Depth Map Upsampling via Compressive Sensing

Longquan Dai*, Haoxing Wang*, Xing Mei* and Xiaopeng Zhang*
* *NLPR, Institute of Automation Chinese Academy of Sciences*
Email:{lqdai,haoxwang,xmei,xpzhang}@nlpr.ia.ac.cn

***Abstract*—We propose a new method to enhance the lateral resolution of depth maps with registered high-resolution color images. Inspired by the theory of Compressive Sensing (CS), we formulate the upsampling task as a sparse signal recovery problem. With a reference color image, the low-resolution depth map is converted into suitable sampling data (measurements). The signal recovery problem, defined in a constrained optimization framework, can be efficiently solved with variable splitting and alternating minimization. Experimental results demonstrate the effectiveness of our CS-based method: it competes favorably with other state-of-the-art methods with large upsampling factors and noisy depth inputs.**

***Keywords*-depth map; compressive sensing; upsampling;**

I. Introduction

In recent years, a wide range of devices have been developed to measure the 3D information in the real world, such as laser scanners, structured-light systems, time-of-flight cameras and passive stereo systems. The depth maps (range images) captured with most active sensors usually suffer from relatively low resolution, limited precision and significant sensor noise. Therefore, effective depth map post-processing techniques are essential for practical applications such as scene reconstruction and 3D video production.

Inspired by the theory of Compressive Sensing [1], [2], we try to recover the upsampled depth map in a sparse signal reconstruction process. We first compute a set of measurement data from the low-resolution depth map. The measurement data near depth discontinuities are generated with a cellular automaton algorithm. Then we reconstruct the depth signal in an optimization model, with constraints on measurements, smoothness and representation sparseness. An efficient numerical method is provided to solve the model with linear complexity in the number of the image pixels. Experimental results show that, by solving the problem in a CS-based framework, our algorithm can produce high quality depth results with relatively low resolution depth maps. And it shows stable performance under noisy conditions.

The idea of enhancing a depth map with a coupled color image is not new. Existing methods can be roughly classified as either filtering-based methods [3], [4], [5], [6] or optimization-based methods [7], [8]. Filtering-based methods employ color information with various edge-preserving filters [9], [10]. Kopf et al. [3] use a joint bilateral filter to refine the upsampled depth results. Yang et al. [4] instead initialize a cost volume and iteratively smooth each cost slice with a bilateral filter. Sub-pixel accuracy is achieved with an interpolation scheme. Huhle el al. [6] rely on nonlocal means filters (NLM) for depth denoising and upsampling.

More closely related to our work are the optimization-based methods [7], [8]. In [7], Diebel and Thrun construct a two-layer Markov Random Field model for depth map upsampling. The color information of neighboring pixels is encoded as edge weights of the graph. Recently, Park et al. [8] improve this model by including a multi-cue edge weighting scheme and a NLM energy term, which turns out to be very effective for preserving fine structures and depth discontinuities. Our method differs from these methods in that we formulate the model with l_1 sparseness and total variation constraints, which shows more robust behavior against noise and low sampling rates.

II. CS-based Upsampling Model

CS builds upon a fundamental fact that many signals can be represented or approximated with only a few coefficients in a suitable basis [1], [2], [11]. Consider a high-resolution depth map $\mathbf{d} \in \mathbb{R}^n$ in column vector form, it can be linearly represented with an orthonormal basis $\Psi \in \mathbb{R}^{n\times n}$ and a set of coefficients $\mathbf{x} \in \mathbb{R}^n$: $\mathbf{d} = \Psi\mathbf{x}, \mathbf{x} = \Psi^T\mathbf{d}$. The map $\mathbf{d}$ is linearly measured m times ($m \ll n$), which leads to a set of measurements $\mathbf{y} \in \mathbb{R}^m$ with a measurement matrix $\Phi \in \mathbb{R}^{m\times n}$: $\mathbf{y} = \Phi\mathbf{d}$. The CS theory tries to recover depth map $\mathbf{d}$ from measurements $\mathbf{y}$ with the sparsest vector $\mathbf{x}$:

$$\begin{aligned} &\min_{\mathbf{d}} \|\Psi^T\mathbf{d}\|_1 \\ &s.t. \quad \|\mathbf{y} - \Phi\mathbf{d}\|_2 < \epsilon \end{aligned} \tag{1}$$

where ϵ is a bound for the underlying noise.

