Depth Image Enhancement Using Local Tangent Plane Approximations

Supplementary Materials

Kiyoshi Matsuo†　　　　　Yoshimitsu Aoki‡
†Hokuyo Automatic Co., LTD.　　‡Keio University
k-matsuo@hokuyo-aut.co.jp　　aoki@elec.keio.ac.jp

1. Conceptual diagrams

In this section, we show some conceptual diagrams of our depth enhancement method.
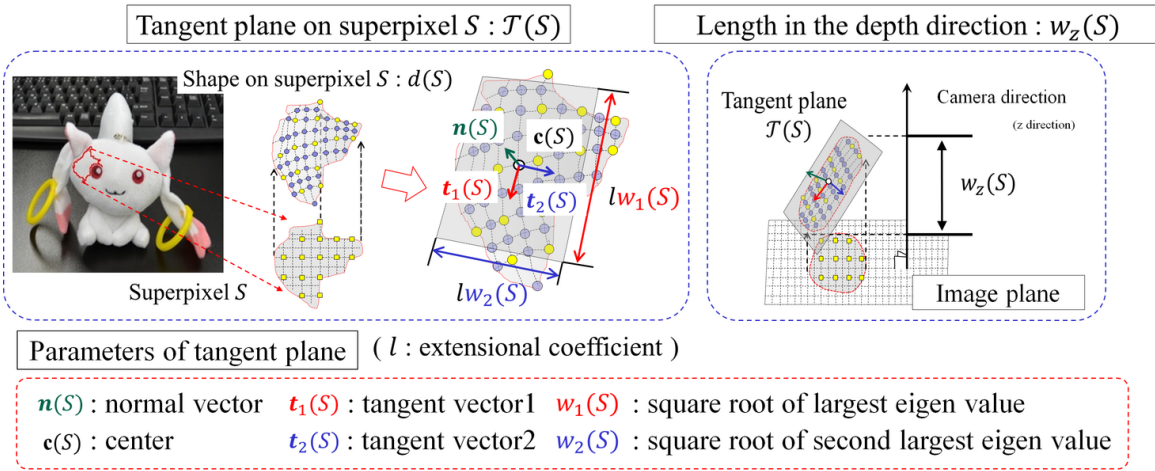

Figure 1. Conceptual diagrams of tangent plane on superpixel and length in the depth direction.

In Figure 1, conceptual diagrams of the tangent plane on a superpixel and of its length in the depth direction are shown. The length in the depth direction is used to extract steep tangent planes.
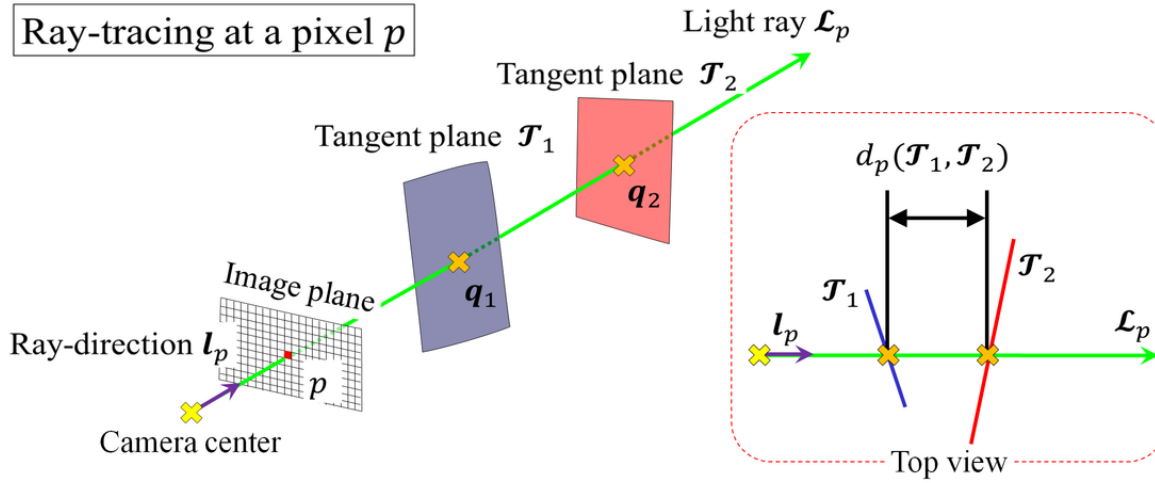

Figure 2. Conceptual diagram of distance at a pixel $d_p(\mathcal{T}_1, \mathcal{T}_2)$.

In Figure 2, a conceptual diagram of the distance $d_p(\mathcal{T}_1, \mathcal{T}_2)$ between two tangent planes $\mathcal{T}_1$ and $\mathcal{T}_2$ at a pixel p is shown. We use this distance to evaluate spatial relationships between the tangent planes on superpixels.

2. Orientation correction results

In this section, we show the results of our orientation correction. In the following results, each orientation of tangents on superpixels is shown as normal maps. For purpose of comparison, we also show uncorrupted normal maps and corrupted normal maps caused by noise from high-resolution depth images. Here, we applied PCA to the high-resolution shapes on the neighboring 14×14 pixels of each pixel. In following results, our normal correction method recovered similar normal maps as uncorrupted normal maps.
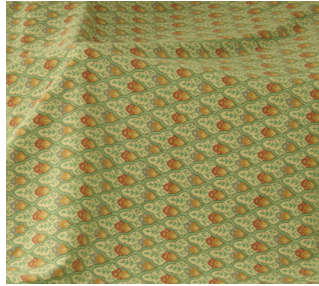

Figure 3. Orientation correction results represented as normal maps. Color images of the scenes, normal maps of smooth high-resolution depth images, normal maps of noisy high-resolution depth images, normal maps calculated by PCA and our color heuristic PCA on each superpixel, and the corrected normal maps by our proposed method. (noise coefficient $k = 3.0 \times 10^{-6}$, magnification factor $\times 4 \times 4$)

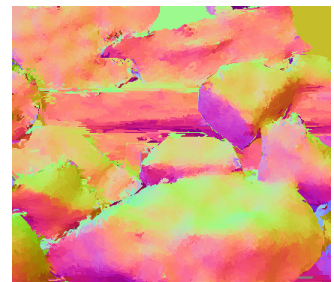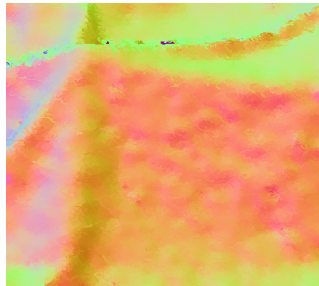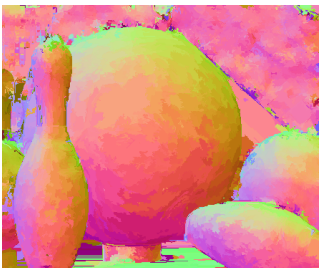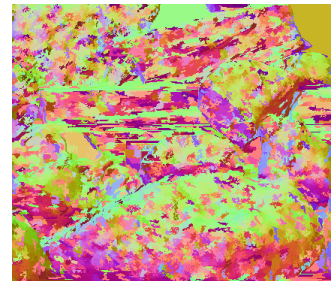| (e) *Bowling2*	(f) *Cloth1*	(g) *Midd2*	(h) *Rocks1*
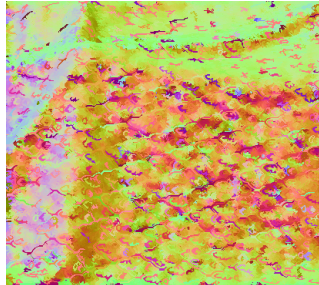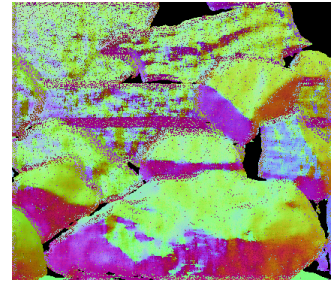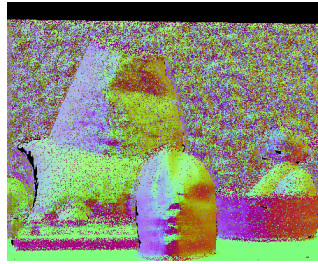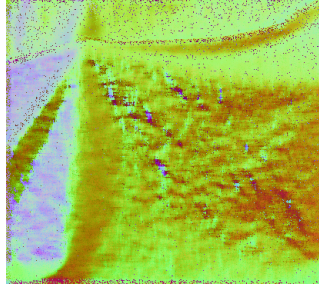
Figure 4. Orientation correction results represented as normal maps. Color images of the scenes, normal maps of smooth high-resolution depth images, normal maps of noisy high-resolution depth images, normal maps ca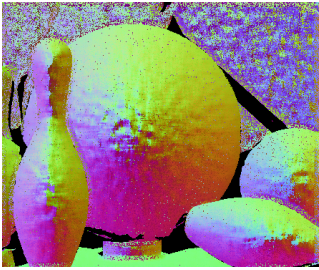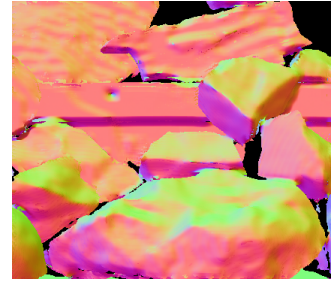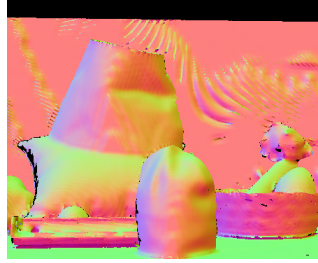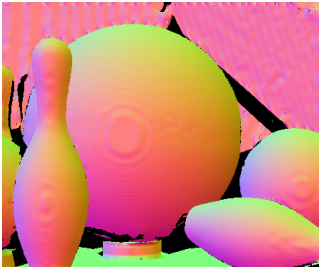lculated by PCA and our color heuristic PCA on each superpixel, and the corrected normal maps by our proposed method. (noise coefficient $k = 3.0 \times 10^{-6}$, magnification factor $\times 4 \times 4$)

