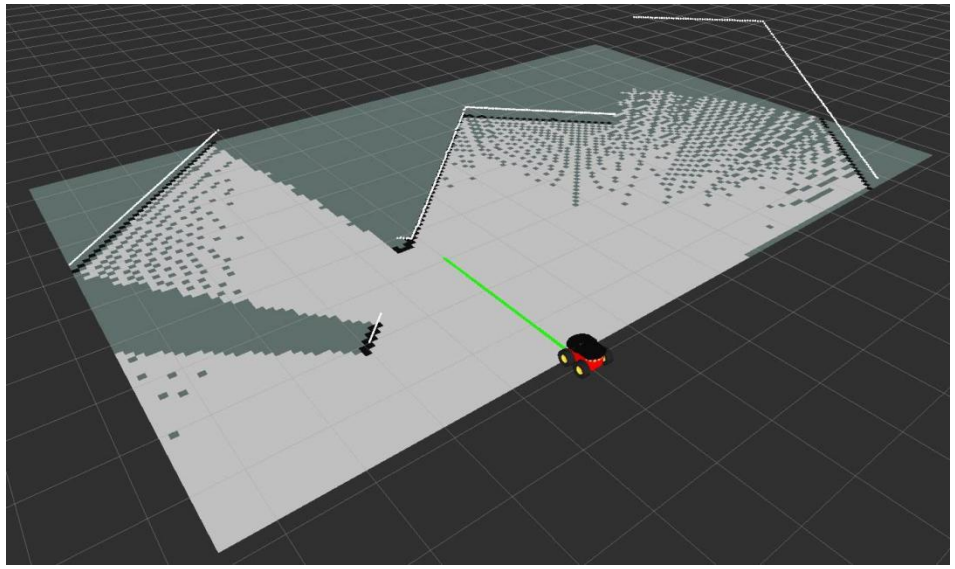
Variational Fusion of Time-of-Flight and Stereo Data Using Edge Selective Joint Filtering

Baoliang Chen, Cheolkon Jung, Zhendong Zhang
School of Electronic Engineering
Xidian University, China

Depth Estimation

- **Depth**: Represent 3D structure in scenes
- **Depth estimation**: Challenging problem in computer vision
- Tools for depth estimation: Stereo vision systems, Time-of-Flight (ToF) camera, light-coded camera (Microsoft Kinect).
- **Applications**: Robot vision, automatic navigation, tracking and action recognition



Passive Stereo Matching

- Advantage: High-resolution depth estimation, effective for textured scenes
- Disadvantage: Difficult for boundaries, homogeneous regions, and repetitive patterns.

Active Depth Imaging by TOF

- Advantage: Independent of surface texture.
- Disadvantage: Low-resolution and systematic errors such as flying pixels, inter reflections and high noise - especially in regions with low infrared reflectance.

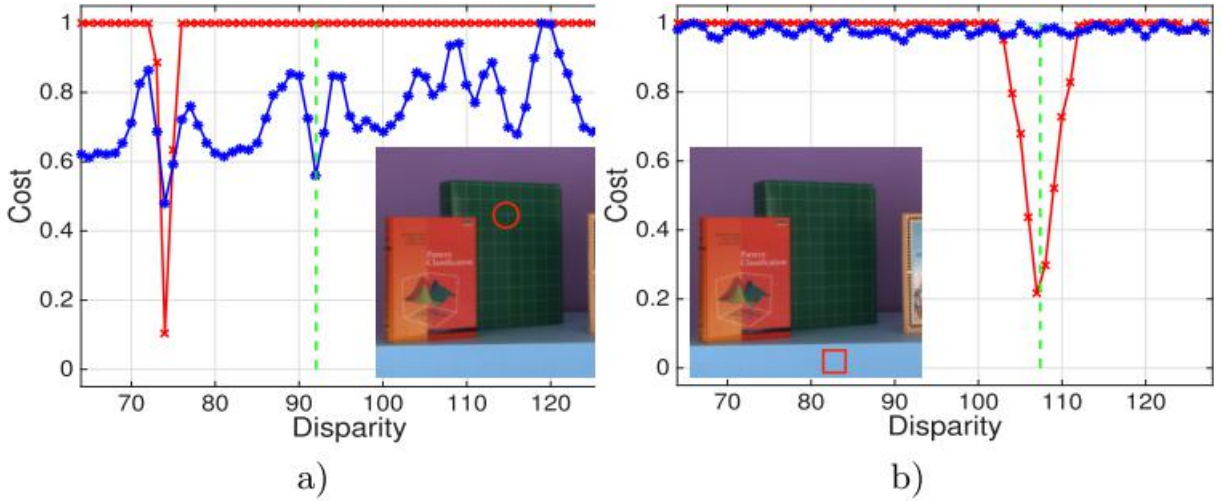
Characteristics of TOF and stereo data: Complementary

It is required to fuse them together.