We incorporate an additional total variation (TV) term for smoothing the depth map while still preserving discontinuities. The TV term is defined in ℓ_1 norm:

$$\|\mathbf{d}\|_{TV} = \sum_{i=1}^{n} (|\nabla_h(\mathbf{d}(i))| + |\nabla_v(\mathbf{d}(i))|) \tag{2}$$

where ∇_h, ∇_v denote the local horizontal and vertical gradients for pixel $\mathbf{d}(i)$ respectively. Thus we convert our final model into an unconstrained optimization problem:

$$\min_{\mathbf{d}} \alpha\|\mathbf{d}\|_{TV} + \beta\|\Psi^T\mathbf{d}\|_1 + \frac{1}{2}\|\mathbf{y} - \Phi\mathbf{d}\|_2^2 \tag{3}$$

where parameter α, β control the weights of the two regularization terms.

978-1-4799-2190-4/13 $31.00 © 2013 IEEE
DOI 10.1109/ACPR.2013.11

For Ψ and Φ, we follow the patterns defined in [12]: Ψ represents a Daubechies Wavelet basis, while Φ samples the high-resolution depth map with canonical pixel basis. To fight against the high mutual coherence between Ψ and ϕ [13], pixels around depth discontinuities should be selected as sampling data points. Since depth borders are not known before upsampling, we infer their positions and the corresponding measurements with auxiliary information, as detailed in the next section.

III. Sampling Data Generation

This section describes how to generate the sampling data from a low-resolution depth map D_l and a registered high-resolution color image I_h. The sampling position information is denoted as a mask image M_h: $M_h(i,j) = 1$ indicates pixel (i,j) is selected as a sampling point, otherwise $M_h(i,j) = 0$. The sampling values are stored in a high resolution depth map D_h. The measurement matrix Φ and the measurements $\mathbf{y}$ can be trivially constructed from M_h and D_h. Without losing any generality, the upsampling factor for both horizontal and vertical directions is set to be U. A pixel $(i,j) \in D_l$ corresponds to a $U \times U$ patch in the high resolution image space.

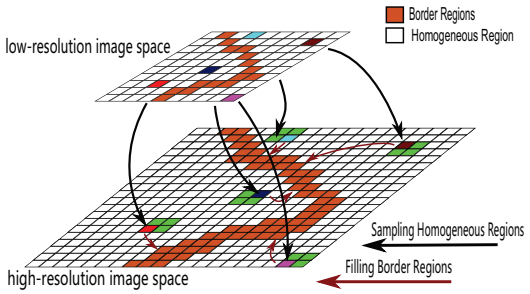


Figure 1. An illustration of the sampling data generation process. Sampling data in homogeneous regions and border regions are generated with two different schemes.

An illustration example of our sampling method is given in Figure 1. We first detect homogeneous regions and border regions in the original depth map D_l with a simple thresholding scheme. A pixel $\mathbf{p} = (i,j) \in D_l$ is classified as 'homogenous' if its depth value $D_l(\mathbf{p})$ satisfies the following condition, otherwise it falls into the 'border' region:

$$|D_l(\mathbf{p}) - D_l(\mathbf{q})| < \lambda, \forall \mathbf{q} \in N(\mathbf{p}) \quad (4)$$

where $N(\mathbf{p})$ is the 4-connected neighborhood of pixel $\mathbf{p}$, and λ is a depth threshold value. We then map this region information to the high resolution image space. M_h, D_h in homogeneous and border regions are computed successively.

