Closed-form Solution of Simultaneous Denoising and Hole Filling of Depth Image

Yong Sun Kim Ph.D.
CG/Vision Research Group
Electronics and Telecommunications Research Institute (ETRI)

Motivation

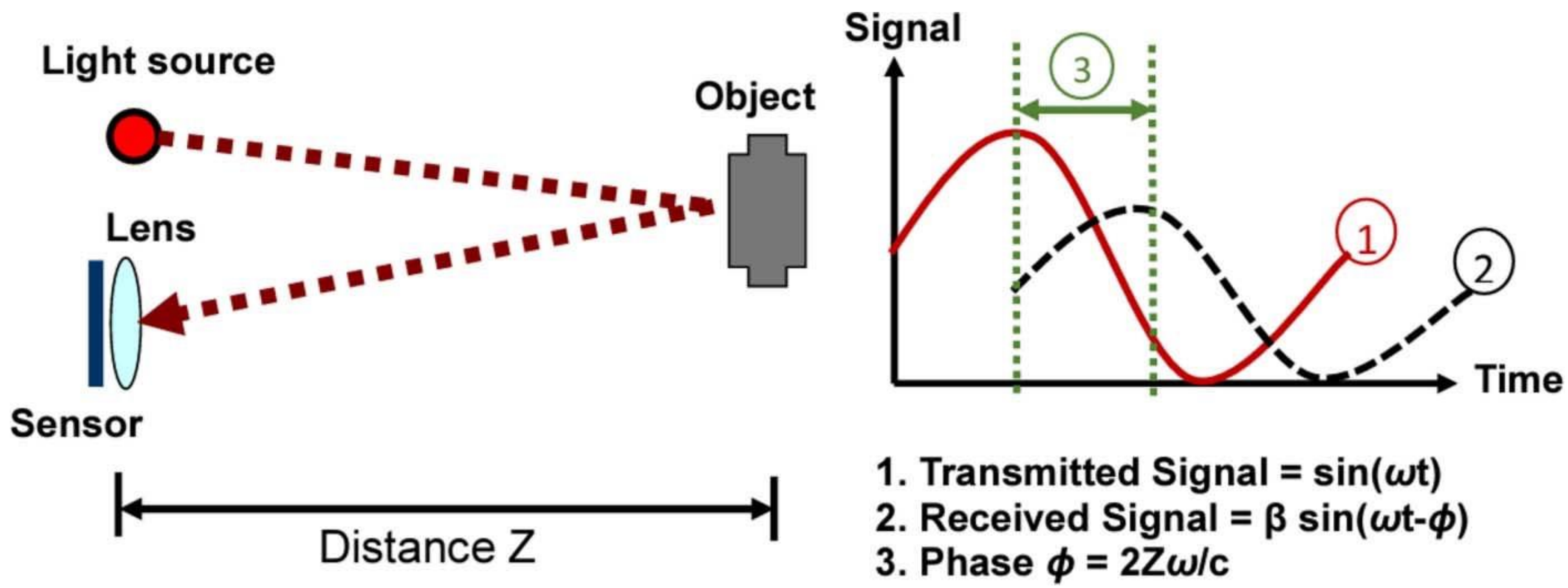
- Depth information
  - Applications: Computer Vision, Virtual Reality/Augmented Reality
  - Acquisition methods
    - Passive stereo, Structured light, Time-of-Flight
  - Problems of low resolution, noise, holes
  - ToF Depth: accurate and real-time depth sensing
- Objective
  - To restore ToF depth image by removing noise and holes

Previous Work

- Assuming high structural correlation between high-resolution color image and low-resolution depth image
- Filter-based approach
  - Joint bilateral upsampling
  - Noise aware filter for depth upsampling
  - Clearly improves depth resolution but may cause texture copy artifacts
- Optimization-based approach
  - Energy function composed of a data term and a regularization term
  - Regularization term: non-local means filtering, autoregressive model, region adaptive joint bilateral kernel
  - Needs an initial depth image, may cause artifact when edges of color and depth do not coincide

ToF Depth

- Can be acquired by measuring travel time of light as phase delay between the emitted and the received infrared light



- ToF Depth camera provides infrared image together with depth image

[19] C. S. Bamji, P. O'Connor, T. Elkhatib, S. Mehta, B. Thompson, L. A. Prather, D. Snow, O. C. Akkaya, A. Daniel, A. D. Payne, T. Perry, M. Fenton, and V.-H. Chan, "A 0.13um CMOS system-on-chip for a 512 x 424 time-of-flight image sensor with multi-frequency photo- demodulation up to 130 MHz and 2 GS/s ADC," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 303-319, Jan. 2015.

