Depth from a Light Field Image with Learning-based Matching Costs

Finalist of the Depth Estimation Challenge at LF4CV &
Submitted to IEEE TPAMI (under review)

Hae-Gon Jeon[1], Jaesik Park[2], Gyeongmin Choe[1], Jinsun Park[1]
Yunsu Bok[3], Yu-Wing Tai[4], In So Kweon[1]

[1]KAIST [2]Intel labs [3]ETRI [4]Tencent

Goal of the Proposed Method



Problem1:
Severe vignetting



Problem2:
Severe noise

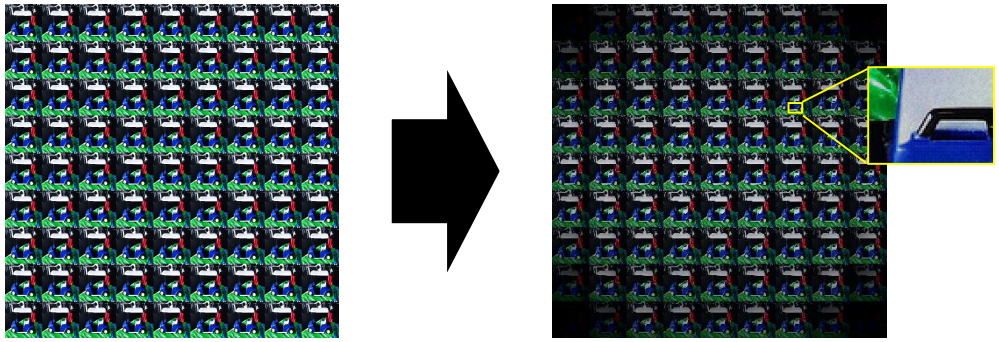
1. Hard to find accurate correspondence in radiometric distortions and severe noise
⇒ Using various hand-craft matching cost

2. Which one is correct matching cost?
⇒ Predicting the correct matching cost using two random forests

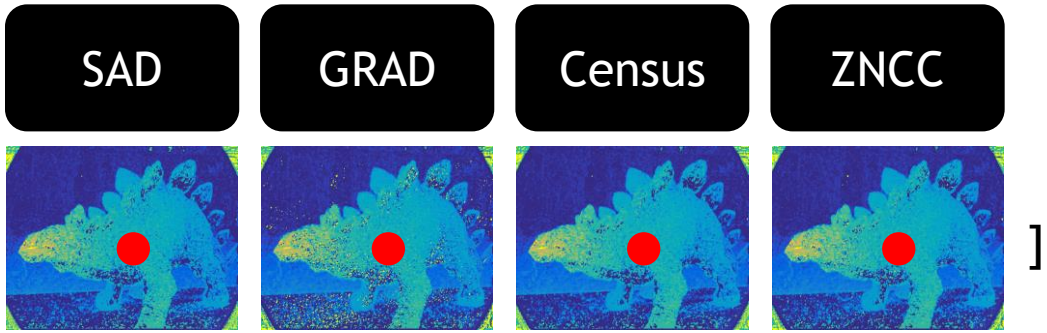
3. Does it work well in real world light-field images?
⇒ Realistic dataset generation based on an imaging pipeline of the Lytro camera

Overview of the Proposed Method

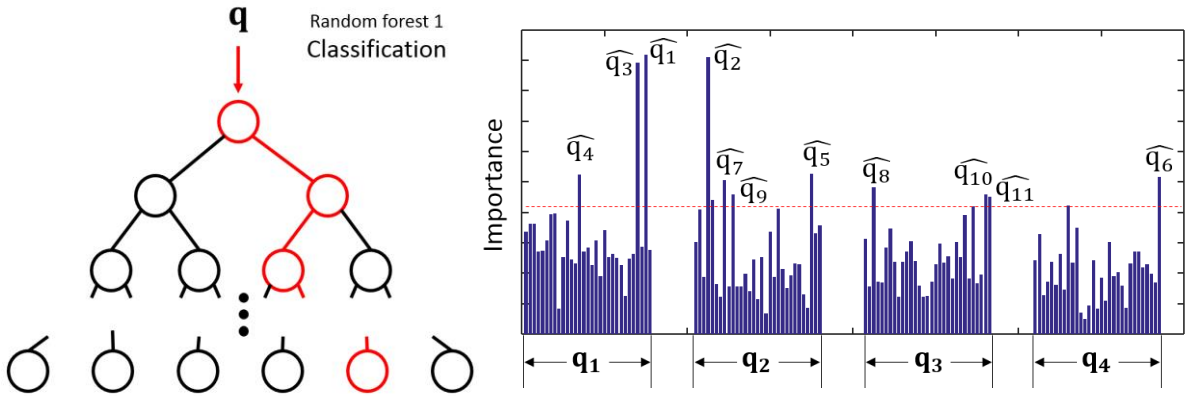
1. Realistic Light Field Image Generation;
Emulating an imaging pipeline of Lytro camera



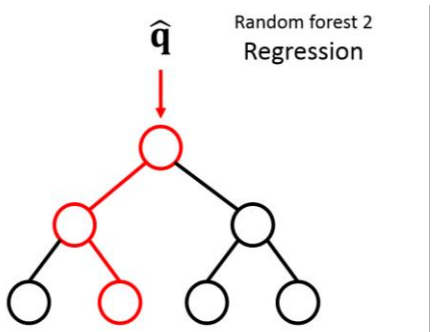
2. Making Cost Volumes using Phase Shift;
Overcoming inherent degradation of light-field images caused by a lenslet array