The procedure described above works well for small or moderate upsampling factors. However, when U reaches 8 or even larger, the generated sampling points would be too sparse in the high resolution image space, which fails to meet the minimum measurement requirements [14]. We provide a simple hierarchical solution for large upsampling factors. Large U is decomposed into a set of small factors: $U = U_1 \times U_2 \cdots U_m$. Then starting from the low resolution depth map, the sampling data generation process is performed m times with small factors to get the final high resolution M_h and D_h. In practice, $U = 8, 16$ are decomposed as $2\times4, 4\times4$ respectively, such that the number of the times m is kept as low as possible.

A. Sampling Homogeneous Regions

For a homogenous pixel $(i,j) \in D_l$, its depth value is directly mapped to a $U \times U$ homogenous patch in D_h as follows:

$$D_h(i*U+s, j*U+t) = D_l(i,j) \quad s,t = 1, \cdots, U \quad (5)$$

From this patch, we randomly select one or several samples with uniform distribution, and set their M_h values to 1. As stated in [13], this random selection helps to lower the mutual coherence between Ψ and ϕ.

B. Sampling Border Regions

For border pixels in D_l, their depth values are not reliable due to the downsampling process, and directly mapping these pixels to D_h would introduce significant sampling errors. We instead try to fill these regions in D_h with homogenous depth values computed in the previous step. The color image I_h should be considered in the filling process. This problem can be posed as an inpainting problem with a reference color image, and it shares some similarities with the occlusion handling problem in traditional stereo depth estimation [15].

Here we provide a border region filling method based on the classic Cellular Automata (CA) [16]. CA usually work on a regular grid of cells, with finite states and local transition rules, which are suitable for many image processing applications [17]. Our solution is based on the CA model proposed by Vezhnevets and Konouchine [18]. Their model can propagate two labels to the full image. We employ this model for depth propagation. We extend the local transition rules to respect the color distribution and the edges in I_h, such that the propagation doesn't generate incorrect depth boundaries.

For each pixel $\mathbf{p}$, four state variables $S_\mathbf{p} = (D_\mathbf{p}, \Theta_\mathbf{p}, C_\mathbf{p}, E_\mathbf{p})$ are stored: depth value, local 'transition strength', normalized color information, and edge information. $\Theta_\mathbf{p}$ is bounded to $[0,1]$, while $C_\mathbf{p}$ and $E_\mathbf{p}$ are extracted from color image I_h. In the initial step, $\Theta_\mathbf{p} = 1$ for all the pixels with valid depth values, otherwise $\Theta_\mathbf{p} = 0$. We apply the Canny filter to detect edges in I_h. If $\mathbf{p}$ lies on an intensity edge, $E_\mathbf{p} = 1$, otherwise $E_\mathbf{p} = 0$. After initialization, we collect all the pixels in the border regions as a set P. The CA-based region filling algorithm updates $S_\mathbf{p}(\forall \mathbf{p} \in P)$ in an iterative process. An iteration from time t to $t+1$ is

Algorithm 1 CA-based Border Region Filling Algorithm

Input:
State variables S^t
Output:
State variables S^{t+1}

1: **for** $\forall \mathbf{p} \in P$ **do**
2: $D_{\mathbf{p}}^{t+1} = D_{\mathbf{p}}^{t}$
3: $\Theta_{\mathbf{p}}^{t+1} = \Theta_{\mathbf{p}}^{t}$
4: **for** $\forall \mathbf{q} \in N(\mathbf{p})$ **do**
5: **if** $E(\mathbf{p}) == 0$ and $E(\mathbf{q}) == 1$ **then**
6: continue
7: **end if**
8: **if** $f(C_{\mathbf{p}}, C_{\mathbf{q}}) \cdot \Theta_{\mathbf{q}} > \Theta_{\mathbf{p}}$ **then**
9: $D_{\mathbf{p}}^{t+1} = D_{\mathbf{p}}^{t}$
10: $\Theta_{\mathbf{p}}^{t+1} = f(C_{\mathbf{p}}, C_{\mathbf{q}}) \cdot \Theta_{\mathbf{q}}$
11: **end if**
12: **end for**
13: **end for**
14: **return** S^{t+1}

