Edge-Guided Depth Image Super-Resolution Based on KSVD

BINHUI LIU AND QIANG LING, (Senior Member, IEEE)
Department of Automation, University of Science and Technology of China, Hefei 230027, China

Corresponding author: Qiang Ling (qling@ustc.edu.cn)

ABSTRACT This paper proposes an edge-guided super-resolution algorithm for single frame depth images based on K singular value decomposition (KSVD). Compared with conventional algorithms, the proposed algorithm has two key contributions. Firstly it suppresses the jagged edge effect of up-sampled depth images by KSVD, which learns a complete dictionary to describe the mapping between the jagged edges and corresponding smooth ones. Secondly it improves the joint bilateral filter based on connectivity. The improved filter can not only preserve the sharpness of the edges during the interpolation, but also suppress noise. The proposed algorithm has been extensively tested on the Middlebury dataset and compared with some existing state-of-the-art methods. Both quantitative and qualitative experimental results show its performance superiority.

INDEX TERMS Depth image, super-resolution, KSVD, improved joint bilateral filter, sparse representation.

I. INTRODUCTION

Depth information is of great importance in the field of computer vision. However, due to hardware limitation, the resolution of depth images generated by ordinary devices, such as Kinect and time-of-flight (TOF) cameras, is relatively low. So super-resolution is strongly expected to improve the visibility and practicability of depth images.

In the early research on image super-resolution, some interpolation methods and excellent filters were proposed for the super-resolution of depth images. A commonly used cubic spline image interpolation method was proposed in [1], which, however, suffers from the edge blurring problem. The Bayesian method is also applied for super-resolution [2], but it only adopts a statistic hypothesis of the input, rather than learn the priori probability of all depth images. These traditional interpolation methods often lose details in the texture areas, especially at edges. 3D applications, which rely on depth images, are often very sensitive to such loss of edge details of depth images. Therefore, maintaining sharp edges is the major goal for depth image super-resolution algorithms, even at the cost of accuracy degradation. Reference [3] achieves good super-resolution results based on depth image sequences which are obtained under various

small offsets of the camera in the same scene. However, there are significant limitations in [3] because camera movement is not allowed in most practical applications.

Current image super-resolution algorithms can be roughly divided into two main types according to their input.

1) The first class is color-image based super-resolution which utilizes the registered color and depth image pairs in the same scene. The basic assumption of this kind of methods is that the depth values of the pixels with similar colors are likely close. However, this assumption is not always true.

2) The other class is single frame depth image super-resolution which needs to make full use of the texture structure of the original depth images. Single depth image super-resolution estimates a high-resolution image from a degraded and noisy low-resolution image. Compared with color-image based super-resolution, single depth image super-resolution suffers less reference information which makes the super-resolution task more challenging. But due to their low hardware cost and low computational complexity, this class of methods have attracted more and more attention.

With the recent development of machine learning, various neural-network based super-resolution algorithms have been proposed and achieved good results. A super-resolution convolutional neural network was proposed in [4]

The associate editor coordinating the review of this manuscript and approving it for publication was Eduardo Rosa-Molinar.

(SRCNN). Reference [5] presented a super-resolution deep convolutional network (VDSR) which uses CNN to learn the end-to-end mapping relationship between low-resolution and high-resolution images. Reference [6] introduced a new color-guided coarse-to-fine convolutional neural network framework that implements super-resolution of depth images by learning an ideal filter. Although neural network-based methods have achieved significantly good performance in color image super-resolution [5], they may not be simply implemented to depth image super-resolution and are often limited in practical applications due to their high computational complexity. It is exactly the goal of this paper to develop a super-resolution algorithm, which can well balance between the super-resolution performance and the computational complexity.

The rest of the paper is organized as follows. The related work is introduced in Section II. Section III describes the framework and implementation of our super-resolution algorithm in details. Section IV presents some qualitative and quantitative experimental results. Finally, some concluding remarks are provided in Section V.

II. RELATED WORK
A. SINGLE DEPTH IMAGE SUPER-RESOLUTION
In reality, it is often required to do super-resolution for a single depth image. In that case, the main reference information is the image texture structure. However, some aforementioned algorithms [1], [2] ignore the image texture structure and uses the same interpolation principle to estimate pixels of both the edges and smooth regions, which results in artifacts in the reconstructed depth image. As shown in [6], reconstruction error is very prominent around the edges when the bicubic up-sampling is directly applied to the depth image.

Sparse representation was firstly implemented to color image super-resolution in [7] and achieved good results. Its key idea is that the low resolution image block and the corresponding high resolution one share the same sparse representation coefficients. Similarly, [8] and [9] applied the sparse representation method to depth image super-resolution. Then in [10], the global pixel value prediction problem faced by depth image super-resolution reconstruction was transformed into two problems, including binary edge map reconstruction and edge-guided interpolation. For edge map reconstruction, Markov random field theory was applied to describe the correlation of neighborhood edge patches, and obtain the global optimal estimation of the random field by solving the minimum Gibbs energy. However, its time cost is unacceptable due to its high computational complexity. In [11], an approximate solution to the optimal field state under the minimum energy was proposed to speed up the reconstruction. It is believed that there are a large number of similar edge structures in the images of different scales. Thus, [13] constructed a pyramid composed of edge maps at different scales to find similar training edge patches and learn the end-to-end mapping between the edges of each scale. However,

the training edge samples taken from up-sampled images are often jagged and of low quality.