Reliable Fusion by Confidence Measures (ECCV 2016)

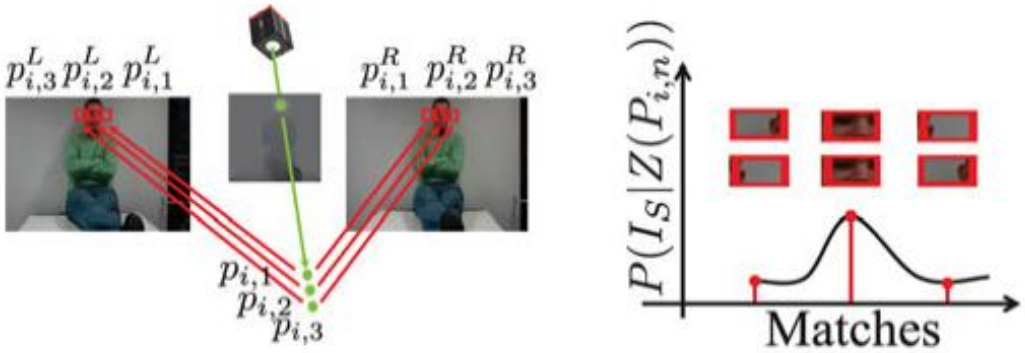
- Reliable confidence measures are extracted for both TOF and stereo depth data.
- Two depth maps are fused by enforcing the local consistency based on the confidence of the two data.

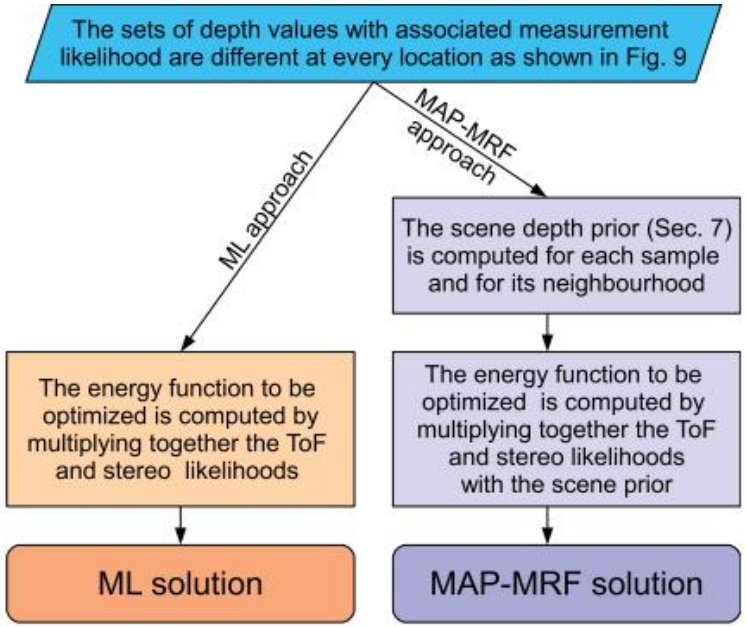
Stereo confidence accounts for the relationship between the pointwise matching costs and the cost obtained by the semi-global optimization.



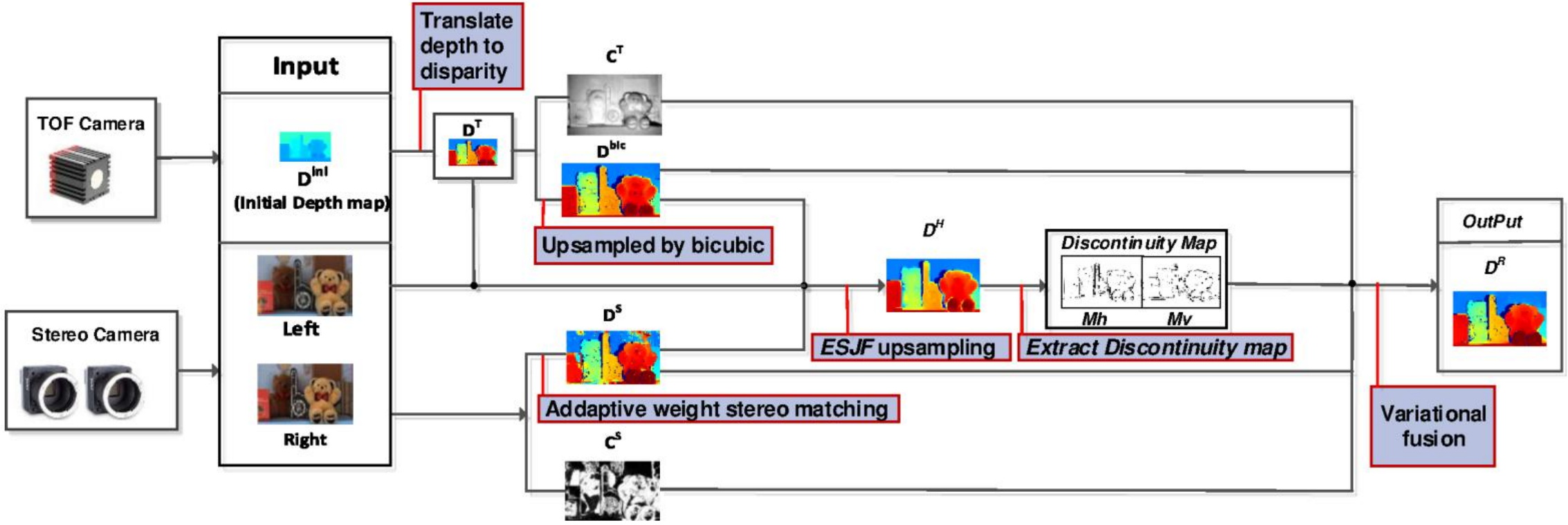
Probabilistic TOF and Stereo Data Fusion (TPAMI2015)

- **New inter-pixel and intra-pixel measurements models**.
- Measurement models are fused in two different probabilistic approaches: **Maximum Likelihood (ML)** and **global maximum-a-posteriori (MAP)-MRF**.





