

Guided Depth Upsampling via A Cosparse Analysis Model

Xiaojin Gong, Jianqiang Ren, Baisheng Lai, Chaohua Yan
Dept. of Information Science and Electronic Engineering
Zhejiang University
Hangzhou, Zhejiang, China
{gongxj, rjq, laibs, 3110101609}@zju.edu.cn

Hui Qian
College of Computer Science
Zhejiang University
Hangzhou, Zhejiang, China
qianhui@zju.edu.cn

Abstract—This paper proposes a new approach to upsample depth maps when aligned high-resolution color images are given. Such a task is referred to as guided depth upsampling in our work. We formulate this problem based on the recently developed sparse representation analysis models. More specifically, we exploit the cosparsity of analytic analysis operators performed on a depth map, together with data fidelity and color guided smoothness constraints for upsampling. The formulated problem is solved by the greedy analysis pursuit algorithm. Since our approach relies on the analytic operators such as the Wavelet transforms and the finite difference operators, it does not require any training data but a single depth-color image pair. A variety of experiments have been conducted on both synthetic and real data. Experimental results demonstrate that our approach outperforms the specialized state-of-the-art algorithms.

Keywords-Guided depth upsampling; cosparse analysis model; multi-modal data fusion

I. INTRODUCTION

Guided depth upsampling in this work refers to up-sampling a depth map while an aligned high-resolution image is taken as guidance. Such a task is desirable in various computer vision applications. For instance, in traditional stereo vision, some techniques [1] reconstruct a dense disparity map by upsampling sparse measurements that are reliably obtained from point-wise correspondence matching. These techniques circumvent matching ambiguities occurred in homogenous or repetitively textured regions, so that reconstruction quality is improved. Another line of exemplary applications surges with the advent of active range sensing technologies. State-of-the-art ranging sensors, such as Velodyne HDL lidars [2], Time-of-Flight (ToF) cameras [3] or Microsoft Kinects [4], are capable of producing high quality range information in real time. However, depth maps obtained by them are still low in resolution, especially compared to high-resolution visual images. Therefore, the ranging sensors are commonly used in conjunction with conventional cameras, intriguing a group of studies on guided depth enhancement [5], [6], [7].

As introduced above, guided depth upsampling aims to generate high-resolution depth maps by integrating sparse range data with visual information. Therefore, it belongs to a multi-modal data fusion problem. This problem is conducted

relying on an observation that depth discontinuities often co-occur with color or intensity changes [5]. A variety of approaches have been developed so far to exploit such dependencies in order to enhance depth maps. Roughly speaking, early methods are mainly based on filtering or Markov Random Field (MRF) techniques. The former uses joint bilateral filters and their variations [8] [6] to integrate color information for depth enhancement. MRF-based methods [5], [9], [10], [11] infer depth via optimizing an energy function that consists of two or more terms: one term evaluates depth consistency with known sparse measurements and the others regularize first- or higher-order depth smoothness according to color information.

Meanwhile, since depth upsampling is closely related to intensity image super-resolution, sparse representation (SR) techniques [12] that are prevalent in super-resolution have also been employed. For example, Li *et al.* [13] jointly train dictionaries for depth and color image patches and reconstruct high-resolution depth maps in terms of SR of learned dictionaries. Hawe *et al.* [1] achieve depth map super-resolution by exploiting SR in the Wavelet domain and a particular sampling strategy guided by intensity edges. Both of these two methods rely on the SR synthesis models. Very recently, cosparse analysis models [14], [15] were proposed and their effectiveness in image reconstruction has been successfully demonstrated. A pioneer work of applying the analysis models to guided depth upsampling is reported by Kiechle *et al.* [16]. It learns a depth-intensity bimodal analysis operator offline and applies the bimodal cosparse analysis model to reconstruct high-resolution depth maps.

Our work takes advantage of the sparse representation analysis models as well. More specifically, the proposed approach explores the cosparsity of analysis operators performed on a depth map, together with data fidelity and color guided smoothness constraints for upsampling. Instead of learning operators [16], we employ the well known analytic ones, such as the Wavelet transforms and the finite difference operators, for our guided depth upsampling. It implies that the approach requires no training data but a low-resolution depth map and an aligned high-resolution color image. It is therefore can be freely applied to either uniformly or non-uniformly sampled low-resolution depth maps, for instance,

depth maps obtained by Kinects and sparse range data collected by 3D lidars.

II. COSPARSE ANALYSIS MODELS

In contrast to the sparse representation synthesis models [17], [12] that have been extensively studied for decades, cosparse analysis counterparts have started to be investigated very recently. Therefore, we briefly review them in this section and stress on the differences.