3. Surface refinement using local geometries

In this section, we show a mathematical property of smoothing the normal direction components of 3D points on each local tangent plane. The explicit definition of smoothed point $\tilde{\mathbf{x}}$ of point $\mathbf{x}$ in local coordinate R is

$$\tilde{\mathbf{x}} = \frac{1}{W} \sum_{\mathbf{x}' \in d(R)} w_{\text{Gauss}}(|\mathbf{x} - \mathbf{x}'|_1)(\mathbf{x}', \mathbf{n}(R))\mathbf{n}(R) + (\mathbf{x}, \mathbf{t}_1(R))\mathbf{t}_1(R) + (\mathbf{x}, \mathbf{t}_2(R))\mathbf{t}_2(R), \quad (1)$$

where w_{Gauss} is the Gaussian weight function, $d(R)$ is the shape in region R, $\mathbf{n}(R)$ is the normal vector, $\mathbf{t}_1(R)$ and $\mathbf{t}_2(R)$ are the orthonormal tangent vectors, and W is the total weighted sum.

When all points of the shape $d(R)$ are on the tangent plane $\mathcal{T}(R)$ (i.e. the points constitute a plane), each point $\mathbf{x}'$ is represented as $\mathbf{x}' = \mathbf{c}(R) + a_1(\mathbf{x}')\mathbf{t}_1(R) + a_2(\mathbf{x}')\mathbf{t}_2(R)$. Here, $\mathbf{c}(R)$ is the center point of the tangent plane $\mathcal{T}(R)$. Therefore,

$$
\begin{aligned}
\tilde{\mathbf{x}} &= \frac{1}{W} \sum_{\mathbf{x}' \in d(R)} w_{\text{Gauss}}(|\mathbf{x} - \mathbf{x}'|_1)(\mathbf{x}', \mathbf{n}(R))\mathbf{n}(R) + (\mathbf{x}, \mathbf{t}_1(R))\mathbf{t}_1(R) + (\mathbf{x}, \mathbf{t}_2(R))\mathbf{t}_2(R) \\
&= \frac{1}{W} \sum_{\mathbf{x}' \in d(R)} w_{\text{Gauss}}(|\mathbf{x} - \mathbf{x}'|_1)(\mathbf{c}(R), \mathbf{n}(R))\mathbf{n}(R) + (\mathbf{x}, \mathbf{t}_1(R))\mathbf{t}_1(R) + (\mathbf{x}, \mathbf{t}_2(R))\mathbf{t}_2(R) \\
&= (\mathbf{c}(R), \mathbf{n}(R))\mathbf{n}(R) + (\mathbf{c}(R) + a_1(\mathbf{x})\mathbf{t}_1(R), \mathbf{t}_1(R))\mathbf{t}_1(R) + (\mathbf{c}(R) + a_2(\mathbf{x})\mathbf{t}_2(R), \mathbf{t}_2(R))\mathbf{t}_2(R) \\
&= (\mathbf{c}(R), \mathbf{n}(R))\mathbf{n}(R) + (\mathbf{c}(R), \mathbf{t}_1(R))\mathbf{t}_1(R) + (\mathbf{c}(R), \mathbf{t}_2(R))\mathbf{t}_2(R) + a_1(\mathbf{x})\mathbf{t}_1(R) + a_2(\mathbf{x})\mathbf{t}_2(R) \\
&= \mathbf{c}(R) + a_1(\mathbf{x})\mathbf{t}_1(R) + a_2(\mathbf{x})\mathbf{t}_2(R) \\
&= \mathbf{x}.
\end{aligned}
$$

Hence if the shape $d(R)$ equals the tangent plane $\mathcal{T}(R)$, this smoothing does not change the shape $d(R)$.

4. Additional information for evaluations

In this section, we show additional information about our evaluations.

4.1. Noise coefficient of real sensors

For reference, noise coefficient values of some real sensors are shown.

Hokuyo URG-04LX [1]	Hokuyo UXM-30LX-EW [2]	SICK LMS200 [3]
1.5×10^{-6} [6]	5.0×10^{-6} [2]	2.3×10^{-6} [6]

Table 1. Noise coefficients k of real sensors.

4.2. Additional results for quantitative comparisons

We show the additional experimental results that used some images of the Middlebury datasets [8]. Figure 5 and Figure 6 show the MAEs using different magnification factors and fixed noise coefficients 3.0×10^{-6} and 5.0×10^{-6}, and MAEs and completion rates using different noise coefficients and a fixed magnification factor $\times 4 \times 4$. In these graphs, the results of the proposed method, bilinear interpolation, MRF [4], PWAS [5], and method in [7] are shown using solid red, dashed watery, dashed purple, dashed blue, and dashed green lines, respectively. In noisy cases, when the noise coefficient k was more than two, our method outperformed the other methods in terms of accuracy. And in all cases, our method had the higher completion rates than the tangent based method in [7].