Proposed Method



Step 1: Confidence Measure

- **Stereo disparity** by adaptive support-weight (CVPR 2005) : D^S
- **Confidence measure:** C^S

$$\text{AML:}\quad P[d^R(p) = d^s(p)] \propto \frac{1}{\sum_d e^{-\frac{(c(d)-c_1)^2}{2\sigma_{AML}^2}}}$$

- **TOF upsampling by bicubic interpolation:** D^T

$$D^T = \frac{b \times f}{D^{init}} (b: baseline \ \ f: \text{focal length})$$

- **Confidence measure**

$$\sigma_z = \frac{c}{4\pi f_{mod}} \frac{\sqrt{I/2}}{A} \longrightarrow \sigma_d = bf \frac{\sigma_z}{z^2 - \sigma_z^2}$$

A: *Amplitude value* *I*: *Intensity* *C*: *Speed of light*

  - Flat scene areas: Gaussian function with standard deviation
  - Depth-discontinuity areas

$$C^{D^{ini}}(p) = \exp[-\frac{var(p)}{2\sigma_v^2}] * \frac{\sigma_{max} - \sigma_d}{\sigma_{max} - \sigma_{min}}$$

Step 2: Depth Up-sampling by Edge Selective Joint Filtering (ESJF)

To up-sample D^{ini} , we use I, D^{S} and D^{bic}. Three types of edges exist in them:

1) Edges exist in the depth map but don't exist in its color image;
2) Edges exist in the color image but its depth map is smooth;
3) Both depth map and color image have the same edges.

ESJF meets three cases to successfully preserve depth edges in depth upsampling :

$$\mathrm{D}^{\mathrm{H}}(p)=\frac{\sum_{\mathrm{q}\downarrow\in\mathrm{N}(\mathrm{p})} W_{d^{bic}} * W_{I} * \mathrm{D}^{\mathrm{T}}(\mathrm{q}_{\downarrow})}{\sum_{\mathrm{q}\downarrow\in\mathrm{N}(\mathrm{p})} W_{d^{bic}} * W_{I}}$$

For W_I :

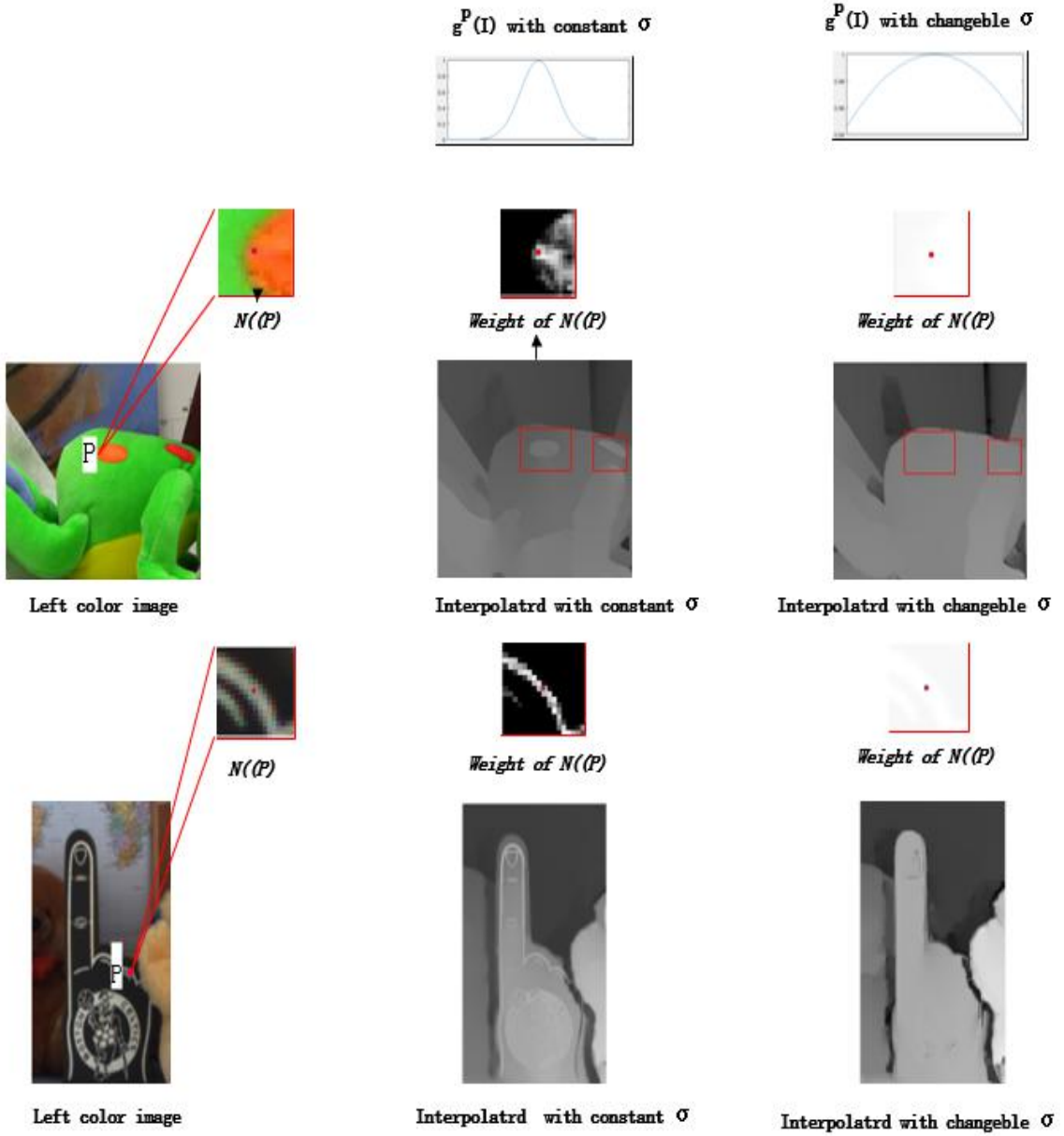
$$W_I = e^{-\frac{\|I(p)-I(q)\|^2}{\sigma_I(p)^2}}$$