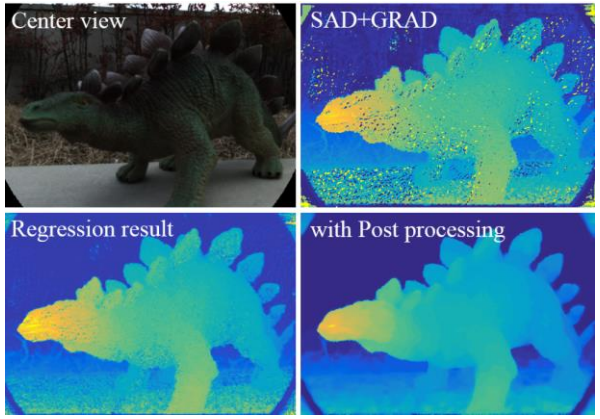


3. Random Forest 1 - Classification;
Selecting dominant matching costs

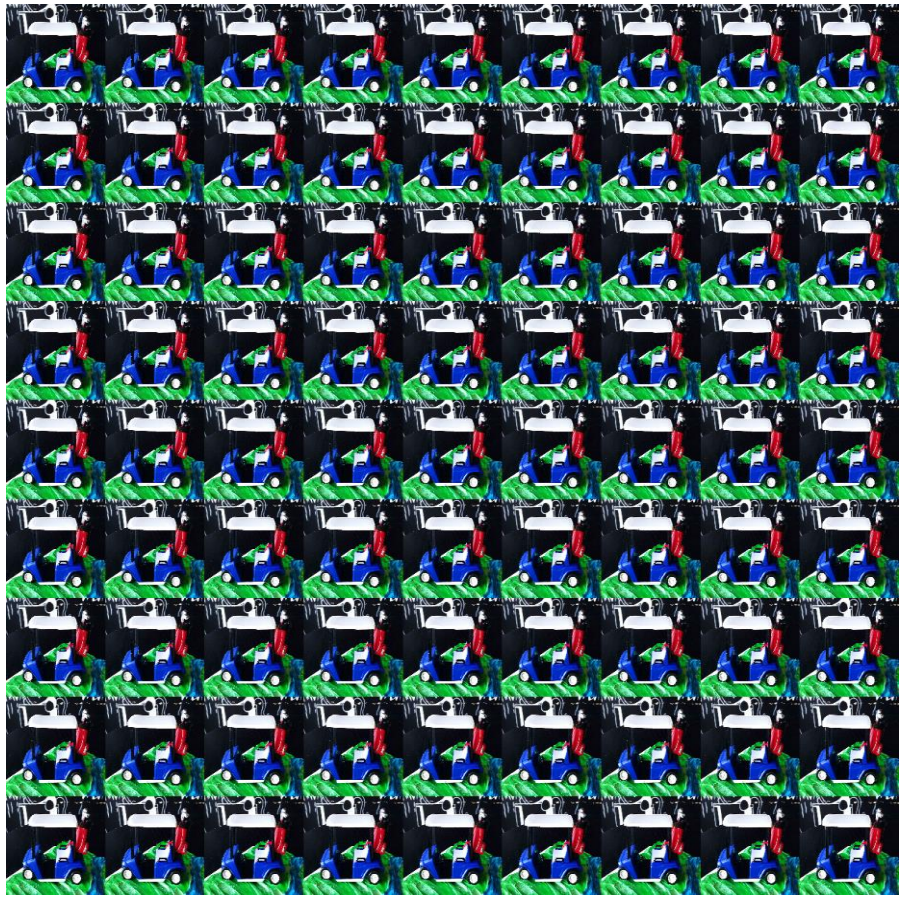


4. Random Forest 2 - Regression;
Predicting a disparity value with sub-pixel precision



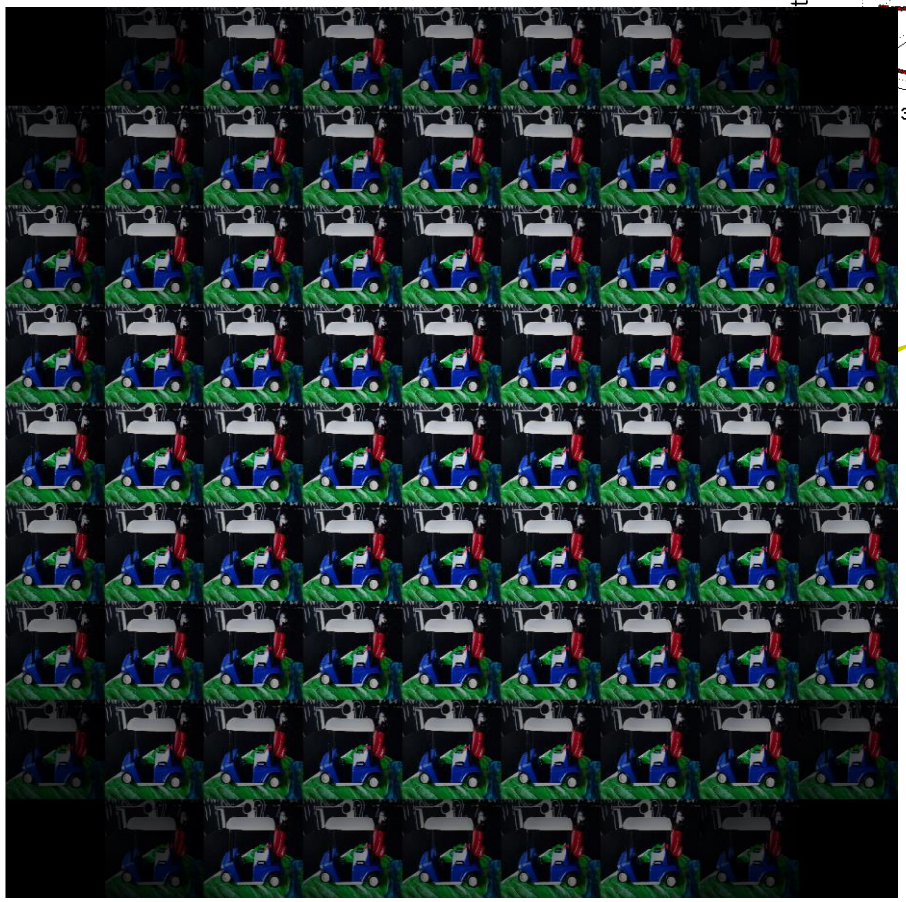


Data Generation Vignetting Map

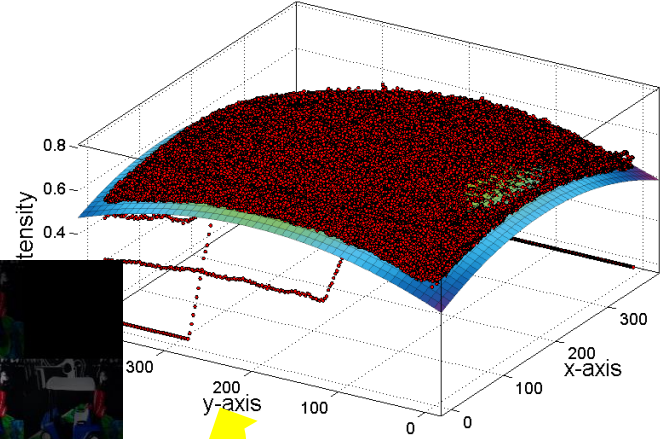


Noise-free multi-view images

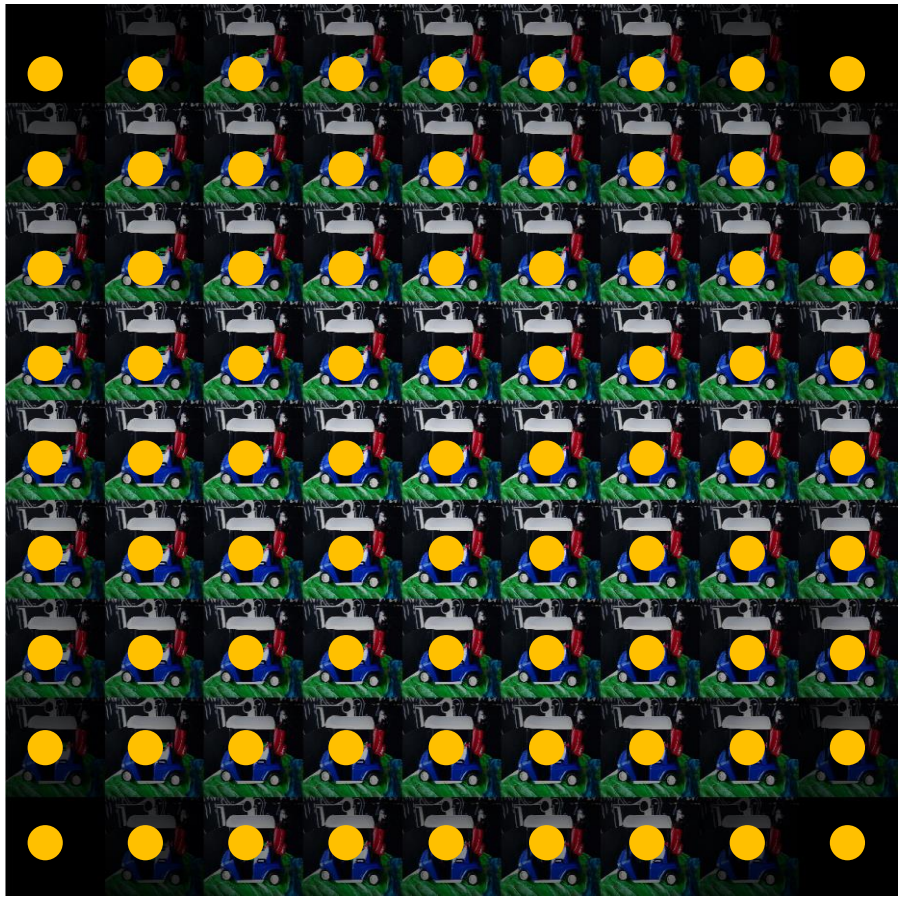
×



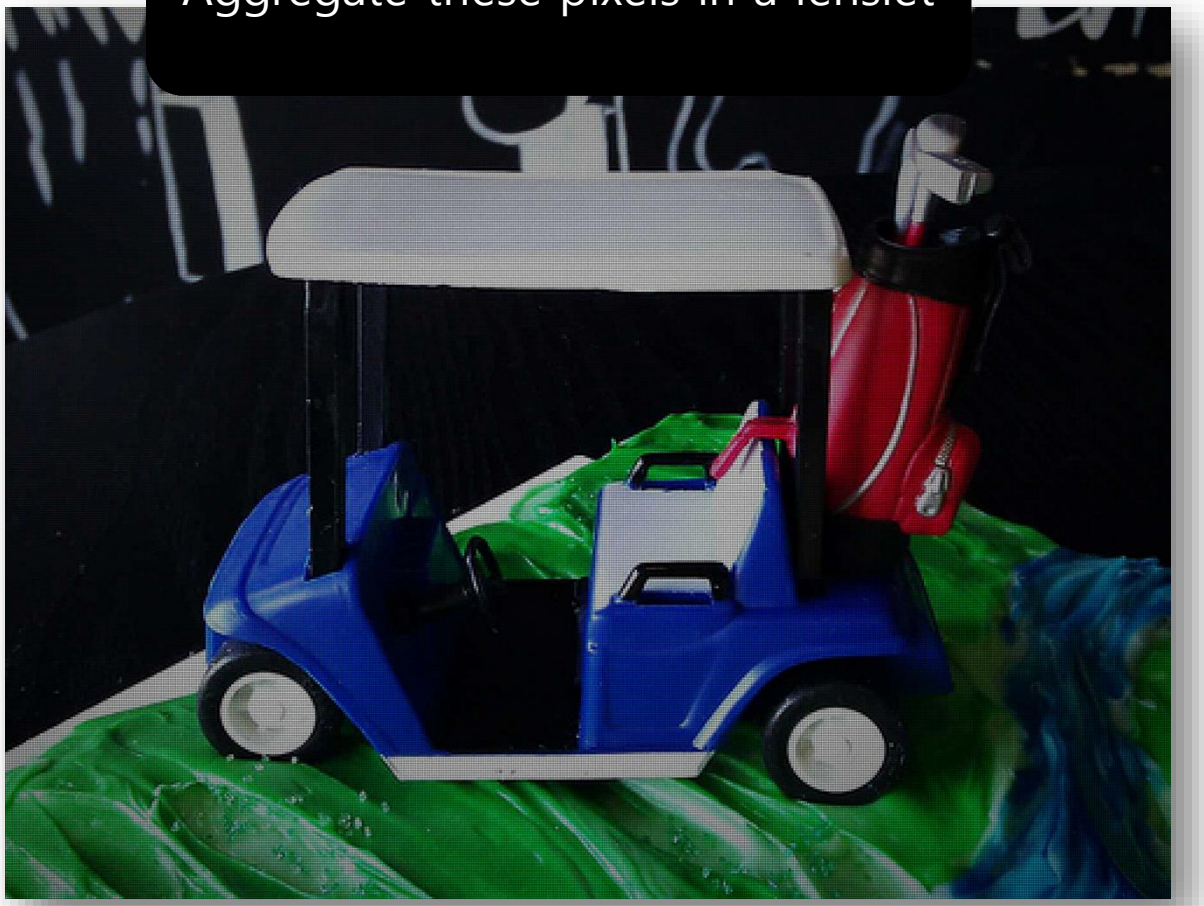
Vignetting map from averaged white plane images