B. EDGE-PRESERVING FILTER
The main goal of depth image super-resolution is to maintain sharp edges. Therefore many excellent edge-preserving filters are applied to depth image super-resolution. Guided filter was proposed in [14] as an edge-preserving smoothing operator. Although the guided filter can hold edges well and is easy to calculate, it is sometimes degraded by artifacts. With the reference information in both the space domain and the intensity domain, bilateral filter [15] can achieve good edge retention characteristics. Reference [16] uses a joint bilateral filter [17] to ensure the super-resolution performance of the input images obtained in dark or noisy environments based on additional color images. Reference [16] also uses an iterative structure to perform super resolution with large up-sampling factor, i.e., 8. Reference [12] proposed a least-square optimization which combines several weighting factors together with non-local means filtering to maintain sharp depth boundaries based on an auxiliary high-resolution RGB image.

Inspired by the above edge-guided super-resolution methods, this paper proposes a depth image super-resolution algorithm based on KSVD. Our algorithm is made up of smooth edge reconstruction and guided filtering based on edge connectivity. We first extract the jagged edges from a depth image and the corresponding smooth edges from its corresponding high-resolution depth image as learning samples, and use the KSVD algorithm to learn a complete dictionary describing the mapping between the jagged edges and the smooth edges. That is, any given jagged edge can be smoothed using the learned dictionary. Then we further improve the joint bilateral filter based on edge connectivity and apply it to the edge-guided interpolation. The improved filter can not only maintain the edge sharpness during interpolation, but also further suppress noise.

III. PROPOSED METHOD
An overview of our algorithm is shown in Figure 1. It is made up of two parts, including smooth edge reconstruction and guided filtering based on edge connectivity. The binary edge map contains only some basic structural information, such as lines and angles, which are highly sparse. Therefore, the sparse coding method can be implemented to restore high quality edges. However, the sparse coding method needs to train an over-complete dictionary from a set of images and then recover the edges using the feature atoms of the over-complete dictionary.

Real edge information is important for accurately distinguishing different objects in view. To acquire a high-quality edge map, the input low resolution depth image D_l is firstly magnified to the expected size by the bicubic interpolation algorithm. Then a shock filter [18] is applied to preliminarily suppress the jagged effect on edges caused by interpolation and produces a depth image D'_l. Afterward, an edge image
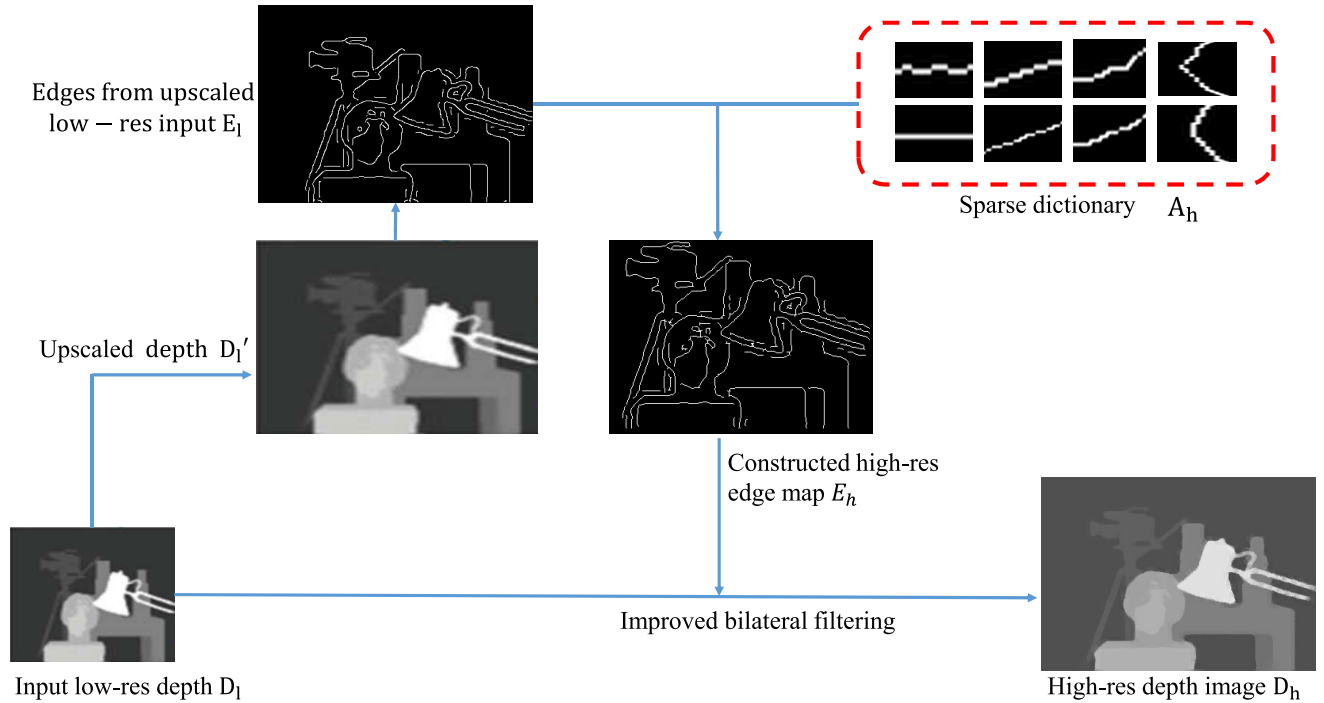
E_l is extracted from the filtered depth map D'_l. Usually, E_l still has a sawtooth effect. From E_l, a high-resolution smoothed edge image E_h is reconstructed using the already learned mapping dictionary A_h by KSVD which describes the mapping between jagged edges and smooth edges. After the edge map E_h is obtained, a novel bilateral filter is utilized to reconstruct a high resolution depth image D_h. Acquisition of sharp edge and artifact elimination are two main challenges of depth image super-resolution. The key contributions of this method are

