D})$: Data term to follow mapped depth value

$E_{\mathrm{s}}(\mathbf{D})$: Desired discontinuity of upsampled depth map

$$E_{\mathrm{s}}(\mathbf{D}) = \sum_p \sum_{q \in \mathcal{N}(p)} w_{pq} \left(\mathbf{D}(p) - \mathbf{D}(q)\right)^2$$

Color affinity

$$w_c = \exp -\left(\sum_{I \in YUV} \frac{(\mathbf{I}(p) - \mathbf{I}(q))^2}{2\sigma_{\mathbf{I}}^2} \right)$$

Saliency Edge [1]

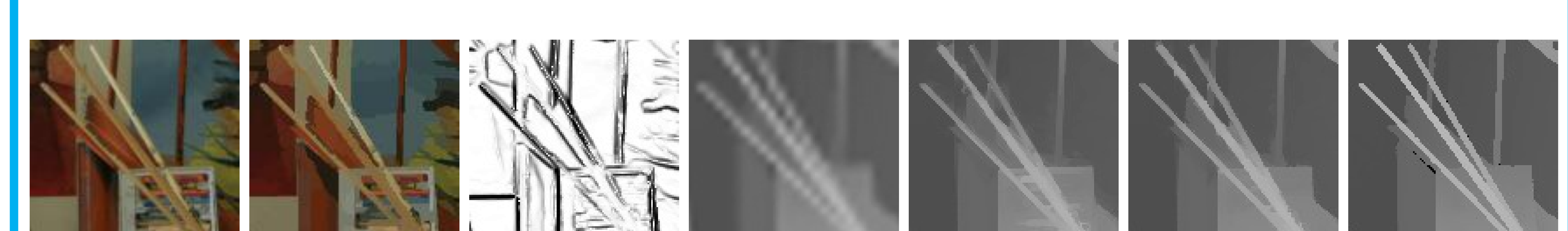
$$w_e = \frac{1}{\sqrt{s_x(p)^2 + s_x(q)^2 + 1}}$$

Segmentation [2]

$$w_s = \begin{cases} 1 & \text{if } \mathbf{S}_{\mathrm{co}}(p) = \mathbf{S}_{\mathrm{co}}(q) \\ t_{\mathrm{se}} & \text{otherwise} \end{cases}$$

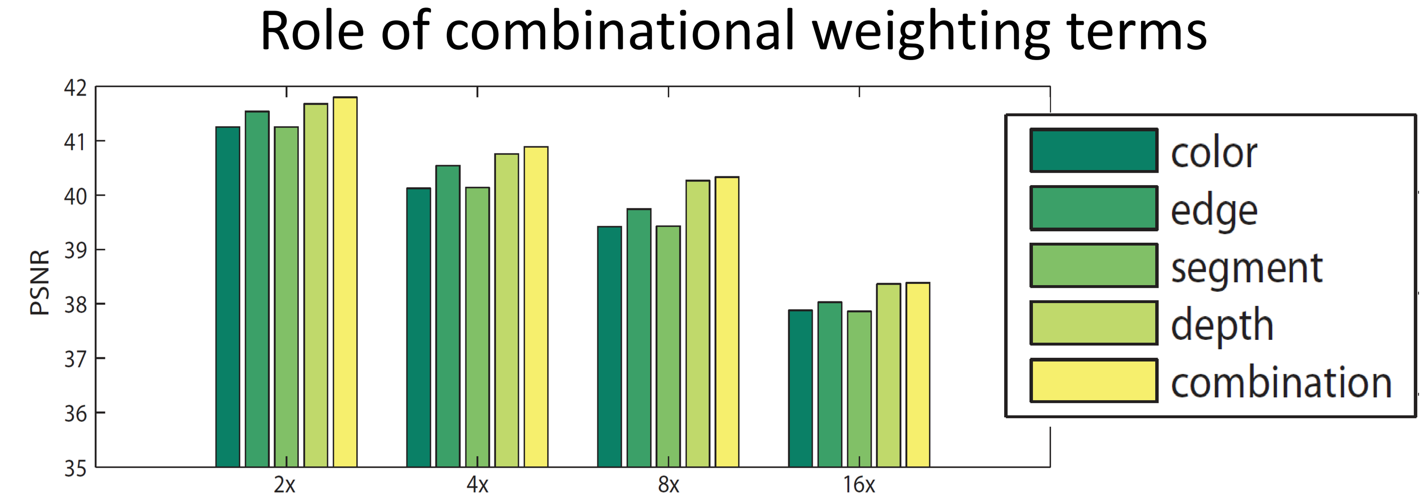
Guided depth map

$$w_d = \exp -\left(\frac{(\mathbf{D}_g(p) - \mathbf{D}_g(q))^2}{2\sigma_g^2} \right)$$



Color image | Segmentation | Edge | Guided depth | Without guided depth | With guided depth | Ground truth