shown in Algorithm 1, where f is a monotone decreasing function bounded to [0,1]:

$$f(C_1, C_2) = 1 - \frac{\|C_1 - C_2\|_2}{\sqrt{3}} \tag{6}$$

In each iteration, the transition strength $\Theta_{\mathbf{p}}$ is updated with the neighboring color information. The pixels lying on intensity edges are only allowed to propagate depth information along the edge (Line $5-7$ in Algorithm 1). When no more pixel changes its state in the iteration, the algorithm stops, and the output state variables are used to update D_h and M_h. The pixels in P lying at intensity edges are all selected as sampling points. Then a subset of the remaining pixels are randomly selected with a uniform distribution.

IV. Numerical Solution

In this section, we provide a first-order numerical solution for the optimization problem defined in (3). A major difficulty in minimizing (3) is that both the TV term and the sparseness term are non-differential ℓ_1 regularizers. We decompose the original problem into three subproblems. with variable-splitting and quadratic penalty techniques. For each subproblem, efficient solution is available. Therefore, the original problem can be solved in an alternating minimization framework [19].

We introduce two auxiliary vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, such that $\mathbf{d}$ can be decoupled from the two terms:

$$\begin{aligned} &\min_{\mathbf{d},\mathbf{u},\mathbf{v}} \alpha\|\mathbf{u}\|_{TV} + \beta\|\mathbf{v}\|_1 + \frac{1}{2}\|\mathbf{y} - \Phi\mathbf{d}\|_2^2 \\ &s.t. \quad \mathbf{u} = \mathbf{d}, \mathbf{v} = \Psi^T\mathbf{d} \end{aligned} \tag{7}$$

By including two quadratic penalty terms, this problem can be approximated by following problem with a unconstrained optimization problem:

$$\min_{\mathbf{d},\mathbf{u},\mathbf{v}} E(\mathbf{d}, \mathbf{u}, \mathbf{v}) \tag{8}$$

where $E(\mathbf{d}, \mathbf{u}, \mathbf{v}) = \alpha\|\mathbf{u}\|_{TV} + \frac{\alpha\gamma}{2}\|\mathbf{u} - \mathbf{d}\|^2 + \beta\|\mathbf{v}\|_1 + \frac{\beta\delta}{2}\|\mathbf{v} - \Psi^T\mathbf{d}\|^2 + \frac{1}{2}\|\mathbf{y} - \Phi\mathbf{d}\|_2^2$ and parameter γ and δ control the approximation penalty for $\mathbf{u}$ and $\mathbf{v}$ respectively.

Problem (8) can be solved in an alternating minimization framework as follows:

1. For fixed $\mathbf{v}, \mathbf{d}$, the problem $\min_{\mathbf{u}} \|\mathbf{u}\|_{TV} + \frac{\gamma}{2}\|\mathbf{u} - \mathbf{d}\|^2$ can be efficiently solved with a Split Bregman algorithm from [20].

2. For fixed $\mathbf{u}, \mathbf{d}$, The subproblem $\min_{\mathbf{v}} \|\mathbf{v}\|_1 + \frac{\delta}{2}\|\mathbf{v} - \Psi^T\mathbf{d}\|_2^2$ can solved with simple one-dimensional shrinkage: $\mathbf{v} = \max\left(\Psi^T\mathbf{d} - \frac{1}{\delta}, 0\right) sgn(\Psi^T\mathbf{d})$

3. For fixed $\mathbf{u}, \mathbf{v}$, The subproblem $\min_{\mathbf{d}} \frac{\alpha\gamma}{2}\|\mathbf{u} - \mathbf{d}\|^2 + \frac{\beta\delta}{2}\|\mathbf{v} - \Psi^T\mathbf{d}\|^2 + \frac{1}{2}\|\mathbf{y} - \Phi\mathbf{d}\|_2^2$ is a least square problem which promises a closed-form solution: $\mathbf{d} = M(\alpha\mathbf{u} + \beta\Psi\mathbf{v} + \mathbf{y})$ where $M = (\alpha\gamma I + \beta\delta I + \Phi^T\Phi)^{-1}$ is a diagonal matrix.