1) The high-resolution fine edge structure obtained by the shock filter and the sparse representation is smooth and natural, which enables the subsequent edge-guided interpolation more precise.
2) The improved filter based on the edge connectivity can well maintain the edge sharpness.
3) High-quality samples from any type of high-resolution depth images are taken to construct the training set. Due to such off-line dictionary training, the reconstruction speed is fast.

A. THE CONSTRUCTION OF THE EDGE MAP

1) TRAINING SET FOR KSVD

For a given high resolution image, its corresponding low resolution image can be acquired by bicubic down-sampling. We first extract the smooth edges and the jagged edge patch pairs from the depth image sets of both high and low resolutions. Next, for the extracted edge patch pairs, the first-order and second-order differential operators are performed in both the horizontal and the vertical directions to extract gradient features $Y_l = \{P_l\}$. In general, depth images have relatively less texture than ordinary color images. Thus the original feature matrix composed of the gradient feature extracted from the binary edge image is very sparse. For higher learning efficiency, principal component analysis (PCA) is performed to greatly reduce the dimension of the original training data while retaining the main features. The KSVD algorithm is used to learn the main features after dimensionality reduction and produce an over-complete dictionary, which can describe the mapping between the jagged edges and the smooth edges. For learning based super-resolution reconstruction methods, the richer learning samples are contained in the edge database, the higher stability and accuracy of the mapping established by the dictionary learning can be expected.

2) KSVD DICTIONARY LEARNING ALGORITHM

Common dictionary learning algorithms include K-means algorithm, K-SVD algorithm and greedy depth dictionary learning. Here we use K-SVD algorithm for dictionary training. The main idea of dictionary learning is to use a dictionary matrix $A \in R^{m \times K}$, which contains K atoms, to sparsely represent the original samples $Y \in R^{m \times n}$ where m represents the attribute number of the samples and n is the number of samples. Ideally $Y = AX$, where $X \in R^{K \times n}$ is the sparse coding matrix and $x_i(i = 1, 2, \ldots, n)$ is the the i-th column vector of X. The requirement of $Y = AX$ can be formulated into the following optimization,

$$\min_{A,X} \|Y - AX\|^2,$$
$$s.t. \ \forall i, \|x_i\|_0 \leq T_0, \tag{1}$$

Algorithm 1 KSVD Dictionary Learning

Input: Original samples $Y \in R^{m \times n}$, Dictionary D, Sparse Matrix X and the maximum training iteration number N

1: **Initialization:** Randomly choose K column vectors from Y as the initial atoms of Dictionary D.
2: Let $j = 0$;
3: **while** $j \leq N-1$ and the error matrix E_k does not converge **do**
4:      Use the dictionary $D^{(j-1)}$, which is obtained in the previous step, to sparsely encode Y, and obtain $X^{(j)} \in R^{K \times n}$;
5:      Update the dictionary $D^{(j)}$ column by column, whose columns are denoted as $d_1, d_2, \cdots, d_K$};
6:      Calculate the error matrix E_k with the updated $D^{(j)}$ by $E_k = Y - \sum d_i x_T^j$;
7:      Extracting the set of index w_k of the k-th non-zero row vector x_T^j of the sparse matrix;
8:      Remove column vectors according to the above index w_k from E_k and get E'_k;
9:      Do singular value decomposition $E_k = UHV^T$ and take the first column of U to update the k-th column of the dictionary, i.e., $d_k = U(:, 1)$. Define $x_T^{k'} = H(1, 1)V(:, 1)$ and replace x_T^k with $x_k^{k'}$;
10:    $j = j + 1$;
11: **end while**

Output: Dictionary D, Sparse matrix X.

