IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 12, 2017, accepted June 1, 2017, date of publication June 19, 2017, date of current version August 8, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2715981

Spatially Adaptive Tensor Total Variation-Tikhonov Model for Depth Image Super Resolution

GANG ZHONG[1], SEN XIANG[2], PENG ZHOU[1], AND LI YU[1]
[1]Huazhong University of Science and Technology, Wuhan 430074, China
[2]School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430070, China

Corresponding author: Gang Zhong (zhonggang@hust.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61231010 and Grant 61202301 and in part by the National High Technology Research and Development Program under Grant 2015AA015903.

ABSTRACT Depth images play an important role in 3-D applications. However, due to the limitation of depth acquisition equipment, the acquired depth images are usually in limited resolution. In this paper, a spatially adaptive *tensor total variation-Tikhonov* model is proposed to solve this problem. The tensor total variation regularization is adopted to maintain sharp edges that reflect latent discontinuities in the real world, while the Tikhonov regularization ensures that depth changes smoothly inside objects. Furthermore, a fused edge map is proposed to indicate edge regions and balance both regularization terms. In edge regions, tensor total variation regularization is predominant, thus *edge blurring* artifacts are suppressed. In non-edge regions, Tikhonov regularization plays a more important role to suppress *staircasing* artifacts. Specifically, texture edges are removed in the fused edge map, and *texture copying* artifacts are avoided. Experimental results demonstrate the effectiveness and superiority of the proposed framework. Moreover, the proposed method yields much sharper edges and a lower percentage of bad pixels.

INDEX TERMS Depth image super resolution, tensor total variation regularization, Tikhonov regularization.

I. INTRODUCTION

The recent years have witnessed the great development and popularity of 3D-based applications such as depth-based image rendering [1], 3DTV [2] and object reconstruction [3]. In these applications, it is of critical importance to acquire depth information. The most common methods of depth acquisition are from stereo matching and from special depth cameras. It is a long history to use stereo matching to obtain depth images with high resolution and high accuracy in computer vision. However, it is difficult to recover depth information for textureless areas and the computational load is quite heavy. Recently, thanks to the development of hardware, it is more common to use special devices, i.e., structure light and time of flight depth cameras, to obtain depth information. Unfortunately, due to the inherent limitation of physical sensors, the captured depth images are often in low resolution, which makes depth image super resolution quite necessary.

In the past years, many super resolution methods have been proposed to increase the resolution of depth images. Inspired by the development of super resolution for natural images [4], [5], early researchers solved the super resolution problem by fusing multiple depth measurements into one [6]. However, methods of this kind require a static scene thus limit their applications. Besides, the calibration and registration of multiple images are also complicated. A more general and flexible case is to upsample depth image from a single low resolution one. Zhong *et al.* [7] recovered high resolution depth image by exploiting the local smoothness property of the depth image itself, while some other researchers upsampled low resolution depth image by incorporating an external database of high resolution depth examplars [8]–[10].

Notice that both the color image and the depth image contain information about edges in the real world. There is a strong relationship between depth discontinuities and color edges [11]. By exploiting this property, many color guided depth super resolution methods have been proposed [12]. Generally, these methods can be generally classified into two categories: *filter* based and *optimization* based.

Filter-based methods interpolate unknown pixels with an interpolating filter, where the weights are designed carefully. Yang *et al.* [13] generated high resolution depth images by iteratively refining the input low resolution depth image with a bilateral filter. Kopf *et al.* [14] produced the high resolution solution by leveraging the guidance color image as a prior. Min *et al.* [15] increased low resolution depth image by using a weighted mode filter based on a joint histogram of

depth image and color image. Kim *et al.* [16] proposed a fast upsampling method based on a common edge region generating from the color and depth images. Pixels in the common edge region are detected and selected by minimizing a cost function, while other pixels are estimated by using bilinear interpolation. Liu *et al.* [17] found that the Euclidean distance may be disturbed by texture edges, and proposed to use geodesic distance to weight pixels nearby. More recently, an edge preserving filter named guided image filter was introduced in [18] and further extended in [19] by exploiting local gradient information in the depth map.

Optimization based methods formulate the super resolution problem as an optimization problem, where *a prior* reflecting the latent property of the desired results is utilized. Diebel and Tuhrun [20] first used Markov random field to model depth image and obtained the upsampling depth image by solving an energy function. The function consists of two quadratic potentials, measuring the reconstruction constraint and the depth smoothness prior. Jung et al. [21] improved this Markov random field model by introducing a confidence map, with which the prior model and the energy function are updated. In recent years, some other properties have also been used. Park *et al.* [22], [23] adopted a nonlocal structure regularization to maintain fine details and structures. Ferstl *et al.* [24] used total generalized variation to recover polynomial order results. Yang *et al.* [25] proposed to use an adaptive color guided autoregressive model exploiting nonlocal similarity in color image to recovery high quality depth images. In [26], Liu *et al.* adopted a robust penalty function to handle the inconsistency between color edges and depth discontinuities.

