



ELSEVIER

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Depth map upsampling using compressive sensing based model



Longquan Dai, Haoxing Wang, Xiaopeng Zhang*

National Laboratory of Pattern Recognition, Institute of Automation Chinese Academy of Sciences, 95 Zhongguancun East Road, Beijing, China

ARTICLE INFO

Article history:

Received 20 November 2013

Received in revised form

24 November 2014

Accepted 30 November 2014

Communicated by Zhouchen Lin

Available online 18 December 2014

Keywords:

Depth map

Compressive sensing

Upsampling

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 that solves an underdetermined system. 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 by 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.

© 2014 Elsevier B.V. All rights reserved.

1. 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, especially for 3D face recognition [1] and 3D object recognition [2,3].

In this paper, we present a method to enhance the spatial resolution of a depth map with a registered high-resolution color image. Our method is based on two key assumptions: first, neighboring pixels with similar colors are likely to have similar depth values; second, just like most natural images, an ideal depth map without noise corruption has large smooth regions and relatively few discontinuities, and therefore can be approximated with a sparse representation in some transform domain such as multiscale wavelets. Although the first assumption has been extensively explored in recent depth post-processing work [4–9], relative less attention has been given to the second assumption [10,11].

Inspired by the theory of compressive sensing [12,13], 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, and no

filtering techniques are involved in the process. 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 rest of the paper is organized as follows. Related work is reviewed in Section 2. Section 3 provides a brief introduction to the CS theory, whereas our CS-based upsampling model is presented in Section 4. After that, in Section 5, we describe how to generate the sampling data for the model, and we provide a numerical solution in Section 6. Section 7 reports the experimental results and discusses how to register a low resolution depth map and its companion high resolution color image as well as the influence of sampling pattern. At last, conclusions are given in Section 8.

2. Related work

As stated in Section 1, 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 [5–8] or optimization-based methods [4,9].

Filtering-based methods employ color information with various edge-preserving filters [14,15]. Kopf et al. [5] use a joint bilateral filter to refine the upsampled depth results. Yang et al. [6] instead initialize a cost volume and iteratively smooth each cost slice with a bilateral filter. Sub-pixel accuracy is achieved with an interpolation

* Corresponding author. Tel./fax: +86 10 82544696.

E-mail addresses: lqdai@nlpr.ia.ac.cn (L. Dai),Haox.wang@nlpr.ia.ac.cn (H. Wang), xpzhang@nlpr.ia.ac.cn (X. Zhang).

scheme. Huhle et al. [8] rely on nonlocal means filters (NLM) for depth denoising and upsampling. One advantage of filtering-based methods is that they can be easily parallelized on graphics hardware [7,8]. However, to find enough support for each pixel, large filtering kernels are often used, or the filters have to be performed iteratively, which might lead to over-smoothed depth results.

The methods which are more closely related to our algorithm are the optimization-based methods [4,9]. In [4], 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. [9] improve this model by including a multi-cue edge weighting scheme and an NLM energy term, which turns out to be very effective for preserving fine structures and depth discontinuities. To make the problem tractable, both methods use quadratic cost functions, which can be solved using standard numerical methods such as conjugate gradient. Our method differs from these methods in which we formulate the model with l_1 sparseness and total variation constraints, which shows the more robust behavior against noise and low sampling rates.

Recently, some researchers have explored sparse representations for depth map processing [10,11]. Tošić et al. [10] use sparse coding techniques [16] to learn a dictionary from Middlebury disparity data sets. This dictionary is exploited in a MRF model, which brings accuracy improvements for stereo depth estimation and range image denoising. Hawe et al. [11] propose a CS-based depth estimation method from sparse measurements. They show that, by taking only 5% of the disparity data as measurements, their method can recover the full disparity map with high accuracy, which is quite impressive. An essential point of their method is that the pixels lying at depth boundaries should be selected as sampling points, otherwise the reconstruction accuracy would be seriously affected. Unfortunately, such information is unavailable in our low-resolution depth inputs. We provide a novel method to generate measurements at these sampling positions with a registered reference color image, which proves crucial for the upsampling accuracy. Moreover, we employ a different CS model with better regularization ability.

3. CS theory and underdetermined linear system

CS theory finds an optimal solution \mathbf{x}^n from the observed data $\mathbf{y} \in \mathbb{R}^m$ by reducing the problem to solving an underdetermined linear system. In mathematical terms, the observed data \mathbf{y}^m is connected to the signal \mathbf{x}^n of interest via

$$\Phi \mathbf{x} = \mathbf{y} \quad (1)$$

where $m < n$, \mathbf{x} is the s -sparse vector which only has s nonzero components and the measurement matrix $\Phi \in \mathbb{R}^{m \times n}$ models the linear measurement process. Traditional wisdom of linear algebra suggests that the number m of measurements must be at least as large as the signal length n . Indeed, if $m < n$, the classical linear algebra indicates that the underdetermined linear system Eq. (1) has infinite solutions. In other words, without additional information, it is impossible to recover \mathbf{x} from \mathbf{y} in the case $m < n$. However, with additional sparsity assumption, it is actually possible to reconstruct the sparse vector \mathbf{x} from underdetermined measurements $\mathbf{y} = \Phi \mathbf{x}$ because many real-world signals are sparse. Even though they are acquired with seemingly too few measurements, exploiting sparsity enables us to solve the resulting underdetermined systems of linear equations. More importantly, there are many efficient algorithms for the reconstruction [17–19].

Specifically, CS theory reconstructs \mathbf{x} as a solution of following combinatorial optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0$$

$$\text{s.t. } \Phi \mathbf{x} = \mathbf{y} \quad (2)$$

where $\|\mathbf{x}\|_0$ denotes the number of nonzero entries of a vector. However, the minimization problem is nonconvex and NP-hard. It thus is intractable for a modern computer. An alternative method is ℓ_1 minimization, which can be interpreted as the convex relaxation ℓ_0 minimization.

$$\begin{aligned} \min_{\mathbf{x}} \|\mathbf{x}\|_1 \\ \text{s.t. } \Phi \mathbf{x} = \mathbf{y} \end{aligned} \quad (3)$$

One major shortcoming of above considerations is that they do not carry over to the complex setting such as the contaminated measurements \mathbf{y} . As a remedy, we can directly extend the ℓ_1 minimization (3) to a more general ℓ_1 minimization taking measurement error into account, namely,

$$\begin{aligned} \min_{\mathbf{x}} \|\mathbf{x}\|_1 \\ \text{s.t. } \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon \end{aligned} \quad (4)$$

It is worth noting that the solution of Eq. (4) is strongly linked to the output of the ℓ_1 denoising, which consists in solving, for some parameter $\beta \geq 0$

$$\min_{\mathbf{x}} \beta \|\mathbf{x}\|_1 + \frac{1}{2} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 \quad (5)$$

Expecting Eq. (3) can restore any \mathbf{x} for any Φ is unreasonable. Instead, CS theory only proves that for any integer $n > 2s$, there exists a measurement matrix $\Phi \in \mathbb{R}^{m \times n}$ with $m = 2s$ rows such that every s sparse vector $\mathbf{x} \in \mathbb{R}^n$ can be recovered from its measurement vector $\mathbf{y} \in \Phi \mathbf{x}$ as a solution of Eq. (3). However, finding out the measurement matrix Φ is a remarkably intriguing endeavor. To date, it is still an open problem to construct explicit matrices which are provably optimal in a compressive sensing setting. One novelty of the work is just providing a method to construct the measurement matrix Φ for depth upsampling.

The depth upsampling problem can be reduced to the problem of using the underdetermined linear system Eq. (1) to find an optimal solution \mathbf{x}^n from few measurements \mathbf{y}^m because depth upsampling aims to restore a high-resolution depth map from a low-resolution depth map and the values of the low-resolution depth map can be viewed as the samplers of the high-resolution depth map. Unfortunately, depth maps are not usually sparse in the canonical (pixel) basis. But they are often sparse after a suitable transformation, for instance, a wavelet transform or discrete cosine transform. This means that we can write $\mathbf{x} = \Psi \mathbf{z}$, where \mathbf{z}^n is a sparse vector and $\Psi^{m \times n}$ is a unitary matrix representing the transform. Recalling $\mathbf{y} = \Phi \mathbf{x}$, depth upsampling finds a solution $\mathbf{x} = \Psi^T \mathbf{z}$ from the underdetermined linear system Eq. (6).

$$\Phi \Psi^T \mathbf{z} = \mathbf{y} \quad (6)$$

Obviously, the underdetermined linear system (6) is similar to the underdetermined linear system (1) and the similarity leads us to guess that CS theory also deals with this kind of underdetermined linear system. Indeed, studying Eq. (6) is the origin of CS theory. Without any doubt, CS theory can efficiently solve it [12,13,20] by using the optimization problem (3).