Role of combinational weighting terms



$E_{\mathrm{NLM}}(\mathbf{D})$: Oriented non-local means term for fine structure

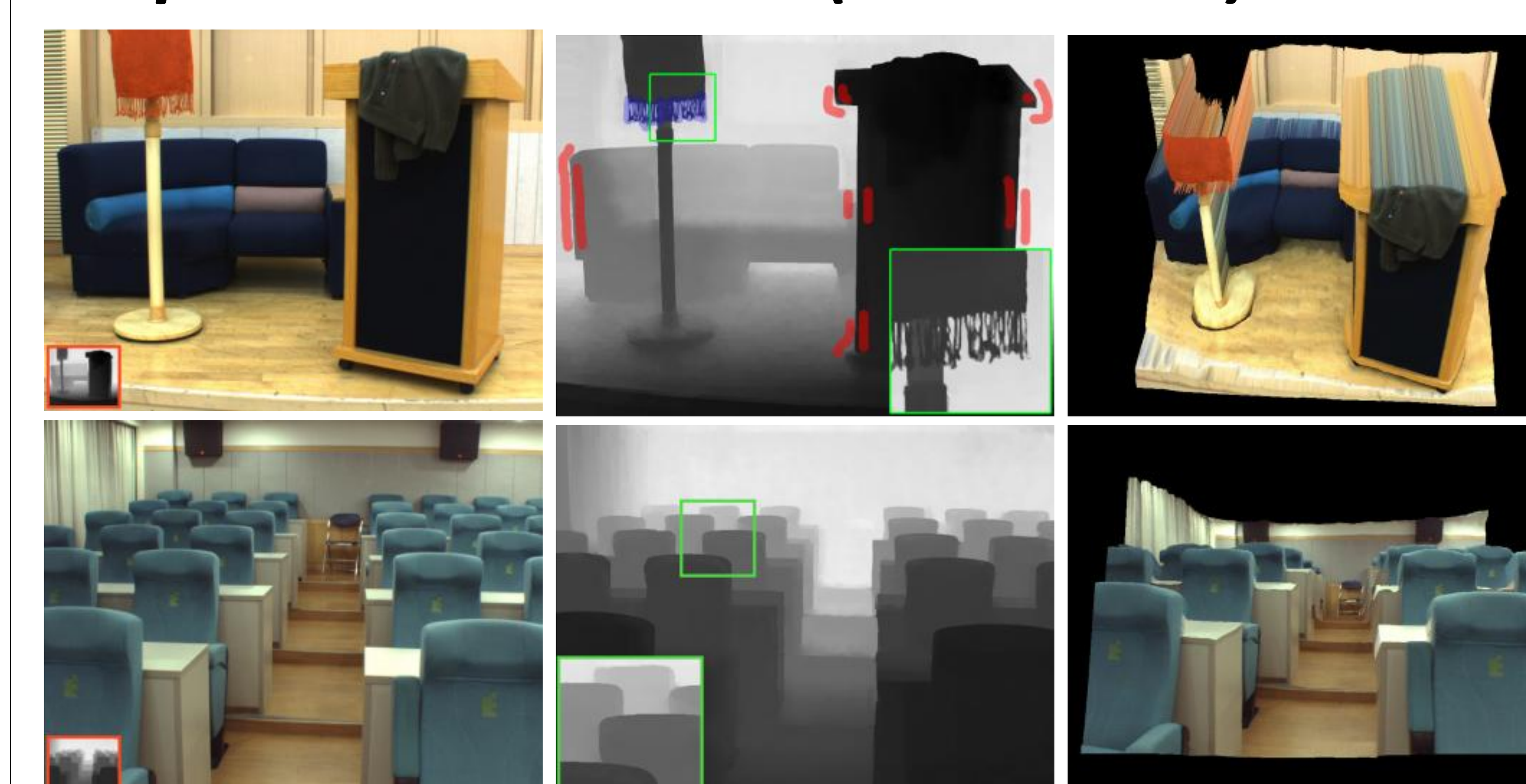
$$\kappa_{pq} = \frac{1}{2}\left(\exp(-(p-q)^T \Sigma_p^{-1}(p-q)) + \exp(-(p-q)^T \Sigma_q^{-1}(p-q))\right)$$

$$\Sigma_p = \frac{1}{|\mathcal{A}|} \sum_{p' \in \mathcal{A}(p)} \nabla I(p') \nabla I(p')^T.$$

Experimental Results (Synthetic)

	Art				Books				Mobius			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
Bilinear	0.56	1.09	2.10	4.03	0.19	0.35	0.65	1.24	0.20	0.37	0.70	1.32
MRFs [3]	0.62	1.01	1.97	3.94	0.22	0.33	0.62	1.21	0.25	0.37	0.67	1.29
Bilateral [4]	0.57	0.70	1.50	3.69	0.30	0.45	0.64	1.45	0.39	0.48	0.69	1.14
Guided [5]	0.66	1.06	1.77	3.63	0.22	0.36	0.60	1.16	0.24	0.38	0.61	1.20
Ours	0.43	0.67	1.08	2.21	0.17	0.31	0.57	1.05	0.18	0.30	0.52	0.90

Experimental Results (Real world)



Color Image | Ours | Rendered view

References

[1] P. Bhat, C. L. Zitnick, M. F. Cohen, and B. Curless. Gradientshop: A gradient-domain optimization framework for image and video filtering. *In SIGGRAPH*, 2010.
[2] A. Vedaldi and B. Fulkerson. VLFeat, http://www.vlfeat.org/, 2008.
[3] J. Diebel and S. Thrun. An application of markov random fields to range sensing. In *NIPS*, 2005.
[4] Q. Yang, R. Yang, J. Davis, and D. Nist´er. Spatial-depth super resolution for range images. In *CVPR*, 2007.
[5] K. He, J. Sun, and X. Tang. Guided image filtering. In ECCV, 2010.