Generally, depth image reflects the latent geometry structure of a scene, where objects are separated by discontinuities and the depth inside objects usually changes smoothly. In this paper, a tensor total variation (TTV) is adopted firstly, to not only preserve sharp edges but also align those edges to the guided image. However, due to the inconsistency between color edges and depth discontinuities, *texture copying* artifacts are very likely to appear if a guided color image is directly utilized. Besides, in smooth regions, the *staircasing* artifacts exist due to the total variation nature of TTV. In order to handle these problems, a fused edge map is then proposed. The fused edge map is generated from the low resolution depth image and its corresponding high resolution color image, indicating discontinuities and balancing the tensor total variation regularization term and Tikhonov regularization term. With the fused edge map, the tensor total variation regularization is predominant in edge regions and sharp edges are preserved. In non-edge regions, Tikhonov regularization plays a more important role, making the depth changing smoothly and suppressing *staircasing* artifacts. Furthermore, texture edges in non-edge regions are removed thus *texture copying* artifacts are avoided.

The contributions of this paper are threefold:

1. Mathematical analysis of the tensor total variation in depth image upsampling is first proposed and how it aligns the edges in the recovered high resolution depth image to the guided image is presented.

2. A fused edge map is proposed to indicate discontinuities from low resolution depth image and the corresponding high resolution color image. With this fused edge map, *texture copying* artifacts introduced by texture edges will be avoided.

3. A spatially adaptive tensor total variation-Tikhonov model is proposed for depth image super resolution. Specially, a strong strength of tensor total variation regularization is enforced in edge regions while a strong strength of Tikhonov regularization is enforced in non-edge regions. In this way, a piecewise smooth high resolution depth image with sharp edges is recovered. In addition, we use a first order primal dual algorithm to solve this convex but not differentiable problem.

The rest of this paper is organized as follows: in Section II we introduce the mathematical model for depth image super resolution and several classical regularization terms. In Section III, we analyze the property of tensor total variation and propose to use a fused edge map to avoid some undesirable artifacts, after that the overall spatially adaptive TTV-Tikhonov model is introduced and a numerical solution is obtained. In order to demonstrate the effectiveness of the proposed method, we test our method on both synthetic scene and Middlebury datasets in Section IV. Finally, we conclude our work in Section V.

II. BACKGROUND

A. MATHEMATICAL MODEL

In this paper, we propose to recover the high resolution depth image u_h via solving the following energy function:

$$u_h = \arg\min_u \left\{ \frac{\mu}{2} f(u, u_0) + \Phi(u) \right\} \tag{1}$$

where f denotes the data fidelity function with u_0 being the observed depth image. The second term Φ is a regularization term and μ is a parameter balancing both terms. The first term is designed to make the solution be consistent with the observed image after degradation. Similar to [24], the data term in the proposed model is defined as follows:

$$f(u, u_0) = \int_\Omega w \, \|u - u_0\|_2^2 \, dx \tag{2}$$

where $w \in \{0, 1\}$ indicates the availability of the depth value in the corresponding position and $\Omega \subset \mathbb{R}^2$ is the image domain.

The regularization term Φ encodes the prior information about the desired solution and solves the ill-posed nature of the super resolution problem [5]. Many regularization terms have been proposed for natural images in recent years. For example, Tikhonov regularization exploits the piecewise smooth property while total variation regularization favors piecewise constant results. More recently, sparsity, non-local similarity and some other properties are also exploited in [27]–[30]. In the following paragraphs, we will list some relevant regularization terms.

B. TIKHONOV REGULARIZATION

The most well-known Tikhonov regularization [31] is defined as

$$Tikhonov(u) = \int_{\Omega} \|\Gamma u\|_2^2 \, dx \tag{3}$$

where u is the restored image. The commonly used Tikhonov matrix Γ is the first or second order derivative operator.

Tikhonov regularization has been widely used in many scenes [32], [33] due to its mathematical and computational simplicity. In spite of these advantages, Tikhonov regularization penalizes large gradients and favors smooth results, which is undesirable in some applications.