4. CS-based upsampling model

We build our upsampling model upon a fundamental fact that many signals can be represented or approximated with only a few coefficients in a suitable basis. 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} :

$$\begin{aligned} \min_{\mathbf{d}} \quad & \|\Psi^T \mathbf{d}\|_0 \\ \text{s.t.} \quad & \mathbf{y} = \Phi \mathbf{d} \end{aligned} \quad (7)$$

This l_0 minimization problem is known to be NP-hard even for approximate solutions [21]. If $\|\cdot\|_0$ is approximated with a convex term $\|\cdot\|_1$, the resulting problem can be posed as a linear program [22]. For practical applications, the measurement constraints are usually relaxed due to additive noise. These approximations lead to the following l_1 regularization problem:

$$\begin{aligned} \min_{\mathbf{d}} \quad & \|\Psi^T \mathbf{d}\|_1 \\ \text{s.t.} \quad & \|\mathbf{y} - \Phi \mathbf{d}\|_2 < \epsilon \end{aligned} \quad (8)$$

where ϵ is a bound for the underlying noise.

Sparsity as a prior could not guarantee that Eq. (8) could produce regularized results as it can only help us identify a solution from the infinite possible solutions of the underdetermined linear system (6). Fig. 1(b) shows the upsampling result produced by Eq. (8). We can observe that the depth surfaces suffer from fluctuation artifacts which significantly lower the upsampling quality. To solve the problem, 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:

$$\|\mathbf{d}\|_{TV} = \sum_{i=1}^n (|\nabla_h(\mathbf{d}(i))| + |\nabla_v(\mathbf{d}(i))|) \quad (9)$$

where ∇_h, ∇_v denote the local horizontal and vertical gradients for pixel $\mathbf{d}(i)$ respectively. In practice, we find that l_1 norm TV (also known as anisotropic TV [23]) produces sharper boundaries than l_2 norm for our data sets. After adding this term to the objective function in Eq. (8), we convert our final model into the following unconstrained optimization problem:

$$\min_{\mathbf{d}} \alpha \|\mathbf{d}\|_{TV} + \beta \|\Psi^T \mathbf{d}\|_1 + \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{d}\|_2^2 \quad (10)$$

where parameters α, β control the weights of the two regularization terms. The regularized result of Eq. (10) is demonstrated in Fig. 1(c). Compared with the result of Eq. (8), the nasty artifacts are removed entirely and therefore the depth surfaces in Fig. 1(c) are rather satisfactory.

For Ψ and Φ , we follow the patterns defined in [11]: Ψ represents a Daubechies Wavelet basis, while Φ samples the high-resolution depth map with canonical pixel basis. The term Φ is important for our depth upsampling application as Φ should satisfy the minimum measurement requirement [24] deduced from CS theory. Strictly speaking, CS theory could not be directly applied to our model without mathematical proofs because we add an extra TV term to the standard CS model and thus our CS-based upsampling model (10) does not coincide with the standard CS model. However, we find

that our CS-based upsampling model shares the similar behavior with the standard CS model because both of them solve the underdetermined linear system. Thus the factors that influence the final results of the standard CS model should also influence the resulting quality of our CS-based upsampling model. In the sequel, we will conduct experiments to verify the assumption and employ the conclusions of CS theory to discuss how to construct a satisfactory Φ for accurate recovery.

5. Sampling data generation

This section describes how to construct the measurement matrix Φ used for depth upsampling (*i.e.* how to generate the sampling data from a low-resolution depth map D_l and a registered high-resolution color image I_h) as the measurement matrix Φ is a specific matrix that should satisfy the minimum measurement requirements. In the following paragraphs, the sampling position information is denoted as a mask image M_h . If the pixel (i,j) is selected as a sampling point, $M_h(i,j) = 1$; otherwise $M_h(i,j) = 0$. The sampling values are stored in a high resolution depth map D_h . Note that D_h is used for sampling purpose only, and it is not the final output of our upsampling algorithm. The measurement matrix Φ and the measurements \mathbf{y} can be easily constructed from M_h and D_h . Without losing any generality, the upsampling factor for both horizontal and vertical directions is set as U . A pixel $(i,j) \in D_l$ corresponds to a $U \times U$ patch in the high resolution image space.

5.1. Minimal number of measurements

Both the compressive sensing problem Eq. (1) and our depth upsampling problem Eq. (6) consist in reconstructing an s -sparse vector from an underdetermined linear system. Although the sparsity assumption hopefully helps in identifying the original vector, it is unreasonable to expect that we can restore original vector from the observed measurements with arbitrary number. Indeed, there is a lower bound for the number of measurements. In other words, the number of measurements must be greater than the minimal number of measurements; otherwise, no one could identify the original vector. More importantly, the lower measurement number implies the larger upsampling rates for depth upsampling.

In literature, CS theory has proved that the mutual coherence $\mu(\Phi, \Psi)$ between the sensing matrix Φ and the representation basis Ψ determines the minimal number of measurements and $\mu(\Phi, \Psi) \in [1, \sqrt{n}]$.

$$\mu(\Psi, \Phi) = \sqrt{n} \cdot \max_{1 \leq k, j \leq n} |\langle \psi_k, \phi_j \rangle| \quad (11)$$

Then, for a fixed signal $f^n \in \mathbb{R}^n$ whose coefficients are at most S nonzero entries in the basis Ψ , the CS theory [20] guarantees that



Fig. 1. The upsampling results of Eqs. (8) and (10). (a) Is the ground truth, (b) demonstrates the 8X upsampling result of Eq. (8), (c) shows the 8X upsampling result of Eq. (10). Compared (c) with (b), we can observe that there are nasty artifacts on the depth surfaces of (b).

select m measurements in the Φ domain uniformly at random; if

$$m \geq C \cdot \mu^2(\Psi, \Phi) \cdot S \cdot \log n \quad (12)$$

for some positive constant C , the solution of compressive sensing is exact with overwhelming probability. More specifically, it is shown that the probability of success exceeds $1 - \delta$ if

$$m \geq C \cdot \mu^2(\Psi, \Phi) \cdot S \cdot \log\left(\frac{n}{\delta}\right) \quad (13)$$

At last, we note that both the lower bound and the restoration quality for our depth upsampling problem are determined by the sensing partner of the sensing matrix Φ as the representation basis Ψ are selected as the Daubechies Wavelet basis in advance.

To fight against the high mutual coherence between the measurement matrix Φ and orthogonal representation bases Ψ [25], it is natural and reasonable to randomly map the depth values of low resolution depth map D_l into the high resolution depth map D_h . Specifically speaking, in depth upsampling situation, a pixel $(i, j) \in D_l$ corresponds to a $U \times U$ patch in D_h . From this patch, we randomly select one sample with uniform distribution, and set their M_h values to 1 as follows:

$$D_h(i*U+s, j*U+t) = D_l(i, j) \quad s, t = 1, \dots, U \quad (14)$$

5.2. Our sampling method

Although the random sampling scheme could reduce the mutual coherence and therefore decreases the lower bound of measurements, the sampling scheme could not keep depth edges in the high resolution depth map as sharp as depth edges in the low resolution depth map. Instead, we manually select pixels around depth discontinuities as sampling data points and add random sampling positions in the homogeneous region to decrease the mutual coherence $\mu(\Psi, \Phi)$. As discussed previously, this procedure is essential for the accuracy of the model. Since depth borders are not known before upsampling, we will infer their positions and the corresponding measurements with the auxiliary information of the registered color photo.

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:

$$|D_l(\mathbf{p}) - D_l(\mathbf{q})| < \lambda, \quad \forall \mathbf{q} \in N(\mathbf{p}) \quad (15)$$

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. Our sampling method is quite different from Hawe et al.’s work [11]. Their method relies on intensity edges for border detection, since no depth information is available. We instead use D_l for rough region detection, and then incorporate edge information to compute M_h and D_h only in border regions, which helps to distinguish intensity edges and actual depth borders.

The procedure described above works well for small or moderate upsampling factors. However, when U reaches 16 or even larger, the generated sampling points would be too sparse in the high resolution image space to meet the minimum measurement requirements [24]. We provide a simple hierarchical solution for large upsampling factors. Large U is decomposed into a set of small factors: $U = U_1 \times U_2 \dots U_m$. 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, $U = 16$ are decomposed as 4×4 respectively, such that the number of the times m is kept as low as possible.

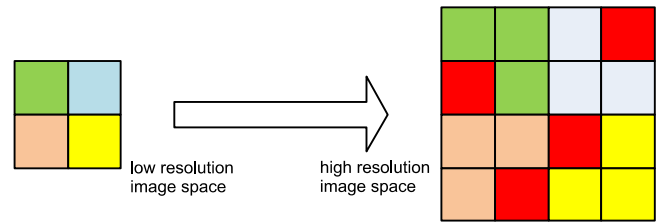


Fig. 2. An illustration of the sampling data generation process for homogeneous regions. Each pixel in the low resolution depth map (left image) corresponds to a 2×2 patch in the high resolution depth map (right image), where the pixels with the same color on either side denote the one-to-one mapping. The red color pixels in each patch are randomly chosen and we assign the corresponding depths in the low resolution depth map to them. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

5.2.1. 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:

$$D_h(i*U+s, j*U+t) = D_l(i, j) \quad s, t = 1, \dots, U \quad (16)$$

From this patch, we randomly select one or several samples with uniform distribution, and set their M_h values to 1. As stated in [25], this random selection helps to lower the mutual coherence between Ψ and Φ .