Let us consider a set of measurements $\mathbf{y} \in \mathbb{R}^m$, which are sampled from an original signal $\mathbf{x} \in \mathbb{R}^d$ and are contaminated by noise \mathbf{v} . That is, $\mathbf{y} = \Phi\mathbf{x} + \mathbf{v}$, where $\Phi \in \mathbb{R}^{m \times d}$ is a sampling matrix, and $m < d$. In order to reconstruct \mathbf{x} , a synthesis model pursues a sparse representation $\mathbf{x} = \Psi\mathbf{z}$ with respect to a redundant dictionary $\Psi \in \mathbb{R}^{d \times n}$ ($n \geq d$). $\mathbf{z} \in \mathbb{R}^n$ is the sparse coefficient vector obtained by the following problem

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{s.t.} \quad \mathbf{y} = \Phi\Psi\mathbf{z} + \mathbf{v}. \quad (1)$$

Instead, an analysis model suggests that the analyzed vector $\Omega\mathbf{x}$ is expected to be sparse, where $\Omega \in \mathbb{R}^{p \times d}$ is a redundant *analysis operator* ($p \geq d$). Upon this, the original signal \mathbf{x} is recovered via

$$\min_{\mathbf{x}} \|\Omega\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = \Phi\mathbf{x} + \mathbf{v}. \quad (2)$$

When both the dictionary Ψ and the analysis operator Ω are square and invertible, the synthesis and the analysis models are the same with $\Omega = \Psi^{-1}$. Otherwise, there is no straightforward relation between them.

By checking the above linear transformations, we find out that the representation $\Omega\mathbf{x}$ cannot be very sparse if the rows of Ω are linearly independent. The reason is that at least $p - d$ of the coefficients of $\Omega\mathbf{x}$ should be non-zeros if $\mathbf{x} \neq 0$ [14]. Therefore, instead of focusing on nonzero elements, the analysis model emphasizes sparse representation on the number of zeros, which is referred to as *cosparsity*. That is,

$$\text{Cosparsity:} \quad \ell := p - \|\Omega\mathbf{x}\|_0. \quad (3)$$

Generally speaking, the cosparse analysis model can be viewed as a sparse synthesis model encoded with some structure. The structure of the signal \mathbf{x} is encoded by its *cosupport*, which is defined as the index set of the zero entries and denoted by

$$\Lambda := \{j \mid \langle \omega_j, \mathbf{x} \rangle = 0\}, \quad (4)$$

where ω_j is the j -th row of Ω .

When we use the above-defined analysis model to reconstruct the original signal, a necessary number of measurements must be given. Assuming that the sampling matrix Φ and the analysis operator Ω are mutually independent, the minimum number m should satisfy the following condition in order to guarantee the uniqueness of a ℓ -cosparse solution.

$$m \geq 2 \cdot \max_{|\Lambda| \geq \ell} \dim(\mathcal{W}_\Lambda), \quad (5)$$

where $\mathcal{W}_\Lambda = \text{Null}(\Omega_\Lambda)$ is the nullspace of Ω_Λ , and $|\Lambda|$ is the cardinality of Λ (refer to [14] for more details). More fortunately, when other constraints are taken into account for upsampling, we expect to recover \mathbf{x} with even less measurements.

III. GUIDED DEPTH UPSAMPLING

The proposed approach for depth upsampling is based on the cosparse analysis model, and meanwhile color information is integrated as guidance. In this section, we first introduce the formulation of our guided depth upsampling. Then, a numerical scheme relying on the Greedy Analysis Pursuit (GAP) [14] algorithm is presented to solve the formulized problem.

A. Problem Formulation

Assume that we are given a set of depth measurements \mathbf{D}_L , together with an aligned high-resolution color image \mathbf{I}_H . Our aim is to recover a depth map \mathbf{D}_H to be of the same resolution as the color image. Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$, respectively, be the vectorized high-resolution depth map and the vectorized sparse depth measurements, where $n = h \times w$ is the number of image pixels. Then, our guided depth upsampling is achieved by minimizing the following objective function:

$$E(\mathbf{x}) = \lambda_1 E_C(\Omega, \mathbf{x}) + \lambda_2 E_D(\mathbf{y}, \mathbf{x}) + \lambda_3 E_S(\mathbf{x}), \quad (6)$$

where $E_C(\Omega, \mathbf{x})$ is a term targeting at a cosparse representation, $E_D(\mathbf{y}, \mathbf{x})$ evaluates the fidelity between the measurements and the estimated depth values, and $E_S(\mathbf{x})$ stands for a term assessing the smoothness of the recovered depth map. λ_1 , λ_2 , and λ_3 are regularization parameters balancing the three terms. The details of each term are presented below.

1) *Cosparse Analysis Term*: Given an analysis operator Ω , the cosparse analysis term aims to achieve a sparse representation of $\Omega\mathbf{x}$. Hence, it is defined as

$$E_C(\Omega, \mathbf{x}) = \|\Omega\mathbf{x}\|_0. \quad (7)$$

A critical point concerned in this term is the chosen of Ω . Several researches [18], [15], [19] have been studied on the learning of analysis operators. Learned operators are considered to be of better performance in image super-resolution than analytic ones such as Wavelets [20]. However, their disadvantages are also obvious. For instance, it takes time to learn an analysis operator and performance of the learned operator highly depends on training sets. Moreover, due to limitations of computational resources, operators are often learned in a patch-wise way and are applied locally to image patches [15].