Proposed Method

- Recover ToF depth image by minimizing the proposed energy function defined as a combination of a filtering term and a reconstruction term

- Utilizing infrared data instead of color image
  - ToF depth and infrared data from the same sensor
  - No calibration error, No parallax effect
  - Infrared values highly related to materials of objects and distances

Filtering Term

- To remove additive white Gaussian noise, weighted averaging filtering such as bilateral filter or non local means shows good performance
  - Derived from minimizing weighted sum of squared error as

$$\arg\min_{\bar{D}(\mathbf{x}_i)} \sum_{\mathbf{x}_j \in \mathrm{N}_i} w(\mathbf{x}_i, \mathbf{x}_j) \left\| \bar{D}(\mathbf{x}_i) - D(\mathbf{x}_j) \right\|^2 \longrightarrow \bar{D}(\mathbf{x}_i) = \frac{\sum_{\mathbf{x}_j \in \mathrm{N}_i} w(\mathbf{x}_i, \mathbf{x}_j) D(\mathbf{x}_j)}{\sum_{\mathbf{x}_j \in \mathrm{N}_i} w(\mathbf{x}_i, \mathbf{x}_j)}$$

- The proposed multilateral filter weight for valid depth pixels with spatial weight, depth weight and infrared weight

$$w(\mathbf{x}_i, \mathbf{x}_j) = K_S(\|\mathbf{x}_i - \mathbf{x}_j\|) \cdot \big(K_D(\|D(\mathbf{x}_i) - D(\mathbf{x}_j)\|) + \varepsilon\big) \cdot \big(K_R(\|R(\mathbf{x}_i) - R(\mathbf{x}_j)\|) + \varepsilon\big)$$

  - If center pixel is invalid depth, D = 0,

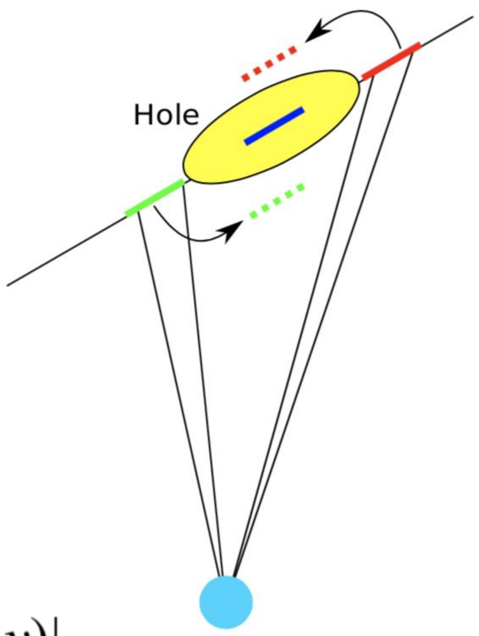
$$w(\mathbf{x}_i, \mathbf{x}_j) = K_S(\|\mathbf{x}_i - \mathbf{x}_j\|) \cdot \big(K_R(\|R(\mathbf{x}_i) - R(\mathbf{x}_j)\|) + \varepsilon\big)$$

  - Introducing small value ε to boost filtering impulsive noise which can hardly be removed using the conventional bilateral kernel

- Define the filtering term by aggregating all pixels and representing in a quadratic matrix form

$$\bar{\mathbf{D}} = \mathbf{WD} \longrightarrow E_F = \frac{1}{2}(\bar{\mathbf{D}} - \mathbf{WD})^T (\bar{\mathbf{D}} - \mathbf{WD})$$