Data Generation Lenslet Image Generation

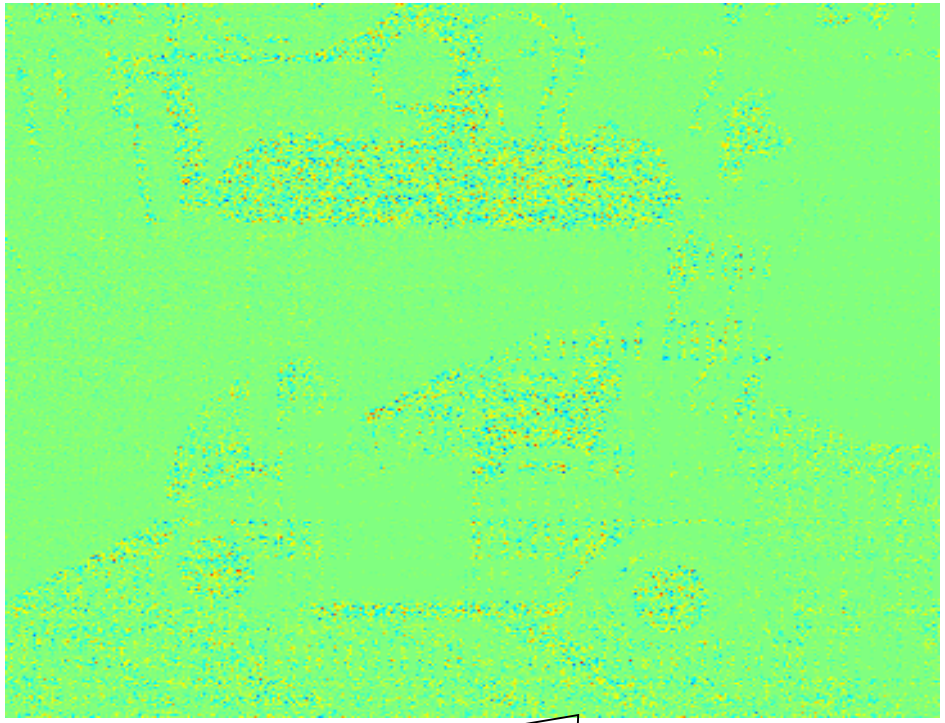


Sub-aperture image with vignetting map



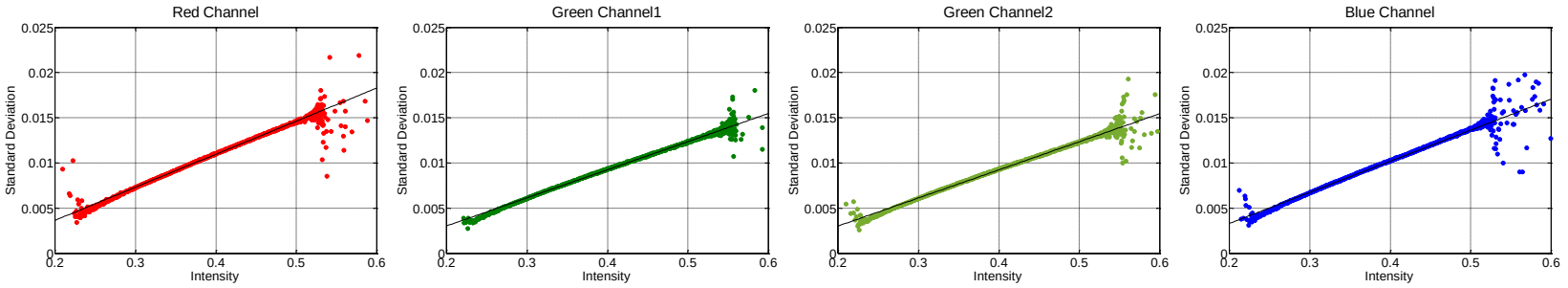
Data Generation Add Noise





Convert color image to raw image

Noise level estimation of each color channel

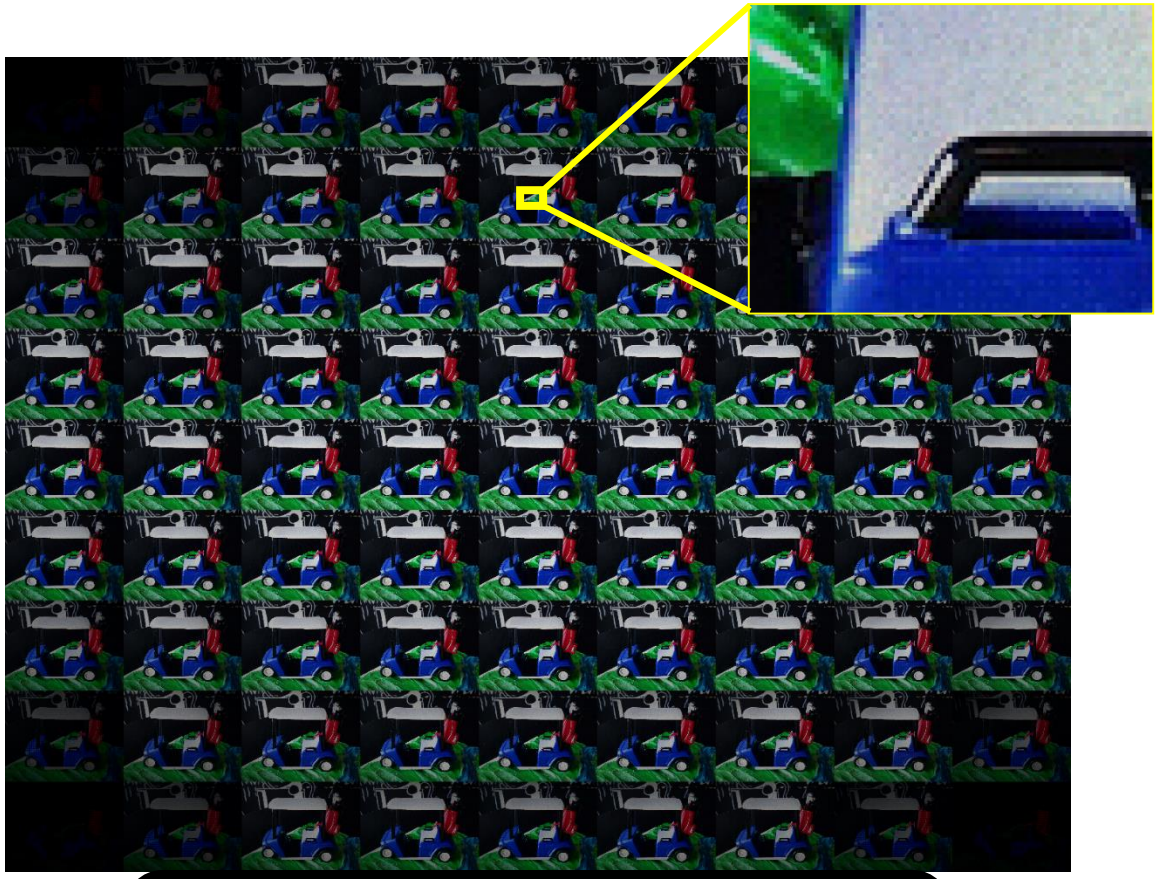


Y. Schechner et al., "Multiplexing for optimal lighting", IEEE TPAMI 2007

Data Generation Realistic Sub-aperture Image Generation

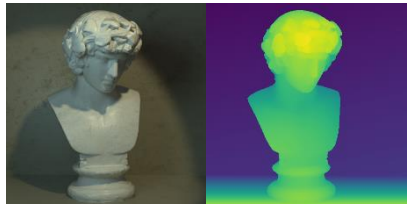


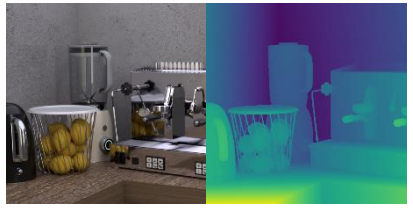
Normal image

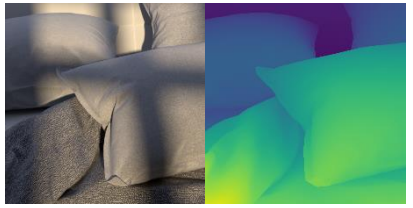


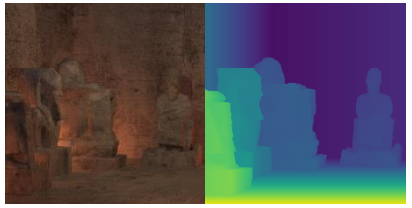
Rearrange pixels at each lenslet to each sub-aperture image

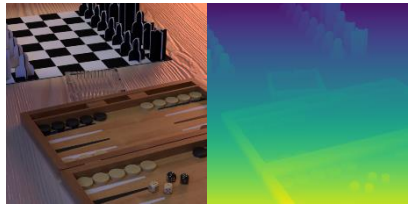
Training Set http://hci-lightfield.iwr.uni-heidelberg.de/


Antinous, Range: [-3.3, 2.8]

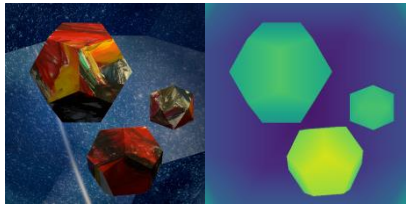

Boardgames, Range: [-1.8, 1.6]

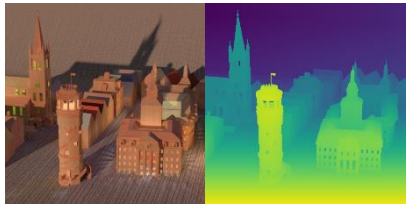

Dishes, Range: [-3.1, 3.5]


Greek, Range: [-3.5, 3.1]


Kitchen, Range: [-1.6, 1.8]

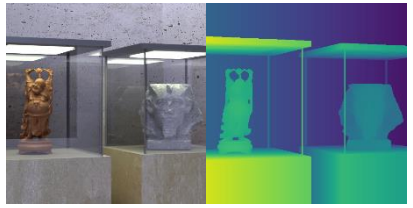

Medieval2, Range: [-1.7, 2.0]


Museum, Range: [-1.5, 1.3]

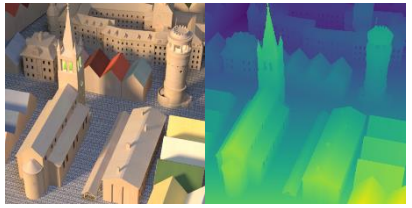

Pens, Range: [-1.7, 2.0]

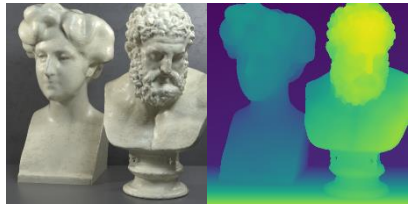


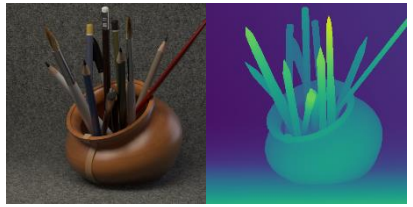
Pillows, Range: [-1.7, 1.8]


Platonic, Range: [-1.7, 1.5]

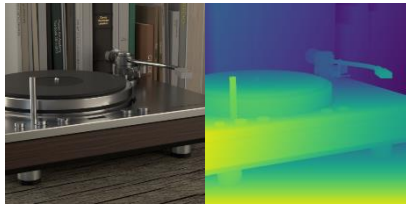

Rosemary, Range: [-1.8, 1.8]


Table, Range: [-2.0, 1.6]

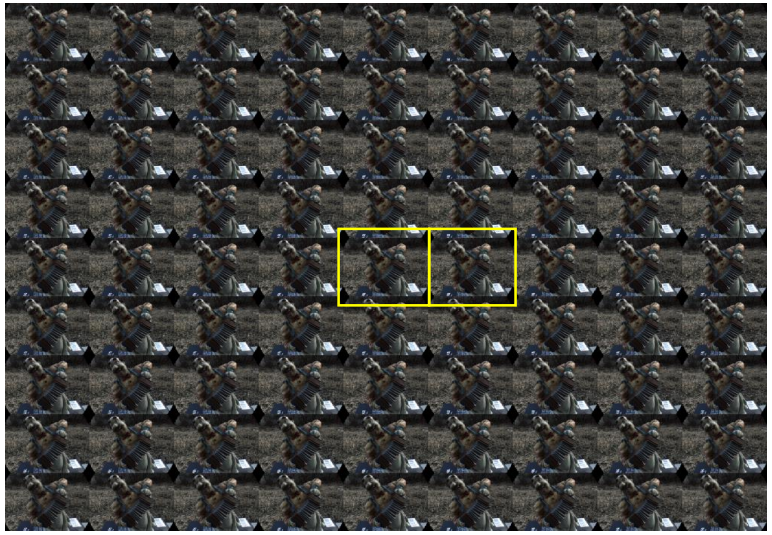

Tomb, Range: [-1.5, 1.9]


Tower, Range: [-3.6, 3.5]


Town, Range: [-1.6, 1.6]

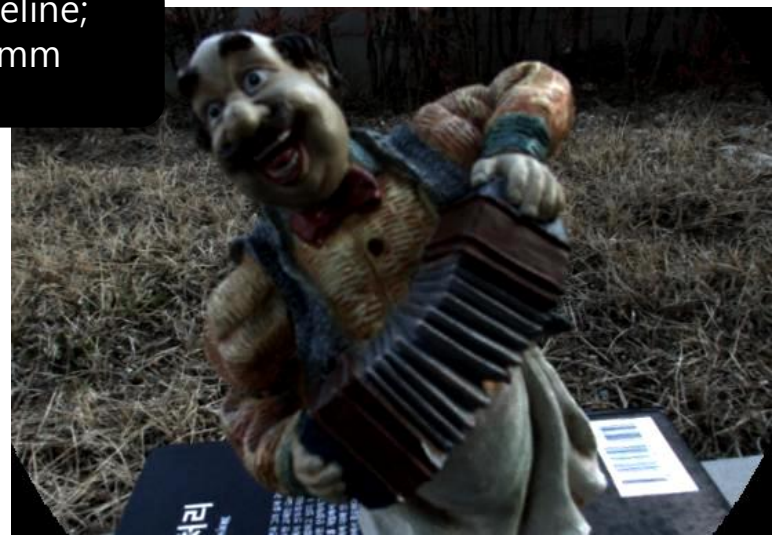

Vinyl, Range: [-1.6, 1.2]

Cost Volumes Phase Shift



Sub-aperture images

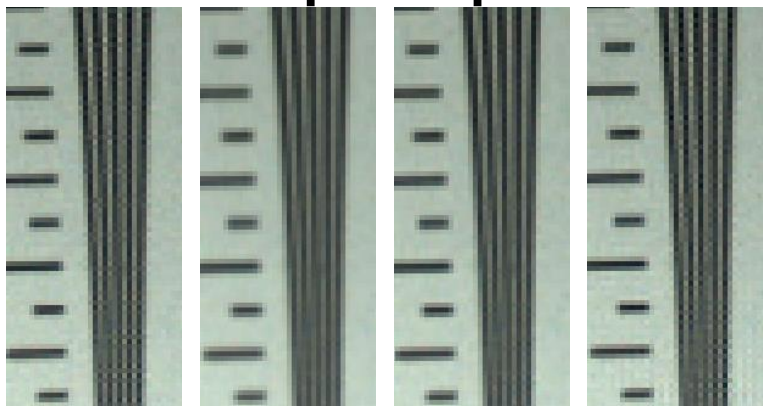
Very narrow baseline;
Physically 0.45mm
Within 1px