Fig. 2 shows the sampling data generation process for homogeneous regions. We can observe that each pixel in the low resolution depth map (left image) corresponds to a 2×2 patch in the high resolution depth map (right image), where the pixels with the same color on either side denote the pixel-patch mapping. In the data generation process, we randomly chose a red pixel in each patch and assign the corresponding depth in the low resolution depth map to it.

5.2.2. Sampling border regions

For border pixels in D_h , 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 [26].

Here, we provide a border region filling method based on the classic Cellular Automata (CA) [27]. CA usually works on a regular grid of cells, with finite states and local transition rules, which are suitable for many image processing applications [28]. In mathematical terms, a cellular automaton is a triplet $A = (S, N, \delta)$, where S is a non-empty state set. N is the neighborhood system, and $\delta: S^N$ Our solution is based on the CA model proposed by Vezhnevets and Konouchine [29]. Their model can propagate two labels to the full image. We employ this model for depth propagation and extend the local transition rules to respect the color distribution and the edges in I_h , such that the propagation does not generate incorrect depth boundaries. Specifically, for each pixel \mathbf{p} , our method stores a four state variable $S_{\mathbf{p}} = (D_{\mathbf{p}}, \Theta_{\mathbf{p}}, \vec{C}_{\mathbf{p}}, E_{\mathbf{p}})$, where $D_{\mathbf{p}}$ denotes the depth of pixel \mathbf{p} , $\Theta_{\mathbf{p}}$ is the ‘transition strength’ of pixel \mathbf{p} , $\vec{C}_{\mathbf{p}}$ stands for the feature vector of pixel \mathbf{p} (i.e. the three dimensional vector of pixel \mathbf{p} ’s color of I_h) and $E_{\mathbf{p}}$ records I_h color edges detected by Canny filter. Without loss of generality, we assume $\Theta_{\mathbf{p}}$ is bounded to $[0, 1]$.

Initially, we assign $\Theta_{\mathbf{p}} = 1$ for all the pixels with valid depth values, otherwise $\Theta_{\mathbf{p}} = 0$. Moreover, if \mathbf{p} lies on a color 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 manner according to the rules listed in Algorithm 1 which renews $S_{\mathbf{p}}$ from time t to $t+1$, where f is

a monotone decreasing function bounded to [0,1]:

$$f(\vec{C}_{p_1}, \vec{C}_{p_2}) = 1 - \frac{\|\vec{C}_{p_1} - \vec{C}_{p_2}\|_2}{\sqrt{3}} \quad (17)$$