C. TOTAL VARIATION REGULARIZATION

In order to preserve sharp edges, many other regularization terms have been proposed. The most well-known one is total variation (TV) [34], which is defined as:

$$TV(u) = \int_{\Omega} \|\nabla u\|_2 dx \tag{4}$$

Compared with Tikhonov regularization, the high variations of u in (4) is not over-penalized. Therefore, sparse gradients are allowed in the solutions and sharp edges can be preserved. However, it is known that the total variation regularization favors sparse gradients and leads to a piecewise constant solution, resulting *staircasing* artifacts in smooth regions [35].

D. TOTAL GENERALIZED VARIATION REGULARIZATION

Total generalized variation (TGV) [36], proposed by Bredies *et al.*, is a generalization of total variation aiming to reduce *staircasing* artifacts by incorporating higher-order derivatives. The TGV assumes that the signal is piecewise polynomials. An order of k favors results which is a polynomial of degree less than k. The higher k is, the higher order smoothness of the solution will be. However, the computational cost will highly increase too. In many situations, an order of 2 produces visually appealing results. The primal definition of the second order TGV is formulated as:

$$TGV_{\alpha}^2(u) = \min_{v} \left\{ \alpha_1 \int_{\Omega} |\nabla u - v| dx + \alpha_0 \int_{\Omega} |\varepsilon(v)| dx \right\} \tag{5}$$

where v is an auxiliary vector field, α_0 and α_1 are fixed positive parameters. $\varepsilon(v) = (\nabla v + \nabla v^T)/2$ denotes the symmetrized derivative. For a more detailed analysis and discussion of TGV, we refer the readers to [36] and references therein for more details. Due to the well performance of recovering signals allowing jump discontinuities while avoiding *staircasing* artifacts, the TGV regularization has been used in many applications [24], [37], [38]. Especially, in [24], Ferstl *et al.* used the color guided total generalized variation regularization to recover piecewise affine surfaces.

III. SPATIALLY ADAPTIVE TTV-TIKHONOV MODEL

The above mentioned regularization terms utilize only a single image. However, in depth image upsampling, an auxiliary color image is usually applicable. Notice that discontinuities in depth image often coincide with color edges. Many color guided depth image super resolution methods have been proposed. In this section, a guided tensor total variation exploiting both the edge information in the guided image and the edge preserving property of total variation is first introduced and analyzed. However, using the color image to guide tensor total variation regularization directly suffers from the *texture copying* and *staircasing* artifacts. These artifacts are demonstrated and analyzed by using an analytic *Chess* scene later. After that, a fused edge map is proposed to solve these problems. Finally, the overall spatially adaptive tensor total variation-Tikhonov model is obtained and a numerical solution is derived to solve this convex but not differentiable problem.

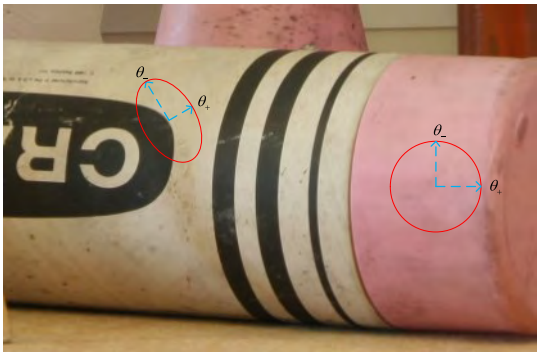
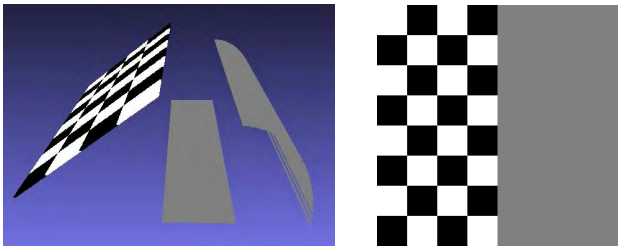


FIGURE 1. Some sample tensors to demonstrate the latent structure of an image.



(a) (b)

FIGURE 2. The analytic *Chess* scene. (a) Synthesized view of *Chess*. (b) Color image of *Chess*.

A. TENSOR AND TENSOR TOTAL VARIATION

It has been observed that the color edges and depth discontinuities usually appear together at boundaries of objects. A high resolution color image provides more accurate edge information such as the location and direction than the low resolution depth one. In order to exploit such information from the guided image while preserving the depth discontinuities, we propose to use a guided tensor total variation regularization term defined as:

$$TTV(u) = \int_{\Omega} \|T \nabla u\|_1 dx \tag{6}$$

where T is a symmetric, positive definite diffusion tensor capturing local information such as edge direction and strength.