---

where $\|\cdot\|$ stands for the 2-norm of a vector, $\|\cdot\|_0$ represents the 0-norm of a vector, i.e., the number of non-zero elements of the concerned vector, and T_0 is a given threshold.

In Optimization 1, the objective function minimizes the error between the dictionary mapping and the original sample, i.e., restoring the original sample as precisely as possible, and the sparsity of X does not exceed T_0. This optimization problem has two optimization variables, A and X. The KSVD algorithm uses an alternate fixed optimization strategy to find the optimal A and X. The whole framework of KSVD is presented in Algorithm 1.

For the original sample set Y, the KSVD algorithm acquires the dictionary matrix A and the corresponding sparse coding matrix X with a given sparsity and the number of feature atoms K.

In the experiment, the low-resolution jagged edges and the corresponding high-resolution smooth edges are combined into patch pairs to constitute their sample matrix Y_l and Y_h. The KSVD algorithm is used to perform dictionary learning on Y_l to obtain a mapping dictionary A_l and corresponding sparse coding matrix X_l according to the following constraints.

$$Y_l = A_l X_l$$
$$Y_h = A_h X_l$$
$$A_h = Y_h X_l^T (X_l X_l^T)^{-1} \qquad (2)$$

For any given low resolution depth image, we extract gradient feature set $Y_l = \{P_l\}$ as described above, and the sparse coding matrix for the dictionary is calculated by the orthogonal matching pursuit (OMP). In the case where the training data set is large enough, i.e., the learned dictionary A_h is over-complete, the predicted smooth edge block should satisfy $Y_h = A_h X_l$. Then the high quality edge map E_h can be reconstructed by stacking patches obtained from Y_h.

B. IMPROVED BILATERAL FILTER

Bilateral filter, a filter with edge-preserving and denoising, is widely used in image super-resolution reconstruction. Its convolution kernel is the product of a spatial Gaussian kernel and another one related to intensity. The mathematical description of the bilateral filter weight is given below.

$$w_{p,q} = exp(-\frac{\|p, q\|^2}{2\sigma_s^2}) exp(-\frac{\|G_p, G_q\|^2}{2\sigma_r^2}), \qquad (3)$$

where p and q denote the positions of two pixels, whose intensity are G_p and G_q, respectively, and $w_{p,q}$ stands for the kernel weight related to the space standard deviation σ_s and the intensity standard deviation σ_r.

By introducing a range kernel, $w_{p,q}$ is revised into

$$w_{p,q} = exp(-\frac{\|p, q\|^2}{2\sigma_s^2}) exp(-\frac{\|G_p, G_q\|^2}{2\sigma_r^2}) f(E_h, p, q), \quad (4)$$

where $f(\cdot, \cdot, \cdot)$ is the range kernel, and defined as:

$$f(E, u, v) = \begin{cases} 1 & \text{if } u \text{ and } v \text{ are at the same side of } E, \\ 0 & \text{otherwise.} \end{cases} \qquad (5)$$

Based on the connectivity between surrounding pixels and the center pixel p, the high resolution depth image D_h can be interpolated using the above improved bilateral filter. For each pixel p in D_h, we take

$$D_h(p) = 1/k_p \sum_{q \in \Omega_p} w_{p,q} D_l(q_\downarrow), \qquad (6)$$

where k_p is a normalizing factor, Ω_p is a neighborhood window centered at pixel p, q is an adjacent pixel in the neighborhood window, $w_{p,q}$ is obtained from (3), $q_\downarrow$ denotes the corresponding position of q in the low resolution image D_l.

With the guidance provided by the high resolution edge map, only pixels at the same side of edges are considered when performing interpolation. Therefore the edge structures can be effectively preserved.