In contrast to learned operators, analytic operators are able to be utilized globally to an entire depth map. Operators such as the wavelet transforms, the finite difference operators, and the curvelet and shearlet transforms [21] have been

successfully used in signal processing applications. Thus, in this work, we choose analytic operators for upsampling. We also defend that, with the use of guided color information, the upsampling performance of applying analytic analysis operators is competitive to those using learned counterparts [16].

2) *Data Fidelity Term*: The data fidelity term evaluates the errors between the sparse measurements and the corresponding depth values that are recovered. It is designed as

$$E_D(\mathbf{y}, \mathbf{x}) = \|\mathbf{y} - \Phi\mathbf{x}\|_2^2, \quad (8)$$

where Φ represents the sampling matrix as before.

It needs mentioning that, when \mathbf{x} is a vector of an entire depth map, very large size matrices have to be constructed to perform the matrix-vector multiplications of $\Omega\mathbf{x}$ and $\Phi\mathbf{x}$, which are not affordable in memory. Practically, both Ω and Φ are implemented by functions. It in essence means that Ω stands for the process of applying the analysis transform and Φ represents the sampling procedure. We stick to the notations in the form of matrix-vector multiplications for simplicity and consistency with others' work [14].

3) *Smoothness Term*: The smoothness term is based upon a prior that depth maps are piecewise smooth. Constraints such as minimizing total variations (TV) are often used for preserving smoothness [1]. In this work, we prefer to design the term as follows:

$$E_S(\mathbf{x}) = \sum_i \sum_{j \in \mathcal{N}(i)} w_{ij} \|x_i - x_j\|_2^2, \quad (9)$$

where x_i is a pixel in \mathbf{x} , and $\mathcal{N}(i)$ denotes the 4-connected neighborhood of x_i in the corresponding depth map \mathbf{D}_H . w_{ij} is a weight that integrates color information for regularization. It is defined by

$$w_{ij} = \exp\left(-\frac{\|\mathbf{I}_i - \mathbf{I}_j\|_2^2}{2\sigma^2}\right), \quad (10)$$

in which \mathbf{I}_i and \mathbf{I}_j are color or intensity values of the pixels registered to \mathbf{x}_i and \mathbf{x}_j , and σ is a standard deviation. This form is also the first-order smoothness constraint commonly used in the Markov Random Field framework [5]. It performs better than the TV norm [1], especially when upsampling factor is large.

B. Numerical Scheme

As one may notice, the formulated problem in Eq.(6) degenerates to a MRF-based energy function when $\lambda_1 = 0$, which can be solved efficiently by the well-known conjugate gradient (CG) algorithm [5]. However, when λ_1 is not zero, the cosparse analysis term $\|\Omega\mathbf{x}\|_0$ leads the entire problem to be a NP-complete one. One way to approximately solve it is turning it to $\|\Omega\mathbf{x}\|_1$, a L_1 -minimization problem. Whereas, in this work, we adopt a simple approach, the Greedy Analysis Pursuit (GAP) algorithm, to find the optimized solution.

GAP imitates Orthogonal Matching Pursuit (OMP) [17] with a form of Iterative Reweighted Least Squares (IRLS) [22], [23]. It starts from a cosupport $\hat{\Lambda}_0$ that is initialized to be the whole set. An initial estimation of the depth map is obtained via

$$\hat{\mathbf{x}}_0 = \arg \min_{\mathbf{x}} \lambda_1 \|\Omega\mathbf{x}\|_2^2 + \lambda_2 E_D(\mathbf{y}, \mathbf{x}) + \lambda_3 E_S(\mathbf{x}), \quad (11)$$

which is a least square optimization problem solved by conjugate gradient in this work. Some elements outside the cosupport set are first detected according to the values of $\Omega\hat{\mathbf{x}}_0$ and are removed from the set. With an updated cosupport, \mathbf{x} is re-estimated for removing more elements. By this means, GAP iteratively reduces the cosupport set till it reaches ℓ -cosparsity or another stop criterion is satisfied.

The details of GAP-based guided depth upsampling is illustrated in Algorithm 1.

Algorithm 1: GAP for Guided Depth Upsampling

Input: The analysis operator Ω ,
the sampling matrix Φ ,
the sparse measurements \mathbf{y} ,
the target cosparsity ℓ ,
and a selection factor $t \in (0, 1]$.

Output: Optimized $\hat{\mathbf{x}} = \hat{\mathbf{x}}_k$.