Flipping adjacent views

Our solution

Phase shift =>
1/100 pixel precision



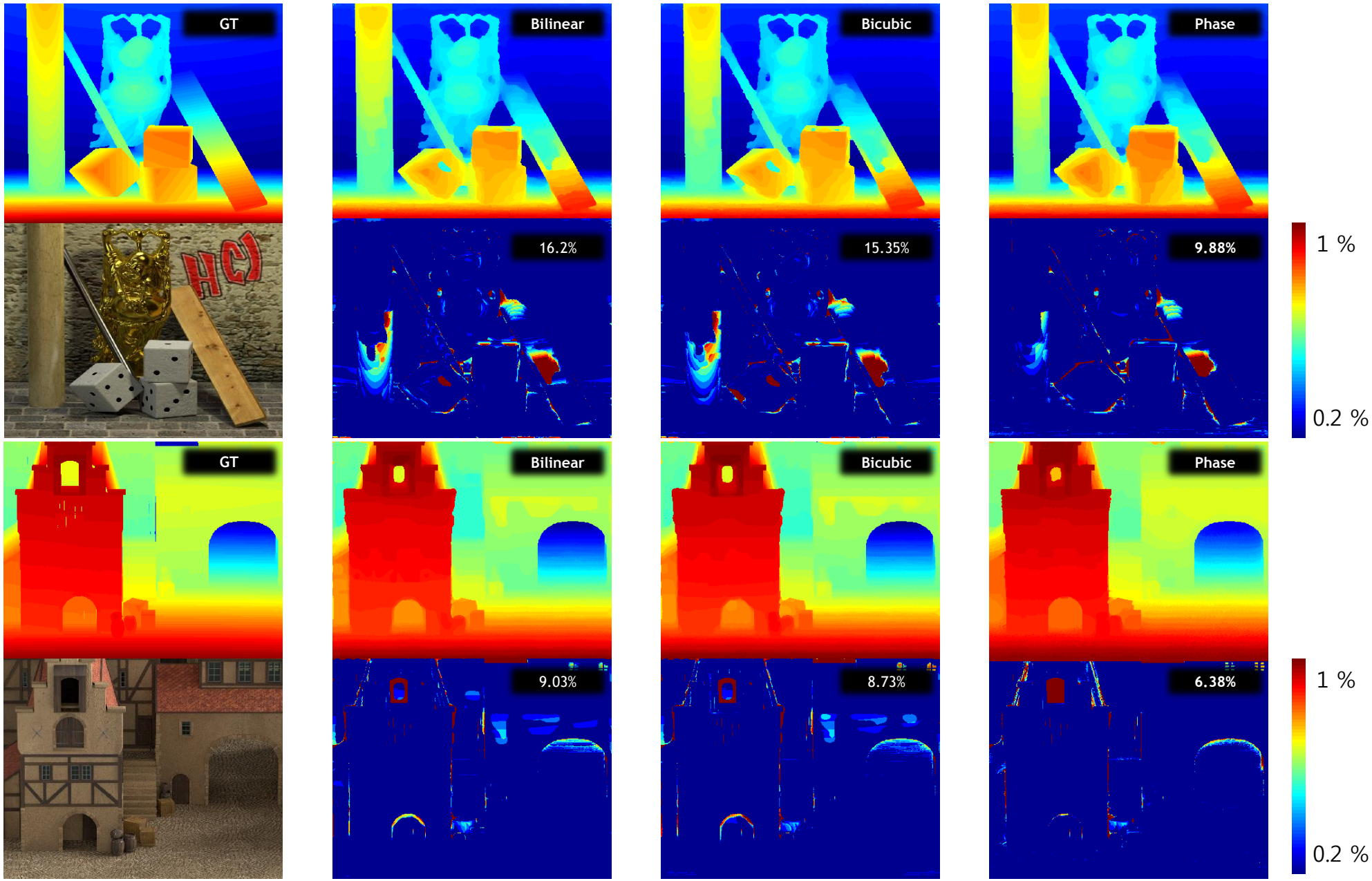
Original Bilinear Bicubic Phase



Averbuch and Keller, "A unified approach to FFT based images registration", IEEE TIP 2003

Cost Volumes Phase Shift

Jeon et al., "Accurate Depth Map Estimation from a Lenslet Light Field Camera", CVPR 2015



Cost Volumes Matching Costs

Sum of Absolute Difference (SAD)

+ Robust to image noise; act as averaged filter

$$C_A(\mathbf{x},l)=\sum_{\mathbf{s}\in V}\sum_{\mathbf{x}\in R_\mathbf{x}}\min(|I(\mathbf{s}_c,\mathbf{x})-I(\mathbf{s},\mathbf{x}+\Delta\mathbf{x}(\mathbf{s},l))|,\tau_1)$$

Zero-mean Normalized Cross correlation (ZNCC)

+ Compensate for differences in both gain and offset

$$C_Z(\mathbf{x},l)=\sum_{\mathbf{s}\in V}\frac{\sum_{\mathbf{x}\in R_\mathbf{x}}(I(\mathbf{s}_c,\mathbf{x})-\mu_1)(I(\mathbf{s},\mathbf{x}+\Delta\mathbf{x}(\mathbf{s},l))-\mu_2)}{\sqrt{\sum_{\mathbf{x}\in R_\mathbf{x}}(I(\mathbf{s}_c,\mathbf{x})-\mu_1)^2\sum_{\mathbf{x}\in R_\mathbf{x}}(I(\mathbf{s},\mathbf{x}+\Delta\mathbf{x}(\mathbf{s},l))-\mu_2)^2}}$$

Census Transform (Census)

+ Tolerate radiometric distortions

$$C_C(\mathbf{x},l)=H(\chi(I(\mathbf{s}_c,\mathbf{x}),I(\mathbf{s},\mathbf{x}+\Delta\mathbf{x}(\mathbf{s},l)))$$

$$\chi(\mathbf{x})=\begin{cases}1, & \text{if } p>I(\mathbf{x})\\ 0, & \text{otherwise,}\end{cases}$$

Sum of Gradient Difference (GRAD)

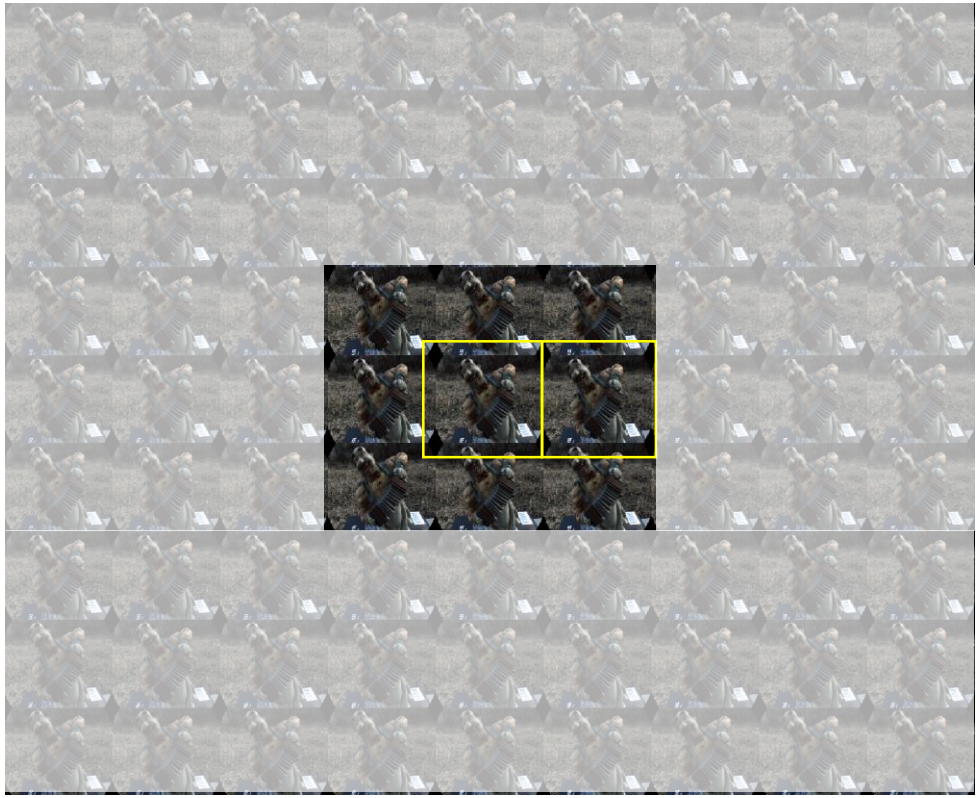
+ Synergy with other matching costs
+ imposing higher weights at edge boundaries

$$C_G(\mathbf{x},l)=\sum_{\mathbf{s}\in V}\sum_{\mathbf{x}\in R_\mathbf{x}}\beta(\mathbf{s})\min\left(\mathrm{Diff}_x(\mathbf{s}_c,\mathbf{s},\mathbf{x},l),\tau_2\right)+\left(1-\beta(\mathbf{s})\right)\min\left(\mathrm{Diff}_y(\mathbf{s}_c,\mathbf{s},\mathbf{x},l),\tau_2\right)$$