C. CONNECTIVITY DETERMINATION

A 4-neighborhood finding algorithm is implemented to determine whether two pixels p and q are at the same side of an edge. To demonstrate that algorithm, an example with the patch size of 5×5 is shown in Figure 2. No matter what an edge looks like, only one connected region is used to perform interpolation and keep the depth image as sharp as possible. Two cases are considered below.
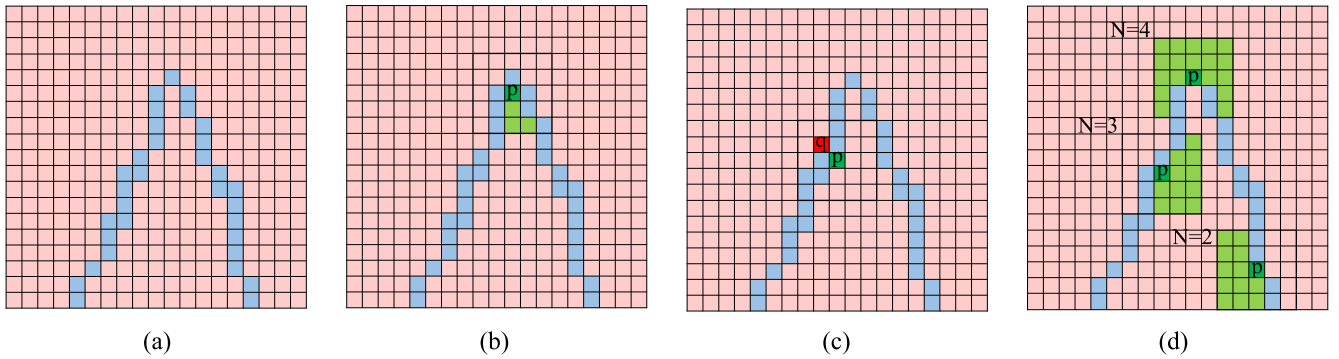
FIGURE 2. (a) Original edge pattern. (b) Illustration of the case that *p* is not an edge pixel. S_4^p consists of green pixels. (c) A special case of two pixels which could be mistakenly classified as at the same side of the edge simply based on a 8-connected pixel set. (d) Illustration of the case that *p* is on the edge.

1) ***p* is not an edge pixel.**

As shown in Figure 2(b), the center pixel *p* is not on the edge. A set of 4-connected pixels of *p*, denoted by S_4^p, is then obtained by the traversal search within the patch. Therefore,

$$f(E, p, q) = \begin{cases} 1 & \text{if } q \in S_4^p, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

The 4-connected pixel set S_4^p, instead of a 8-connected pixel set, is chosen to avoid the situations shown in Figure 2(c), where obviously pixel *p* and *q* are on the different sides, but mistakenly classified as $q \in S_4^p$ based on a 8-connected pixel set.

2) ***p* is an edge pixel**

If *p* is on the edge, it is actually ambiguous to decide whether *p* and *q* are at the same side or not. Based on *N*, the number of non-edge pixels in $S_4(p)$ which denotes the 4-neighborhood set of *p*, it can be further divided into 5 cases, i.e., $N = 0, 1, 2, 3, 4$. For completeness, if there is no pixel at the same side with *p* in case 1) or N = 0 here, we simply use the corresponding pixel value generated by the bicubic interpolation as the up-sampled value of *p*.

For cases of $N \neq 0$, we first define *t*, which satisfies the following optimization equation (8),

$$\max_t \quad Card(S_4^t),$$
$$s.t. \quad t \in S_4^p,$$
$$t \text{ is non} - \text{edge pixel}, \quad (8)$$

where $Card(S_4^t)$ denotes the cardinality of the set S_4^t. Then, S_4^p is defined as,

$$S_4^p = S_4^t \cup p, \quad (9)$$

where S_4^p satisfies (7).

If $N = 1$, it is straightforward to acquire S_4^p. Some examples with $N = 2, 3, 4$ are shown in Figure 2(d). Once we determine whether two pixels are at the same side of the edge, the high resolution image can be interpolated using (6).

IV. EXPERIMENTAL RESULTS

This section presents both quantitative and qualitative assessment of the proposed algorithm and existing depth image super-resolution algorithms. In the experiments, Middlebury 2014 [19], a popular depth image dataset, was used as the training set. From Middlebury 2014, 5×5 edge patch pairs were extracted and rotated to produce over 400,000 training samples which can ensure the good completeness of the learned dictionary. To obtain low resolution images, we down-sampled their corresponding high resolution ones. The trained results were verified in the Middlebury 2003 and 2006 datasets [20]. The experimental results confirm the effectiveness of the proposed algorithm. Note that the following qualitative and quantitative experimental results were obtained at the scaling factor of 4.

A. QUALITATIVE RESULTS

To qualitatively evaluate the performance in Figure 3, we firstly provide the ground-truth edge maps of test images "artl", "teddy" and "cones" and their reconstructed ones (with the scaling factor of 4), respectively. As shown in Figure 3(b) and 3(c), the edges of bicubic interpolated images are jagged and have a large number of artifacts. The shock filter can effectively eliminate artifacts.From these images, it can be observed that our reconstructed edges can not only avoid blurred edges, but also help reduce zigzags near edges. Note that both Figure 3(d) and 3(f) have slight edge dilation caused by overlapping between nearby edge patches.

Then Figure 4 and Figure 5 show ground-truth edge maps and depth images along with their zoomed-in details so that we can visually evaluate the obtained results. We see that our proposed algorithm achieves more visually appealing results both on the reconstructed edge maps and depth images.