Initialization:

$k = 0$;
Initialize cosupport: $\hat{\Lambda}_0 = \{1, 2, \dots, p\}$;
Initialize solution:

$$\hat{\mathbf{x}}_0 = \arg \min_{\mathbf{x}} \lambda_1 \|\Omega_{\hat{\Lambda}_0} \mathbf{x}\|_2^2 + \lambda_2 E_D(\mathbf{y}, \mathbf{x}) + \lambda_3 E_S(\mathbf{x}).$$

while $k < p - \ell$ **do**

$k = k + 1$;

Compute $\alpha = \Omega\hat{\mathbf{x}}_{k-1}$;

Find largest entries: $\Gamma_k = \{i : |\alpha_i| \geq t \max_j |\alpha_j|\}$

Update cosupport: $\hat{\Lambda}_k = \hat{\Lambda}_{k-1} \setminus \Gamma_k$;

Update solution:

$$\hat{\mathbf{x}}_k = \arg \min_{\mathbf{x}} \lambda_1 \|\Omega_{\hat{\Lambda}_k} \mathbf{x}\|_2^2 + \lambda_2 E_D(\mathbf{y}, \mathbf{x}) + \lambda_3 E_S(\mathbf{x}).$$

end

IV. EXPERIMENTS

Extensive experiments have been conducted in order to validate the proposed approach. In this section, we first present sets of experiments performed on the Middlebury stereo dataset [24] and provide both quantitative and qualitative comparison to state-of-the-art algorithms. Experimental settings, such as the chosen of analysis operators and the assigned parameters, are also discussed. In the end, we apply our approach to upsample real-world lidar data collected by a Velodyne HDL 64E lidar.

A. Depth Upsampling from Random Samples

This group of experiments is designed to investigate the performance of the cosparse analysis model in contrast to synthesis models. Therefore, we compare our approach to the work done by Hawe *et al.* [1]. In order to achieve quantitative evaluation, experiments are performed on the Middlebury stereo dataset. Synthetic data are generated in a way exactly the same as Hawe’s work. That is, we randomly sample a certain percentage of points from known depth maps, and include measurements on edges detected by Canny filter as well. Aligned color images are taken as guidance in both methods.

The parameters of our algorithm are experimentally assigned throughout all experiments as follows. The weighting factors in Eq.(6) are $\lambda_1 = 0.01$, $\lambda_2 = 1$, and $\lambda_3 = 0.1$. The standard deviation in Eq.(10) is determined to $\sigma = 30$. Considering that depth discontinuities co-occur with intensity changes, we take the cosparsity of an aligned intensity edge map as the target cosparsity ℓ , and set the selection factor $t = 0.6$ in Algorithm 1. Moreover, we investigate two different types of analytic analysis operators, which are Daubechies wavelets (db2-wavelets) and the finite difference operators. For the Daubechies wavelets, decomposition levels from one to four are tested. We denote the corresponding operators, respectively, by Ω_{WT1} , Ω_{WT2} , Ω_{WT3} , and Ω_{WT4} . A finite difference operator that concatenates the horizontal and vertical derivatives and the other extra concatenating the diagonal derivatives are examined as well, which are denoted as Ω_{DIFF} and Ω_{DIFF_DIAG} respectively. Table I illustrates the dimensions of all the operators for each depth map. As introduced in Section II, an operator is of $p \times d$ dimension. d is the same as the resolution of an image and p is different from case to case. Regarding to the extremely high dimensionality, we implement the operators by functions instead of matrices.

	Tsukuba d = 110592	Venus d = 166222	Teddy, Cones d = 168750
Ω_{WT1}	111940	168296	170856
Ω_{WT2}	112963	169342	171918
Ω_{WT3}	113311	169762	172342
Ω_{WT4}	113491	169978	172620
Ω_{DIFF}	220512	331627	336675
Ω_{DIFF_DIAG}	440354	662439	672527

Table I

THE DIMENSION $p \times d$ OF EACH ANALYSIS OPERATOR. d IS THE SAME AS THE RESOLUTION OF AN IMAGE AND IS LISTED ON THE TOP. THE REMAINING VALUES ARE p , WHICH VARIES FROM CASE TO CASE.

To obtain quantitative evaluation, we assess all experimental results in terms of the root mean square error (RMSE) of the upsampled depth maps against to ground truth. As in [16], each depth map is first scaled to a certain disparity range that is required in the Middlebury website for evalua-

tion. Table II reports the evaluation results for experiments upsampling from 5%, 15%, and 25% randomly sampled measurements. In the table, ‘CSC’ stands for the synthesis model based upsampling method as explained in [1]. The others are our approaches while using different analysis operators. The results of ‘CSC’ are obtained by running the code released by the authors, with a large number of iterations to make sure that each case converges. All other parameters are set as reported in their paper. Table II shows that, except for the Ω_{WT1} operator which fails in some cases, all other analysis operators perform better than the synthesis model (CSC). Among all, the Ω_{WT4} operator performs best, followed closely by Ω_{DIFF_DIAG} and Ω_{DIFF} . The results also tell us that, for the same type of operators, the better performance is obtained if the operator is more redundant.

Fig. 1 illustrates some comparative results upsampled from 5% measurements, in which the columns (a) and (b) are the aligned color images and ground truth depth maps, the column (c) shows the sampled data, (d) are the results of CSC, and the last column are obtained by our approach using the Ω_{WT4} operator. From the results we observe that our approach preserves depth edges better. This superiority not only benefits from our color guided smoothness term, but also from the cosparse analysis term to a great extent. The advantage of the cosparse analysis term will be demonstrated further in the next experiment.