In Figure 7, Figure 8, Figure 9, Figure 10, and Figure 11, the absolute interpolation errors at each pixel are shown. In all Figures, the magnification factor is $\times 4 \times 4$ and the noise coefficients are $k = 3.0 \times 10^{-6}$. Our interpolation errors inside each surface were lower than other methods. Our method had regions where depth interpolation was not performed, similar to [7]. These regions are colored black. However, our method had the higher completion rates of over 78%, than the tangent based method in [7]. The tangent based method in [7] does not interpolate regions that have steep tangent planes caused by noise, hence its errors were small but the completion rates were lower.
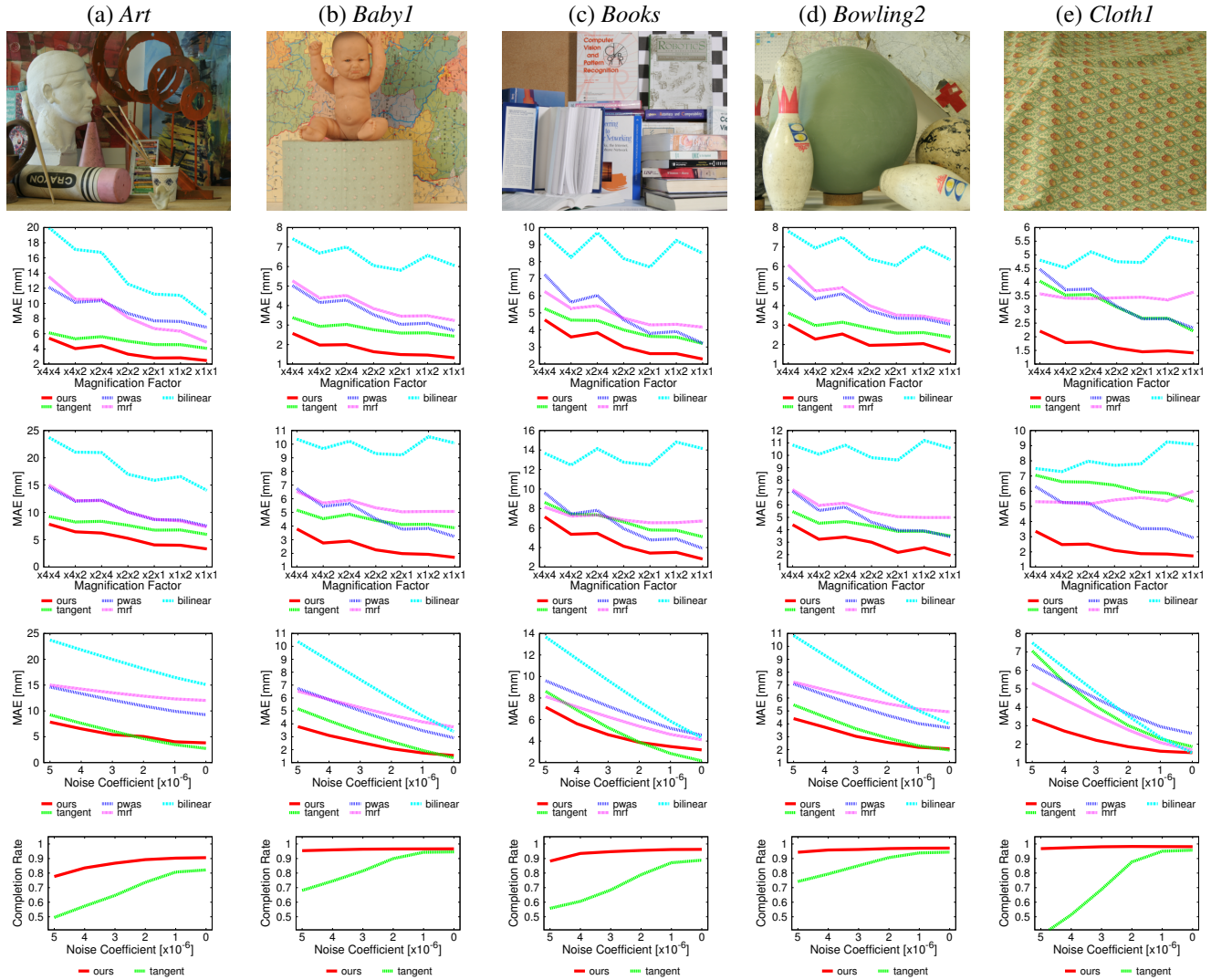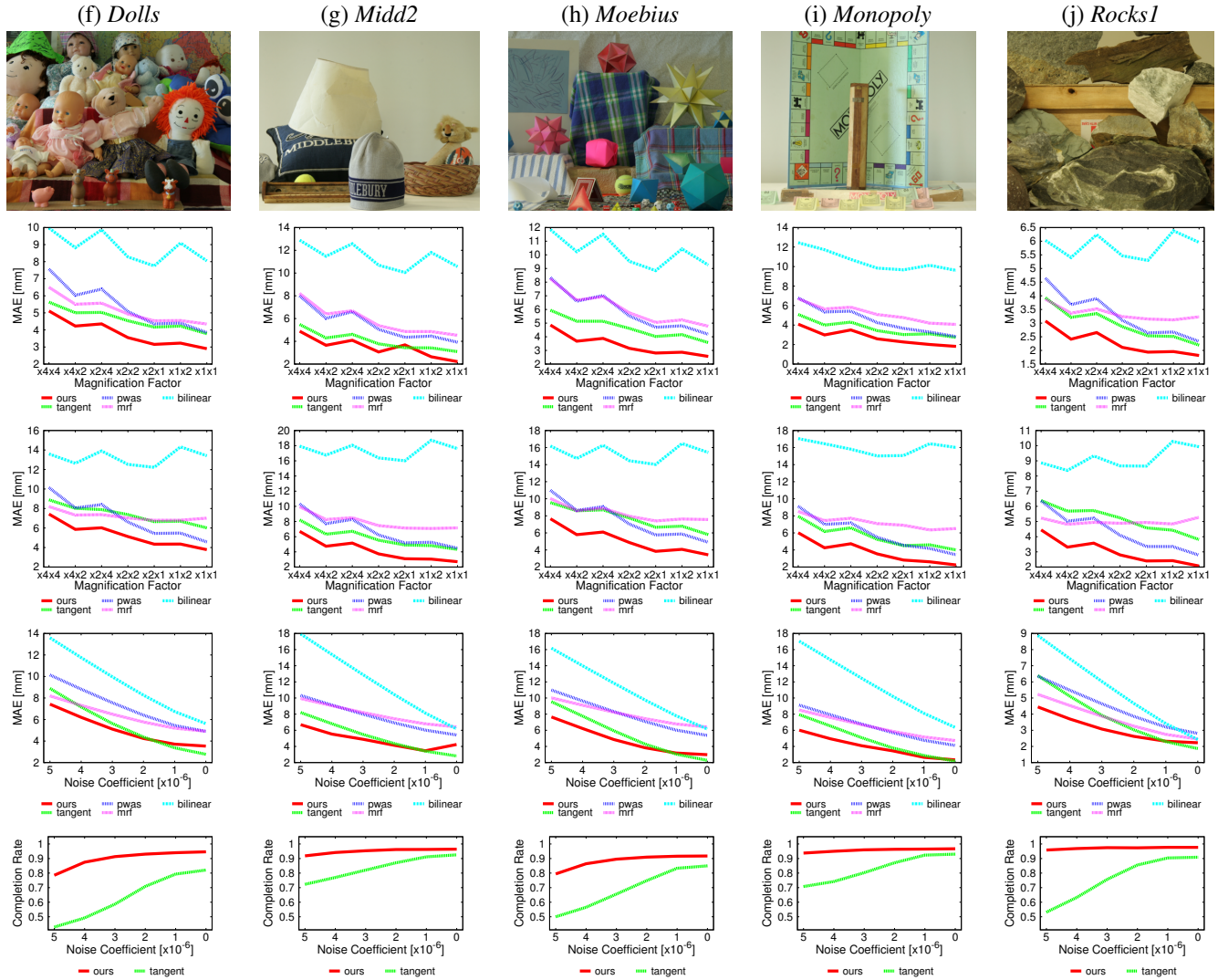
Figure 5. Quantitative comparisons: Color images of scenes, MAEs under low level noise ($k = 3.0 \times 10^{-6}$), MAEs under high level noise (5.0×10^{-6}), and MAEs and completion rate for various noise levels (upsampling rate $= \times 4 \times 4$).
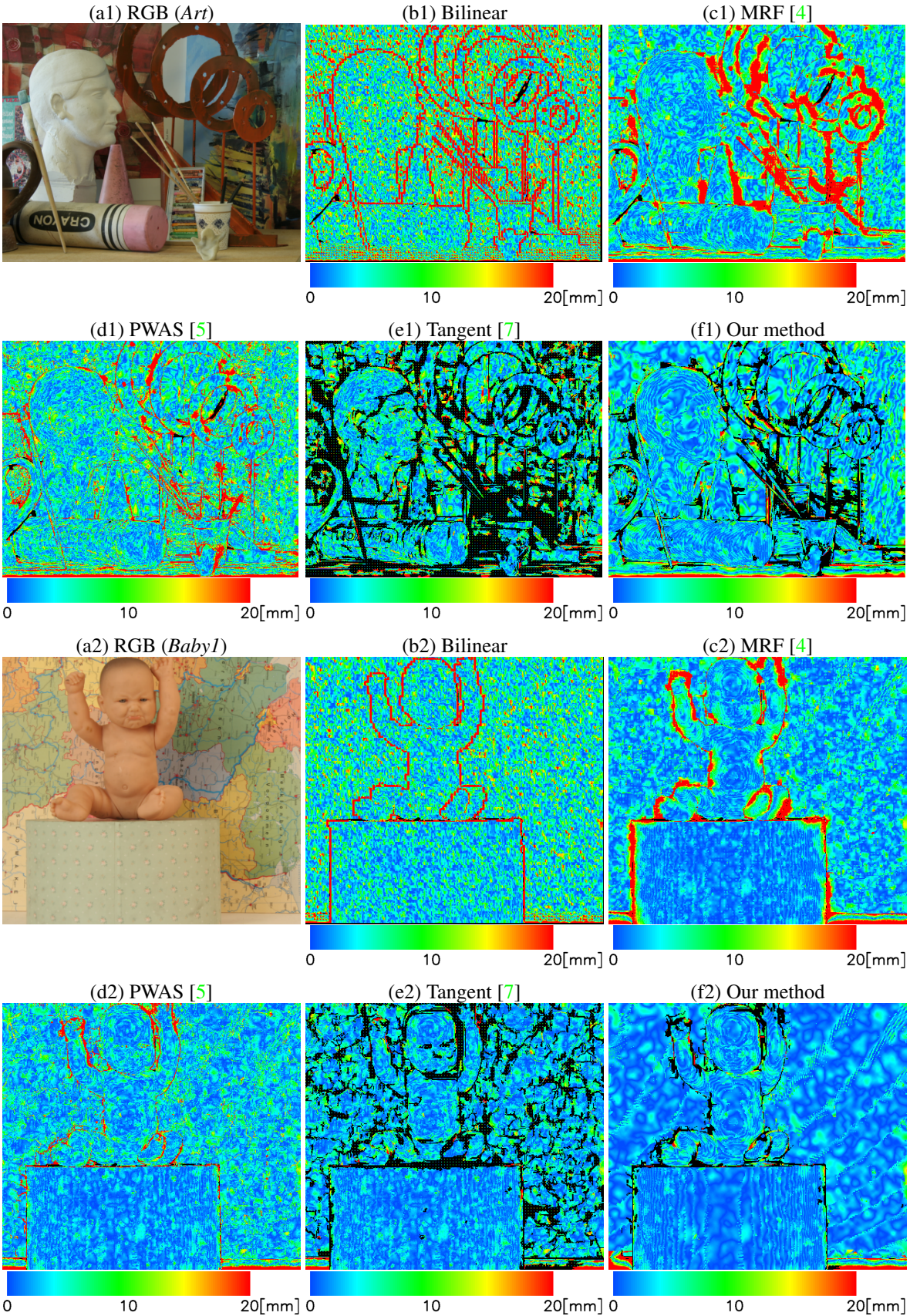
Figure 6. Quantitative comparisons: Color images of scenes, MAEs under low level noise ($k = 3.0 \times 10^{-6}$), MAEs under high level noise (5.0×10^{-6}), and MAEs and completion rate for various noise levels (upsampling rate $= \times 4 \times 4$).