$$\sigma_S(p) = e^{-\frac{var(d^S(p))}{2\sigma_{vs}^2}} \qquad \sigma_T(p) = e^{-\frac{var(d^{bic}(p))}{2\sigma_{vt}^2}}$$

$$\sigma_I(p) = \max(\alpha * \max(\sigma_S(p), \sigma_T(p)), Th)$$

$\sigma_I(p)$: Adjusting the weight of color image:

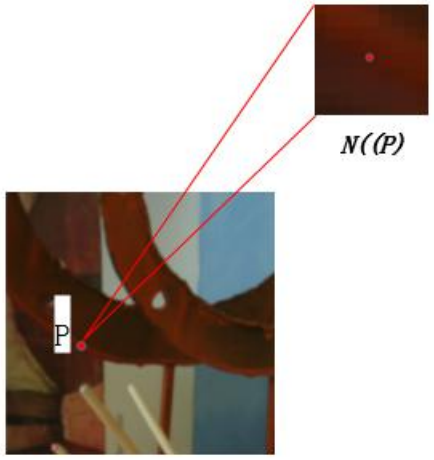
- Edges exist in the color image but its depth map is smooth,texture copying artifacts occur. $\sigma_I(p)$can be adjusted by $var(d^S(p))$ and $var(d^{bic}(p))$ which are close to zero in this case, difference between weights of p and its neighbors decreases,sothat,the texture copy phenomenon can be eliminate.
- If the edges are exist in the disparity map and color map, then it is small and make the final edges more sharp .

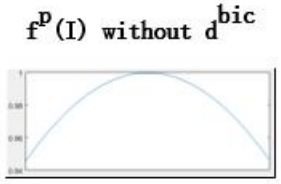


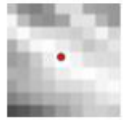
For $W_{d^{bic}}$:

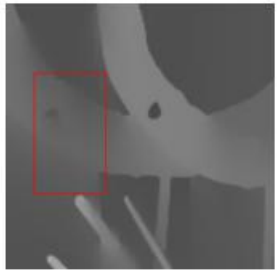
$$W_{d^{bic}} = e^{-\frac{\left\|d^{bic}(p)-d^{bic}(q)\right\|^2}{\sigma_T^2}}$$

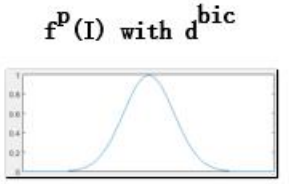
If Edges exist in the depth map but don't exist in its color image, this term preserves edge rather than smoothes it by the color image.