Step 1-3 are iteratively performed until the algorithm converges. For our upsampling problem on 1390×1110 images, stable results can be efficiently achieved within 200 iterations.

V. Experimental Results

We quantitatively test our algorithm on the Middlebury stereo datasets [21], which provide both high resolution color images and ground truth depth maps. The datasets we use are 'Books', 'Dolls', 'Moebius' and 'Plastic'. Parameter $\lambda, \alpha, \beta, \gamma, \delta$ are chosen as $4.0, 1.0, 1.0, 32.0, 32.0$, which are kept constant for all the data sets. The upsampling factor U varies from $2\times$ to $16\times$, which covers the resolution range for most depth sensors. For a given factor U, the ground truth depth map is downsampled by U to create the input depth data. We measure the accuracy of the upsampled results with PSNR.

For comparison, four methods are selected: the bilateral-filtering based method (denoted as Bilateral) [4], two MRF-based methods (denoted as MRF1, MRF2) [7], [8] and the original CS-based method (denoted as CS2) [12]. We provide the results to show that our sampling strategy is more suitable for the specific problem.

We first test the algorithm with 'ideal' low resolution depth maps without noise corruption. The PSNR results for the five methods under various upsampling factors are presented in Figure 2. Our algorithm works better under large upsampling factors. It consistently outperforms other methods with $4\times$, $8\times$ and $16\times$ upsampling.

An interesting feature of our method is that the accuracy doesn't necessarily go down when U increases, which is different from other methods. In fact, for most data sets, the best results are achieved with $4\times$ or $8\times$ upsampling. This

feature can be explained by CS theory and the sampling data generation method: first, we employ a hierarchical sampling scheme for large Us, which means the number of the samples is not seriously affected by U; second, our method generates accurate samples on boundaries with CA-based region filling method, which is mainly controlled by the reference color image; finally, large U values bring more randomness to the selection of the sampling positions, which helps to lower the mutual coherence between the measurement matrix and the representation matrix. All these reasons contributes to the good performance of our algorithm under large upsampling factors. For qualitative comparison, we present some $8\times$ upsampled results computed by CS1, Bilateral and MRF2 methods in Figure 3. It can be seen that our method preserves sharp and accurate depth boundaries during the upsampling process, which demonstrate the effect of the ℓ_1 regularization terms.

We also test the algorithm with noisy measurements. The noise characteristics in practical range sensors usually depends on the distance between the sensor and the scene. To simulate this effect, we employ a conditional Gaussian model from [8]. The PSNR results for the five methods are presented in Figure 4. Figure 5 provides $8\times$ upsampled results computed by CS1, MRF2 and Bilateral methods.

VI. Conclusion

We have presented a new method for depth map upsampling. Based on the theory of Compressive Sensing, our method converts the low resolution depth maps into a set of measurements, and then formulates the upsampling task as an constrained optimization problem. We validate our method with the Middlebury data sets and demonstrate that our method clearly outperforms previous methods.

Acknowledgment

The work is supported by the 2013 Annual Beijing Technological and Cultural Fusion for Demonstrated Base Construction and Industrial Nurture (No. Z131100000113007), and the National Natural Science Foundation of China (Nos. 61331018, 61271431, and 61271430).

References

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, 2006.

[2] D. Donoho, "Compressive sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289–1306, 2006.

[3] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, p. 96, 2007.

[4] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in *CVPR*, 2007.

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

[6] B. Huhle, T. Schairer, P. Jenke, and W. Strasser, "Fusion of range and color images for denoising and resolution enhancement with a non-local filter," *CVIU*, vol. 114, pp. 1336–1345, 2010.

[7] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," in *NIPS*, 2005.

[8] J. Park, H. Kim, Y. Tai, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3d-tof cameras," in *ICCV*, 2011.