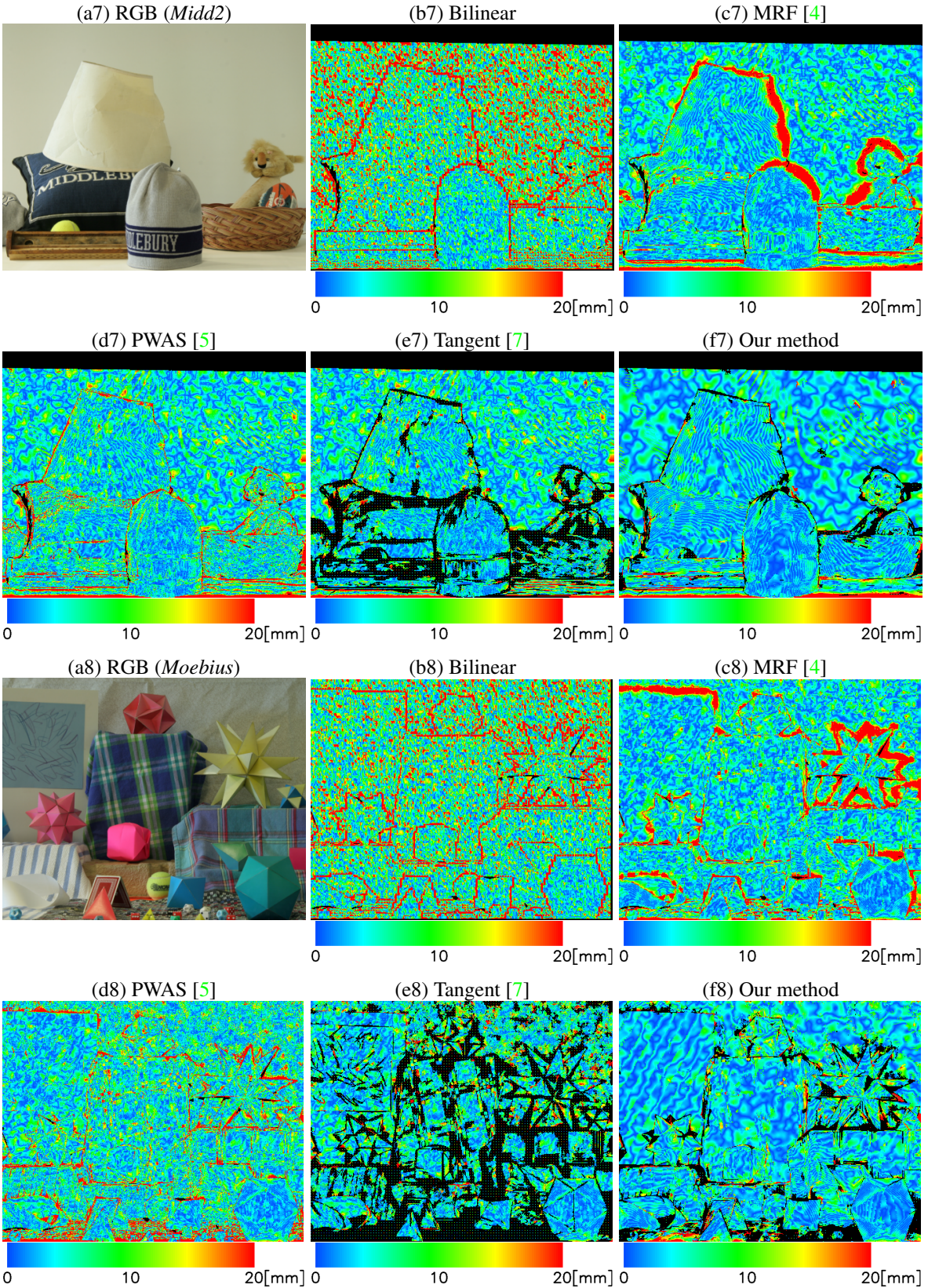
Figure 7. Resulting depth error (Scenes *Art* and *Baby1*). (a) Color image of the Scene; (b) Bilinear interpolation; (c) MRF [4]; (d) PWAS [5]; (e) the method in [7]; (f) our method. Pixels that have no depth value are colored black. (noise coefficient $k = 3.0 \times 10^{-6}$, magnification factor $\times 4 \times 4$).
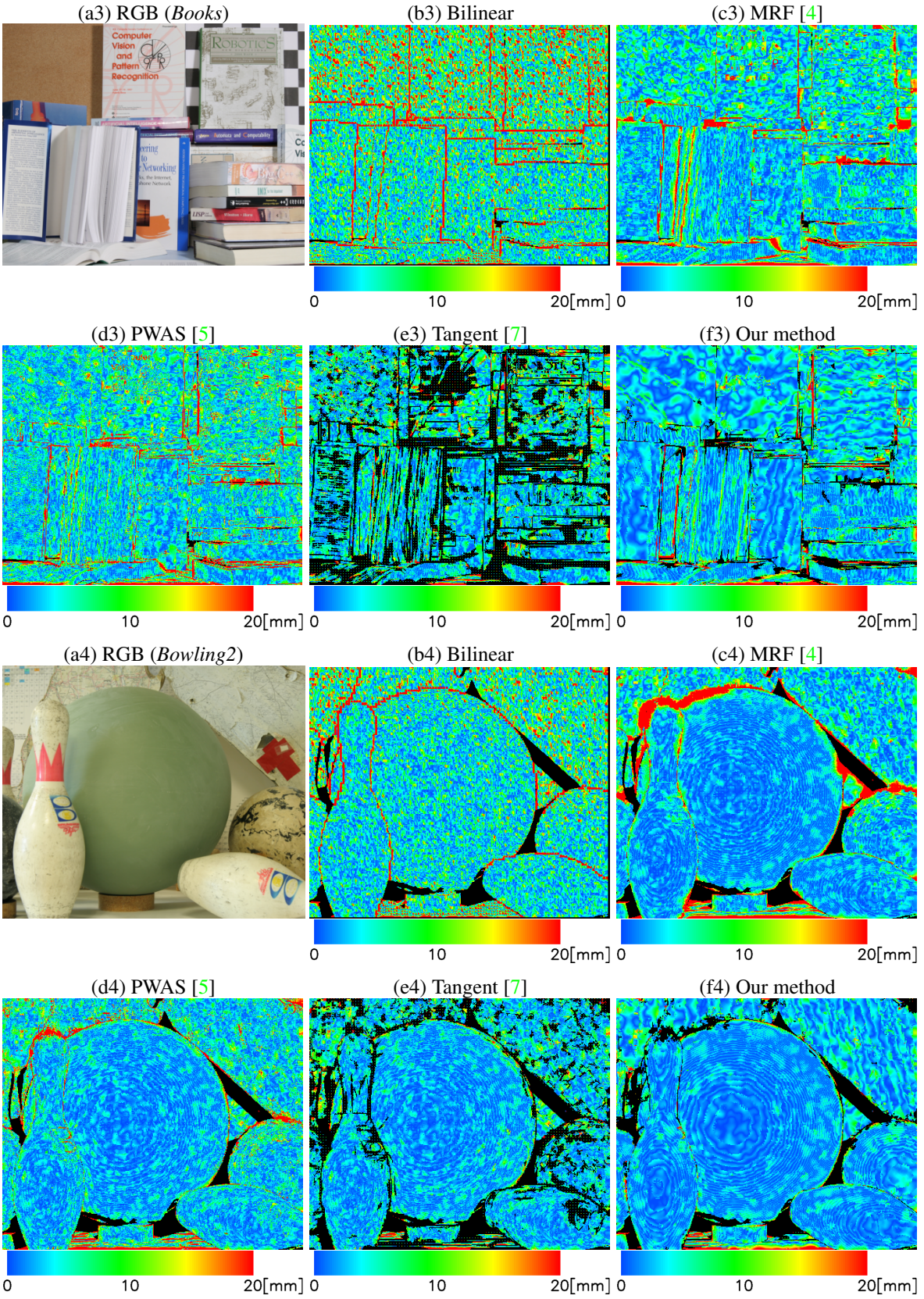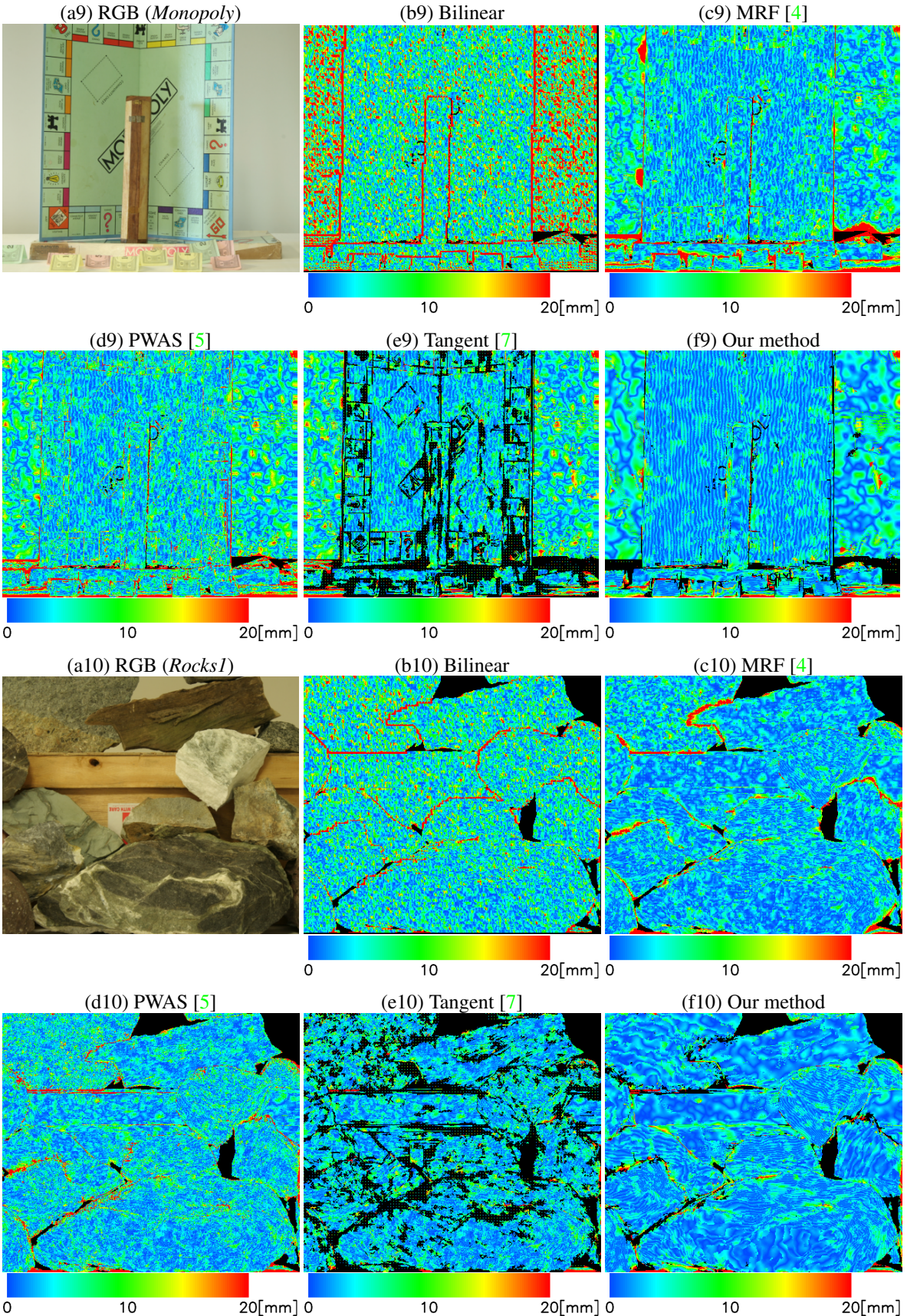
Figure 8. Resulting depth error (Scenes *Books* and *Bowling2*). (a) Color image of the Scene; (b) Bilinear interpolation; (c) MRF [4]; (d) PWAS [5]; (e) the method in [7]; (f) our method. Pixels that have no depth value are colored black. (noise coefficient $k = 3.0 \times 10^{-6}$, magnification factor $\times 4 \times 4$).
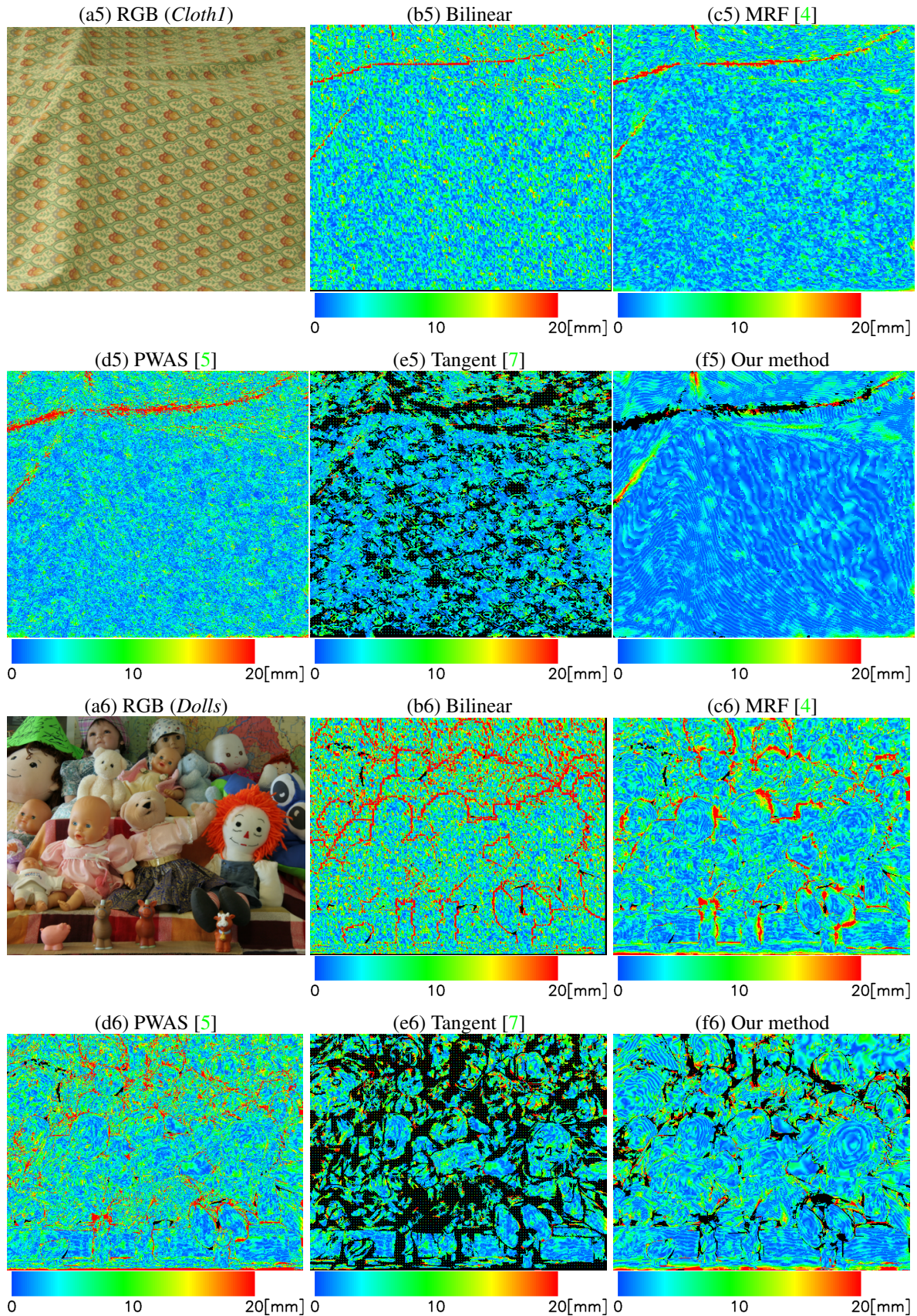
Figure 9. Resulting depth error (Scenes *Cloth1* and *Dolls*). (a) Color image of the Scene; (b) Bilinear interpolation; (c) MRF [4]; (d) PWAS [5]; (e) the method in [7]; (f) our method. Pixels that have no depth value are colored black. (noise coefficient $k = 3.0 \times 10^{-6}$, magnification factor $\times 4 \times 4$).