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  $f(\vec{C}_{\mathbf{p}}, \vec{C}_{\mathbf{q}}) \cdot \Theta_{\mathbf{q}}^t > \Theta_{\mathbf{p}}^t$  then
9:                $D_{\mathbf{q}}^{t+1} = D_{\mathbf{p}}^t$ 
10:               $\Theta_{\mathbf{q}}^{t+1} = f(\vec{C}_{\mathbf{p}}, \vec{C}_{\mathbf{q}}) \cdot \Theta_{\mathbf{q}}^t$ 
11:           end if
12:       end for
13:   end for
14:   return  $S^{t+1}$ 

```

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 (Lines 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 color edges (i.e. the pixel $\mathbf{p} \in E = \{\mathbf{p} | E_{\mathbf{p}} = 1\}$) are all selected as sampling points. Then the subset $P - \{\mathbf{p} | E_{\mathbf{p}} = 0\}$ of the remaining pixels are randomly selected with a uniform distribution.

Using biological metaphor, the pseudocode listed above has an intuitive explanation. Since the depth values are discrete and have L different quantization levels, we can treat the depth assigning process which diffuses the depths from interpolating seeds to interpolated pixels as growth and struggle for domination of L types of bacteria. The culture media is limited to the border regions P . The bacteria start to spread from the seed pixels and try to occupy all the border regions P . We define the battlefield of different bacteria types as the pixel set E and constrain that the warriors in the battlefield should not escape from the battlefield and invade the region of other bacteria types (i.e. the if statement $E_{\mathbf{p}} = 0$ and $E_{\mathbf{q}} = 1$ in the pseudocode does). At each step, each bacteria \mathbf{p} tries to attack its neighbors $N(\mathbf{p})$. The attack force is defined by the attacker's transition strength $\Theta_{\mathbf{p}}$ and the distance between the feature vectors $\vec{C}_{\mathbf{p}_1}, \vec{C}_{\mathbf{p}_2}$ of attacker and defender. If the attack force is greater than defender's strength, the defending pixel is conquered and its depth and strength are changed. The result of these local competitions is that the strongest bacteria occupy the neighboring sites and gradually spread over the border regions P .

An illustrative example of our border regions sampling method is given in Fig. 3. As we stated above, a pixel in the low resolution depth map corresponds to a patch in the high resolution depth map according to the pixel-patch mapping used in Fig. 2. Hence, only using the depth edge information of the low resolution depth map, we could not accurately determine depth edges in the high resolution depth map. Instead, we detect depth edges in the low resolution depth map and map them into the high resolution depth map. In this way, we are able to locate the border regions which

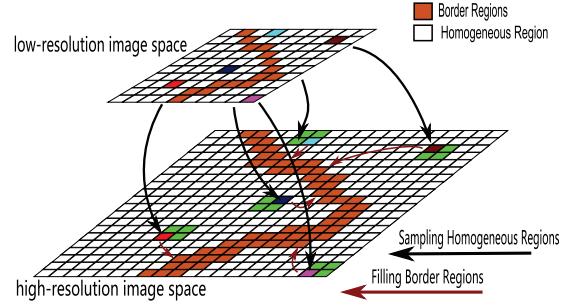


Fig. 3. An illustration of the sampling data generation process for border regions. In the sampling data generation process, we map the border region in the low resolution depth map to the high resolution depth map according to the pixel-patch mapping used in Fig. 2. At the meanwhile, the nearest depths for depth edges in the low resolution depth are also mapped into the high resolution depth map and form the nearest depths for depth edges in the high resolution depth. By choosing these depths as seeds, our CA-based border region filling algorithm is able to interpolate the depths in the border region satisfactorily.

indicate the possible location of depth edges in the high resolution depth map. Note that the depths in the depth edges of the low resolution depth map are neglect because they are not as reliable as the nearest depths in the homogeneous region to determine the depths in the border region. Therefore, we employ the nearest depths in the homogeneous region and our CA-based border region filling algorithm to fill the depths in the border regions. Fig. 3 just shows this sampling data generation process for border regions.

6. Numerical solution

In this section, we provide a first-order numerical solution for the optimization problem defined in Eq. (10). A major difficulty in minimizing Eq. (10) is that both the TV term and the sparseness term are non-differential l_1 regularizes. 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 [30].

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

$$\begin{aligned} \min_{\mathbf{u}, \mathbf{v}} \quad & \alpha \|\mathbf{u}\|_{TV} + \beta \|\mathbf{v}\|_1 + \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{d}\|_2^2 \\ \text{s.t.} \quad & \mathbf{u} = \mathbf{d}, \mathbf{v} = \Psi^T \mathbf{d} \end{aligned} \quad (18)$$

Problem (18) is hard to solve. We use following unconstrained optimization Problem (19), which includes two quadratic penalty terms to Problem (18), to approximate the original problem.

$$\min_{\mathbf{u}, \mathbf{v}} \alpha \|\mathbf{u}\|_{TV} + \frac{\alpha \gamma}{2} \|\mathbf{u} - \mathbf{d}\|^2 + \beta \|\mathbf{v}\|_1 + \frac{\beta \delta}{2} \|\mathbf{v} - \Psi^T \mathbf{d}\|^2 + \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{d}\|_2^2 \quad (19)$$

where parameter γ and δ control the approximation penalty for \mathbf{u} and \mathbf{v} respectively.

Problem (19) can be solved in an alternating minimization framework as follows:

1. For fixed \mathbf{v}, \mathbf{d} , solve the subproblem for \mathbf{u} :

$$\min_{\mathbf{u}} \|\mathbf{u}\|_{TV} + \frac{\gamma}{2} \|\mathbf{u} - \mathbf{d}\|^2 \quad (20)$$

This is a typical anisotropic TV problem, which can be efficiently solved with a Split Bregman algorithm from [23].

2. For fixed \mathbf{u}, \mathbf{d} , solve the subproblem for \mathbf{v} :

$$\min_{\mathbf{v}} \|\mathbf{v}\|_1 + \frac{\delta}{2} \|\mathbf{v} - \Psi^T \mathbf{d}\|^2 \quad (21)$$

The problem is well studied in CS literature. Using the simple one-dimensional shrinkage operator, we can directly write down its analytic solution.

$$\mathbf{v} = \max\left(\Psi^T \mathbf{d} - \frac{1}{\delta}, 0\right) \text{sgn}(\Psi^T \mathbf{d}) \quad (22)$$

3. Finally, for fixed \mathbf{u}, \mathbf{v} , solve the subproblem for \mathbf{d} :

$$\min_{\mathbf{d}} \frac{\alpha\gamma}{2} \|\mathbf{u} - \mathbf{d}\|^2 + \frac{\beta\delta}{2} \|\mathbf{v} - \Psi^T \mathbf{d}\|^2 + \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{d}\|_2^2 \quad (23)$$

This least square problem promises a closed-form solution:

$$\mathbf{d} = K(\alpha\gamma \mathbf{u} + \beta\Psi^T \mathbf{v} + \mathbf{y}) \quad (24)$$

where $K = (\alpha\gamma I + \beta\Psi^T \Phi)^{-1}$ is a diagonal matrix.

Steps 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.

7. Experiments

In this section, we first describe a preprocessing step to register the depth camera and conventional camera as the procedure is an essential step for following experiments. Second, the experiments' parameter configuration and the evaluation index are presented. After that, we compare the upsampling results of different sensing patterns and discuss the functions of different terms in Eq. (10). Last but not least, we conduct extensive experiments, including the synthesized data and the real world data, to illustrate the upsampling ability of our method.

7.1. Depth map registration

Depth maps and their companion color images are captured by different cameras, thus they do not registered well in reality. Let $\mathbf{X}_d = (X, Y, Z, 1)$ denote the 3D homogeneous coordinates of the pixels of a depth map, and $\mathbf{X}_c = (r, c, 1)$ represent the 2D homogeneous coordinates of the companion high-resolution RGB image of the depth map. We have the following projection relationship about \mathbf{X}_d and \mathbf{X}_c :