B. QUANTITATIVE RESULTS

We use root-mean-square error (RMSE), percent of error (PE) and peak signal-to-noise ratio (PSNR) as metrics to quantitatively compare the state-of-art algorithms and our algorithm. PE is calculated as the percentage of pixels whose absolute
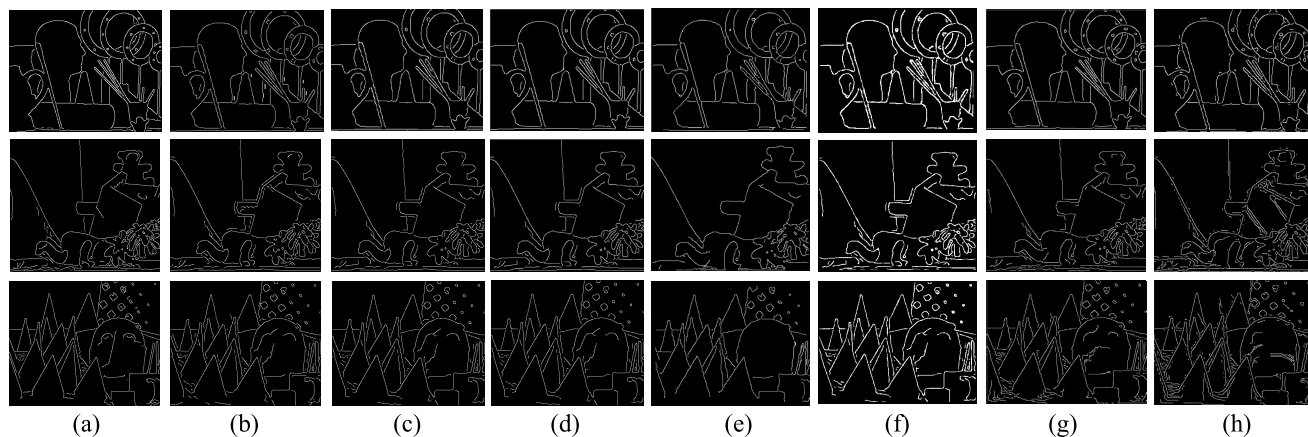
FIGURE 3. Edge maps of "atrl", "teddy" and "cones" at the scaling factor of 4. (a) Ground truth. (b) bicubic. (c) shock filter. (d) our algorithm. (e) Xie [10]. (f) ZHOU [13]. (g) Kim [5]. (h) Zeyde [21].
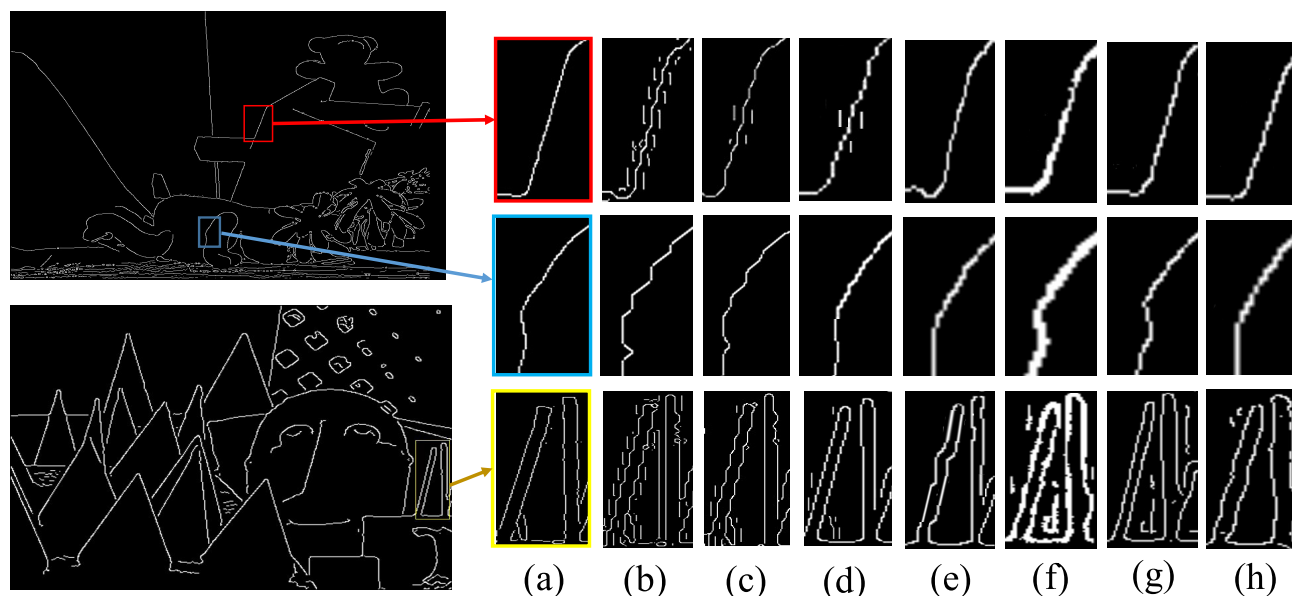