Figure 10. Resulting depth error (Scenes *Midd2* and *Moebius*). (a) Color image of the Scene; (b) Bilinear interpolation; (c) MRF [4]; (d) PWAS [5]; (e) the method in [7]; (f) our method. Pixels that have no depth value are colored black. (noise coefficient $k = 3.0 \times 10^{-6}$, magnification factor $\times 4 \times 4$).

Figure 11. Resulting depth error (Scenes *Monopoly* and *Rocks1*). (a) Color image of the Scene; (b) Bilinear interpolation; (c) MRF [4]; (d) PWAS [5]; (e) the method in [7]; (f) our method. Pixels that have no depth value are colored black. (noise coefficient $k = 3.0 \times 10^{-6}$, magnification factor $\times 4 \times 4$).

5. Additional results for qualitative comparisons

In this section, we show additional results for qualitative evaluations on real RGB-D data. First, original real images captured by a consumer RGB-D camera are shown (Figure 12, Figure 14, Figure 16, and Figure 18). Then, 3D point clouds reconstructed by each method (MRF [4], PWAS [5], tangent-based method in [7], and our method) are shown (Figure 13, Figure 15, Figure 17, and Figure 19)

In Figure 13, Figure 15, and Figure 17, 3D point clouds reconstructed by MRF [4] and PWAS [5] were over-smoothed around surface-boundaries. Contrarily, tangent-based method in [7] and our method could recover sharp boundaries. In Figure 19, resulting images captured by a far viewpoint shows this difference between each method.

A single-colored smooth cushion in Figure 17 was reconstructed as a rough surface by PWAS [5] and tangent-based method in [7]. And our method reconstructed it as a more smooth surface than MRF [4]. As discussed in our paper, the interior regions of surfaces were smoothly recovered by our method.