B. Depth Upsampling from Uniform Samples

Although random sampling together with edge samples can improve performance, it is not able to obtain these measurements in applications such as super-resolution of depth maps captured by Kinects or ToF cameras. Therefore, in this experiment, we explore the performance of our approach when applied to uniformly sampled low-resolution depth maps. Synthetic data are generated as the way in [16]. That is, we first smooth a high-resolution depth map [25] and then downsample it through bicubic interpolation at a certain sampling factor. Considering that bicubic interpolation has edge effect near occlusion, the occluded regions are inpainted by our approach before downsampling.

Among the developed depth map super-resolution techniques, some [5], [16] use guidance information and the others [26] do not. In our evaluation, only those methods using guidance, for instance, bilateral filtering and MRF based techniques, are chosen for comparison. Besides these two typical techniques, we also compare our work to a state-of-the-art method [16], which is based on a cosparse analysis model as well. In contrast to us, it relies on analysis operators learned from a training set. The results of the bilateral filtering method are produced from our own implementation. To achieve the results of MRF, we simply set $\lambda_1 = 0$, $\lambda_2 = 1$, and $\lambda_3 = 1$ in our model. The quantitative results of Kiechle *et al.*’s [16] work are from their paper.

Method \ Scenario	Tsukuba			Venus			Teddy			Cones		
	5%	15%	25%	5%	15%	25%	5%	15%	25%	5%	15%	25%
CSC [1]	0.423	0.455	0.442	0.703	0.603	0.545	2.469	2.070	2.121	2.633	2.266	2.113
Ω_{WT1}	0.425	<u>5.019</u>	0.317	<u>7.343</u>	<u>4.852</u>	0.171	<u>24.335</u>	<u>22.144</u>	1.271	<u>5.092</u>	<u>27.015</u>	1.536
Ω_{WT2}	0.406	0.257	0.187	0.116	0.134	0.105	1.427	1.001	0.967	1.594	1.080	0.886
Ω_{WT3}	0.399	0.243	0.194	0.127	0.041	0.042	1.326	0.918	0.844	1.276	0.909	0.769
Ω_{WT4}	0.395	0.251	0.195	0.094	0.040	0.041	1.115	0.883	0.804	1.042	0.884	0.749
Ω_{DIFF}	0.420	0.248	0.181	0.049	0.041	0.047	1.362	1.019	0.945	1.112	0.954	0.845
Ω_{DIFF_DIAG}	0.451	0.383	0.217	0.105	0.057	0.048	1.057	0.867	0.794	1.060	0.891	0.715

Table II
 RMSE OF EXPERIMENTAL RESULTS THAT ARE UPSAMPLED FROM RANDOMLY SAMPLED DATA PLUS EDGE POINTS. FOR Ω_{WT1} , UNDERLINED VALUES ARE THE FAILED CASES.