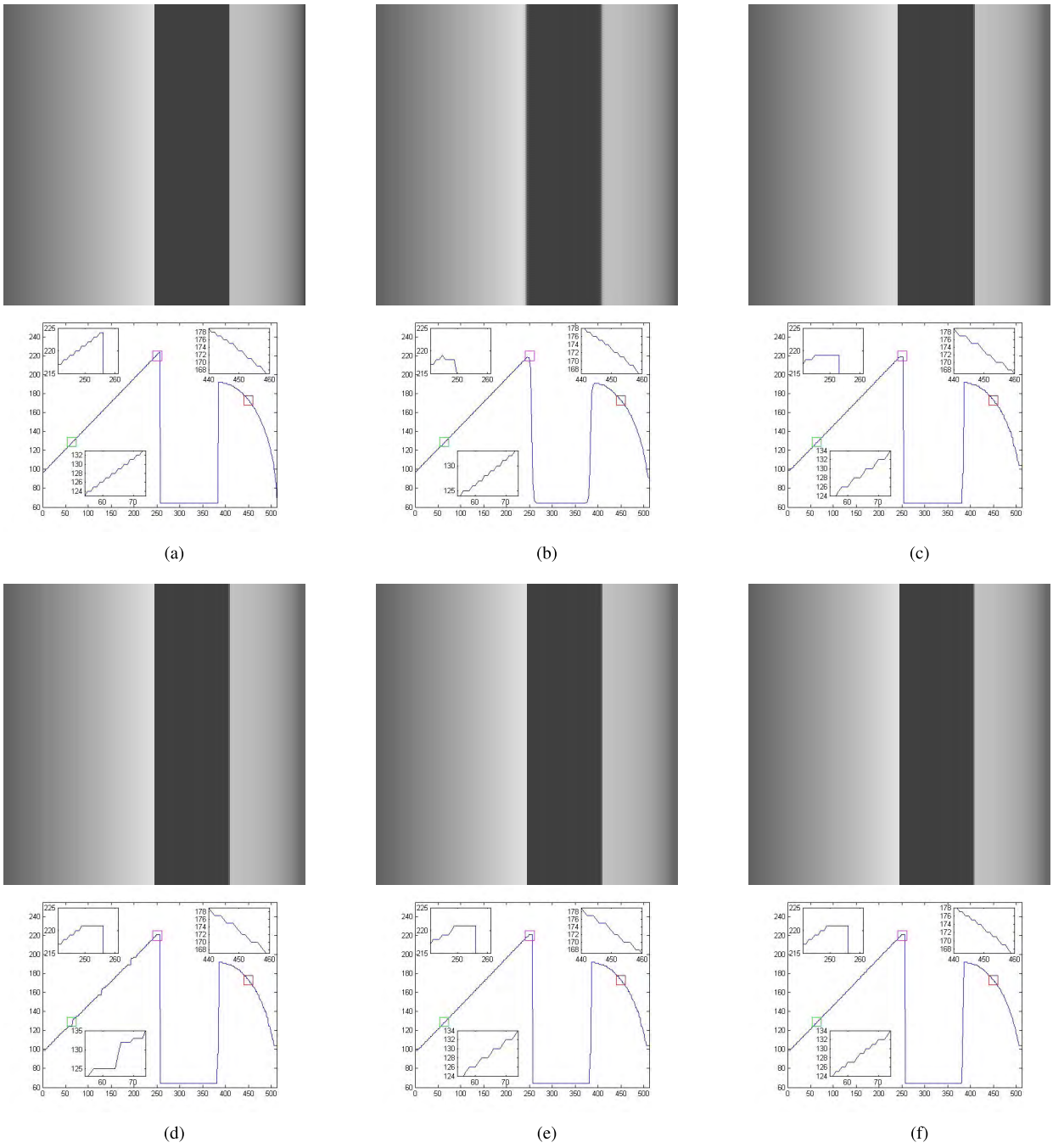


FIGURE 3. Visual comparison of 4× super resolution results on our synthetic scene *Chess*: (a) the groundtruth depth image. Super resolution results of (b) using Tikhonov regularization; (c) using total variation regularization; (d) using color guided tensor total variation regularization; (e) using fused edge map guided tensor total variation regularization; and (f) using the spatially adaptive tensor total variation - Tikhonov regularization.

In this paper, the proposed diffusion tensor is defined as:

$$T = \exp(-\beta|\nabla I_H|^{\gamma})\theta_+\theta_+^T + \theta_-\theta_-^T \tag{7}$$

where $\theta_+ = \nabla I_H / \|\nabla I_H\|_2$ is the normalized direction of the image gradient and θ_- is orthogonal to θ_+. I_H is the guided image and two parameters β, γ are used to adjust the magnitude and the sharpness of T, respectively. In the following we will analyze how this guided tensor total variation regularization aligns depth edges to color ones.