$$\mathbf{X}_c = sK[R \mid t]\mathbf{X}_d \quad (25)$$

where s is a scale factor, K is the intrinsic parameters of the optical camera, R and t are the rotation and translation matrix which describe the rotation and translation of the optical camera and the depth camera.

We can use the well-known calibration method introduced by Zhang [31] to calibrate the parameters of the two cameras. However, the depth camera could not capture textures as the RGB camera does. Instead of using the traditional visible textures, we can use a planar calibration pattern which consists of holes for the reason that these holes can be captured by both cameras. A similar method has been used by Park et al. [9] to register depth maps and color images too. For the two calibrated cameras, we can project any points \mathbf{X}_d which are on the low-resolution depth map onto the high-resolution RGB image by using Eq. (25). In this situation, the scaling term s is the relative resolution between the depth camera and the optical camera, in other words, it is the upsampling rate. For any point x_t which is on the low-resolution depth map with depth value d_t , we can transform the local coordinate of the depth camera $[x_t, d_t, 1]^T$ into the world coordinate \mathbf{X}_d by the following equation:

$$\mathbf{X}_d = P_t^{-1}[x_t \ d_t \ 1]^T \quad (26)$$

where P_t^{-1} is the inverse of the 4×4 projective transformation P_t which converts the world coordinate \mathbf{X}_d into the local coordinate $[x_t \ d_t \ 1]^T$ of the depth camera.

7.2. Parameter configuration and evaluation index

In the sequent experiments, we quantitatively test our algorithm on the Middlebury stereo datasets [32], which provide both high resolution color images and ground truth depth maps, as well as the well-known KITTI Vision Benchmark Suite. Specifically, for the synthesized experiments illustrated in Section 7.4, we use 'Books', 'Dolls', 'Moebius' and 'Plastic' images of the Middlebury datasets whereas we randomly choose three depth images from the KITTI Vision Benchmark Suite to perform real data upsampling. The test platform is a PC with Intel i5 2.8GHz CPU and 4GB memory. The upsampling performance of our method is rather stable when $\lambda, \alpha, \beta, \gamma, \delta$ are in the ranges [2, 5], [0.5, 2], [0.5, 1.5], [25, 35], [25, 35], respectively. For convenience, we consistently keep the parameter setting $\lambda, \alpha, \beta, \gamma, \delta$ of all the data sets used in the experiments unchanged, where $\lambda = 4.0, \alpha = 1.0, \beta = 1.0, \gamma = 32.0, \delta = 32.0$. The upsampling factor U , in our experiments, varies from 2X to 16X, 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.

For performance evaluation, we need to explain the meaning of getting a satisfactory result with a high probability because CS theory only guarantees that we have a high probability (refer to Eqs. (12) and (13)) to obtain an exact result. Performing hundred experiments for an image and accounting the number of satisfactory results is not a rational method to evaluate the performance of our method. Instead of using this direct method, we evaluate the PSNR index of an upsampling result, which quantifies the quality of a result. We measure the quality of the upsampled results with peak signal-to-noise ratio, often abbreviated PSNR, which is used to measure the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation [33]. If our method gets a satisfactory result with a high probability, the upsampling result tends to have a large PSNR index, otherwise, it should have a small PSNR index at most of the time.

7.3. Sampling pattern

A common doubt about our upsampling method is whether the CS theory can be applied to our depth upsampling model as an additional TV term is added to the tradition CS restoration model (Eq. (7)). Frankly speaking, we do not offer a mathematical analysis for the model as the traditional CS theory does and we think that a detailed mathematical analysis is out of the scope of this paper as we only target at providing a novel and practical depth upsampling method. However, the experimental results in the sequel support the idea of that our depth upsampling model behaves as the traditional CS restoration model (Eq. (7)) does.

A major difference between CS model and the traditional interpolation methods is whether the quality of final results is heavily affected by the sampling pattern of observed data. To the best of our knowledge, the sampling pattern of observed data plays a minor role in the interpolation quality for the traditional interpolation methods. In other words, the final results only depend on the number of interpolating pixels. On the contrary, the observed data's sampling pattern completely determines the final results of CS model. *i.e.* for some sampling patterns, the CS model could satisfactorily restore the final result from the observed data. For the other sampling patterns, the CS model fails to complete this task, even the number of observed depths is the same in both situations. The reason is that, as we stated in Section 5.1, the minimal number m of measurements is determined by the mutual coherence $\mu^2(\Psi, \Phi)$, *i.e.*

$$m \geq C \cdot \mu^2(\Psi, \Phi) \cdot S \cdot \log n \quad (27)$$

For uniform grid distribution, $\mu(\Psi, \Phi) = 6$. Whereas, in the random sampling distribution situation, $\mu(\Psi, \Phi) = 2.5$ which is smaller than the mutual coherence value of the uniform grid distribution.

Fig. 4(c) and (d) illustrates the phenomenon for 8X upsampling, where we take two different sampling patterns (or distribution) with the same sampling number. One distribution is the uniform grid, which is usually employed by the traditional interpolation methods. The other one is the random distribution generated by Eq. (14). We can observe that our CS based model is unable to interpolate the missing depths in the edges boundary for the uniform sampling pattern and succeeds to recover the depths for random distribution. The behavior coincides with the prediction of the CS theory. Random distribution is still an inappropriate sampling pattern as it could not keep the sharp depth edges. The shortcoming can be observed in Fig. 4(c). As a remedy, in Section 5.2, we proposed a novel hybrid sampling pattern which randomly takes sample in homogeneous regions and deterministically the depths in border regions. The results is demonstrated in Fig. 4(d). Compared with previous two sampling patterns which are the deterministic sampling pattern (*i.e.* uniform grid) and random distribution (*i.e.* random sampling) respectively, the restoration result Fig. 4(e) of our hybrid sampling pattern is the best among the three sampling patterns.

7.4. Evaluations using the Middlebury stereo datasets

Our algorithm (denoted as CS1) was implemented with Matlab. For comparison, four methods are selected: the bilateral-filtering

based method (denoted as Bilateral) [6], two MRF-based methods (denoted as MRF1, MRF2) [4,9] and the original CS-based method (denoted as CS2) [11]. For the first two methods, we implemented them with Matlab, and the parameters were finely tuned. For CS2 and MRF2, we directly use the source code provided by the authors. One thing needed to be clarified is that CS2 was not designed for the upsampling problem. 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 Fig. 5. Our algorithm works better under large upsampling factors. It consistently outperforms other methods with 2X, 4X, 8X and 16X upsampling. MRF2 also gives overall satisfactory results in most cases, while Bilateral and MRF 1 performs well under low upsampling rates. Compared with the sampling method used in CS2, our sampling data generation method plays an important role in high quality CS-based upsampling.