Reconstruction Term



- Patch-based inpainting schemes such as exemplar-based methods may not find proper patches in a depth image
- Adopts a structure guided depth reconstruction approach $\nabla^2 \overline{D}(x,y) = \frac{\partial G_x}{\partial x}(x,y) + \frac{\partial G_y}{\partial y}(x,y)$ over Ω, $s.t. \overline{D}(x,y)|_{\partial\Omega} = D(x,y)|_{\partial\Omega}$
  - Given depth gradients as a guidance vector, depth values can be obtained using a discretized version of the Poisson solution of guided interpolation

$$\overline{D}(x+1,y) + \overline{D}(x-1,y) + \overline{D}(x,y+1) + \overline{D}(x,y-1) - 4\overline{D}(x,y) = \frac{\partial D_x}{\partial x}(x,y) + \frac{\partial D_y}{\partial y}(x,y). \longrightarrow \mathbf{L}\overline{\mathbf{D}} = \mathbf{G}$$

- Define the reconstruction term in a quadratic matrix form as

$$E_H = \frac{1}{2}(\mathbf{L}\overline{\mathbf{D}} - \mathbf{G})^T(\mathbf{L}\overline{\mathbf{D}} - \mathbf{G})$$

[17] D. Doria and R. J. Radke, "Filling large holes in lidar data by inpainting depth gradients," in *Proc. CVPRW*, 2012.

ETRI

Depth Gradient Inpainting

- Moving Least Squares (MLS) interpolation approach to obtain depth gradients of pixels inside hole regions, utilizing pixel position and infrared data

- MLS can be solved by minimizing the weighted least squares for each pixel as $$\underset{\mathbf{c}(\mathbf{p})}{\arg\min} \sum_{i \in \Pi} \theta(\mathbf{p};\mathbf{p}_i)\left(f(\mathbf{p}_i) - f_i\right)^2$$
- 2nd order polynomial function using infrared data as additional structural information to prevent unwanted smoothing of inpainted depth gradients

$$f(x,y,r) = c_0 + c_1x + c_2y + c_3r + c_4x^2 + c_5xy + c_6y^2 + c_7xr + c_8yr + c_9r^2$$

- The bilateral weighting function

$$\theta(\mathbf{p};\mathbf{p}_\mathrm{i}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{2\sigma_1^2}\right)\cdot\exp\left(-\frac{\|r-r_i\|^2}{2\sigma_2^2}\right)$$

Optimization

- The proposed quadratic energy function

$$E = \frac{1}{2}(\bar{\mathbf{D}} - \mathbf{WD})^T(\bar{\mathbf{D}} - \mathbf{WD}) + \lambda\frac{1}{2}(\mathbf{L}\bar{\mathbf{D}} - \mathbf{G})^T(\mathbf{L}\bar{\mathbf{D}} - \mathbf{G})$$

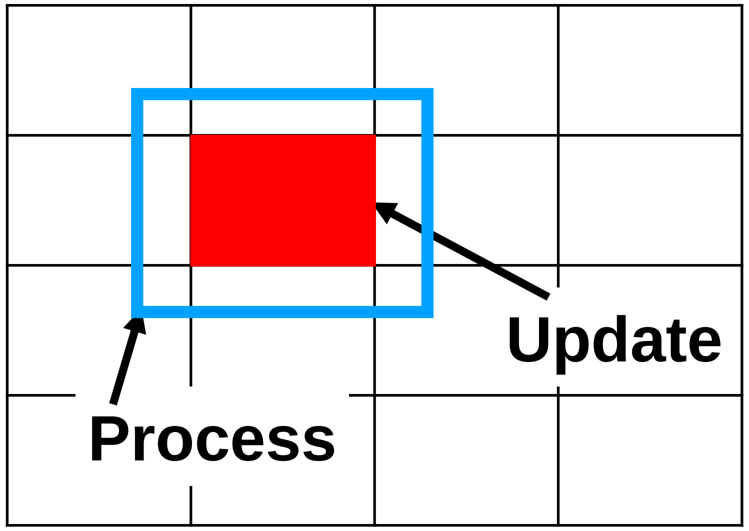
minimized when

$$\frac{\partial E}{\partial \bar{\mathbf{D}}} = (\bar{\mathbf{D}} - \mathbf{WD}) + \lambda\mathbf{L}^T(\mathbf{L}\bar{\mathbf{D}} - \mathbf{G}) = 0$$

- Solving the sparse linear system

$$(\mathbf{I} + \lambda\mathbf{L}^T\mathbf{L})\bar{\mathbf{D}} = \mathbf{WD} + \lambda\mathbf{L}^T\mathbf{G}$$