Generally, let λ_1, λ_2 be the eigenvalues of T, with $\lambda_1 \geq \lambda_2$, and v_1 and v_2 are the corresponding unit eigenvectors. From (7), we have $\lambda_1 = 1$, $\lambda_2 = \exp(-\beta|\nabla I_H|^{\gamma})$, $v_1 = \theta_-$ and $v_2 = \theta_+$. In Fig. 1, we visualize T as an ellipse, where the semi-major axis and semi-minor axis is λ_1 and λ_2. The directions of the semi-major and semi-minor axis are given by v_1 and v_2, respectively. We find that the eigenvalues can describe local structure: $\lambda_1 \approx \lambda_2$ in homogeneous regions, while $\lambda_1 >> \lambda_2$ at edge region and corner region.

In order to analyze the guided tensor total variation regularization, suppose n is an unit vector standing for an arbitrary 2D direction and $\nabla u = An$ where A is the amplitude of the gradient of u. Let $\omega \in (-\pi, \pi]$ denotes the angle between n and v_1. According to the eigendecomposition theorem, (6) can be expressed as

$$\begin{aligned} \|T\nabla u\|_1 &= \sqrt{(T\nabla u)^T(T\nabla u)} \\ &= A\sqrt{(\lambda_1^2 - \lambda_2^2)\cos^2\omega + \lambda_2^2} \end{aligned} \tag{8}$$

The above function implies that the value of tensor total variation is influenced by both the amplitude A and the angle ω. The minimum value is achieved for $\omega = \pm\pi/2$, which means a solution with the same direction as v_2 of the guided image is preferred. Thus edges in the recovered depth image are aligned to the guided ones. Besides, the value of TTV is also influenced by the minor eigenvalue of tensor T, reflecting the local geometry of the guided image.

B. ARTIFACTS ANALYSIS

In order to provide a more clearly analysis of different regularization terms, we make a synthetic *Chess* scene on which different regularization terms are applied. The *Chess* scene consists of a slant planes, a fronted planar plane and a curve plane. The synthesized view and the color image of the scene are shown in Fig. 2. The depth image of this scene is shown in Fig. 3(a). The Tikhonov regularization, TV regularization and color guided TTV regularization are adopted for comparison and the corresponding results are shown in Fig. 3(b)–(d). To clearly compare results of these methods, we plot the 200th line of the every resultant depth image. Especially, the magenta, red and green regions are enlarged to the top-left, top-right and bottom-left corner inside the figure, respectively, to offer a more intuitive comparison of different regularization terms.

We first compare discontinuities recovered by different regularization terms, especially the location and the sharpness. In Fig. 3(b), the *edge blurring* artifacts appear at boundaries. On the contrary, in Fig. 3(c), the sharpness of edges are better recovered due to the edge-preserving property of total variation regularization. Furthermore, by introducing the corresponding color image, a more accurate edge is recovered, as shown in the magenta region in Fig. 3(d).

Although color guided tensor total variation regularization recovers more accurate and sharper discontinuities, some undesired artifacts still exist in the recovered depth image. In this paper we pay more attention to the *staircasing* artifacts and the *texture copying* artifacts. The *staircasing* artifacts [35] are shown in the red region in Fig. 3(c), where the smooth depth becomes piecewise constant when using total variation. In these regions, the intensity also varies smoothly, (6) is degraded into the classic total variation and the *staircasing* artifacts appear in Fig. 3(d). On the contrary, in Fig. 3(b), Tikhonov regularization provides smooth results in these regions, which is further utilized in the proposed model.

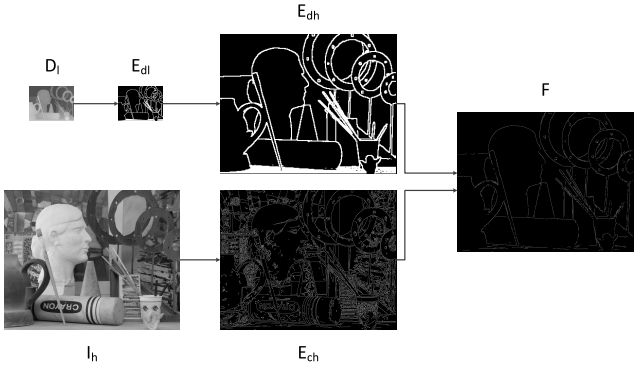


FIGURE 4. Generation of the fused edge map F from low resolution depth image D_l and corresponding high-resultion color image I_h.

When color guided tensor total variation regularization is applied, the minor eigenvalues of the tensors are spatially varying and the tensor total variation is degraded into a spatially weighted total variation: lower weights are around color edges while higher weights are in smooth regions. In this way, a smooth solution with jumps at texture edges will be preferred, resulting the *texture copying* artifacts [39]. As shown in the green regions in Fig. 3(d), the depth is changing smoothly while a texture edge exists. In this situation, a jump edge appears in the recovered depth image.