(a) RGB   (b) Original sensor data
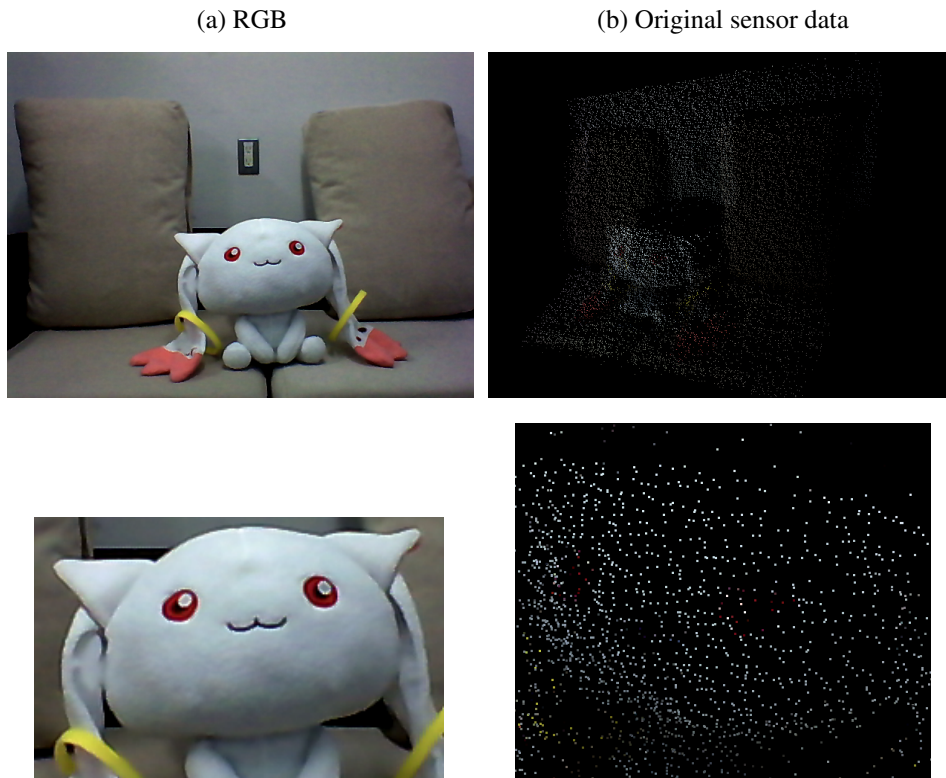
Figure 12. Real sensor data (Scene 1).

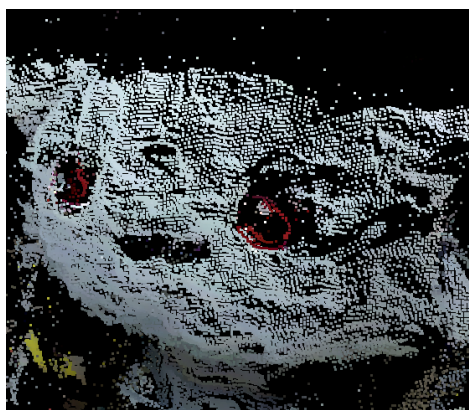(c) MRF [4]
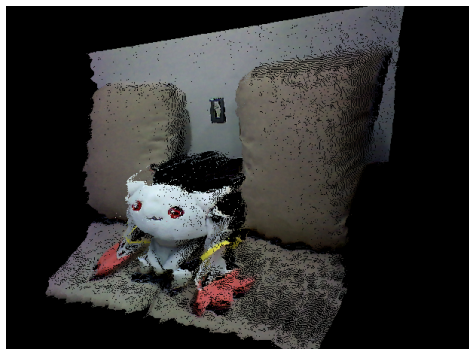
(d) PWAS [5]
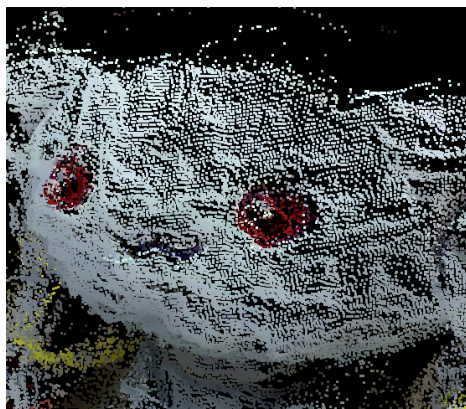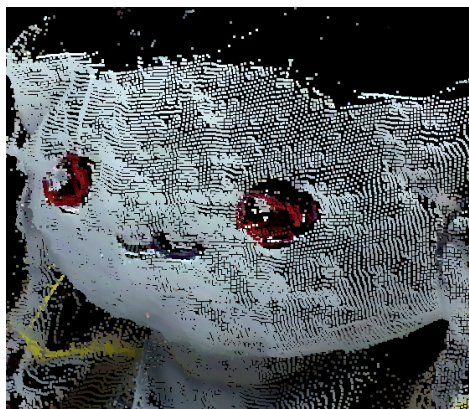
(e) Tangent [7]
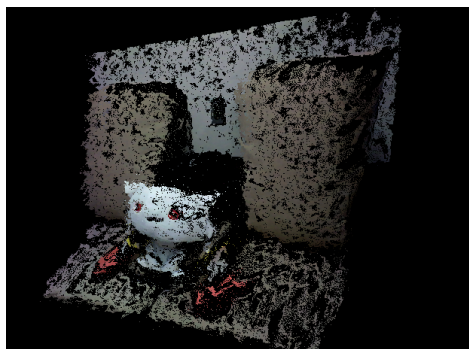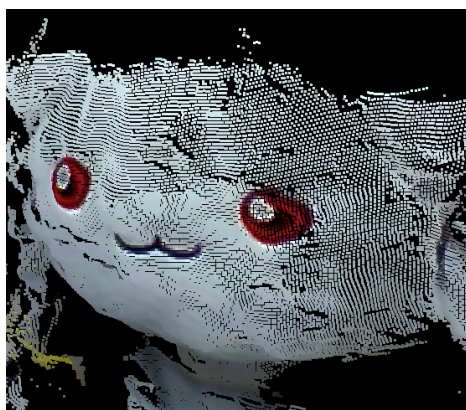
(f) Our method

Figure 13. Visual evaluation on a real sensor data (Scene 1).
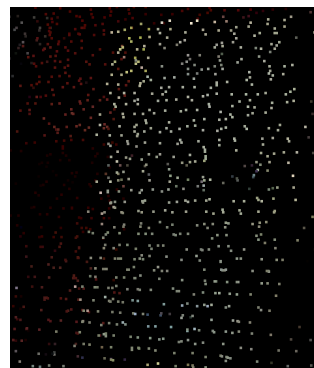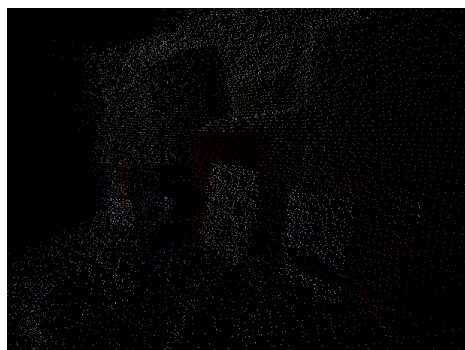
(a) RGB
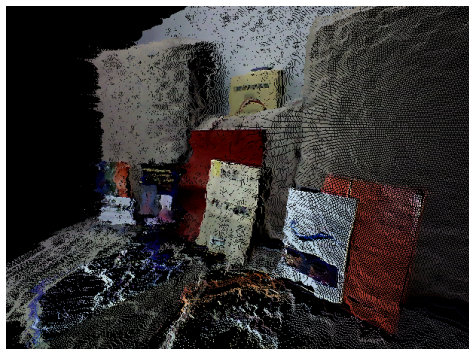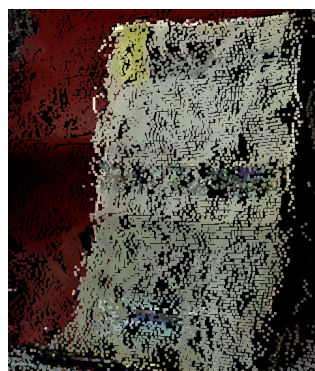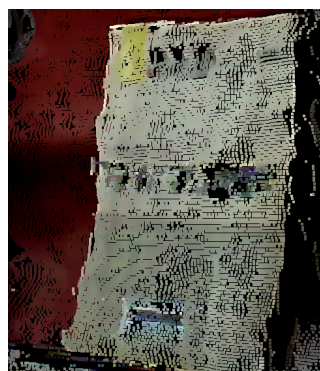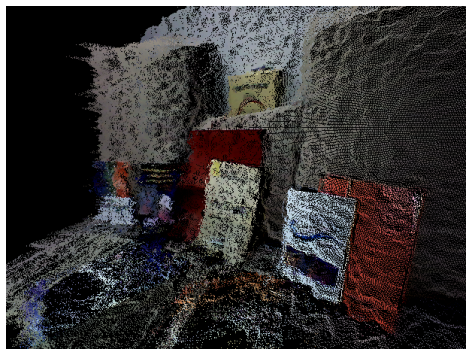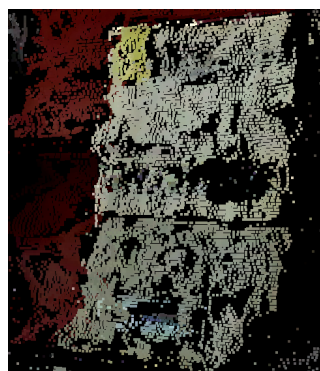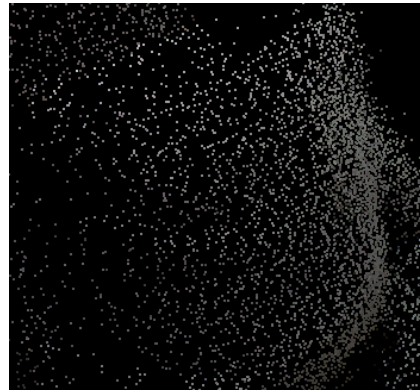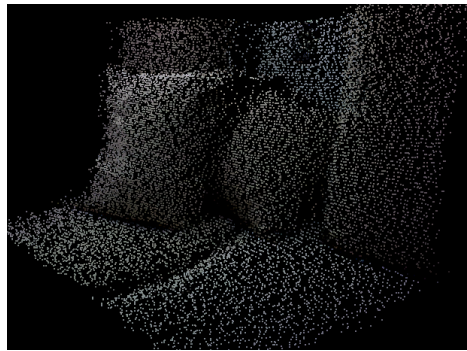
(b) Original sensor data

Figure 14. Real sensor data (Scene 2).