$$\beta(\mathbf{s})=\frac{|s-s_c|}{|s-s_c|+|t-t_c|}$$

H. Hirschmuller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," IEEE TPAMI 2009.

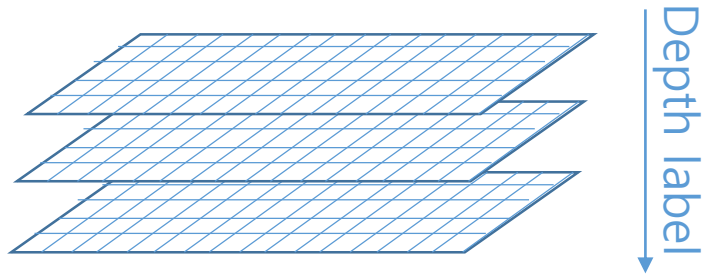
Cost Volumes Matching group1



Sub-aperture images

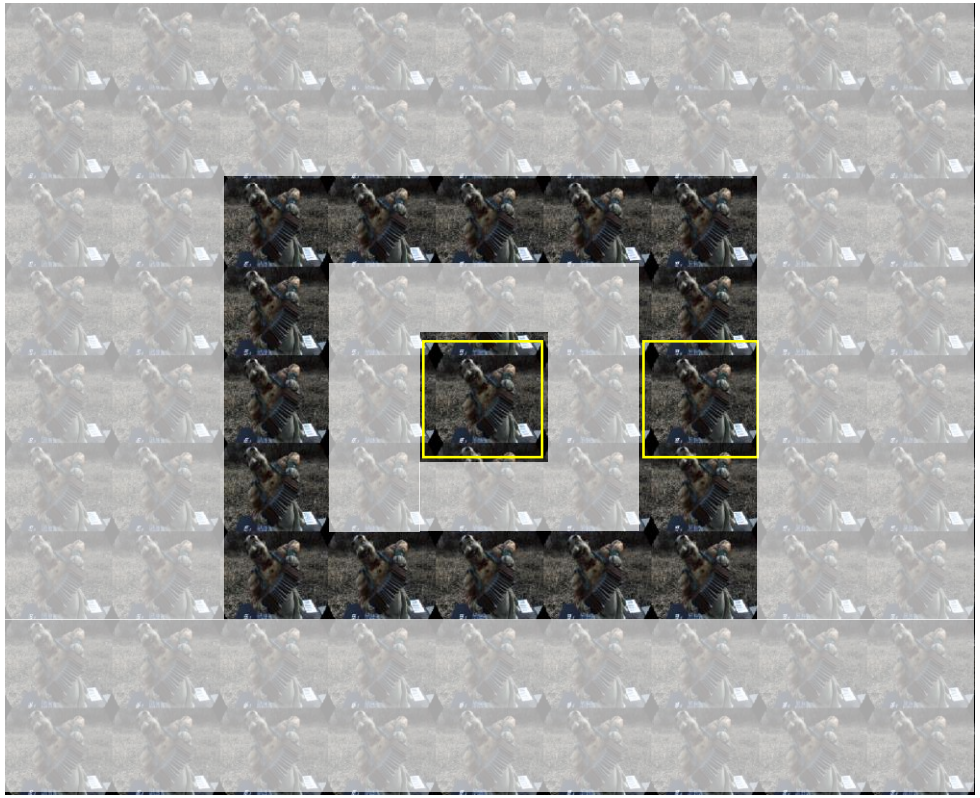
Matching Cost

$$f\left(\ \text{Reference view}\ ,\ \text{Target view}\ \right)$$



Cost volume

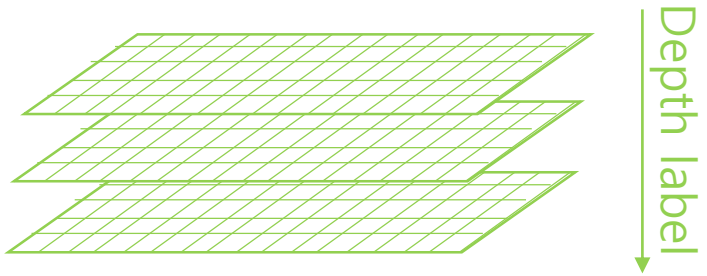
Cost Volumes Matching group2



Sub-aperture images

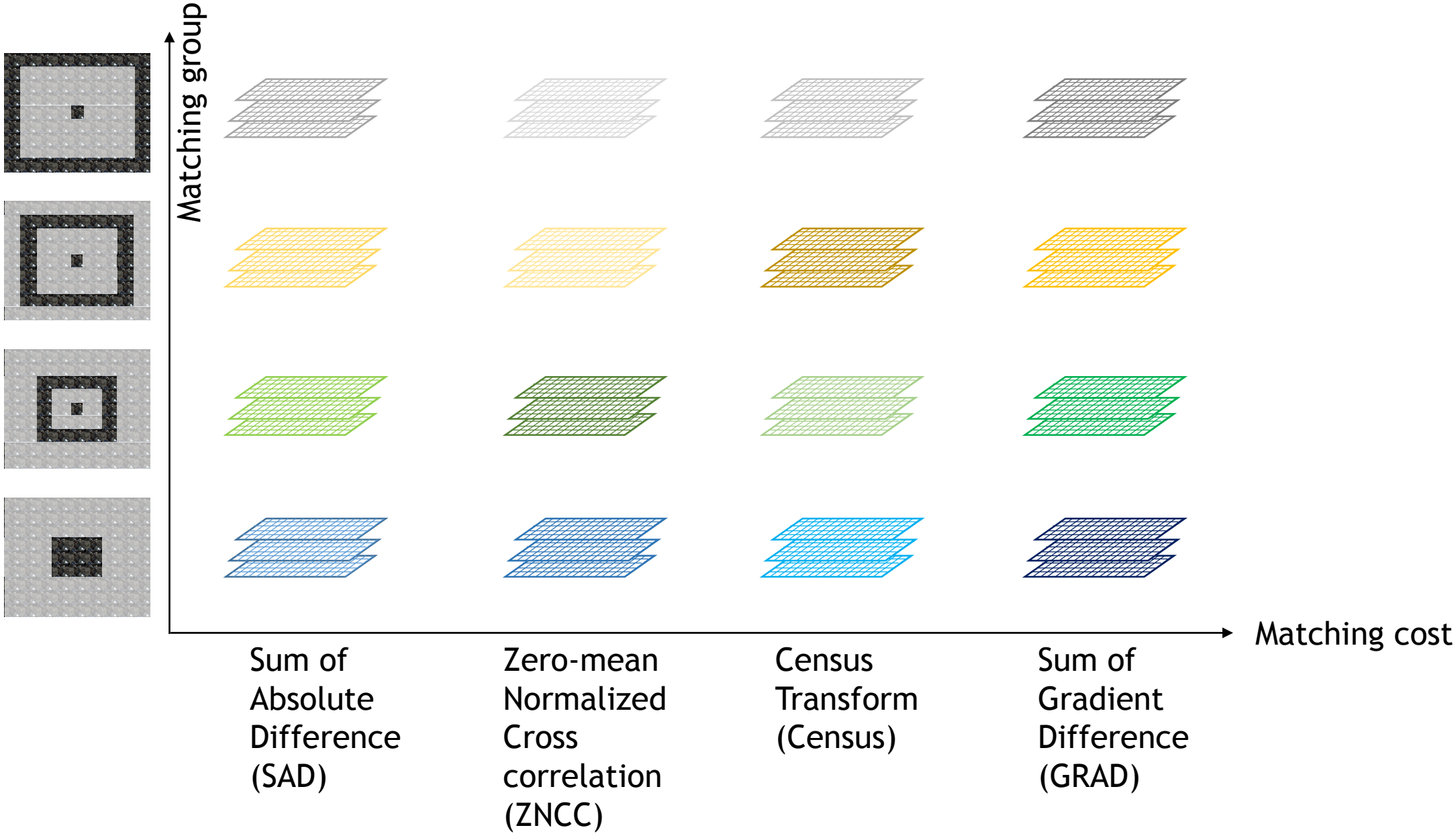
Matching Cost

$$f\left(\ \text{Reference view}\ ,\ \text{Target view}\ \right)$$

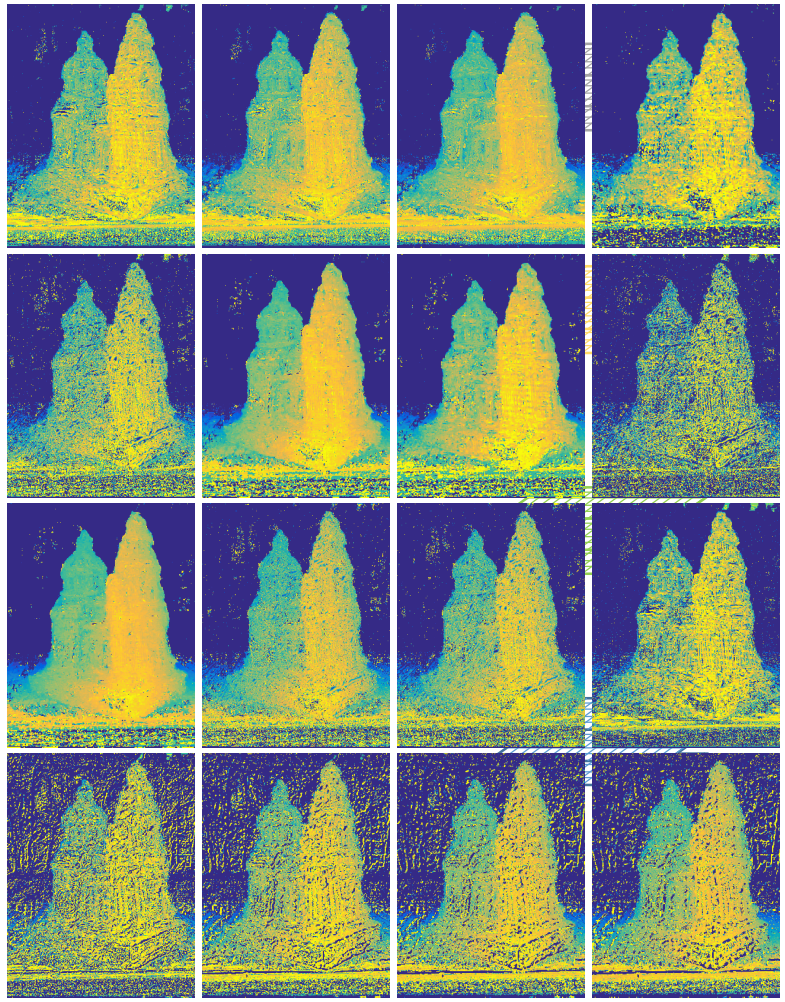


Cost volume

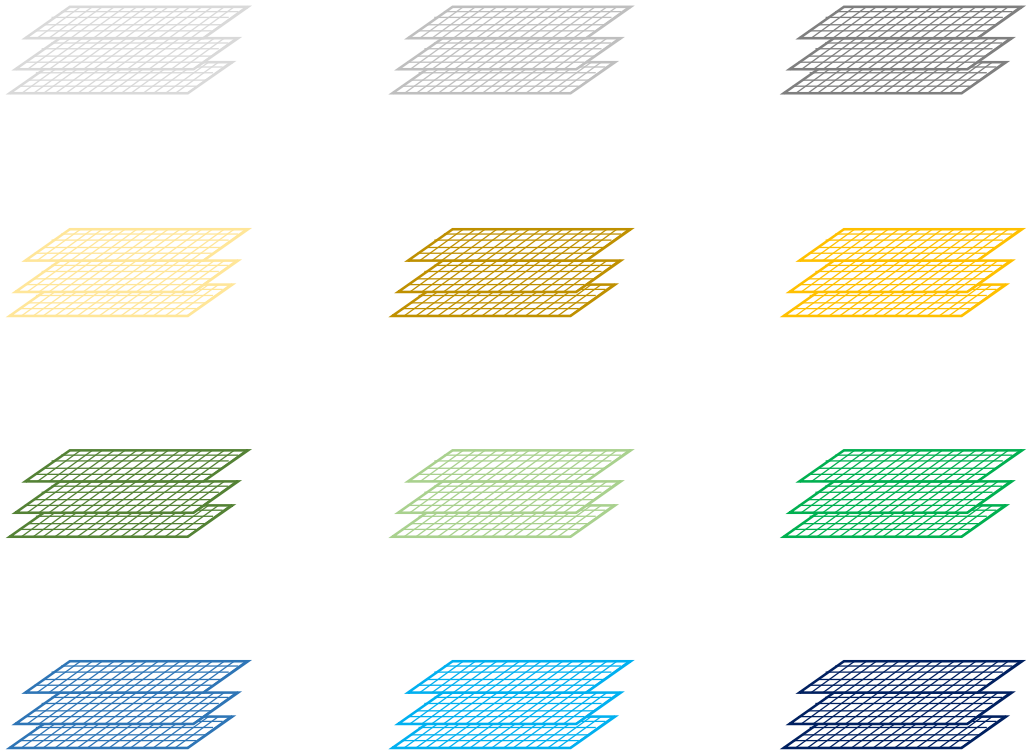
Cost Volumes Computed Cost Volumes



Cost Volumes Computed Cost Volumes

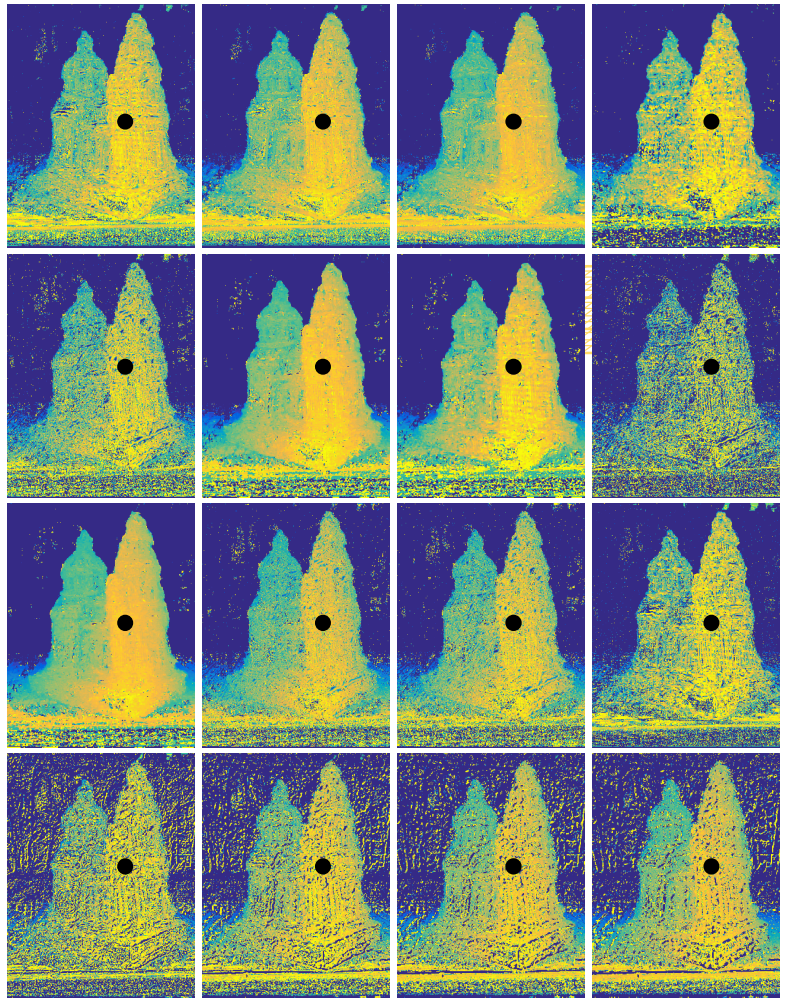


Disparities from each cost volume
via Winner-Takes-All



Cost Volumes Computed Cost Volumes

Campbell et al., "Using Multiple hypotheses to improve depth-maps for multi-view stereo", ECCV 2008