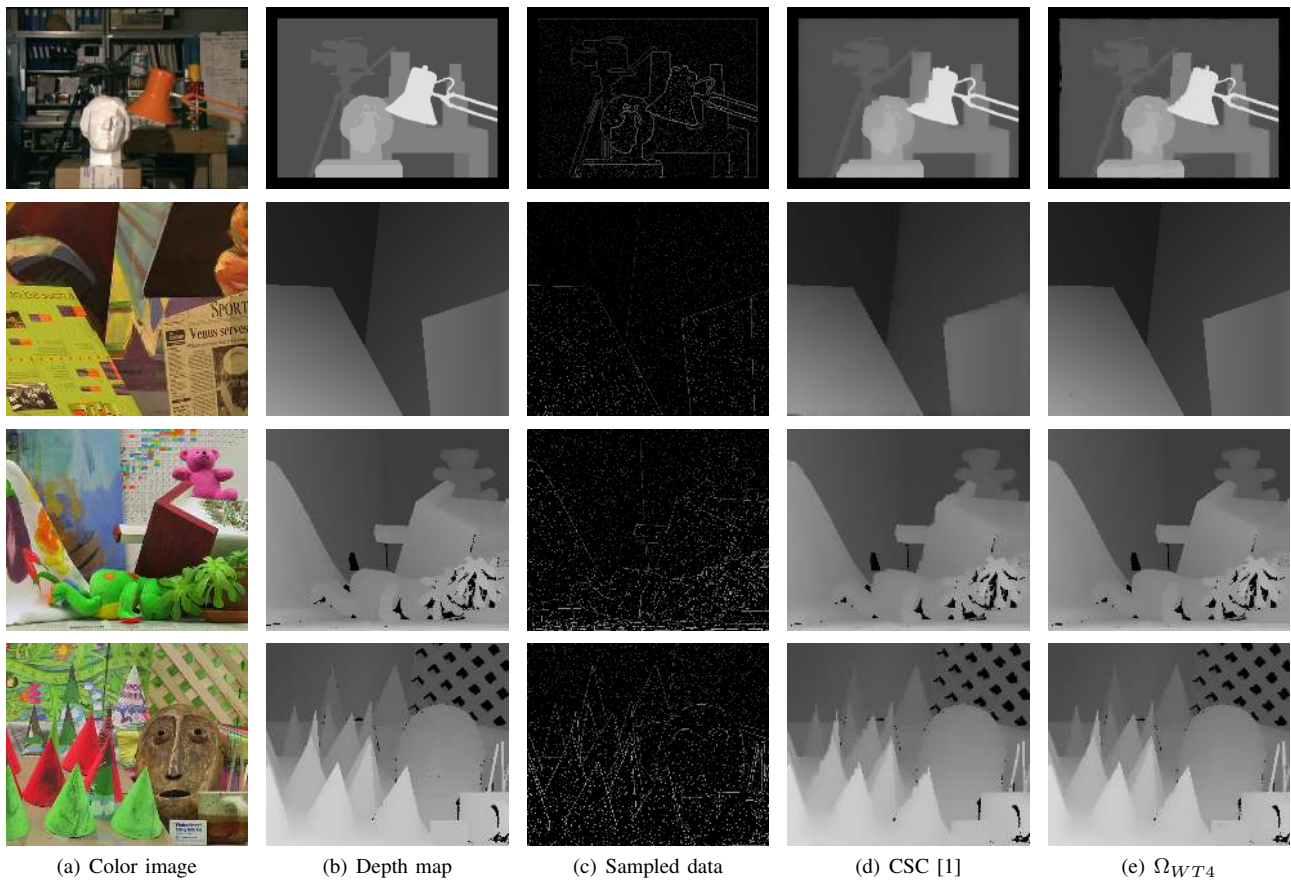
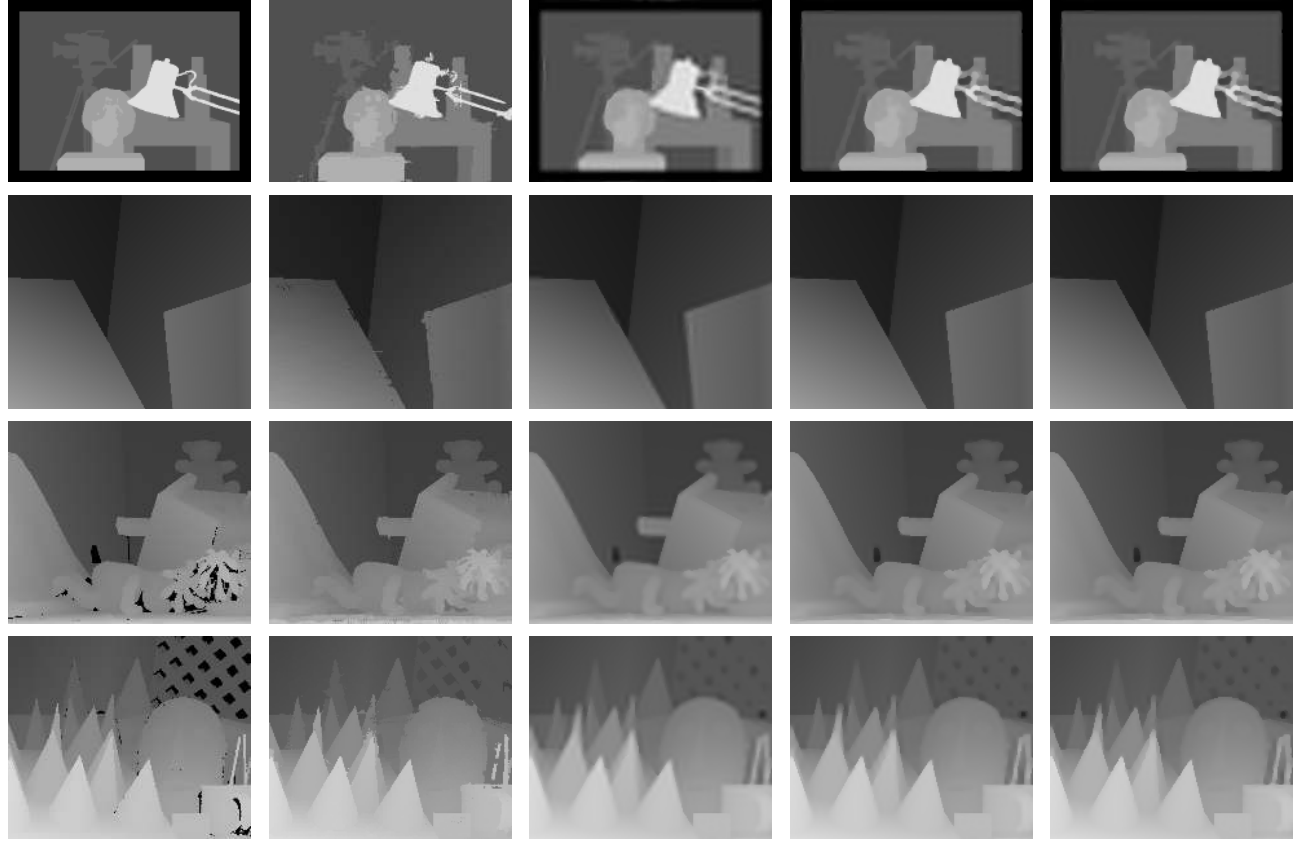


Figure 1. Visual comparison of experimental results upsampled from 5% measurements plus edge points. The four scenarios are 'Tsukuba', 'Venus', 'Teddy', and 'Cones' respectively. The results show that our approach preserve depth edges better. (Zoom in for better view.)