- Block-wise processing
  - Overlapped block to prevent blocking artifacts



Experiments

- Compared with Joint Bilateral Filtering (JBF) and Noise Aware Filtering (NAF)
- The methods are applied iteratively.
- Parameters of the filtering term in the proposed method are commonly used in the existing methods.
- Synthetic data (8bit) for quantitative evaluation
  - Middlebury data (Ground truth): Disparity image as depth data and Color image as infrared data
  - Degraded according to the characteristics of Kinect v2 as a ToF camera
    - Make holes (around object boundary and on reflective/absorptive objects having high/low infrared values) in depth image
    - Add Gaussian noise in depth and infrared images
  - Parameters: window size= 5, sigma_spatial = 1.5, sigma_infrared = 10, sigma_depth = 10

[1] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graphics*, vol. 26, no. 3, p. 96, July 2007.

[2] D. Chan, H. Buisman, C. Theobalt, and S. Thrun, "A noise-aware filter for real-time depth upsampling," in *Workshop Multi-camera Multi- modal Sensor Fusion Algorithms and Applications*, 2008.

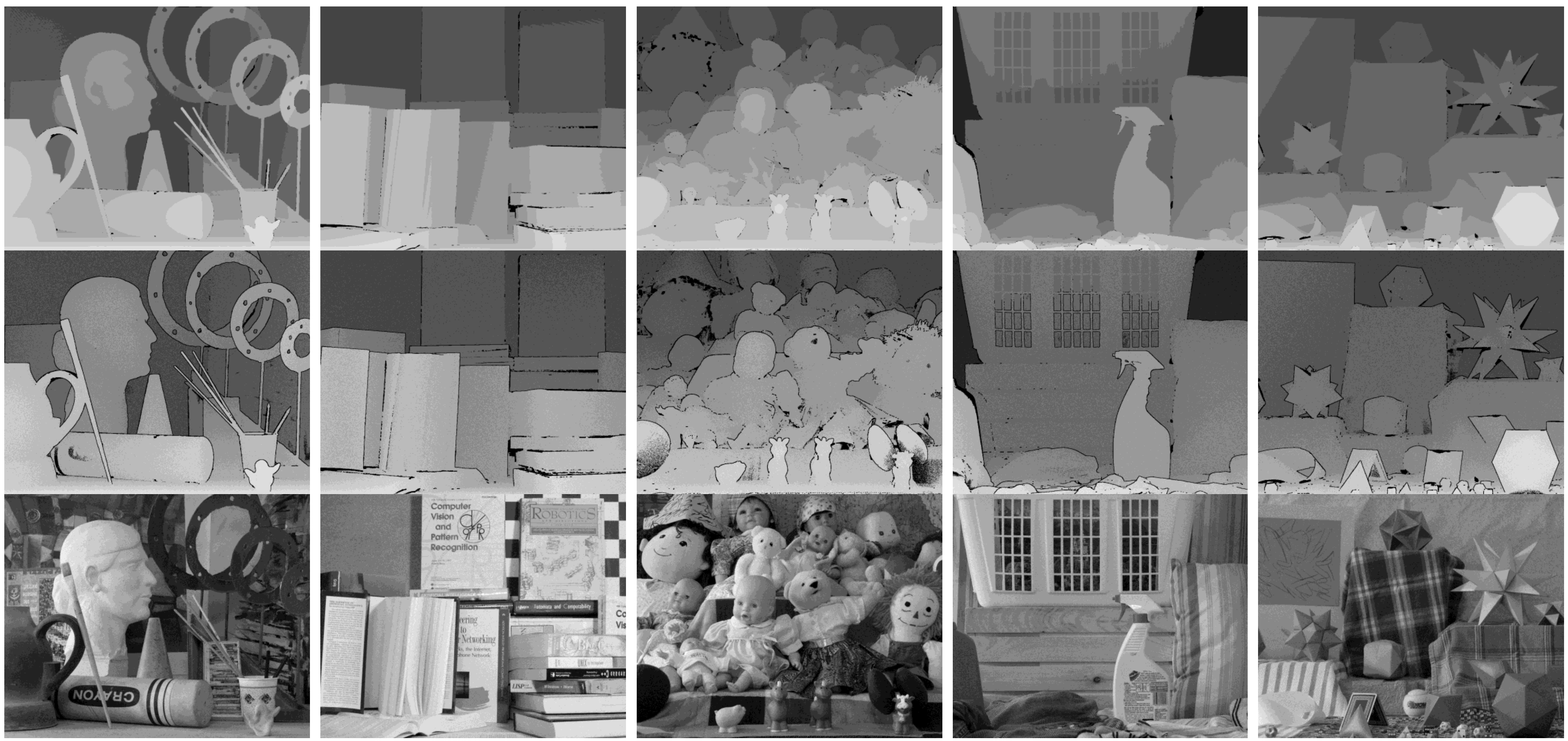
Synthetic Image

- Dataset (Image size: 1390 x 1110)