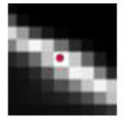


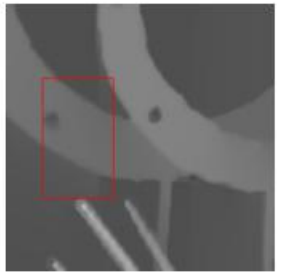










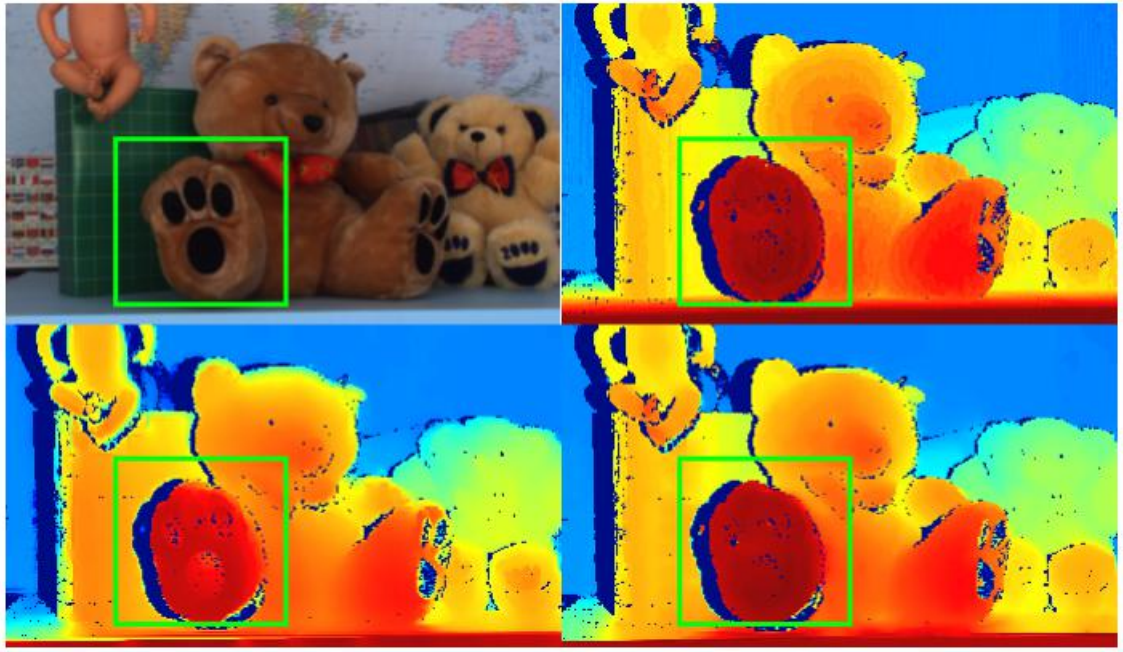


Step 3: Extraction of Discontinuity Maps

Disparities in D^H are not accurate, especially in the **green boxes**.

Reason: Use **only edge information** of stereo disparity D^S.

As the upsampling scale is high(about $\times$ 10),when the scene's depth is gradually changed,details in the final result will not accurate .

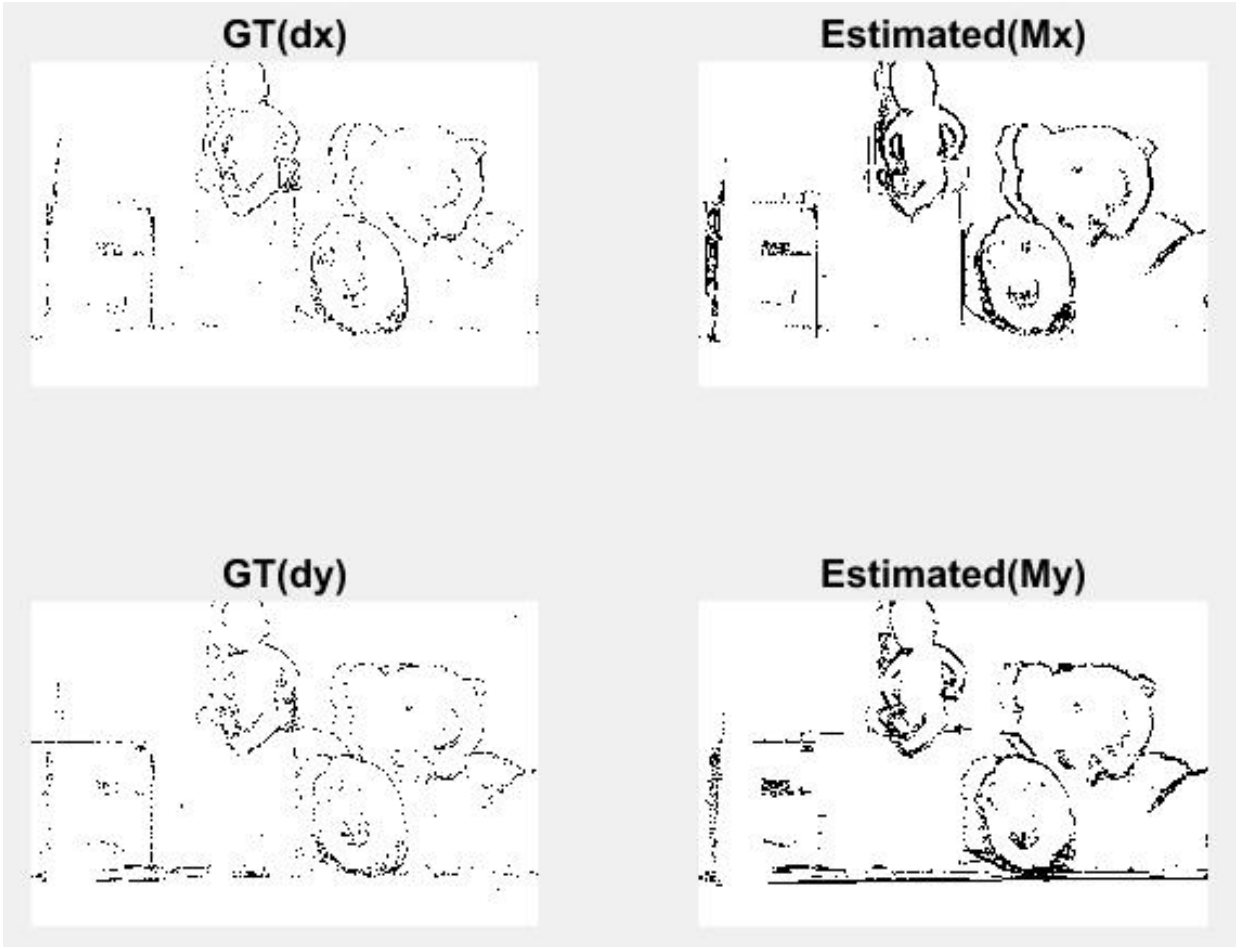


First row: Color image and Ground truth.
Second row: D^H and the final result D^R.