For the qualitative comparison of PSNR, we present some 8X upsampled results computed by CS1, Bilateral and MRF2 methods in Fig. 6. It can be seen that our method preserves sharp and accurate depth boundaries during the upsampling process, which demonstrates the effects of the l_1 regularization terms. We also own this achievement to our sampling method because our method could generate accurate samples on boundaries with CA-based region filling method and the produced boundaries coincide with the color edges of the high-resolution reference RGB photo.

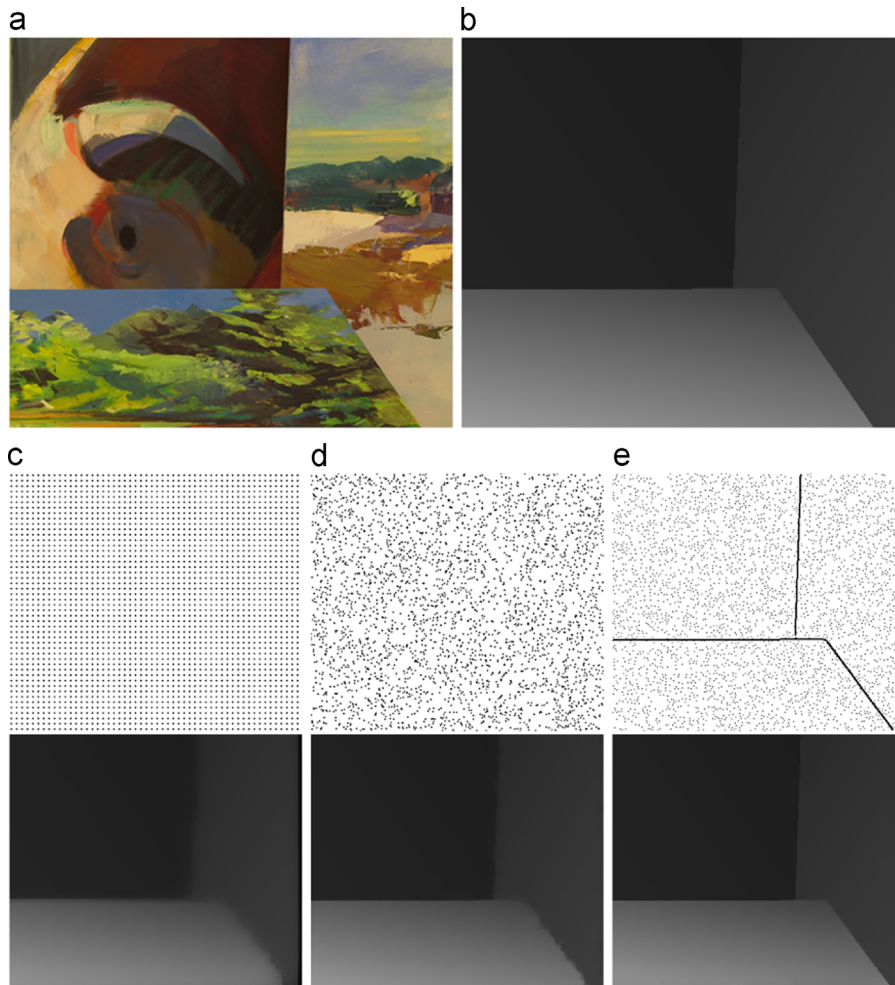


Fig. 4. Sampling Pattern for 8X upsampling. (a) and (b) exhibit the guidance image and ground truth respectively, (c) demonstrates the uniform sampling, (d) illustrates the random sampling, and (e) shows our sampling pattern. The sampling number is the same for the three distributions. We can observe that the restoration quality is very different.

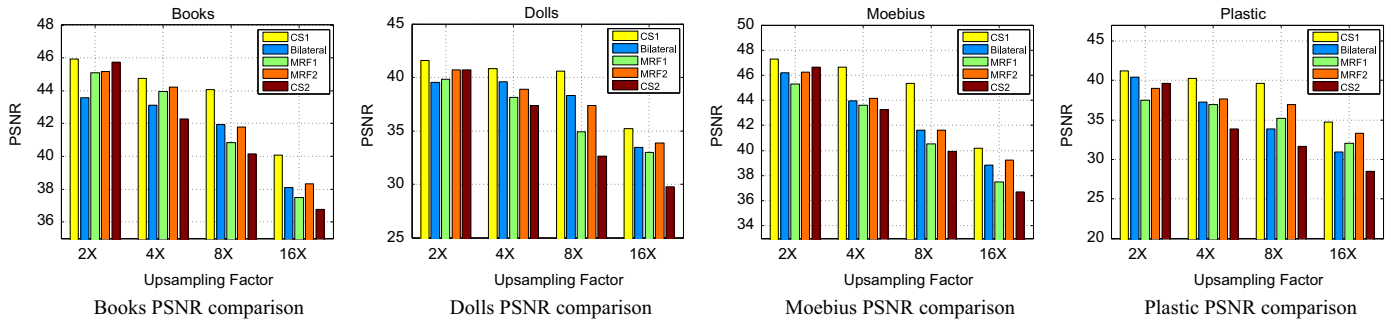


Fig. 5. Upsampling PSNR results with ideal depth measurements. Our method (CS1) consistently outperforms other methods with $4 \times$, $8 \times$ and $16 \times$ upsampling. (a) Books PSNR comparison, (b) Dolls PSNR comparison, (c) Moebius PSNR comparison and (d) Plastic PSNR comparison.

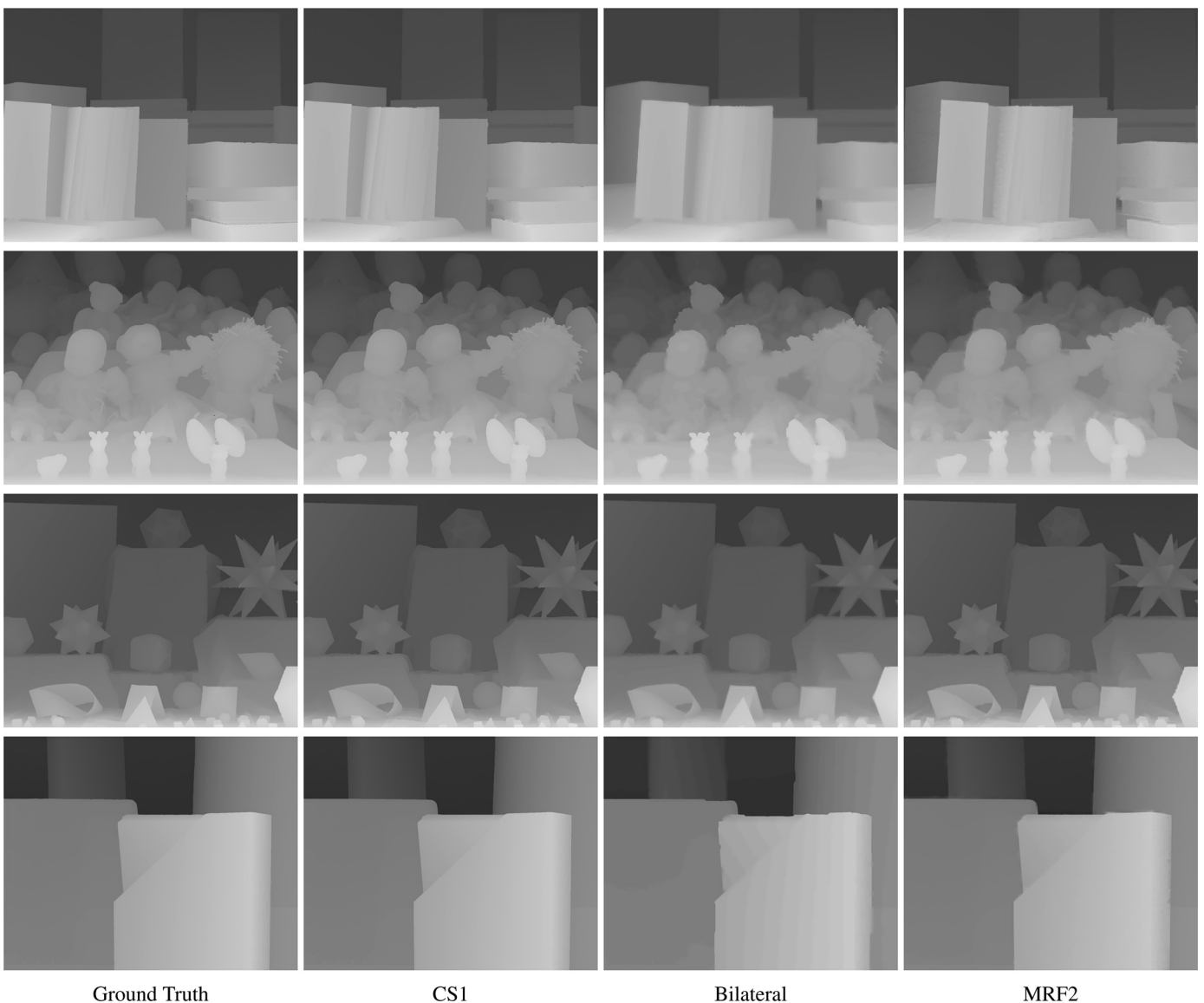


Fig. 6. $8 \times$ upsampled depth maps for 'Books' 'Dolls' 'Moebius' and 'Plastic' data sets. The depth results are computed with CS1, Bilateral and MRF2 respectively.

The experimental results could prove our conclusions in the above sections.

We then test the algorithm with noisy measurements. The noise characteristics in practical range sensors usually depend on the distance between the sensor and the scene. To simulate this effect, we employ a conditional Gaussian model from [9] for noise generation.

The amplitude of the noise varies for each pixel, which is set to be $\pm 20\%$ of the measured depth. The PSNR results for the five methods are presented in Fig. 7. Our method is robust against noise with little accuracy loss, and it still outperforms other methods in most cases. The performance of MRF2 and Bilateral drop dramatically due to the noise. This can also be verified in Fig. 8, which provides some $8X$

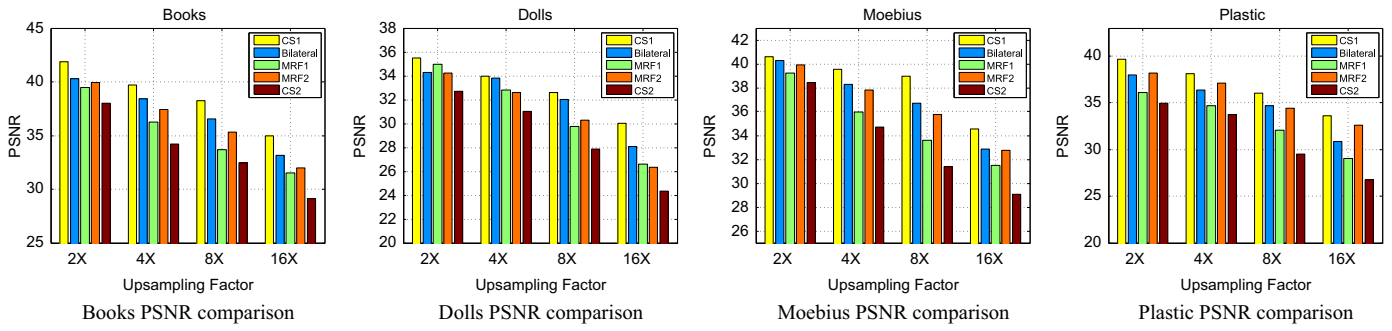


Fig. 7. Upsampling PSNR results with noisy measurements. Our method (CS1) still outperforms other methods in most cases. It shows robust behavior in noisy conditions.

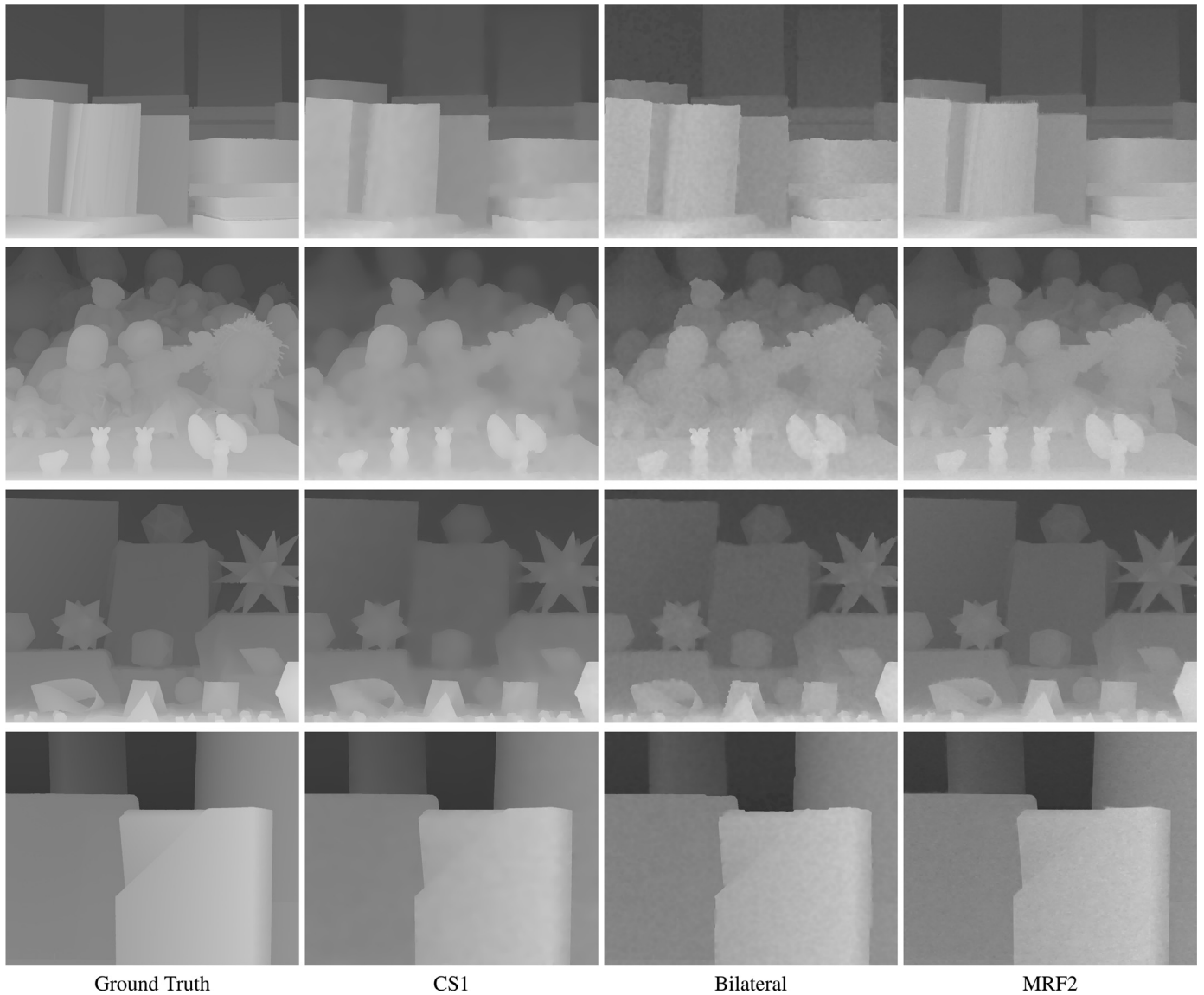


Fig. 8. $8 \times$ upsampled depth maps with noisy measurements for 'Books' 'Dolls' 'Moebius' and 'Plastic' data sets. The depth results are computed with CS1, Bilateral and MRF2 respectively.

upsampled results computed by CS1, MRF2 and Bilateral methods respectively.

Our method has a great set of advantages. On the border regions, our cellular automata sampling method can produce sharp and accurate edges. Therefore, we could obtain edge-preserving results. On the homogeneous regions, our sampling method could reduce

the mutual coherence between the measurement matrix Φ and the unitary matrix Ψ by random sampling. So our CS-based upsampling algorithm outperforms other methods. Although it is abnormal that the performance of our method is nearly same for lower upsampling factors such as 2X, 4X, 8X, the behavior is reasonable according to CS theory and the behavior does not imply fewer samples can keep

reconstruction accuracy because the accuracy of our method decrease drastically at 16X upsampling factor and this tendency will be kept for larger U . For all the data sets, our algorithm produces the final results in 100 seconds, which is almost the same to the running time of CS2, but a bit slower than the MRF-based methods.

7.5. The performance characteristic of our method

Unlike traditional upsampling methods whose performances consistently decrease while U increasing, the accuracy of our method is not sensitive to the small upsampling rate U . Indeed, it tends to keep the accuracy unchanged at the lowest upsampling rate. The inflection point usually is $U=16$ for the Middlebury stereo datasets. It is an interesting feature which is unique to our method. In fact, for most data sets, the best results are achieved at low upsampling rates. For 'Books' and 'Dolls', the PSNR results for 2X and 8X are still quite close. In this section, we give an explanation for this feature. Since we divide the domain of depth map into homogeneous region and border region and take different sampling patterns, we discriminate the upsampling behavior at different areas and provide corresponding explanation.

For border region upsampling, we use cellular automata to detect the color edges of the guidance image and map them onto the high resolution depth map. Undoubtedly, the accuracy of detection will decrease while the upsampling rate becomes large. However, the error is not proportional to upsampling rate. From Table 1, we can observe that the PSNR indices of border regions are kept unchanged for small upsampling rates. Only for 16X upsampling, the PSNR indices have a sharp reduction. The reason is that for low upsampling rate, the cellular automata could detect the depth edges accurately under the guidance of the registered color photo. For 16X upsampling rate, the distances between seeds are very large and there are many color edges between them. Thus our cellular automata algorithm fails to decide which one is the correct depth edge.

For homogeneous region upsampling, we randomly map the low resolution depth map into the high resolution depth map and then employ our CS based upsampling method to interpolate the missing depths. Specifically, we can only offer an heuristic explanation for the upsampling behavior of homogeneous regions because our model is out of the scope of traditional CS theory and we do not offer an in-deep mathematical analysis for our CS-based upsampling model as traditional CS theory does for the standard CS model. However, in the experiments, we find that the mutual coherence $\mu(\Psi, \Phi)$ proposed for disclosing the reconstruction condition of traditional CS theory is also a good evaluation index for our CS-based upsampling model. Specifically, CS theory [20] guarantees that fix a signal $f^n \in \mathbb{R}^n$ whose coefficients are at most S nonzero entries in the basis Ψ ; select m measurements in the Φ domain uniformly at random; if

$$m \geq C \cdot \mu^2(\Psi, \Phi) \cdot S \cdot \log\left(\frac{n}{\delta}\right) \quad (28)$$

for some positive constant C , the probability of successfully obtain an exact solution to (18) exceeds $1-\delta$. Note that CS theory does not ensure that we always get a satisfactory result; it only guarantees that the probability of getting a satisfactory result is $1-\delta$ with respect to m measurements. According to our sampling method, the mutual coherence $\mu(\Psi, \Phi)$ tends to decrease with the increasing upsampling factor U because large U value brings more randomness to the selection of the sampling positions, which helps to lower the mutual coherence between the measurement matrix and the representation matrix. Moreover, the smaller the coherence is, the fewer samples are needed. However, compared with the mutual coherence $\mu(\Psi, \Phi)$, the number m of sampling pixels (or measurements) decreases drastically along with the increasing upsampling rate U . Therefore, when $U=16$, the number m of measurements tends to lower than the expected sampling number which guarantees that we can obtain a satisfactory result with a high probability $1-\delta$. Although we can obtain the most samples when $U=2$, $\mu(\Psi, \Phi)$ also becomes larger in this case. In contrast, samples become fewer, but,