(a) Ground truth (b) Bilateral Filter (c) MRF (d) Ω_{WT4} (e) Ω_{DIFF_DIAG}

Figure 2. Visual comparison of experiments that are upsampled with a scaling factor 8. (Zoom in for better view.)

Method \ Scenario	Tsukuba			Venus			Teddy			Cones		
	2×	4×	8×	2×	4×	8×	2×	4×	8×	2×	4×	8×
Kiechle <i>et al.</i> [16]	0.255	0.487	0.753	0.075	0.129	0.156	0.702	1.347	1.662	0.680	1.383	1.871
Bilateral Filter [8]	0.392	0.540	0.830	0.129	0.195	0.342	0.486	0.682	0.987	0.691	0.944	1.419
MRF [5]	0.479	0.646	0.855	0.142	0.181	0.255	0.461	0.580	0.780	0.664	0.851	1.263
Ω_{WT4}	0.147	0.390	0.708	0.089	0.121	0.171	0.310	0.456	0.664	0.347	0.681	1.159
Ω_{DIFF}	0.154	0.385	0.738	0.045	0.118	0.175	0.234	0.474	0.710	0.282	0.676	1.166
Ω_{DIFF_DIAG}	0.074	0.339	0.722	0.029	0.113	0.172	0.213	0.478	0.769	0.247	0.659	1.167

Table III
RMSE OF EXPERIMENTS THAT ARE UPSAMPLED WITH A SCALING FACTOR OF 2, 4, AND 8 RESPECTIVELY.

Method \ Scenario	Tsukuba			Venus			Teddy			Cones		
	2×	4×	8×	2×	4×	8×	2×	4×	8×	2×	4×	8×
Kiechle <i>et al.</i> [16]	0.47	1.73	3.53	0.09	0.25	0.33	1.41	3.54	6.49	1.81	5.16	9.22
Bilateral Filter [8]	0.89	1.19	1.87	0.15	0.21	0.43	1.58	3.21	5.40	1.39	1.88	3.31
MRF [5]	3.35	5.83	9.23	0.44	0.73	1.51	2.44	3.80	8.00	3.93	6.24	11.69
Ω_{WT4}	0.31	1.78	4.54	0.31	0.22	0.40	1.26	2.18	5.76	1.23	3.15	7.45
Ω_{DIFF}	0.19	1.32	4.75	0.03	0.13	0.30	0.57	2.16	5.79	0.66	2.55	7.39
Ω_{DIFF_DIAG}	0.04	0.79	3.17	0.01	0.07	0.16	0.53	2.48	5.82	0.42	2.05	6.27

Table IV
THE PERCENTAGE OF BAD PIXELS OF THE EXPERIMENTS THAT ARE UPSAMPLED WITH A SCALING FACTOR OF 2, 4, AND 8 RESPECTIVELY.



Figure 3. Results of the guided depth upsampling performed on a Velodyne 64E Lidar dataset. The first two rows, respectively, are high-resolution color images and sparse 3D lidar points. The third row contains low-resolution maps obtained by registering 3D points to the images. The fourth row presents upsampled depth maps, and the fifth shows the rendered dense 3D point clouds obtained from our upsampling results (sky regions are removed). For comparison, the sparse rendered 3D point clouds are also provided in the last row.

Visual comparison of these methods behaved with a upscaling factor of 8 are illustrated in Fig. 2. RMSE of more experiments are listed in Table III. Moreover, Table IV reports the percentage of bad pixels, whose error is greater than one pixel. Considering the performance of our analysis operators, we only report the results of Ω_{WT4} , Ω_{DIFF} , Ω_{DIFF_DIAG} here. The comparisons show that, with the cospase analysis term, we achieve sharper depth edges than the MRF based technique and comparable to Bilateral filtering. Among all, our approach with the Ω_{DIFF_DIAG} operator performance best, followed by Ω_{WT4} .

C. Experiments on Lidar Data

Finally, we apply our approach to upsample lidar data in order to finally achieve dense 3D reconstruction. Experiments are conducted on the KITTI vision benchmark suite [27]. This dataset consists of sparse lidar data and high-resolution color images, which are simultaneously collected by a Velodyne HDL 64E lidar and a video camera in real road scenarios. With known sensor parameters, lidar points

are first registered to images so that low-resolution depth maps are obtained, as shown in the first three rows of Fig. 3. Our approach is then applied to upsample these low-resolution depth maps. The upsampled results are demonstrated in the fourth row, followed by the rendered high-resolution 3D point clouds. Meanwhile, we also provide the sparse rendered 3D point clouds in the last row for visual comparison. From the results we see that our approach increases the resolution of 3D point clouds successfully.