Ground Truth

Depth Image

Infrared (Gray) Image

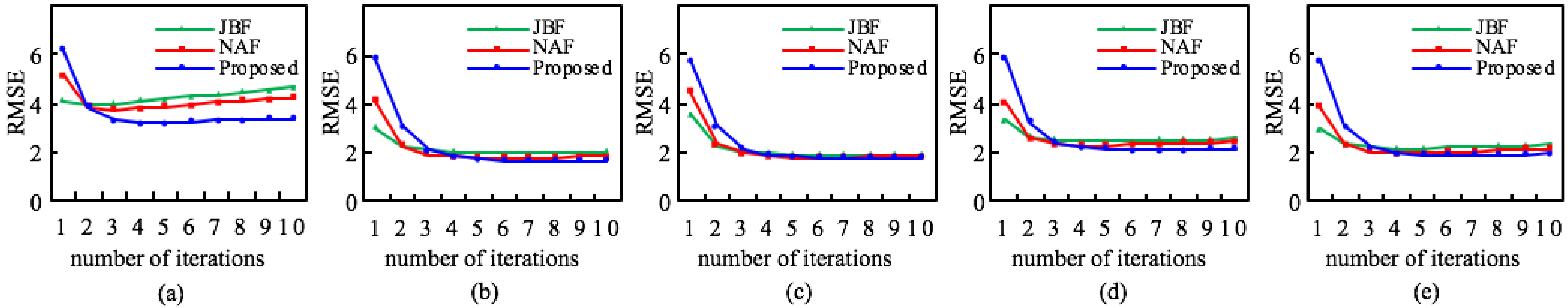


(a) Art (b) Books (c) Dolls (d) Laundry (e) Moebius

Synthetic Image

- Quantitative evaluation: Root Mean Squared Error (RMSE) between the ground truth and the recovered depth image

- RMSE curves along the number of iterations

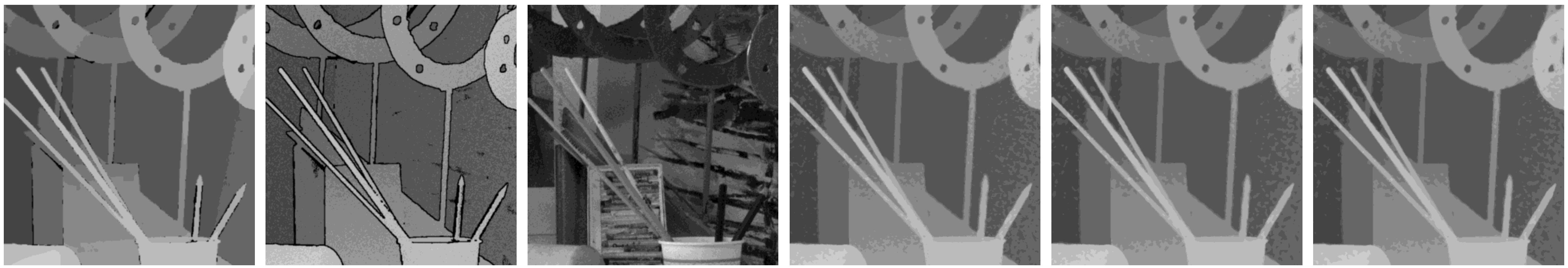


- The best RMSE values among 10 iterations

Data	Input	JBF	NAF	Proposed
Art	27.70	3.91	3.74	3.18
Books	21.02	1.96	1.79	1.66
Dolls	35.21	1.87	1.78	1.75
Laundry	23.28	2.45	2.29	2.10
Moebius	23.09	2.16	1.98	1.88