[9] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *ICCV*, 1998.

[10] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *CVPR*, 2005.

[11] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, pp. 21–30, 2008.

[12] S. Hawe and K. D. M. Kleinsteuber, "Dense disparity maps from sparse disparity measurements," in *ICCV*, 2011.

[13] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2845–2862, 2001.

[14] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *IEEE Signal Processing Magazine*, vol. 23, pp. 969–985.

[15] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic inpainting: Joint color and depth completion from stereo images," in *CVPR*, 2008.

[16] J. V. Neumann, *Theory of Self-Reproducing Automata*. Champaign, IL, USA: University of Illinois Press, 1966.

[17] D. Popovici and A. Popovici, "Cellular automata in image processing," in *15th International Symposium on Mathematical Theory of Networks and Systems*, 2002.

[18] V. Vezhnevets and V. Konouchine, "Growcut: Interactive multi-label n-d image segmentation by cellular automata," in *GraphiCon*, 2005.

[19] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Img. Sci.*, vol. 1, pp. 248–272, 2008.

[20] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM J. Img. Sci.*, vol. 2, pp. 323–343, 2009.

[21] H. Hirschmüller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *CVPR*, 2007.

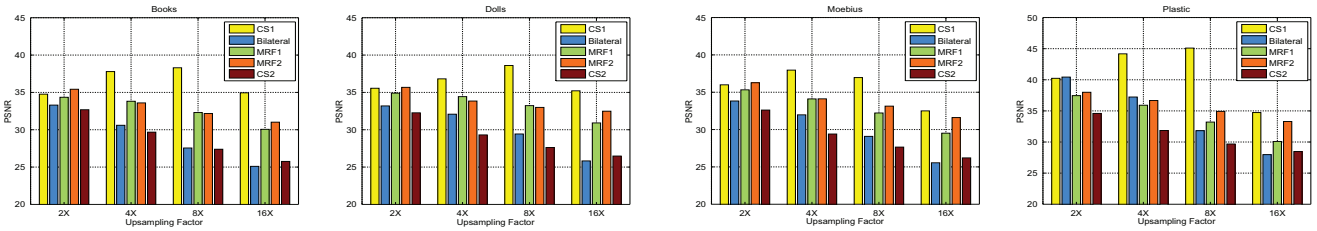
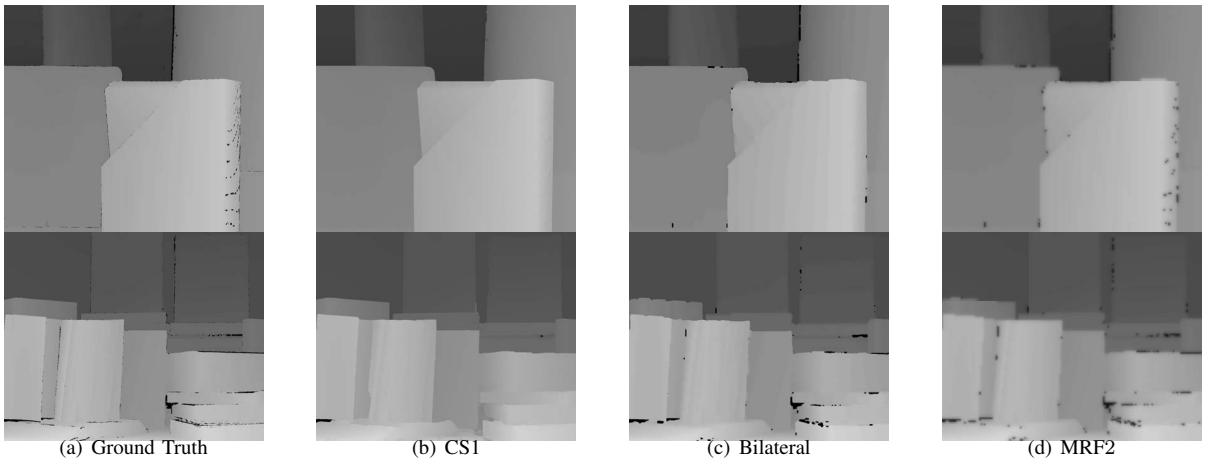


Figure 2. Upsampling PSNR results with ideal depth measurements. Our method (CS1) consistently outperforms other methods with $4\times$, $8\times$ and $16\times$ upsampling.