From the above analysis we find that the *texture copying* artifacts appear due to the incorrect guidance of color edges, thus using color edges to directly guide tensor is inappropriate. In order to suppress these artifacts, we propose to use a fused edge map to guide tensor instead of directly using color image. The fused edge map is calculated from both the color and the depth images. We hope this fused map is more similar to the true structure edges of depth image, eliminating texture edges. By using this map, the unmatched edges can be detected and *texture copying* artifacts will be eased. Furthermore, with the guidance of the fused edge map, smooth regions are also extracted, thus the Tikhonov regularization term can be enforced to alleviate the *staircasing* artifacts.

C. FUSED EDGE MAP

In this paper, the proposed fused edge map is generated as illustrated in Fig. 4. First, a color edge image and a depth edge image are extracted. The color edge image E_{ch} is obtained with Canny operator, while the depth edge image E_{dl} is identified with a neighborhood-statistic scheme. Specifically, given a pixel p and its neighborhood $N(p)$, p is regarded as an edge pixel if the difference between the maximum and minimum depth value in $N(p)$ is greater than a predefined threshold. After that the E_{dl} is upsampled to the same resolution of color image by using nearest interpolation. Finally, the fused edge map F is generated by fusing E_{dh} and E_{ch}.

From Fig. 4 we find that both E_{dh} and E_{ch} contain information of the latent discontinuities in real world. However, E_{dh} provides more accurate information of the appearance of discontinuities but is lack of location information. On the

contrary, E_{ch} has more accurate location information but also too many abundant useless texture edges. In order to utilize information of discontinuities from both the color image and the depth image, the fused edge map is derived according to the following rules:

Rule 1: Color edges and depth edges with similar directions are regarded as true edges and preserved in the fused edge map. Since color image introduces too many abundant edges, direction information is adopted to remove those fake edges in E_{ch}. Supposing θ_d and θ_c are the directions of an edge pixel in E_{dh} and E_{ch}, if $|\theta_d - \theta_c|$ is smaller than a predefined angle θ_{th}, this pixel is regarded as an edge pixel in the fused depth map.

Rule 2: Color edges with no correspondent depth edges are regarded as fake edges and discarded in the fused edge map. This situation happens mostly due to the abundant texture in color image. Using these fake edges will result in the *texture copying* artifacts, as analyzed in III-B.

Rule 3: Depth edges with no correspondent color edges are disregarded in the fused edge map. The edges in low resolution depth images are more likely to be the true boundaries. However, these depth edges are disregarded since the accuracy of the location information degrades rapidly as the magnification increases.

Rule 4: Non-edge regions in both depth and color images are regarded as non-edge regions in the fused edge map.

Following the above rules, we obtain a binary fused edge map F where pixels in edge regions are set to 1 while pixels in non-edge regions are set to 0. From Fig. 4 we can find that most true edges are preserved exactly and many abundant texture edges are removed.

D. SPATIALLY ADAPTIVE TTV-TIKHONOV MODEL

After the fused edge map is obtained, we use this map to adaptive balance the tensor total variation regularization and Tikhonov regularization at different regions. Specifically, the high resolution depth image u_h is obtained by

$$u_h = \arg\min_u \left\{ \int_{\Omega_H} \omega \|u - u_0\|_2^2 \, dx + \int_{\Omega_H} \alpha_1 \|T_F \nabla u\|_1 dx + \int_{\Omega_H} \frac{\alpha_2}{2} \|\nabla u\|_2^2 \, dx \right\} \quad (9)$$

where the $\|T_F \nabla u\|_1$ is the fused edge map guided tensor total variation term and $\|\nabla u\|_2^2$ is the Tikhonov regularization term. α_1 and α_2 are positive weights that balance TTV and Tikhonov regularization terms. $T_F \nabla u$ is defined as

$$T_F \nabla u = \begin{cases} T \nabla u, & F_u = 1 \\ \nabla u, & F_u = 0 \end{cases} \quad (10)$$

F_u indicates the current pixel u in the fused edge map F. Compared with (6), for those texture edges, the original total variation is used thus the weights are equal in the local neighborhood and a smooth result is preferred. Then the *texture copying* artifacts can be alleviate.

In order to suppress the *staircasing* artifacts, the Tikhonov regularization terms is imposed to non-edge regions. We use the fused edge map to separate edge regions and non-edge regions. For those edge regions, total variation is the dominant regularization thus edges can be preserved, for non-edge regions, the Tikhonov regularization strength the depth values to be smooth.