To guide weights for TV, we extract discontinuity maps vertically and horizontally for **variational data fusion** as follows:

$$M_h(p) = \begin{cases} 0 & \left|\partial x(D^H(p))\right| > th\_h \\ 1 & \text{others} \end{cases}$$

$$M_v(p) = \begin{cases} 0 & \left|\partial y(D^H(p))\right| > th\_v \\ 1 & \text{others} \end{cases}$$



Step 4: Variational Fusion

Variational fusion of data fidelity and regularization:

$$\min_{d_R} \mathrm{E}(d_R) = E(d_T) + \lambda E(d_S) + \lambda_2 R_{smooth}$$

λ, : Fixed parameter to adjust data fidelity and smoothness

$$E(d_T) = \sum_p \left[C_T(p) * (\mathrm{d_R}(p) - \mathrm{d_T}(p))^2 \right]$$

$$E(d_S) = \sum_p \left[C_S(p) * (\mathrm{d_R}(p) - \mathrm{d_S}(p))^2 \right]$$

$$\mathrm{R_{smooth}} = \sum_p \left[\mathrm{M_h}(p) |\partial_{\mathrm{x}}(\mathrm{d_R}(\mathrm{p}))|^2 + \mathrm{M_v}(p) |\partial_{\mathrm{y}}(\mathrm{d_R}(\mathrm{p}))|^2 \right]$$

Experimental Results

1) Experimental environment:

- Hardware: PC with CPU G3260 4 GB RAM
- Software: Matlab R2015b, Windows7
- Database: Multimedia Technology and Telecommunications Lab(5 scenes)

  http://lttm.dei.unipd.it/downloads/tofstereo/

2) Evaluation metrics:

- Mean squared errors (MSE)
- Structure similarity (SSIM)

3) Performance comparison:

- Giulio et al.: Reliable fusion by confidence measures (ECCV 2016)
- DalMutto1: Probabilistic TOF and stereo data fusion (TPAMI 2015)
- DalMutto2: Locally consistent TOF and stereo data fusion (ECCV 2012)
- Yang et al.: Fusion of active and passive sensors for fast 3d capture (TPAMI 2010)

Experimental Results: MSE

Our average RMSE is about 15% better than Giulio et al.'s method.

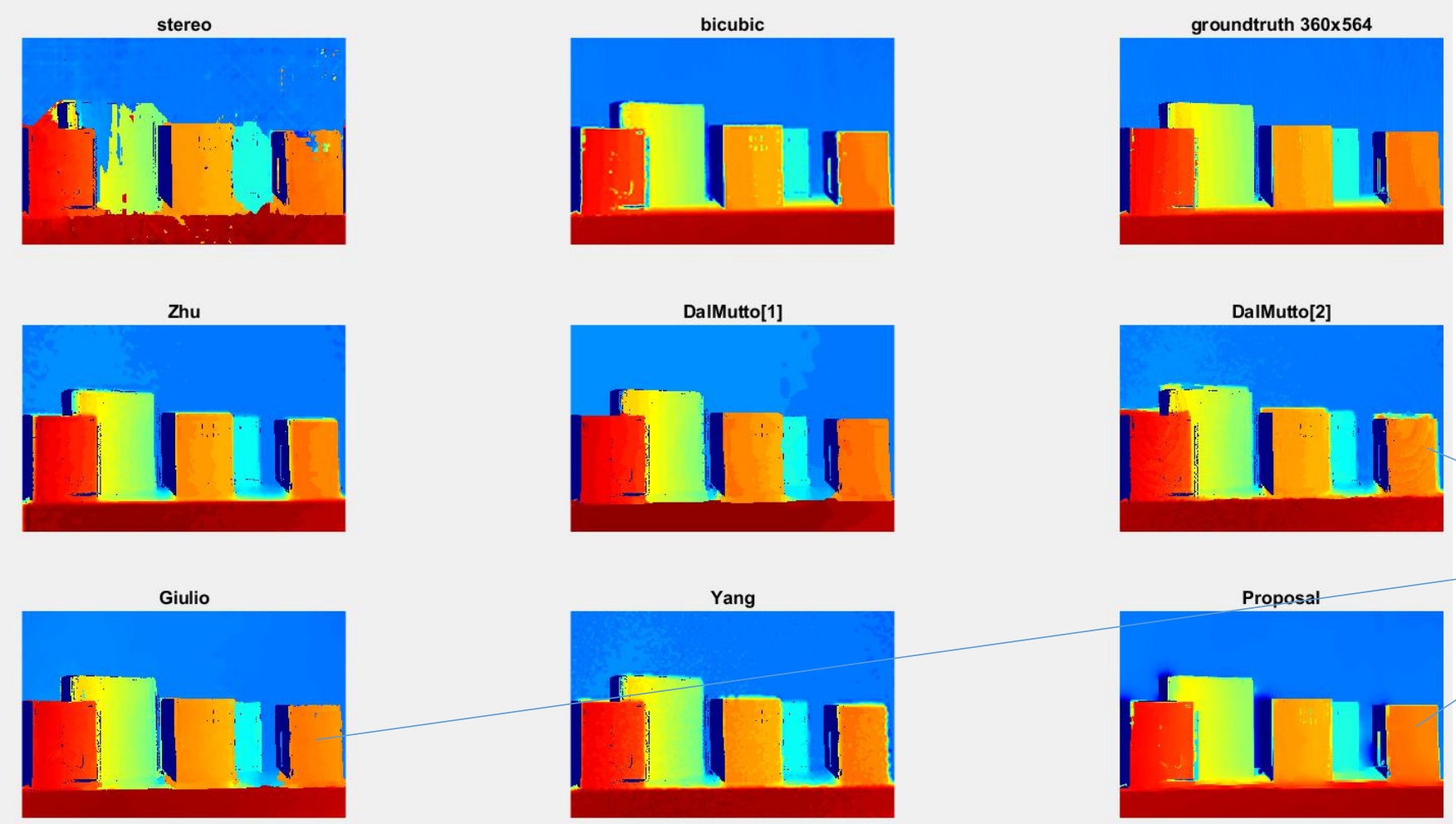
Method	**1**	**2**	**3**	**4**	**5**
TOF	12.25	12.25	15.67	14.86	15.66
Stereo	46.13	32.30	8.92	16.32	12.61
Giulio et al.	6.84	6.84	7.94	7.44	8.24
DalMutto1	8.36	8.36	7.65	8.32	9.95
DalMutto2	8.03	8.02	9.41	9.51	9.44
Yang et al.	7.45	7.44	10.85	10.56	12.36
Proposed	**6.29**	**6.35**	**4.62**	**4.30**	**3.83**