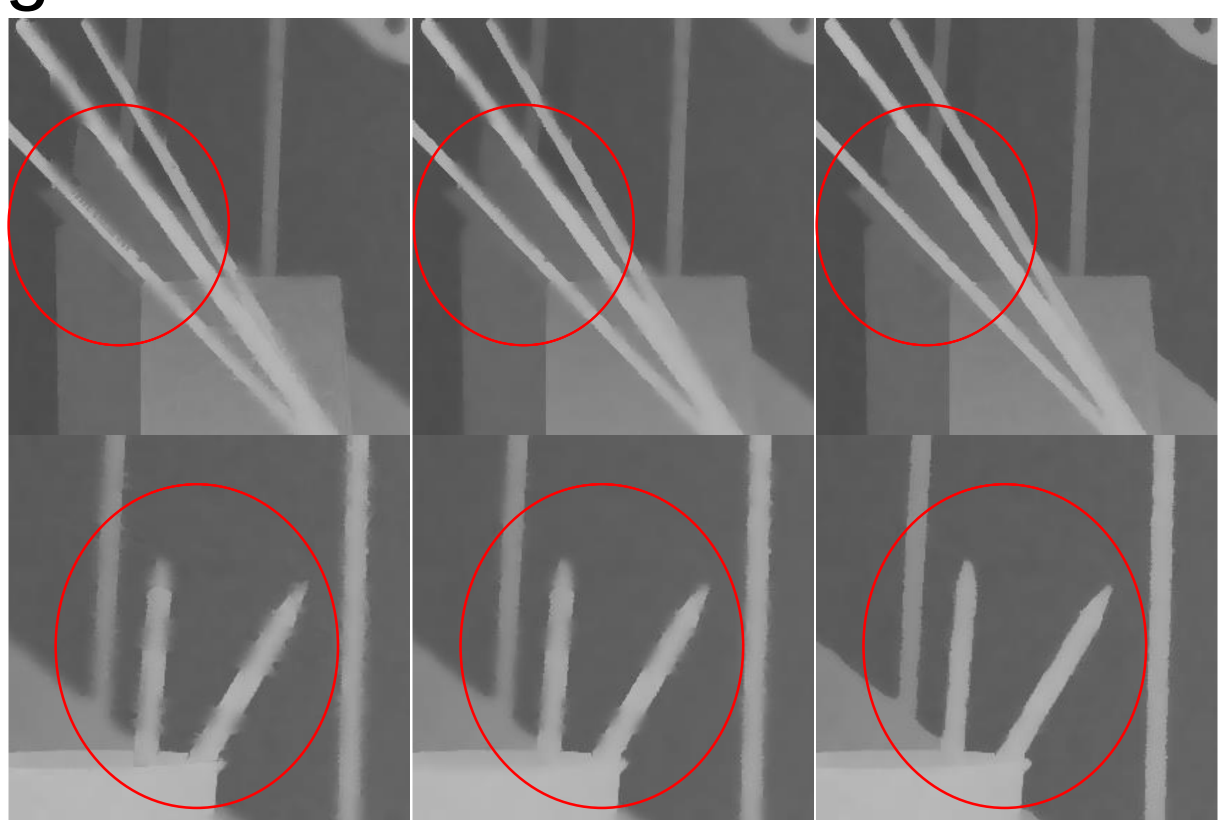
Synthetic Image

- Best results among 10 iterations



GT | Input Depth | Input Infrared | JBF | NAF | Proposed

- Result images after 10 iterations



JBF | NAF | Proposed

Real Image

- Real Kinect v2 data (16bit, 512 x 424)
  - Ground truth with 10,000 frames averaging
  - Parameters: window size = 5, sigma_spatial = 0.5, sigma_infrared = 100, sigma_depth = 10
  - The number of iterations: 5

GT | **Input Depth (RMSE: 302.4mm)** | **Input Infrared**

JBF (RMSE: 49.5mm) | **NAF (RMSE 43.9mm)** | **Proposed (RMSE 37.6mm)**

Conclusion

- Closed form solution for simultaneous denoising and hole filling of ToF depth data by minimizing the quadratic energy function
  - Filtering term utilizing boosted multilateral kernels with selective use of spatial, depth, and infrared information
  - Reconstruction term based on depth gradients obtained by infrared-guided MLS interpolation
  - Provides recovered depth without introducing texture copy and blur artifacts
- Can be applicable to conventional RGBD data as in synthetic dataset
- Large holes and computational complexity

Thank You!