Vectorizing estimated depth labels with a ground truth depth label

Multiple disparity hypotheses

Ground truth

31 53 ⋯ 60 43 ⋯ 55 61 ⋯ 74 | 55

Multiple disparity hypotheses

67 51 ⋯ 58 53 ⋯ 37 66 ⋯ 12 | 55

Multiple disparity hypotheses

25 42 ⋯ 43 49 ⋯ 55 61 ⋯ 57 | 55

76 72 ⋯ 58 66 ⋯ 23 55 ⋯ 56 | 55

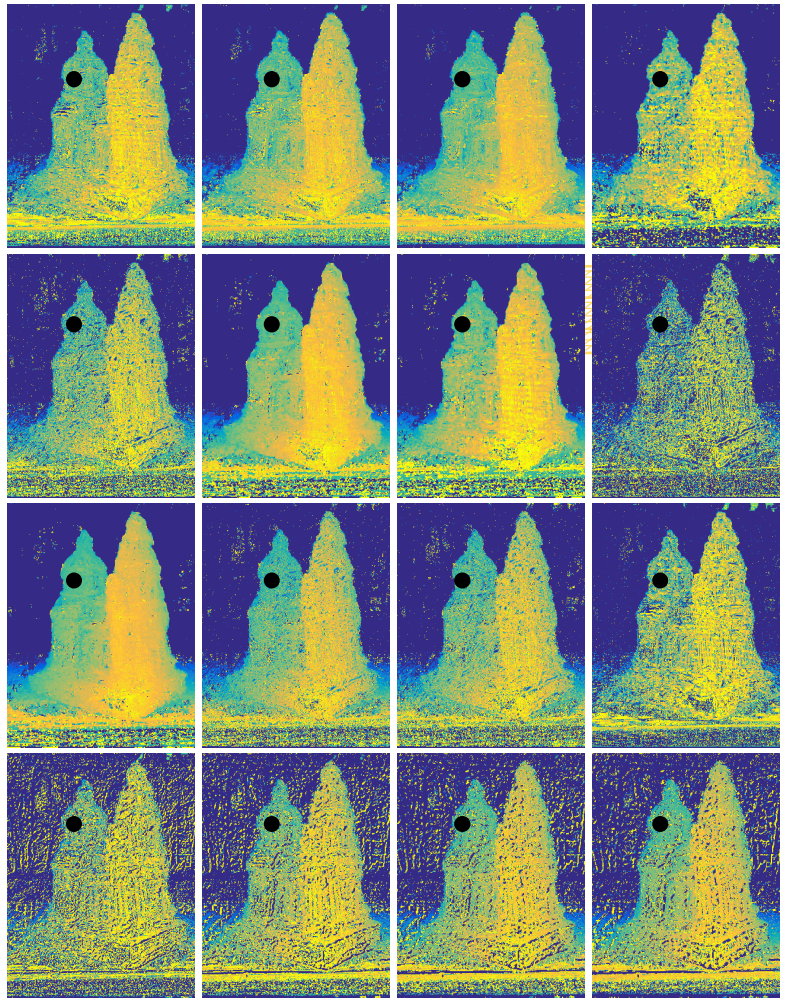
$C_{A-G}(\mathbf{x},l) = \alpha C_A(\mathbf{x},l) + (1-\alpha) C_G(\mathbf{x},l)$ SAD+GRAD

$C_{C-G}(\mathbf{x},l) = \alpha C_C(\mathbf{x},l) + (1-\alpha) C_G(\mathbf{x},l)$ GRAD+Census

$C_{A-C}(\mathbf{x},l) = \alpha C_A(\mathbf{x},l) + (1-\alpha) C_C(\mathbf{x},l)$ Census+SAD

$\alpha \in [0, 1.0]$

Cost Volumes Computed Cost Volumes



Vectorizing estimated depth labels with a ground truth depth label

Multiple disparity hypotheses

Ground truth

25 54 ⋯ 42 48 ⋯ 32 34 ⋯ 11 | 32

Multiple disparity hypotheses

19 20 ⋯ 33 43 ⋯ 37 32 ⋯ 5 | 32

31 42 ⋯ 34 29 ⋯ 12 41 ⋯ 57 | 32

Multiple disparity hypotheses

44 39 ⋯ 17 56 ⋯ 49 43 ⋯ 32 | 32

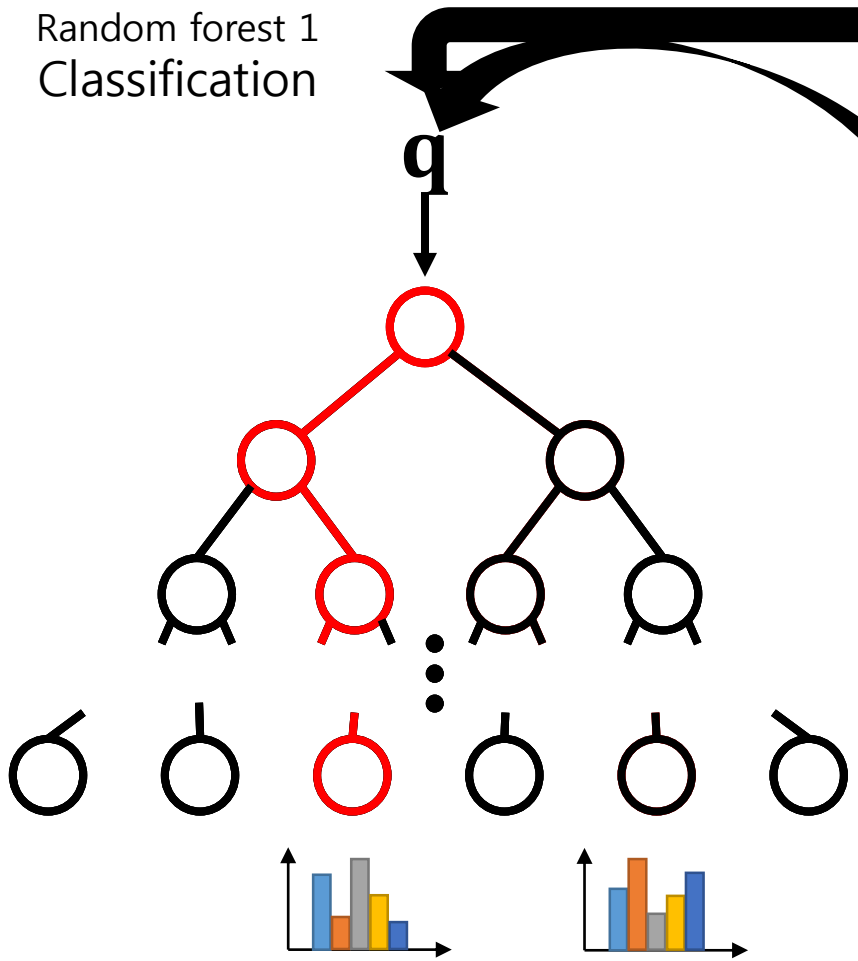
$C_{A-G}(\mathbf{x},l) = \alpha C_A(\mathbf{x},l) + (1-\alpha) C_G(\mathbf{x},l)$ SAD+GRAD

$C_{C-G}(\mathbf{x},l) = \alpha C_C(\mathbf{x},l) + (1-\alpha) C_G(\mathbf{x},l)$ GRAD+Census

$C_{A-C}(\mathbf{x},l) = \alpha C_A(\mathbf{x},l) + (1-\alpha) C_C(\mathbf{x},l)$ Census+SAD

$\alpha \in [0, 1.0]$

Random Forest1 - Classification



25	54	…	42	48	…	32	34	…	11	32
19	20	…	33	43	…	37	32	…	5	32
31	42	…	34	29	…	12	41	…	57	32
44	39	…	17	56	…	49	43	…	32	32

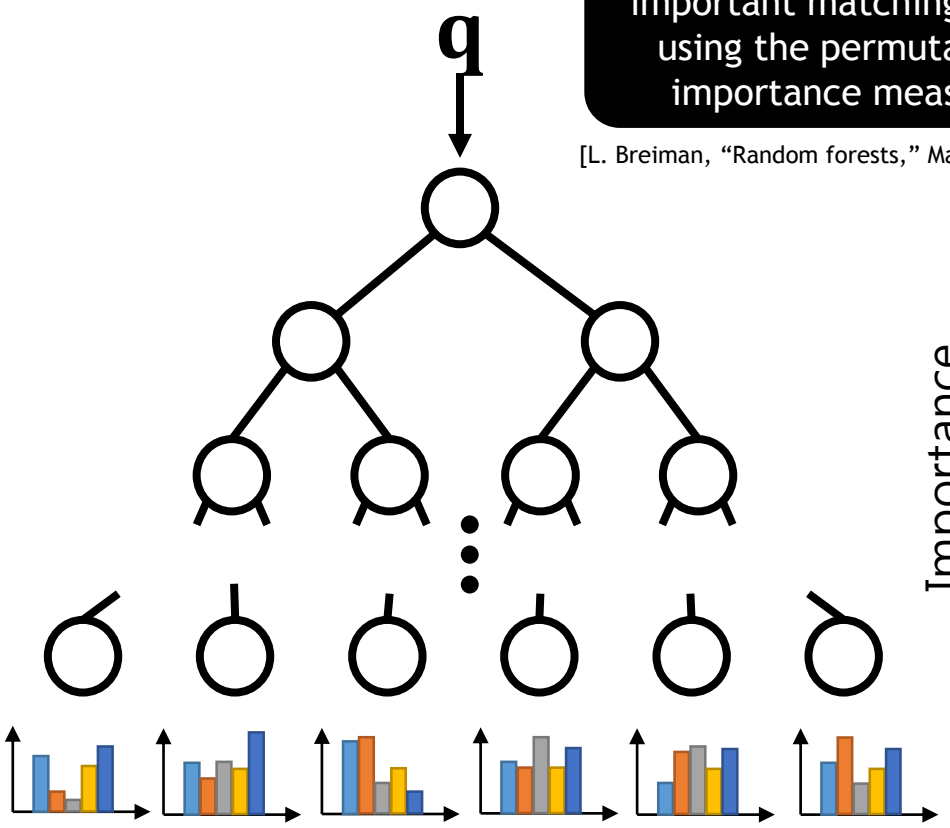
Training a random forest

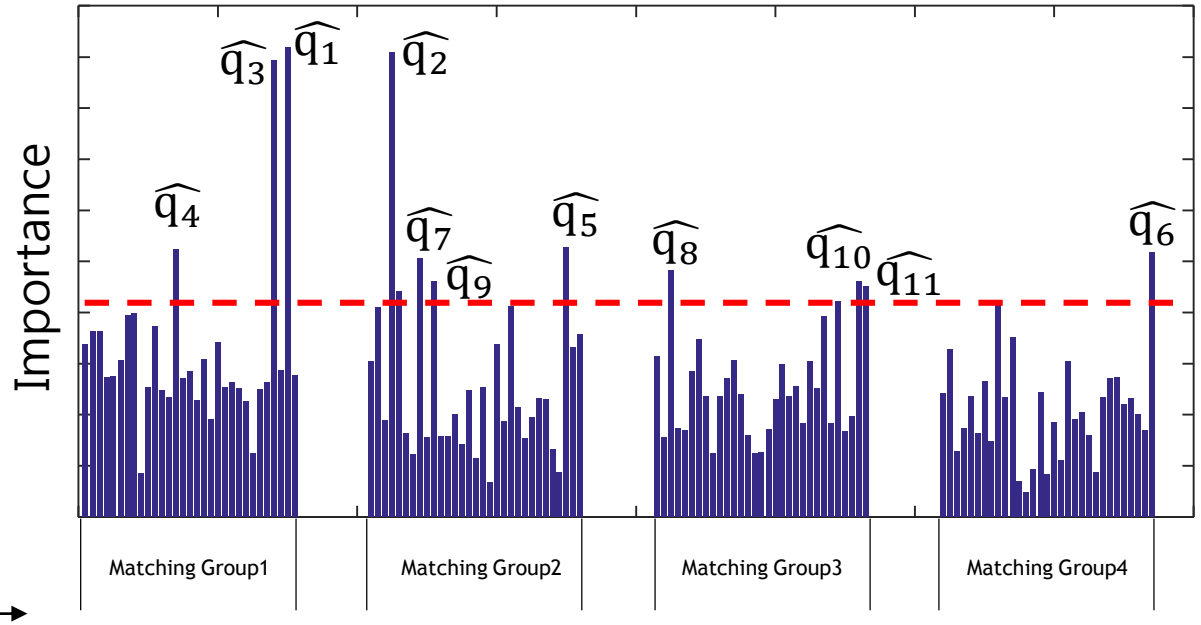
Random Forest1 - Classification

Retrieving a set of important matching costs using the permutation importance measure

[L. Breiman, “Random forests,” Machine learning]