Experimental Results: SSIM

The proposed method achieves the best SSIM value by producing **accurate edge information** in depth.

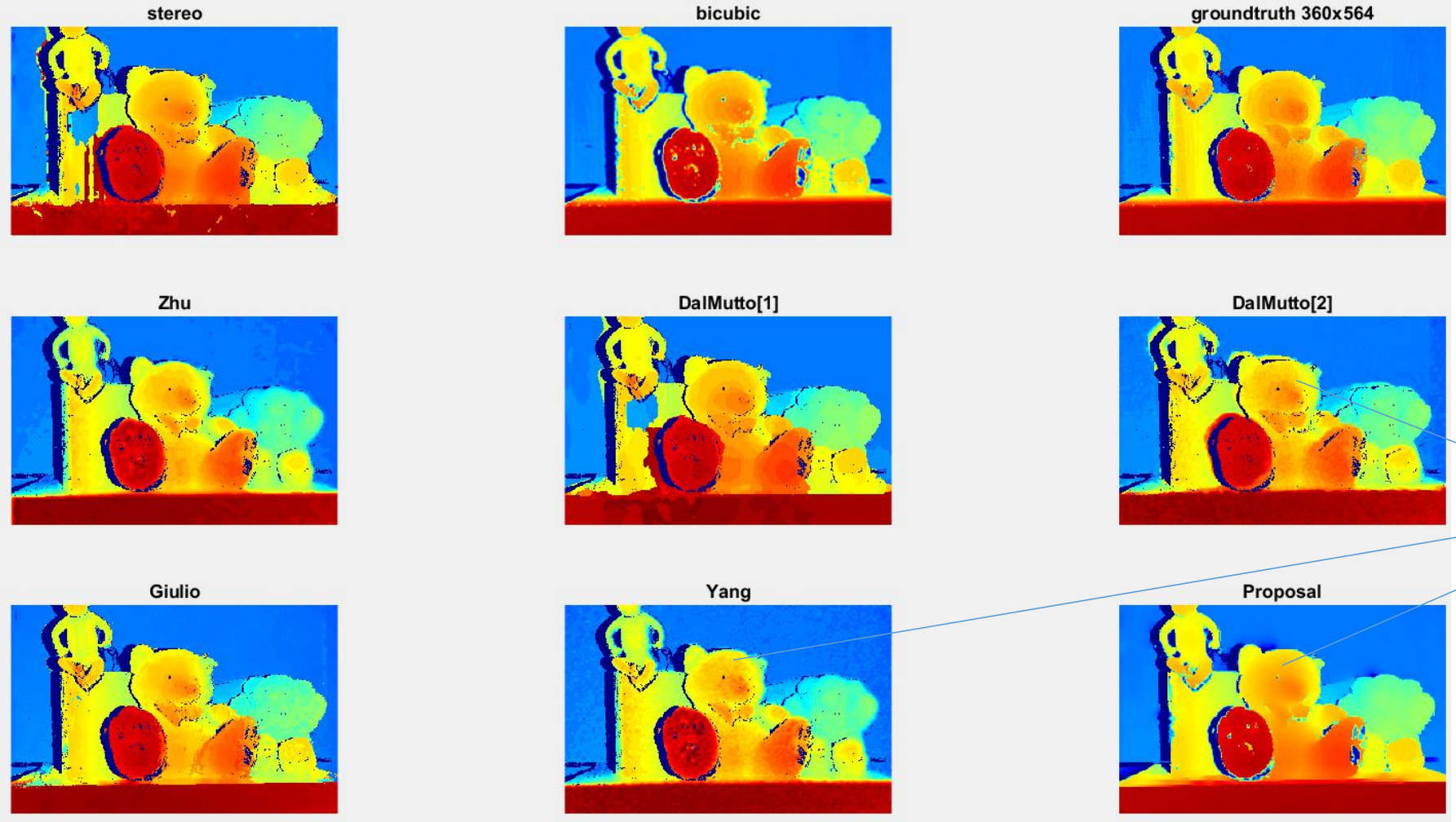
Method	**1**	**2**	**3**	**4**	**5**
TOF	0.84	0.84	0.77	0.79	0.76
Stereo	0.54	0.49	0.80	0.70	0.80
Giulio et al.	0.91	0.91	0.87	0.89	0.89
DalMutto1	0.92	0.91	0.89	0.91	0.91
DalMutto2	0.90	0.89	0.88	0.87	0.89
Yang et al.	0.90	0.90	0.89	0.88	0.88
Proposed	**0.94**	**0.94**	**0.95**	**0.95**	**0.95**