Table 1

The quality evaluation indices at different regions with four upsampling rates. The table illustrates the PSNR indices of the upsampling results of Books, Dolls, Moebius and Plastic. We can observe that the PSNR indices of homogenous regions and border region are rather stable for 2X, 4X and 8X upsampling rates.

	Books		Dolls		Moebius		Plastic	
	Homo region	Border region	Homo region	Border region	Homo region	Border region	Homo region	Border region
2X	53.8519	33.1011	54.3603	31.3568	56.7017	38.2858	55.8319	35.6013
4X	51.1759	31.9870	53.3611	29.5401	55.7788	37.2668	54.3312	34.9886
8X	49.8089	30.3277	50.6128	28.0730	53.6832	34.8609	53.6831	33.6051
16X	45.5557	21.4138	41.4395	24.6849	47.0020	32.7195	50.6382	21.4138

Table 2

The quality evaluation for different parts of our algorithm. The table illustrates the PSNR indices of the upsampling results of Books, Dolls, Moebius and Plastic. We can observe that the PSNR indices of homogenous regions and border region are rather stable for 2X, 4X and 8X upsampling rates. In the missing TV rows, we remove the TV term from our objective function and evaluate the quality of upsampling results. In the missing CA rows, we replace our CA-based border region filling algorithm by the uniform sampling and calculate the PSNR indices of upsampling results.

	Homo region	Border region	Homo region	Border region	Homo region	Border region	Homo region	Border region
Missing TV								
2X	53.0372	33.1011	53.7279	31.0730	55.7442	38.0812	55.4101	34.9456
4X	50.2701	31.7098	52.2636	29.1045	54.8139	37.0478	53.4155	34.2395
8X	49.6819	30.2011	50.3343	27.7391	53.1256	34.4606	52.8909	33.1167
16X	44.6423	21.1348	40.8926	24.3241	46.3314	32.1776	49.6787	20.9351
Missing CA								
2X	53.8519	26.5463	54.3603	25.9694	56.7017	33.3882	55.8319	29.0503
4X	51.1759	24.9265	53.3611	23.7245	55.7788	32.8109	54.3312	26.3912
8X	49.8089	22.0931	50.6128	20.4178	53.6832	28.3978	53.6831	23.2012
16X	45.5557	14.4655	41.4395	13.7329	47.0020	25.6259	50.6382	17.9143

the minimal number of measurements also small. As a result, the probability $1 - \delta$ of getting a satisfactory result is not changed very much for a wide range of upsampling.

It is also interesting to discuss the affect of the TV regularization and our CA-based border region filling algorithm for depth upsampling because they are two novel terms which are introduced to the classical compressive sensing model by us. In order to disclose the affect of them, we remove the TV regularization from our sampling model and evaluate the quality of upsampling results at first. After that, we substitute our CA-based border region filling algorithm with the uniform sampling illustrated in Fig. 4(c) to reveal the function of our CA-based border region filling algorithm. The quantitative results are reported in Table 2. Compared to Table 1, we can easily find that the TV regularization term mainly affects the upsampling quality of homogeneous regions. This is because the PSNR indices are significantly decreased in contrast to the PSNR indices in the border region. Different from the TV regularization term, our CA-based border region filling algorithm only affects the quality of upsampling results in the border region. The reason is that the PSNR indices of homogeneous regions in the missing CA rows are same to the data in Table 1. If we compare the PSNR indices of the border region in the missing TV rows to PSNR indices in the missing CA rows, we can also observe that the CA-based border region filling algorithm heavily affects the final upsampling results. However, it is not reasonable to say that the CA-based border region filling algorithm is more important than the TV regularization because the TV regularization is only designed to smooth the depth values in the homogeneous region and the CA-based border region filling algorithm is only used to keep the sharp edges in the border region. In our algorithm, the two different terms are jointly exploited to improve the upsampling quality in the homogeneous region and border region simultaneously.

7.6. Real world experiments

Lacking of 3D laser scanner, we use the standard database from KITTI Vision Benchmark Suite [34]. All of the data is acquired by a standard station wagon with two color and grayscale video cameras. The accurate ground truth is provided by a Velodyne laser scanner. We take advantage of the laser scanner and companied color image to perform our experiments.

The KITTI database aims to offer people a challenging real-world computer vision benchmarks for autonomous driving. However, the database is not in preparation for stereo evaluation as Scharstein [35] did. It is a big challenge to use this KITTI Vision database for upsampling experiments, because all of the data is captured from Karlsruhe streets and the obtained scenes are usually very large, thus the object structures are relatively small in the pictures. Compared with the artificial in-door pictures used in state-of-the-art methods [4,6,9,11], it suffers from varied sensing noise. This situation demands strong denosing ability of upsampling algorithm.

The depth data acquired by the 3D laser scanner usually irregularly distribute on the high resolution guidance color image domain, thus we need to regulate the laser points and produce a low resolution depth map as the input of our upsampling algorithm. In Section 5, we assume that a pixel $p \in D_l$ corresponds to a $U \times U$ patch in the high resolution image space. Then, if there are laser points in the $U \times U$ patch, we can randomly select a point and assign its value to $p \in D_l$ because these points concentrate in a small region and should have similar depth values. Indeed, only using the depth assignment strategy, we could not guarantee that each $p \in D_l$ has a defined depth value because the depth data acquired by the laser scanner usually has holes. In order to fulfill the holes in the low resolution depth map D_l , we down sample the high resolution color image to obtain a small guidance image whose size is the same as D_l and use our cellular automata algorithm to fulfill the missing values by viewing the holes as edges.

The upsampling technique which can significantly reduce the demand measurement number of captured depth data is illustrated in Fig. 9. In the experiment, we only perform 6X super-resolution because the size of groundtruth is 1390×1110 . For 8X upsampling, the size of downsampled image is only 100×28 . Obviously, the dimension is too small and the downsampled image loses too much information. Moreover, we catch that our algorithm can keep the sharp boundary of depth map and the method could reliably restore the geometric relationship.

8. Conclusion

We had presented a new method for depth map upsampling. Based on the theory of compressive sensing, our method converts

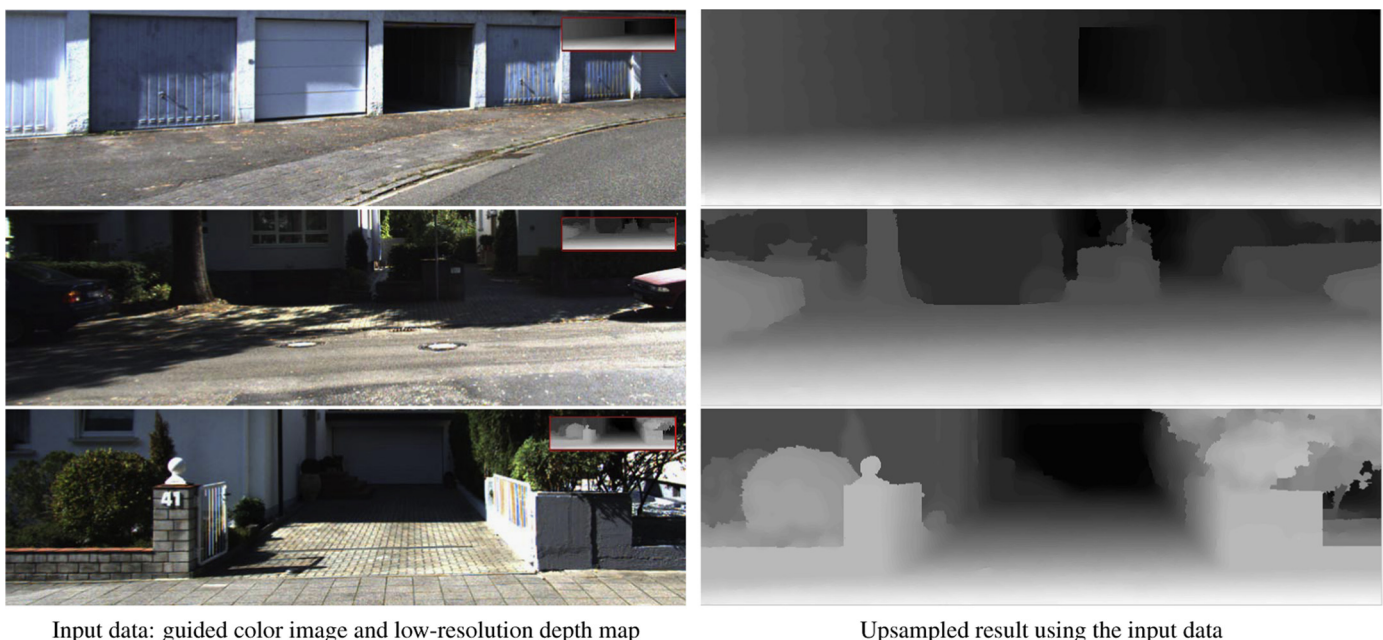


Fig. 9. 6X upsampling results of our method for real world data. The depth-color pairs are shown at their original ratio of size. (a) Input data: guided color image and low-resolution depth map and (b) upsampling result using the input data.