FIGURE 4. Zoomed-in edge maps. (a) Ground truth. (b) bicubic. (c) shock filter. (d) our algorithm. (e) Xie [10]. (f) ZHOU [13]. (g) Kim [5]. (h) Zeyde [21].
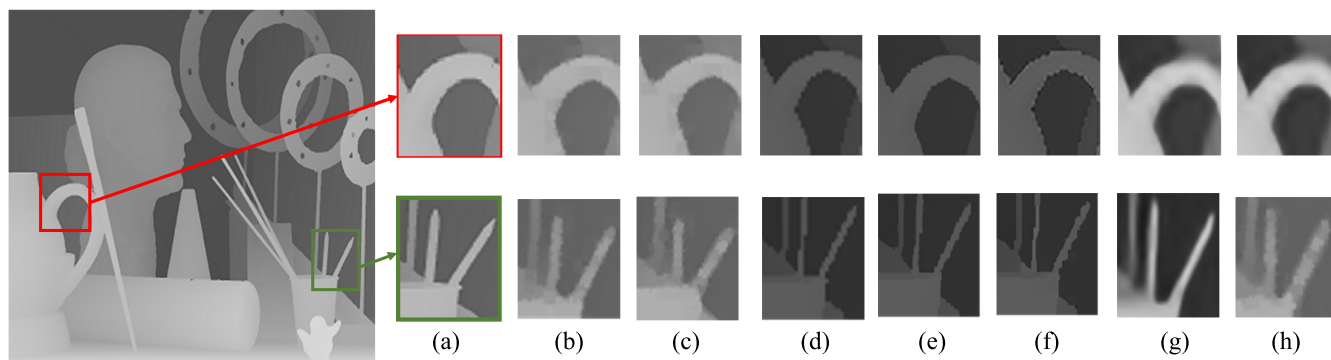
FIGURE 5. Zoomed-in depth images. (a) Ground truth. (b) bicubic. (c) shock filter. (d) our algorithm. (e) Xie [10]. (f) ZHOU [13]. (g) Kim [5]. (h) Zeyde [21].

difference value with respect to ground truth exceeds 1. PSNR characterizes the similarity between images. Although PSNR is widely used in color image super-resolution, it is not very appropriate for depth image because PSNR can be dominated by error pixels on the boundaries in depth images. A depth map with blurred boundaries may get a better PSNR score

TABLE 1. Root-mean-square error (RMSE).

RMSE(x4)	bowling	aloe	cones	Indian	venus	warrior	tsukuba	hand
XIE [10]	1.766	2.583	1.240	0.771	0.617	4.081	3.009	1.926
ZHOU [13]	1.623	2.217	1.214	0.703	0.553	3.316	2.638	1.654
ours	**1.175**	**1.718**	**0.960**	**0.552**	**0.416**	2.319	**2.03**	1.25
baseline	1.839	2.351	1.624	1.218	0.579	2.605	2.333	**1.186**
Kim [5]	2.238	3.245	1.778	0.987	0.845	4.424	3.505	2.174
Bicubic up-sample	1.809	2.169	1.397	0.999	0.887	2.886	2.543	1.759
Zeyde [21]	1.803	2.329	1.338	0.798	0.635	3.62	2.844	1.832

TABLE 2. Percent of error (PE).

PE(x4)	bowling	aloe	cones	Indian	venus	warrior	tsukuba	hand
XIE [10]	2.405	8.299	**2.829**	0.951	**0.505**	2.575	4.239	0.918
ZHOU [13]	**2.365**	8.242	3.154	0.943	0.641	2.593	4.217	0.912
ours	3.339	9.507	4.536	1.420	1.014	5.91	9.068	2.972
baseline	3.962	10.029	4.700	1.755	1.216	6.322	10.211	3.429
Kim [5]	4.232	13.454	6.993	1.795	1.696	8.077	11.34	2.624
Bicubic up-sample	4.892	12.337	7.367	1.786	2.247	6.509	13.429	3.440
Zeyde [21]	6.041	15.751	7.968	2.208	2.447	8.228	14.24	4.036

TABLE 3. Peak signal-to-noise ratio (PSNR).

PSNR(x4)	bowling	aloe	cones	Indian	venus	warrior	tsukuba	hand
XIE [10]	42.12	39.33	46.26	50.38	52.32	35.92	38.56	42.45
ZHOU [13]	43.31	41.22	46.33	51.21	53.53	37.75	39.70	43.55
ours	**46.73**	**43.43**	**48.49**	**53.29**	**55.74**	**40.83**	**41.97**	**46.22**
baseline	44.00	40.77	45.82	51.84	53.55	38.88	39.97	44.66
Kim [5]	40.67	37.76	43.13	48.24	49.59	35.21	37.24	41.40
Bicubic up-sample	42.79	39.18	44.44	48.58	50.36	36.22	37.51	41.48
Zeyde [21]	43.01	40.78	45.60	50.09	52.07	36.95	39.05	42.88

than a depth map with sharp borders and some pixel misalignment. Results of XIE, ZHOU, KIM, and Zeyde were cited from [13]. The test images [20] and parameters of our algorithm are consistent with the ones used by Zhou *et al.* [13]. Therefore, the comparison of these results is fair. As for the input low resolution test images, we obtained them by down-sampling their ground-truth high resolution counterparts.