V. CONCLUSIONS

In this paper, we have presented a new approach for guided depth upsampling. It relies on the cospase analysis models and makes use of analytic operators. Different well known operators have been investigated. The experimental results show that, as opposed to the synthesis cases, more redundant operators are preferred.

Benefitted from the use of analytic operators, the approach requires no training data. It can be applied to either

randomly or uniformly sampled low-resolution depth maps, and achieves high performance even with very low sampling rate. Numerous experiments have been conducted on both synthetic and real data, while comparing to a synthesis model based method [1], an analysis operator learning based one [16], and two other typical guided depth upsampling techniques. Experiments have shown that our approach outperforms these state-of-the-art algorithms. Moreover, due to the underline scheme, our approach can also be applied to guided depth inpainting and denoising directly.

ACKNOWLEDGMENT

This research work was supported in parts by the National Natural Science Foundation of China via grants 61001171, 61071219, 90820306, and the Fundamental Research Funds for the Central Universities.

REFERENCES

- [1] S. Hawe, M. Kleinsteuber, and K. Diepold, "Dense Disparity Maps from Sparse Disparity Measurements," in *ICCV*, 2011, pp. 2126–2133.
- [2] Velodyne, 2013, <http://velodynelidar.com/lidar/lidar.aspx>.
- [3] Primesense, 2013, <http://www.primesense.com/>.
- [4] Microsoft, 2013, <http://www.xbox.com/en-US/kinect/>.
- [5] J. Diebel and S. Thrun, "An Application of Markov Random Fields to Range Sensing," in *NIPS*, 2005.
- [6] J. Dolson, J. Baek, C. Plagemann, and S. Thrun, "Upsampling Range Data in Dynamic Environments," in *CVPR*, 2010.
- [7] X. Gong, J. Liu, W. Zhou, and J. Liu, "Guided depth enhancement via a fast marching method," *Image and Vision Computing*, vol. 31, no. 10, pp. 695 – 703, 2013.
- [8] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-Depth Super Resolution for Range Images," in *CVPR*, 2007.
- [9] J. Zhu, L. Wang, J. Gao, and R. Yang, "Spatial-Temporal Fusion for High Accuracy Depth Maps Using Dynamic MRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 899–909, 2010.
- [10] U. Wong, B. Garney, W. Whittaker, and R. Whittaker, "Camera and LIDAR Fusion for Mapping of Actively Illuminated Subterranean Voids." International Conference on Field and Service Robotics, 2009.
- [11] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon, "High Quality Depth Map Upsampling for 3D TOF Cameras." *ICCV*, 2011.
- [12] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image Super-Resolution Via Sparse Representation," *Image Processing, IEEE Transactions on*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [13] Y. Li, T. Xue, L. Sun, and J. Liu, "Joint example-based depth map super-resolution," in *ICME*, 2012, pp. 152–157.
- [14] S. Nam, M. Davies, M. Elad, and R. Gribonval, "The cosparsity analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30 – 56, 2013.
- [15] S. Hawe, M. Kleinsteuber, and K. Diepold, "Analysis Operator Learning and its Application to Image Reconstruction," *Image Processing, IEEE Transactions on*, vol. 22, no. 6, pp. 2138–2150, 2013.
- [16] M. Kiechle, S. Hawe, and M. Kleinsteuber, "A Joint Intensity and Depth Co-Sparse Analysis Model for Depth Map Super-Resolution," in *ICCV*, 2013.
- [17] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st ed. Springer Publishing Company, Incorporated, 2010.
- [18] B. Ophir, M. Elad, N. Bertin, and M. D. Plumbley, "Sequential Minimal Eigenvalues - An Approach to Analysis Dictionary Learning," in *The 19th European Signal Processing Conference (EUSIPCO)*, Barcelona, Espagne, 2011.
- [19] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A Dictionary-Learning Algorithm for the Analysis Sparse Model," *Signal Processing, IEEE Transactions on*, vol. 61, no. 3, pp. 661–677, 2013.
- [20] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd ed. Academic Press, 2008.
- [21] J.-L. Starck, E. Candes, and D. Donoho, "The Curvelet Transform for Image Denoising," *Image Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 670–684, 2002.
- [22] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Comm. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.
- [23] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. E. Davies, "Greedy-like algorithms for the cosparsity analysis model," 2012.
- [24] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7–42, 2002.
- [25] T. Michaeli and M. Irani, "Nonparametric Blind Super-Resolution," in *ICCV*, 2013.
- [26] O. M. Aodha, N. D. F. Campbell, A. Nair, and G. J. Brostow, "Patch Based Synthesis for Single Depth Image Super-Resolution." in *ECCV*, 2012.
- [27] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," in *CVPR*, 2012.