(c) MRF [4]

(d) PWAS [5]

(e) Tangent [7]

(f) Our method

Figure 15. Visual evaluation on a real sensor data (Scene 2).
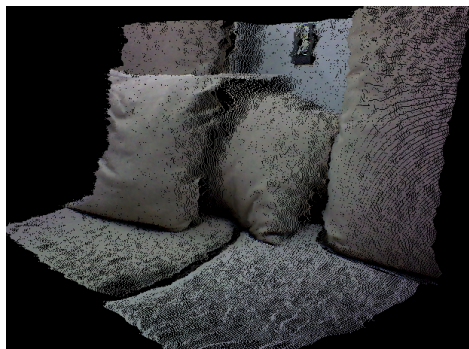
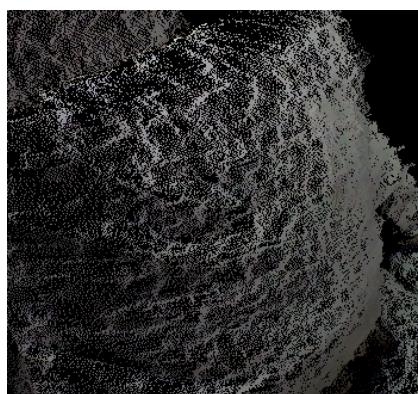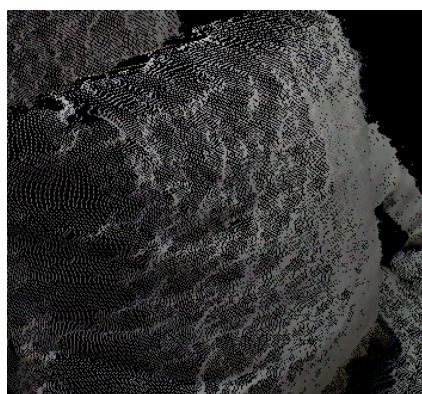(a) RGB   (b) Original sensor data
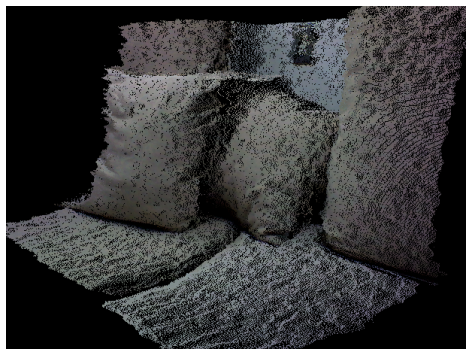

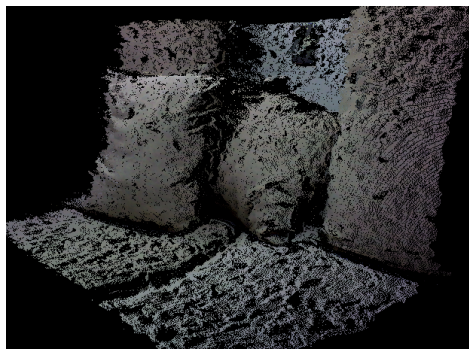Figure 16. Real sensor data (Scene 3).

(c) MRF [4]
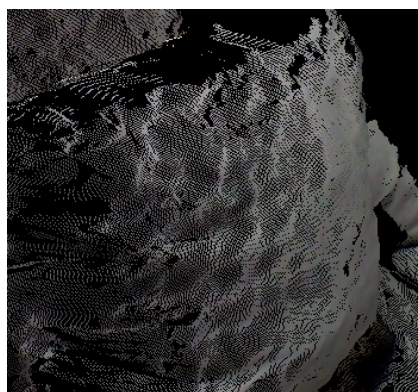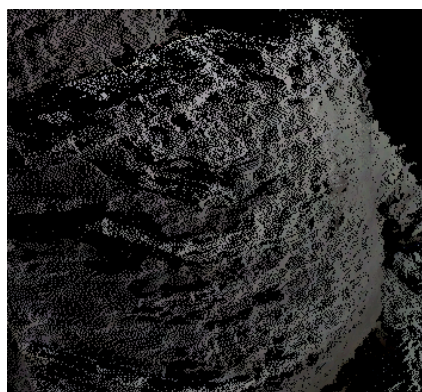
(d) PWAS [5]

(e) Tangent [7]

(f) Our method

Figure 17. Visual evaluation on a real sensor data (Scene 3).
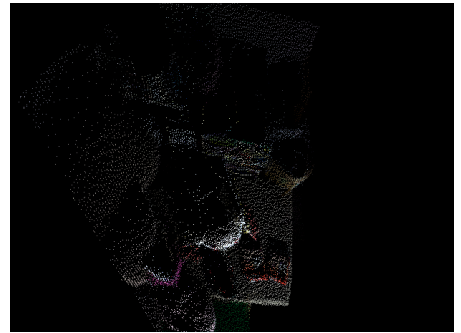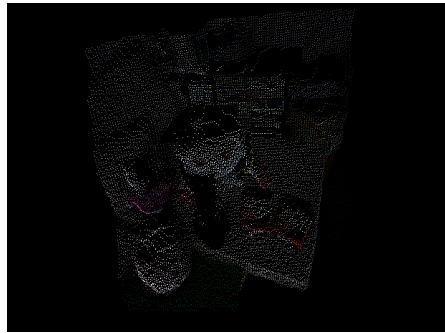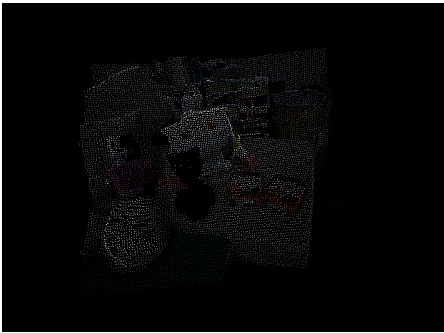
(a) RGB

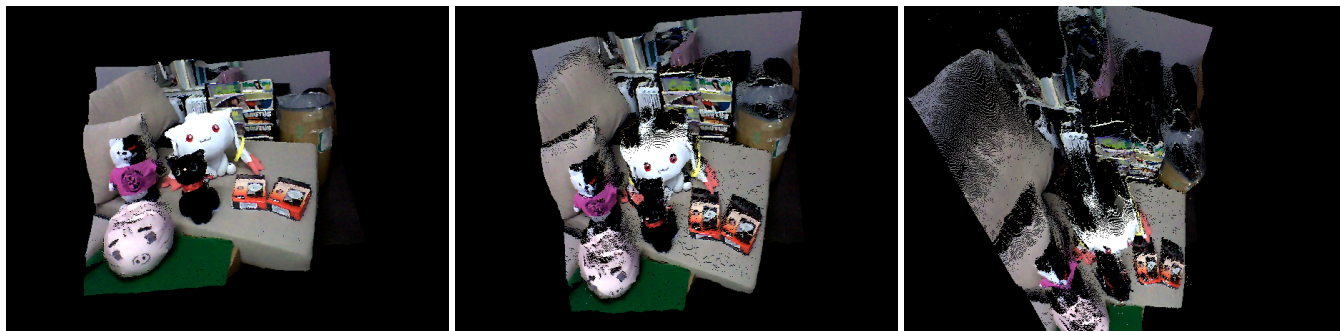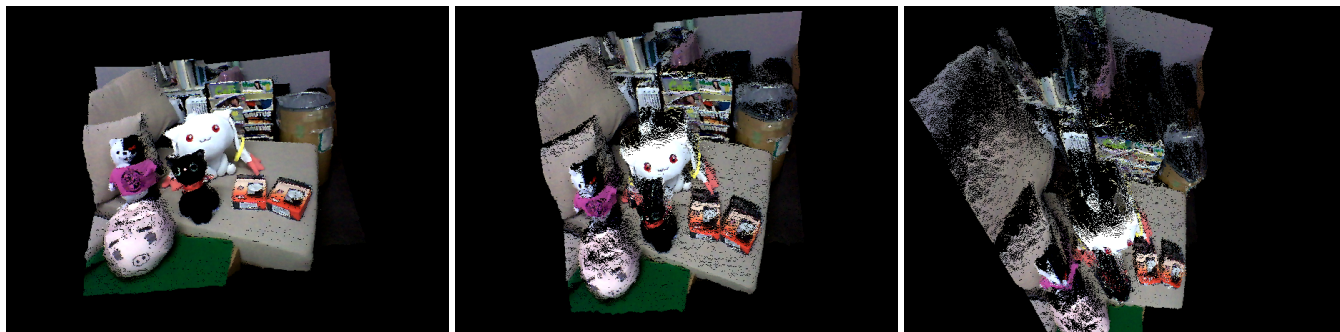(b) Original sensor data