To demonstrate the validity of the proposed algorithm, we evaluated the reconstructed results at the scaling factor of 4 in terms of the above metrics. The experimental results are shown in Table 1-3. To enhance readability, we marked the top three reconstruction algorithms. More specifically, the value in bold is the best, the underlined value stands for the second or the third best.

As shown in Table 1 and Table 3 our algorithm almost outperforms other algorithms on all test images in terms of RMSE and PSNR, which indicates that our reconstructed depth images are the closest to the ground truth. Note that PSNR is dominated by error pixels in the edge regions. As far as PE is concerned, XIE and ZHOU achieve the best results on test images. The effectiveness of our algorithm is verified from two aspects. On the one hand, based on the same edge guidance filter, we determine bicubic edges as the baseline of the proposed algorithm and use the edges of the bicubic up-sampled depth image to guide the interpolation. As shown in Table 2, the proposed algorithm is superior to the bicubic algorithm and the bicubic edge guidance algorithm. On the

other hand, Kim and Zeyde algorithms have not yet reached the baseline, but our edge-guided reconstruction algorithm greatly improves the performance. Kim *et al.* [5] is a CNN based method which achieves pretty good performance in color image super-resolution. This also demonstrates that color image super-resolution method may not be simply applied to depth image super-resolution.

V. CONCLUSION AND FUTURE WORK

This paper proposes a depth image super-resolution algorithm based on KSVD. Our algorithm is made up of two parts, smooth edge reconstruction and guided filtering based on edge connectivity. Our algorithm can produce smooth and natural edge structure of the reconstructed super-resolution depth images due to its shock filtering and sparse dictionary reconstruction, and enforce the subsequent edge-guided interpolation being more precise. Moreover, it improves the bilateral filter based on connectivity and can produce sharp edges. Our algorithm has been extensively tested on the Middlebury dataset and compared with some existing state-of-the-art algorithms. The quantitative and qualitative experimental results confirm the performance superiority of the proposed algorithm.

There still leaves much to be desired in our proposed algorithm. For example, our algorithm performs pixel level edge reconstruction which smooths both the true edge and the false ones caused by up-sampling. However, smoothing false edge does not help improve the quality of depth images or even makes it worse. In the future, we plan to add the distribution prior of the false edges at the training stage and restore more accurate edge maps.

REFERENCES

[1] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.

[2] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Process.*, vol. 3, no. 3, pp. 233–242, May 1994.

[3] M. Elad and A. Feuer, "Super-resolution reconstruction of image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 9, pp. 817–834, Sep. 1999.

[4] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[5] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[6] Y. Wen, B. Sheng, P. Li, W. Lin, and D. D. Feng, "Deep color guided coarse-to-fine convolutional network cascade for depth image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 994–1006, Feb. 2019.

[7] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[8] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.

[9] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 111–126,".

[10] J. Xie, R. S. Feris, and M.-T. Sun, "Edge-guided single depth image super resolution," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 428–438, Jan. 2016.

[11] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, May/Apr. 2002.

[12] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3D-TOF cameras," in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 1623–1630.

[13] D. Zhou, R. Wang, X. Yang, Q. Zhang, and X. Wei, "Depth image super-resolution reconstruction based on a modified joint trilateral filter," *Roy. Soc. Open Sci.*, vol. 6, no. 1, Jan. 2019, Art. no. 181074.

[14] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[15] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Int. Conf. Comput. Vis.*, 1998, pp. 839–846.

[16] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[17] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral up-sampling," *ACM Trans. Graph*, vol. 26, no. 3, 2007, Art. no. 96.

[18] G. Gilboa, N. Sochen, and Y. Y. Zeevi, "Image enhancement and denoising by complex diffusion processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1020–1036, Aug. 2004.

[19] D. Scharstein, H. Hirschmuller, and Y. Kitajima, "High-resolution stereo datasets with subpixel-accurate ground truth," in *Proc. German Conf. Pattern Recognit.* Cham, Switzerland: Springer, 2014, pp. 31–42.

[20] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Madison, WI, USA, vol. 1, Jun. 2003, pp. 195–202.

[21] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.* Berlin, Germany: Springer, 2010, pp. 711–730.

BINHUI LIU received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2019, where he is currently pursuing the M.E. degree. His research interests include computer vision and machine learning.

QIANG LING (Senior Member, IEEE) received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 1997, the M.E. degree from Tsinghua University, Beijing, China, in 2000, and the Ph.D. degree from the University of Notre Dame, Notre Dame, IN, USA, in 2005.

He was a Research Staff Member with Seagate Technology, from 2005 to 2008. He joined the University of Science and Technology of China, in 2008. He is currently a Professor with the Department of Automation. His research interests include networked control systems and signal processing. He is currently serving as an Associate Editor on the IEEE Control Systems Society Conference Editorial Board.

• • •