E. NUMERICAL SOLUTION

The proposed model is convex but not smooth, so simple minimization algorithms are not applicable. In order to solve such TV minimization problems, many methods are proposed in recent years [40]–[42]. In this paper, we use first order primal dual algorithm [43] to solve this problem. Applying the Legendre Fenchel transform to (9), we can reformulate it to the following convex-concave saddle point problem:

$$\min_u \max_{p,q} \alpha_1 \langle T_F \nabla u, p \rangle + \alpha_2 \langle \nabla u, q \rangle + \omega \|u - u_0\|_2^2 \quad (11)$$

where p and q are dual variables. The feasible sets of these variables are defined by

$$\begin{aligned} P &= \{p \in \mathbb{R}^2 | \, \|p_{ij}\| \le 1\} \\ Q &= \{q \in \mathbb{R} | \, \|q_{ij}\| \le 1\} \end{aligned} \quad (12)$$

Choosing steps $\sigma_p, \sigma_q, \tau_u > 0$ and setting initialization value $p^0, q^0 = 0, \bar{u}^n = u^n = u_0$. By iteratively updating primal and dual variables, this model can be solved as follows:

$$\begin{cases} p^{n+1} = P_p\{p^n + \sigma_p \alpha_1 T_F \nabla \bar{u}^n\} \\ q^{n+1} = (1 + \sigma_q \partial F^*)^{-1}(q^n + \sigma_q \sqrt{\alpha_2} \nabla \bar{u}^n) \\ u^{n+1} = \dfrac{u^n + \tau_u((\alpha_1 T_F \nabla)^* p^{n+1} + (\sqrt{\alpha_2}\nabla)^* q^{n+1} + \omega u_0)}{1 + \tau_u \omega} \\ \bar{u}^{n+1} = u^{n+1} + \theta(u^{n+1} - \bar{u}^n) \end{cases} \quad (13)$$

where $(\alpha_1 T_F \nabla)^*$ and $(\sqrt{\alpha_2}\nabla)^*$ have the adjointness property

$$\begin{cases} -u \cdot (\alpha_1 T_F \nabla)^*(p) = (\alpha_1 T_F \nabla)(u) \cdot p \\ -u \cdot (\sqrt{\alpha_2}\nabla)^*(q) = (\sqrt{\alpha_2}\nabla)(u) \cdot q \end{cases} \quad (14)$$

The projection operator P_p and the resolvent operator of F is defined as

$$P_p(\tilde{p}_{i,j}) = \frac{\tilde{p}_{i,j}}{\max(1, |\tilde{p}_{i,j}|)} \quad (15)$$

$$\begin{aligned} x &= (I + \tau \partial F)^{-1}(y) \\ &= \arg\min_x \left\{ \frac{\|x - y\|_2^2}{2\tau} + F(x) \right\} \end{aligned} \quad (16)$$

In (9) we have $F(\sqrt{\alpha_2}\nabla u) = \left\|\sqrt{\alpha_2}\nabla u\right\|_2^2 \Big/ 2$ thus

$$\begin{aligned} q^{n+1} &= (1 + \sigma_q \partial F^*)^{-1}(q^n + \sigma_q \sqrt{\alpha_2} \nabla \bar{u}^n) \\ &= \frac{q^n + \sigma_q \sqrt{\alpha_2} \nabla \bar{u}^n}{1 + \sigma_q} \end{aligned} \quad (17)$$

We update the above iteration until a stopping criterion is reached or the iteration reaches a predefined number. In our method, the relaxation parameter θ is set to 1 during the iteration.