+ Removing unnecessary matching cost
+ Designing a better prediction model





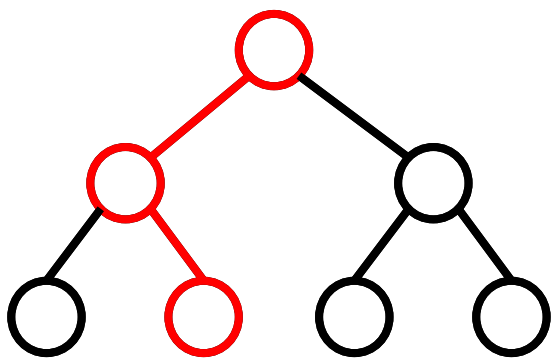
Random Forest2 - Regression

Input of a random forest for regression

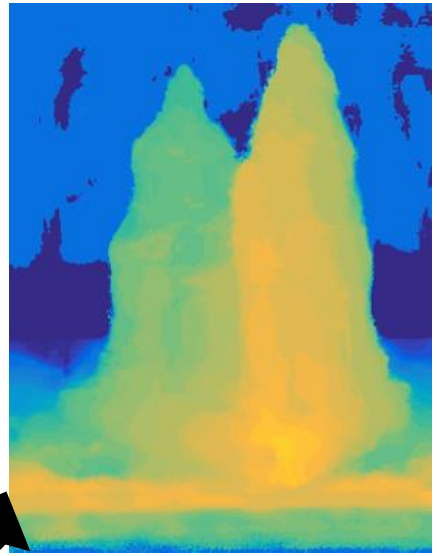
$$\begin{pmatrix} \widehat{q_1} & \widehat{q_2} & \widehat{q_3} & \widehat{q_4} & \widehat{q_5} & \widehat{q_6} & \widehat{q_7} & \widehat{q_8} & \widehat{q_9} & \widehat{q_{10}} & \widehat{q_{11}} \end{pmatrix}$$

$\hat{\mathbf{q}}$

Random forest 2
Regression

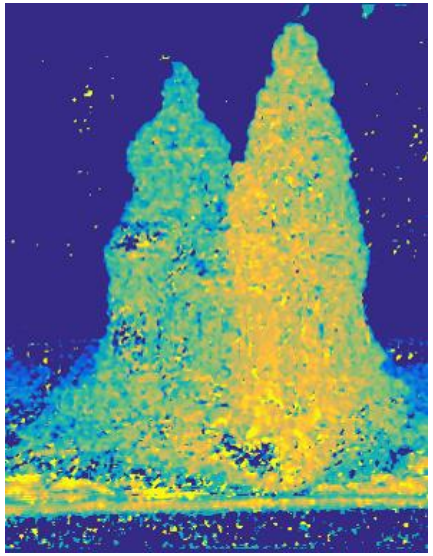


Estimated disparity value with sub-pixel precision



with Weighted Median Filter
[Z. Ma *et al.*, IEEE ICCV 2013]

vs.



SAD+GRAD
[H.-G. *Jeon et al.*, IEEE CVPR 2015]

Benchmark Bad pixel ratio (>0.07px) & Mean square error

Bad pixel ratio

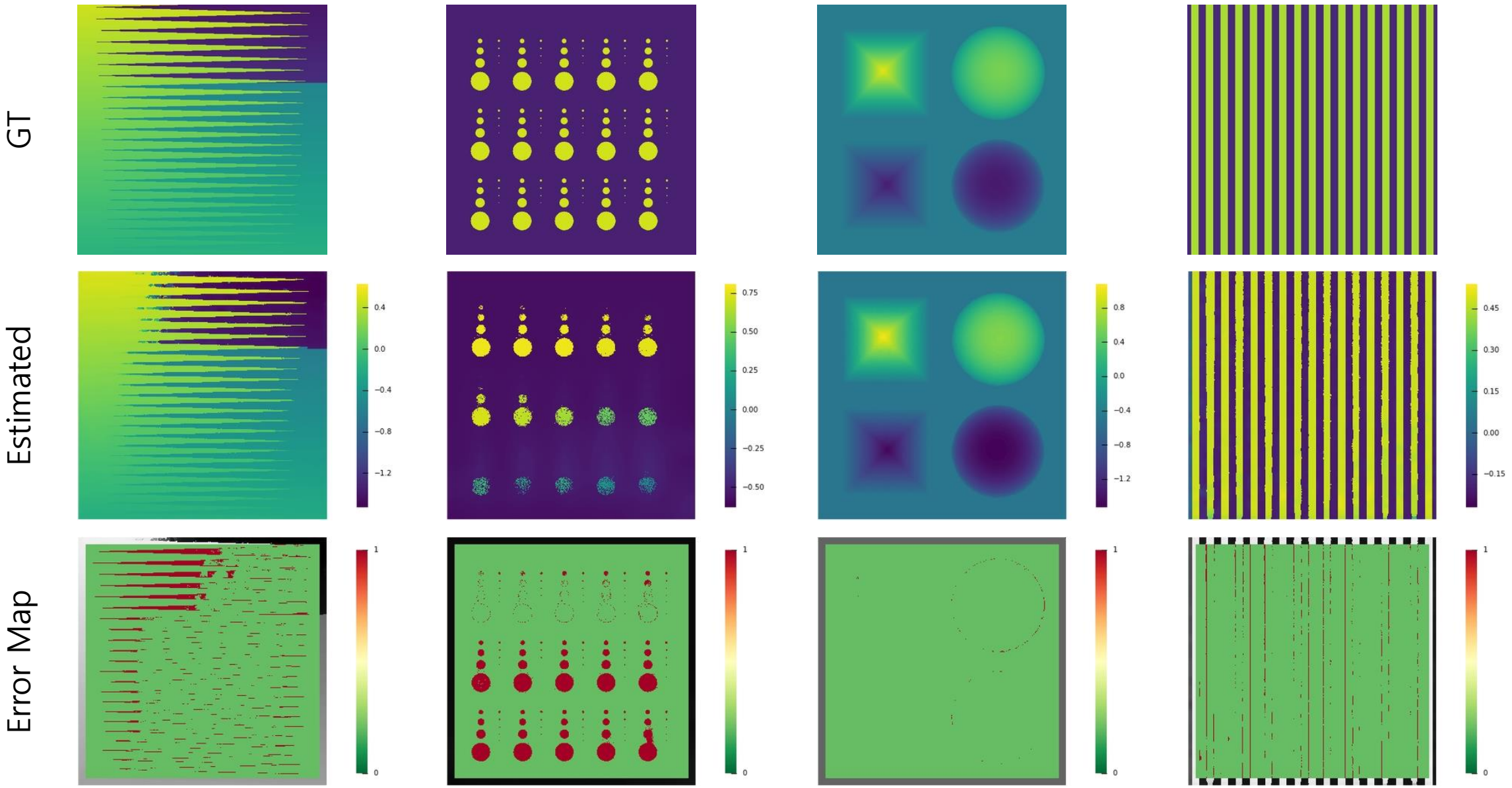
Algorithm	Meta MEDIAN No preview		Meta AVG No preview	
SPO	8.779	4	8.466	1
PS_RF	7.559	2	8.578	2
RM3DE	7.992	3	10.216	3
OBER-cross	6.715	1	10.289	4
OBER	9.021	5	11.677	5
OFSY_330/DNR	11.329	7	12.036	6
FBS	12.982	8	12.960	7
SC_GC	10.206	6	14.299	8
BSL_M	16.457	13	14.926	9
LF_OCC	17.816	15	15.074	10
OMG_occ	14.605	10	15.403	11
*LF	16.146	12	16.193	12
ZCTV	15.151	11	16.323	13
EPN+GC	14.426	9	16.896	14
*LF_OCC	18.451	16	17.579	15
*MV	17.736	14	18.405	16
MVCMv0	21.952	17	21.060	17
*EPI2	22.942	19	22.651	18
*EPI1	22.891	18	24.324	19
EpiNet_f	31.138	20	26.911	20

Mean square error

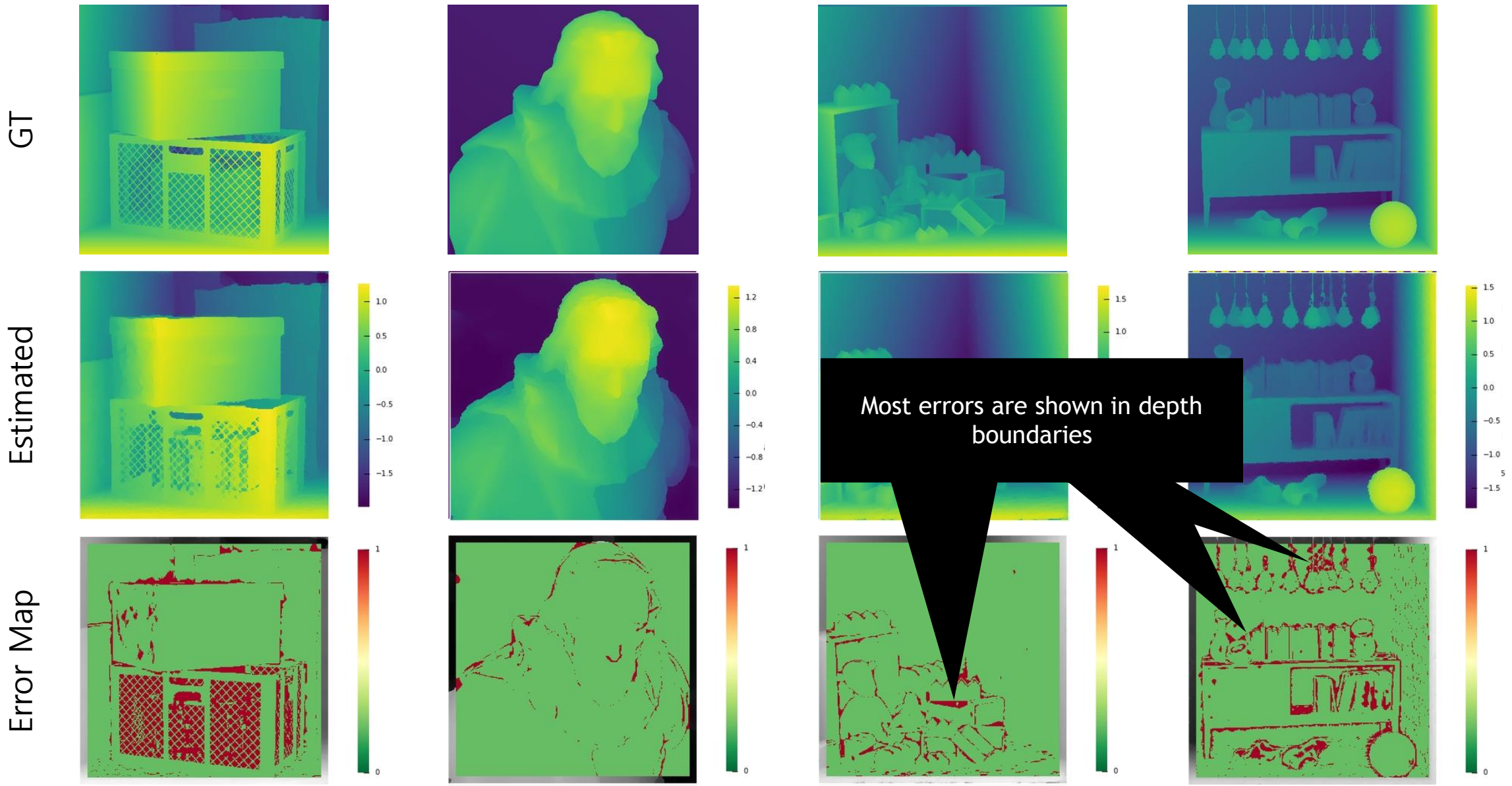
Algorithm	Meta MEDIAN No preview		Meta AVG No preview	
RM3DE	1.455	1	3.922	1
FBS	1.664	2	3.981	3
PS_RF	2.169	3	4.617	7
OBER	2.381	4	4.616	6
OBER-cross	2.547	5	4.010	4
LF_OCC	2.704	6	6.755	13
*LF_OCC	2.803	7	6.690	12
SPO	3.309	8	3.968	2
OMG_occ	3.737	9	6.586	10
*EPI1	3.932	10	5.975	8
SC_GC	3.940	11	6.584	9
EpiNet_f	4.573	12	4.436	5
BSL_M	5.082	13	7.429	15
EPN+GC	5.323	14	17.166	19
*MV	5.349	15	6.611	11
OFSY_330/DNR	5.428	16	7.030	14
ZCTV	5.439	17	8.442	17
*EPI2	5.723	18	8.240	16
*LF	7.963	19	9.128	18
MVCMv0	11.094	20	29.175	20

(2017.05.23)