(a) Ground Truth (b) CS1 (c) Bilateral (d) MRF2

Figure 3. $8\times$ upsampled depth maps for 'Plastic' and 'Books' data sets. The depth results are computed with CS1, Bilateral and MRF2 respectively.

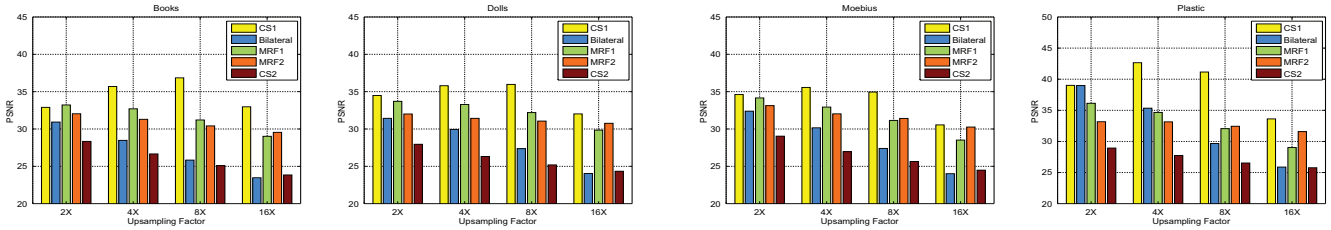
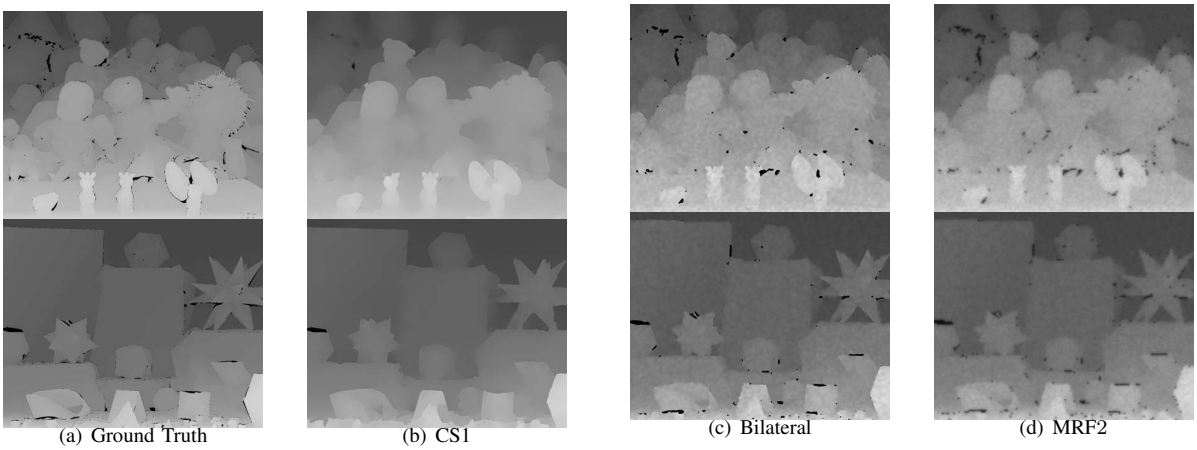


Figure 4. Upsampling PSNR results with noisy measurements. Our method (CS1) still outperforms other methods in most cases. It shows robust behavior in noisy conditions.



(a) Ground Truth (b) CS1 (c) Bilateral (d) MRF2

Figure 5. $8\times$ upsampled depth maps with noisy measurements for 'Dolls' and 'Moebius' data sets. The depth results are computed with CS1, Bilateral and MRF2 respectively.