Experimental Results: *scene 2*



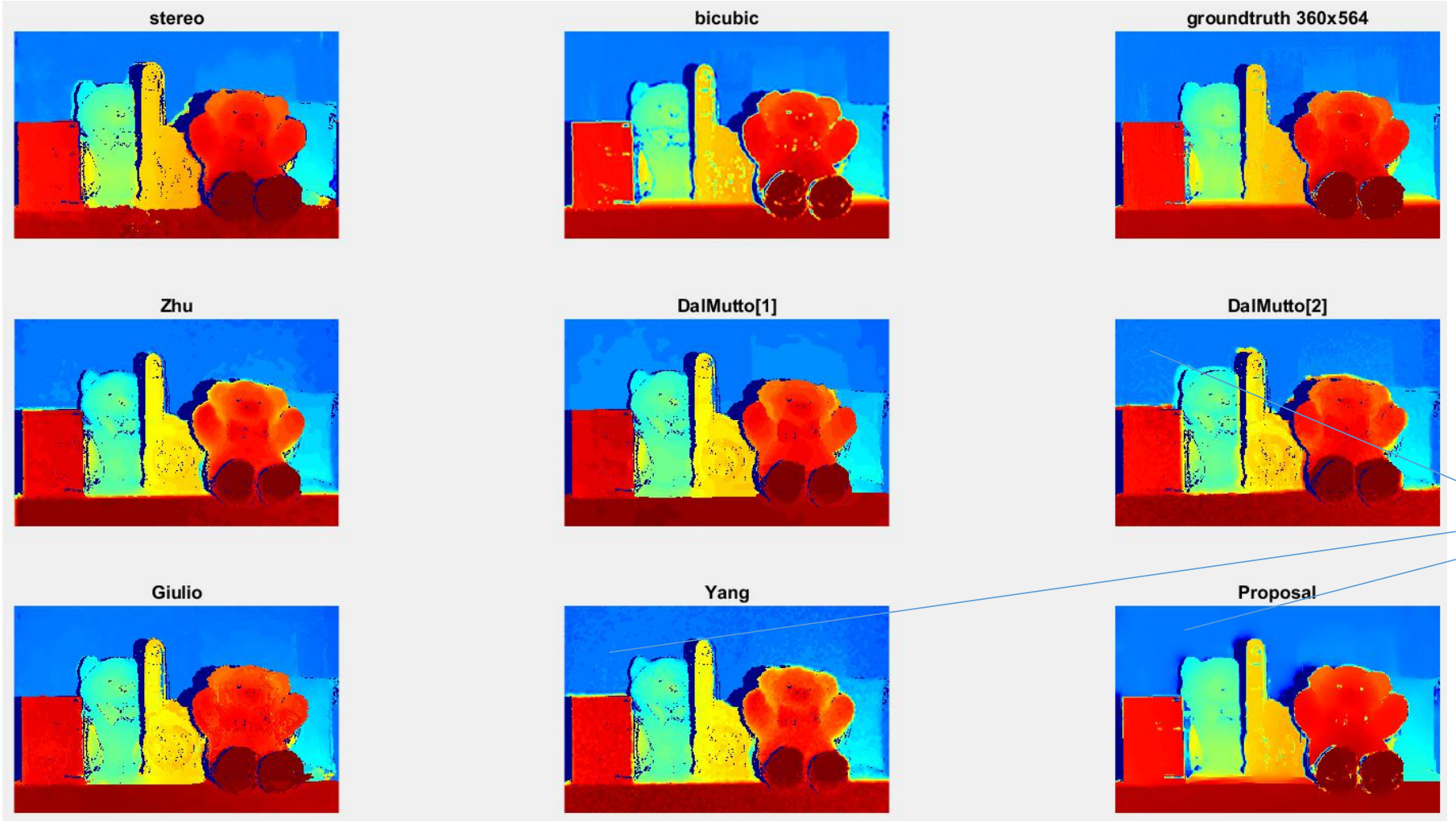
Boundaries in our results are generally sharper along the edge direction

Experimental Results: *scene 3*



In Depth smooth regions,our result is more close to reality,as we preserve more right disparity values from stereo data with our new confidence measurement

Experimental Results: *scene 5*



Less noise generated with using TV as regular term.

Conclusions

- We have proposed **variation fusion of TOF and stereo data** using edge selective joint filtering.

- **Two main contributions** to data fusion:
  - Edge selective joint filter (ESJF) for depth up-sampling: Selectly preseve true edges from TOF,Color and stereo data.
  - Variational data fusion : Eliminate most noise in the smooth region with total variation as regularation.

- Experimental results show the proposed method effectively generates HR depth maps and effectively preserving edges and removing noise with MSE is about 15% reduced Compared with state of the art results and best SSIM value acquired.

THANK YOU!