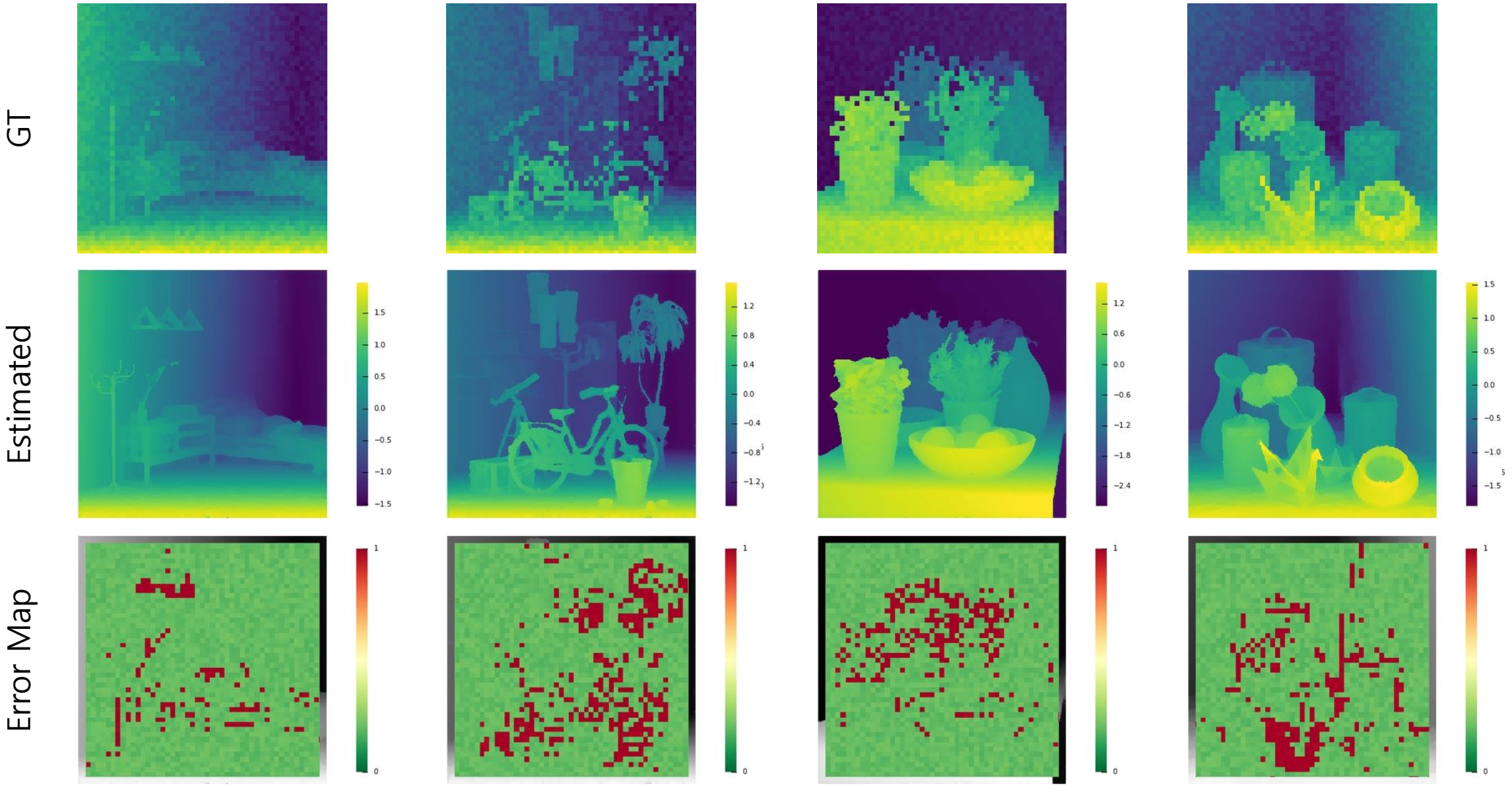
Evaluation Results - Stratified



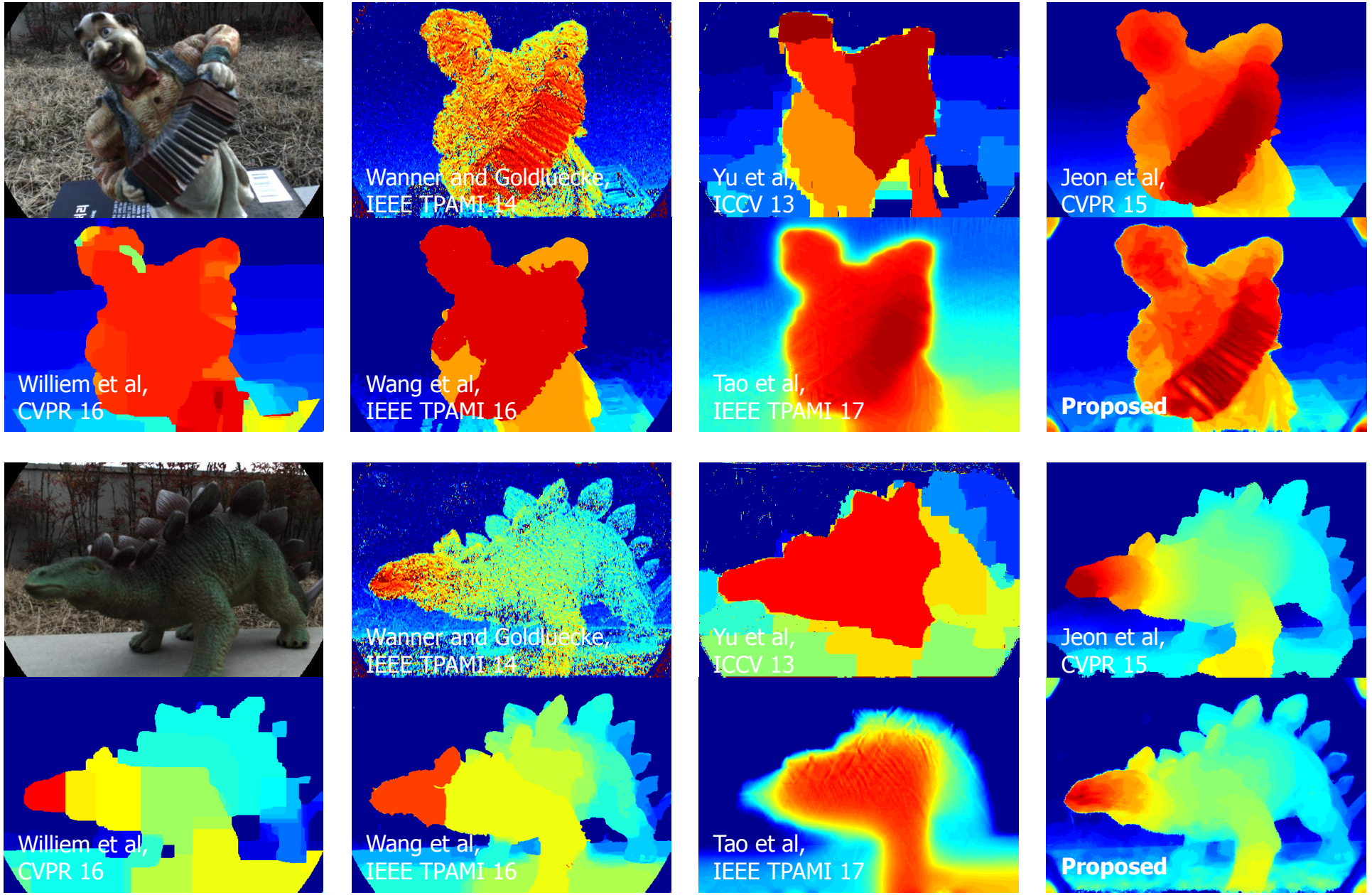
Evaluation Results - Training



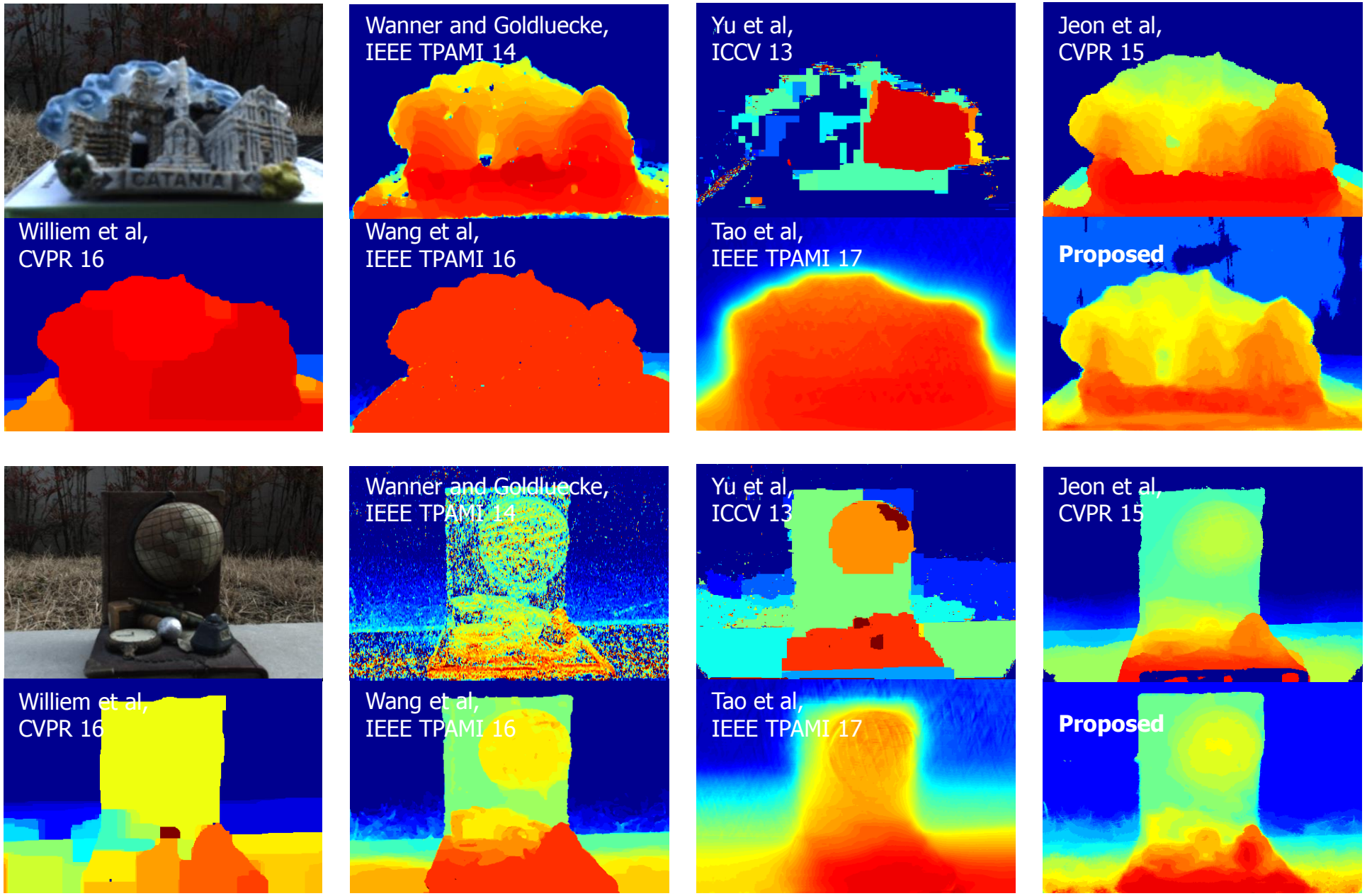
Evaluation Results - Test



Real World Examples – Lytro Illum



Real World Examples – Lytro Illum



Conclusion

Contributions:
- Analysis of the problems of depth estimation using light-field cameras
- Data augmentation that simulates a pipeline of a hand-held light-field camera
- Pixel-wise disparity value prediction using two random forests

Pros: Accurate disparity estimation
+ Handling narrow baseline problem
+ Robust to Image noise
+ Applicable real-world light field images

Cons:
- Heavy computational burden
- Need to minimize disparity error in depth discontinuities
- Requiring for handling textureless regions



Data Generation Add Noise