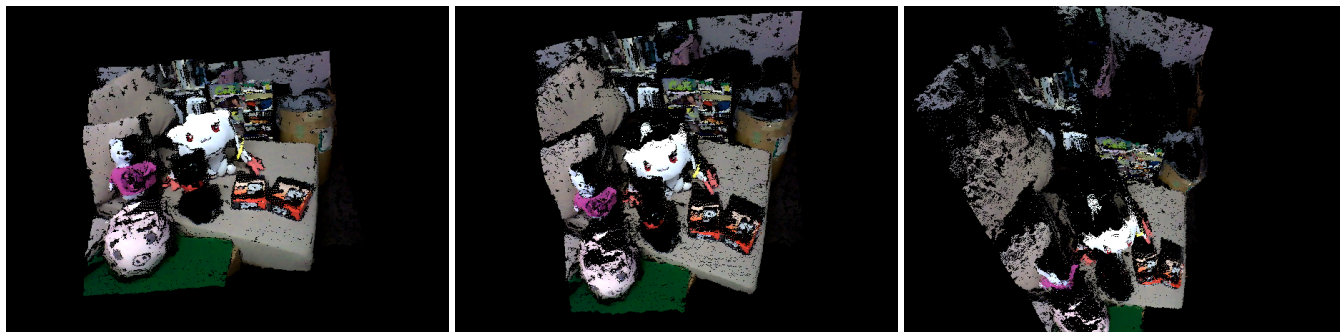Figure 18. Real sensor data (Scene 4).

(c) MRF [4]

(d) PWAS [5]
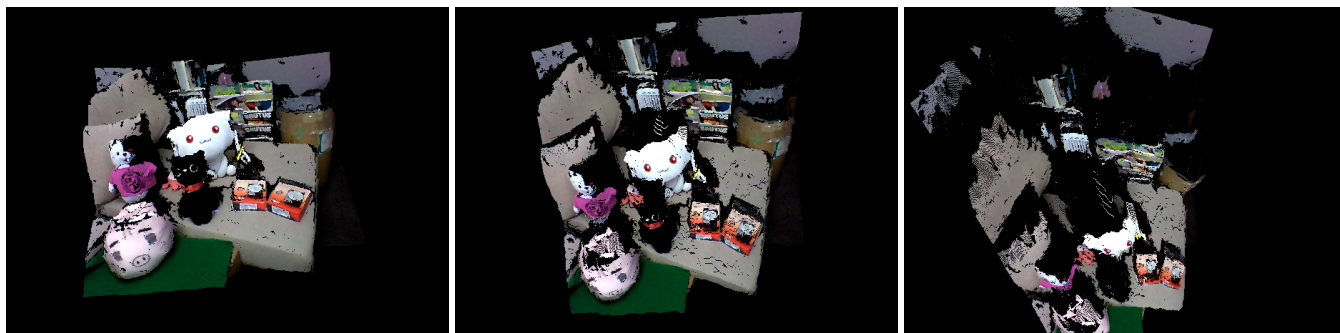
(e) Tangent [7]

(f) Our method

Figure 19. Visual evaluation on a real sensor data (Scene 4).
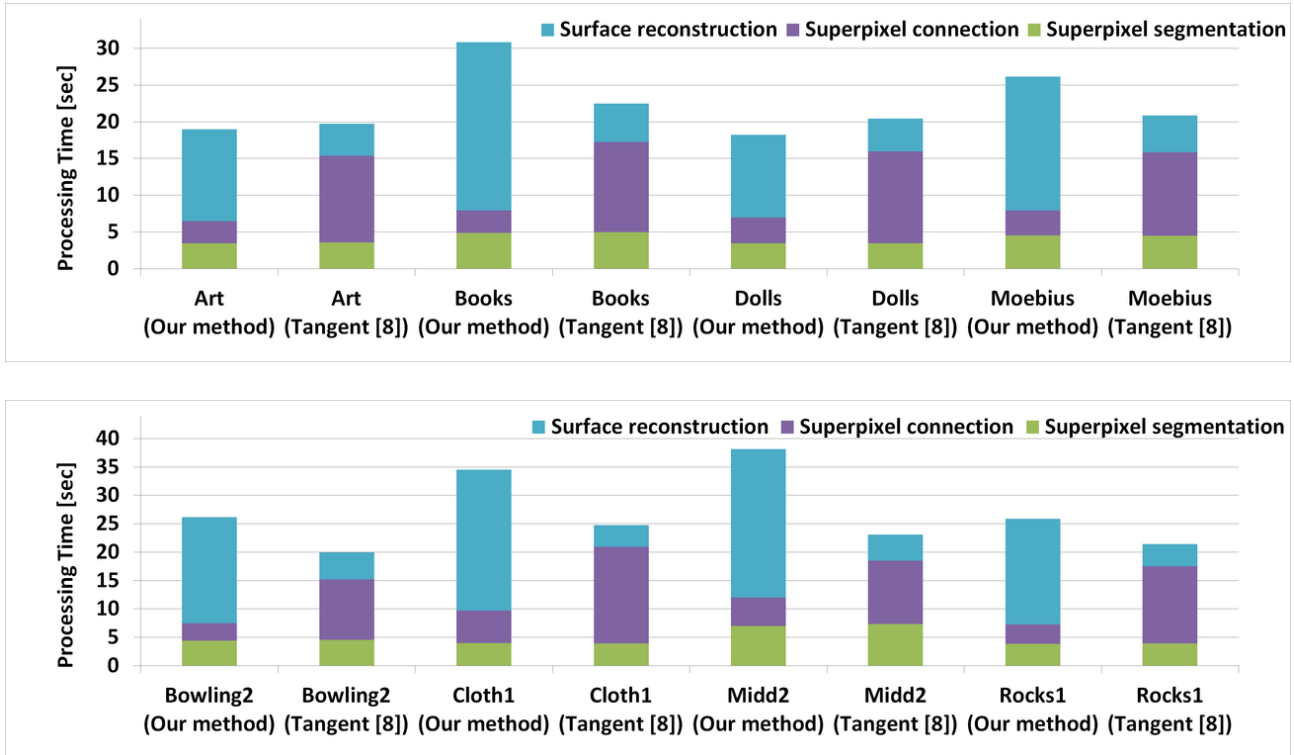
6. Processing time

Figure 20. Processing times (on an upsampling of $\times 4 \times 4$). The bars labeled "tangent" show the processing time of the tangent plane based method [7].

Figure 20 shows the processing times of our method and tangent plane based method in [7]. In this paper, we propose a partially parallelizable accurate and noise robust interpolation procedure but not its parallel implementation. Therefore, we have not optimized the implementation for our experiments. However, we have implemented the proposed method and tangent plane based method in [7] using OpenMP so that we can approximately evaluate their processing times. We ran our experiments to estimate processing times on an Intel Core i7-2600 CPU (3.40 GHz) with 16 GB RAM.

The processing times of surface segmentation of our method were smaller than the tangent plane based method in [7], which were under 14 sec. However, the total processing times were larger. Because in our implementation, our shape refinement process is parallelized in each local coordinate, larger processing time is needed for larger local coordinates. Therefore, for the scenes where large smooth flat surfaces exist, such as *Book*, *Moebius*, *Cloth1*, and *Midd2*, total processing times were large, which were over 25 sec.

References

[1] Scanning laser range finder urg-04lx specifications. http://www.hokuyo-aut.jp/02sensor/07scanner/download/pdf/URG-04LX_spec_en.pdf. 5

[2] Scanning laser range finder uxm-304lx-ew specification. http://www.hokuyo-aut.jp/02sensor/07scanner/download/pdf/UXM-30LX-EW_spec_en.pdf. 5

[3] Sick lms200/211/221/291 laser measurement systems. http://sicktoolbox.sourceforge.net/docs/sick-lms-technical-description.pdf. 5

[4] J. Diebel and S. Thrun. An application of markov random fields to range sensing. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 291–298. MIT Press, Cambridge, MA, 2006. 5, 8, 9, 10, 11, 12, 13, 14, 16, 18, 20

[5] F. Garcia, B. Mirbach, B. Ottersten, F. Grandidier, and A. Cuesta. Pixel weighted average strategy for depth sensor data fusion. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 2805 –2808, sept. 2010. 5, 8, 9, 10, 11, 12, 13, 14, 16, 18, 20

[6] L. Kneip, F. Tache, G. Caprari, and R. Siegwart. Characterization of the compact hokuyo urg-04lx 2d laser range scanner. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 1447–1454, 2009. 5

[7] K. Matsuo and Y. Aoki. Depth interpolation via smooth surface segmentation using tangent planes based on the super-pixels of a color image. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 29–36, Dec 2013. 5, 8, 9, 10, 11, 12, 13, 14, 16, 18, 20, 21

[8] D. Scharstein and C. Pal. Learning conditional random fields for stereo. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1 –8, june 2007. 5