the low resolution depth maps into a set of measurements, and then formulates the upsampling task as a constrained optimization problem with data, smoothness and represent sparseness constraints. We validated our method with the Middlebury data sets, demonstrating that our method clearly outperforms previous methods under large upsampling factors and noisy inputs.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Nos. 61332017, 61331018, 91338202, and 61271430). The authors would like to thank respected anonymous reviewers for their constructive and valuable suggestions for improving the overall quality of this paper.

References

- [1] Y. Lei, M. Bennamoun, M. Hayat, Y. Guo, An efficient 3D face recognition approach using local geometrical signatures, *Pattern Recognit.* 47 (2) (2014) 509–524.
- [2] E. Rodolà, A. Albarelli, F. Bergamasco, A. Torsello, A scale independent selection process for 3D object recognition in cluttered scenes, *Int. J. Comput. Vis.* 102 (1–3) (2013) 129–145, ISSN 0920-5691.
- [3] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, J. Wan, Rotational projection statistics for 3D local surface description and object recognition, *Int. J. Comput. Vis.* 105 (1) (2013) 63–86, ISSN 0920-5691.
- [4] J. Diebel, S. Thrun, An application of Markov random fields to range sensing, in: *Proceedings of Conference on Neural Information Processing Systems*, MIT Press, Cambridge, MA, 2005.
- [5] J. Kopf, M.F. Cohen, D. Lischinski, M. Uyttendaele, Joint bilateral upsampling, *ACM Transa. Graph.* 26 (3), ISSN 0730-0301.
- [6] Q. Yang, R. Yang, J. Davis, D. Nister, Spatial-depth super resolution for range images, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, ISSN 1063-6919, 1–8.
- [7] D. Chan, H. Buisman, C. Theobalt, S. Thrun, A noise aware filter for real-time depth upsampling, in: *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, Marseille, France, 2008.
- [8] B. Huhle, T. Schairer, P. Jenke, W. Strasser, Fusion of range and color images for denoising and resolution enhancement with a non-local filter, *Comput. Vis. Image Underst.* 114 (12) (2010) 1336–1345, ISSN 1077-3142.
- [9] J. Park, H. Kim, Y. Tai, M.S. Brown, I. Kweon, High quality depth map upsampling for 3D-TOF cameras, in: *Proceedings of the 2011 International Conference on Computer Vision*, IEEE Computer Society, Washington, DC, USA, 2011, ISBN 978-1-4577-1101-5, 1623–1630.
- [10] I. Tošić, B.A. Olshausen, B.J. Culpepper, Learning sparse representations of depth, *IEEE J. Sel. Top. Signal Process.* 5 (2011) 941–952.
- [11] S. Hawe, K.D.M. Kleinsteuber, Dense Disparity Maps from Sparse Disparity Measurements, *IEEE Computer Society*, Los Alamitos, CA, USA, ISBN 978-1-4577-1101-5, 2126–2133, 2011.
- [12] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory* 52 (2) (2006) 489–509, ISSN 0018-9448.
- [13] D. Donoho, Compressive sensing, *IEEE Trans. Inf. Theory* 52 (4) (2006) 1289–1306, ISSN 0018-9448.
- [14] C. Tomasi, R. Manduchi, Bilateral Filtering for Gray and Color Images, in: *Proceedings of the Sixth International Conference on Computer Vision*, IEEE Computer Society, Washington, DC, USA, vol. 839, 1998, ISBN 81-7319-221-9.
- [15] A. Buades, B. Coll, J.M. Morel, in: *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, vol. 2, 2005, pp. 60–65, ISSN 1063-6919.
- [16] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vis. Res.* 37 (1997) 3311–3325, ISSN 0042-6989.
- [17] M. Figueiredo, R. Nowak, S. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, *IEEE J. Sel. Top. Signal Process.* 1 (4) (2007) 586–597, ISSN 1932-4553.
- [18] E.T. Hale, W. Yin, Y. Zhang, Fixed-point continuation for ℓ_1 -minimization: methodology and convergence, *SIAM J. Optim.* 19 (3) (2008) 1107–1130, ISSN 1052-6234.
- [19] Y. Wang, W. Yin, Sparse signal reconstruction via iterative support detection, *SIAM J. Imaging Sci.* 3 (3) (2010) 462–491, ISSN 1936-4954.
- [20] E. Candès, M. Wakin, An introduction to compressive sampling, *IEEE Signal Process. Mag.* 25 (2) (2008) 21–30, ISSN 1053-5888.
- [21] S. Muthukrishnan, *Data streams: Algorithms and Applications*, Found. Trends in Theoretical Computer Science, vol. 1, Now Publishers Inc., Hanover, MA, USA, 2005.
- [22] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, *SIAM Rev.* 43 (1) (2001) 129–159, ISSN 0036-1445.
- [23] T. Goldstein, S. Osher, The split Bregman method for L1-regularized problems, *SIAM J. Imaging Sci.* 2 (2) (2009) 323–343, ISSN 1936-4954.
- [24] E. Candès, J. Romberg, Sparsity and incoherence in compressive sampling, *Inverse Probl.* 23 (2007) 969–985, ISSN 0266-5611.
- [25] D.L. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, *IEEE Trans. Inf. Theory* 47 (7) (2006) 2845–2862, ISSN 0018-9448.
- [26] L. Wang, H. Jin, R. Yang, M. Gong, Stereoscopic inpainting: joint color and depth completion from stereo images, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8, ISSN 1063-6919.
- [27] J.V. Neumann, *Theory of Self-Reproducing Automata*, University of Illinois Press, Champaign, IL, USA, 1966.
- [28] D. Popovici, A. Popovici, Cellular automata in image processing, in: *15th International Symposium on Mathematical Theory of Networks and Systems*, 2002.
- [29] V. Vezhnevets, V. Konouchine, GrowCut: interactive multi-label N-D image segmentation by cellular automata, in: *GraphiCon*, 2005.
- [30] Y. Wang, J. Yang, W. Yin, Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, *SIAM J. Imaging Sci.* 1 (3) (2008) 248–272, ISSN 1936-4954.
- [31] Z. Zhang, A flexible new technique for camera calibration, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (11) (2000) 1330–1334, ISSN 0162-8828.
- [32] H. Hirschmüller, D. Scharstein, Evaluation of cost functions for stereo matching, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, 1–8, ISSN 1063-6919.
- [33] Q. Huynh-Thu, M. Ghanbari, Scope of validity of PSNR in image/video quality assessment, *Electron. Lett.* 44 (13) (2008) 800–801, ISSN 0013-5194.
- [34] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? The KITTI vision benchmark suite, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361, ISSN 1063-6919.
- [35] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *Int. J. Comput. Vis.* 47 (1–3) (2002) 7–42, ISSN 0920-5691.



Longquan Dai received his B.S. degree in Electronic Engineering from Henan University of Technology, China, in 2006. He received his M.S. degree in Electronic Engineering from Shantou University, China, in 2010. Currently, he is working toward the Ph.D. degree in Computer Science at Institute of Automation, Chinese Academy of Sciences, China. His research interests lie in computer graphics, computer vision and optimization-based techniques for image analysis and synthesis.



Haoxing Wang is a Ph.D. candidate in the Sino-French Laboratory (LIAMA) and National Laboratory of Pattern Recognition (NLPR) at Institute of Automation, Chinese Academy of Sciences. He received his B.S. and M.S. degrees in geographic information system from Wuhan University of Technology, China in 2006, and Beijing Normal University, China, in 2010, respectively. His research interests include photogrammetry, structure from motion techniques, optimization-based techniques for image analysis and synthesis.



Xiaopeng Zhang received his M.Sc. degree in Mathematics from Northwest University in 1987, and the Ph.D. degree in Computer Science from Institute of Software, Chinese Academy of Sciences (CAS), in 1999. He is a Professor in the Sino-French Laboratory (LIAMA) and (National Laboratory of Pattern Recognition) at Institute of Automation, CAS. His main research interests are computer graphics and pattern recognition. He received the National Scientific and Technological Progress Prize (Second Class) in 2004. Xiaopeng Zhang is a member of ACM and IEEE.