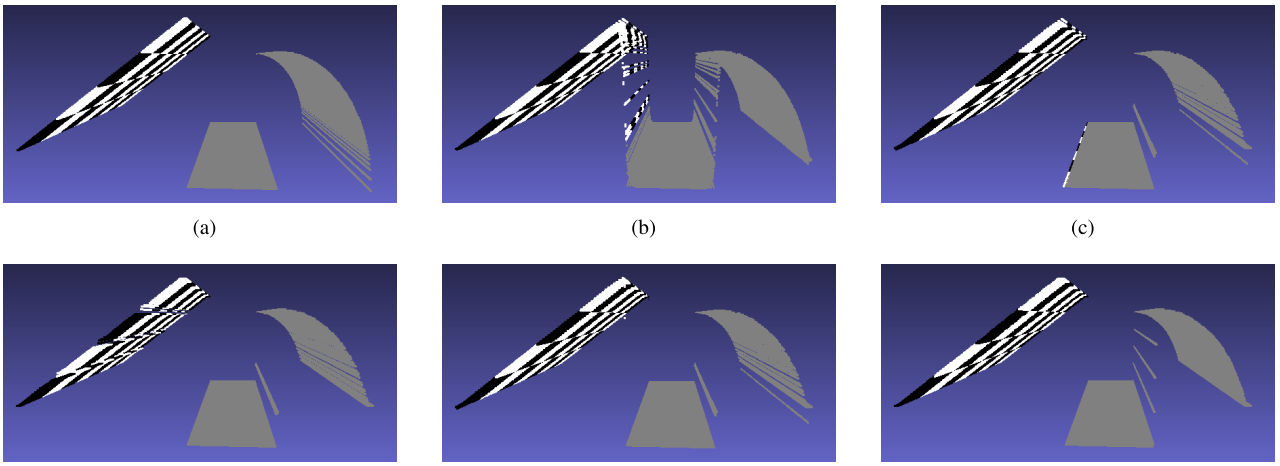


FIGURE 5. Visual comparison of 4× upsampling results on our synthetic scene *Chess*: The synthesize view using recovered depth image of (a) the groundtruth depth image; (b) using Tikhonov regularization; (c) using total variation regularization; (d) using color guided tensor total variation regularization; (e) using fused depth map guided tensor total variation regularization; and (f) using the spatially adaptive tensor total variation - Tikhonov regularization.

TABLE 1. Quantitative results (in MAD) on Middlebury datasets at three subsampling rates. The best results are in bold.

	Art			Books			Moebius		
	2×	4×	8×	2×	4×	8×	2×	4×	8×
Bilinear	0.56	1.09	2.10	0.19	0.35	0.65	0.20	0.37	0.70
MRF[20]	0.62	1.01	1.97	0.22	0.33	0.62	0.25	0.37	0.67
GF[18]	0.66	1.06	1.77	0.22	0.36	0.60	0.24	0.38	0.61
JBFcv[13]	0.57	0.70	1.51	0.30	0.45	0.64	0.39	0.48	0.69
NLM-WLS[22]	0.43	0.67	**1.07**	0.20	0.32	0.55	0.18	0.30	0.52
TGV[24]	0.47	0.69	1.23	0.18	0.27	**0.44**	0.19	0.30	**0.50**
Ours	**0.29**	**0.55**	1.19	**0.12**	**0.26**	0.58	**0.13**	**0.27**	0.60

TABLE 2. Quantitative results (in PBP with error threshold 1) on Middlebury datasets at three subsampling rates. The best results are in bold.

	Art			Books			Moebius		
	2×	4×	8×	2×	4×	8×	2×	4×	8×
Bilinear	5.13	9.61	18.16	2.07	4.11	8.31	2.73	5.42	10.65
MRF[20]	6.59	11.10	18.95	2.50	4.37	8.40	3.51	5.79	11.06
GF[18]	7.33	11.96	20.01	2.39	4.33	8.04	3.27	5.63	10.02
JBFcv[13]	3.67	7.24	21.87	2.87	8.79	10.41	4.97	7.15	12.08
NLM-WLS[22]	3.64	7.30	13.98	1.78	3.35	8.15	1.85	3.39	7.28
TGV[24]	3.73	5.35	10.98	1.60	2.39	**4.89**	2.08	3.57	7.00
Ours	**1.13**	**3.00**	**8.86**	**0.70**	**2.04**	7.53	**0.96**	**2.32**	**6.92**

IV. EXPERIMENTS

In order to demonstrate the effectiveness of our methods, we perform experiments with both synthetic scene and Middlebury datasets. We first compare the proposed adaptive regularization term with the classical regularization terms in synthetic *Chess* scene to demonstrate the superiority. Furthermore, extensive experiments on Middlebury datasets are used to test the proposed method with some state-of-the-art algorithms.

A. SYNTHETIC SCENE

We use the Tikhonov regularization, total variation regularization, color guided tensor total variation, fused edge map guided tensor total variation, and the proposed method for depth image super resolution for the synthetic *Chess* scene and the results are illustrated in Fig. 3. In order to give a clearly comparison of different super resolution methods, the synthesized views are shown in Fig. 5.

It can be observed in Fig. 3(b) that the Tikhonov regularization favors smooth solutions, and surface inside objects are well recovered while the discontinuities are over-smoothed. Thus *jagged* artifacts appear in Fig. 5(b). Due to the edge preserving property of the total variation, sharp discontinuities are better maintained but *staircasing* artifacts are introduced. Besides, note that without color images, the edges in the upsampled depth images are misaligned with color images, as shown in Fig. 5(c).

When a color image is introduced, as analyzed in Section III-B, the location of edges in the interpolated image are well aligned to color images in Fig. 3(d). However,

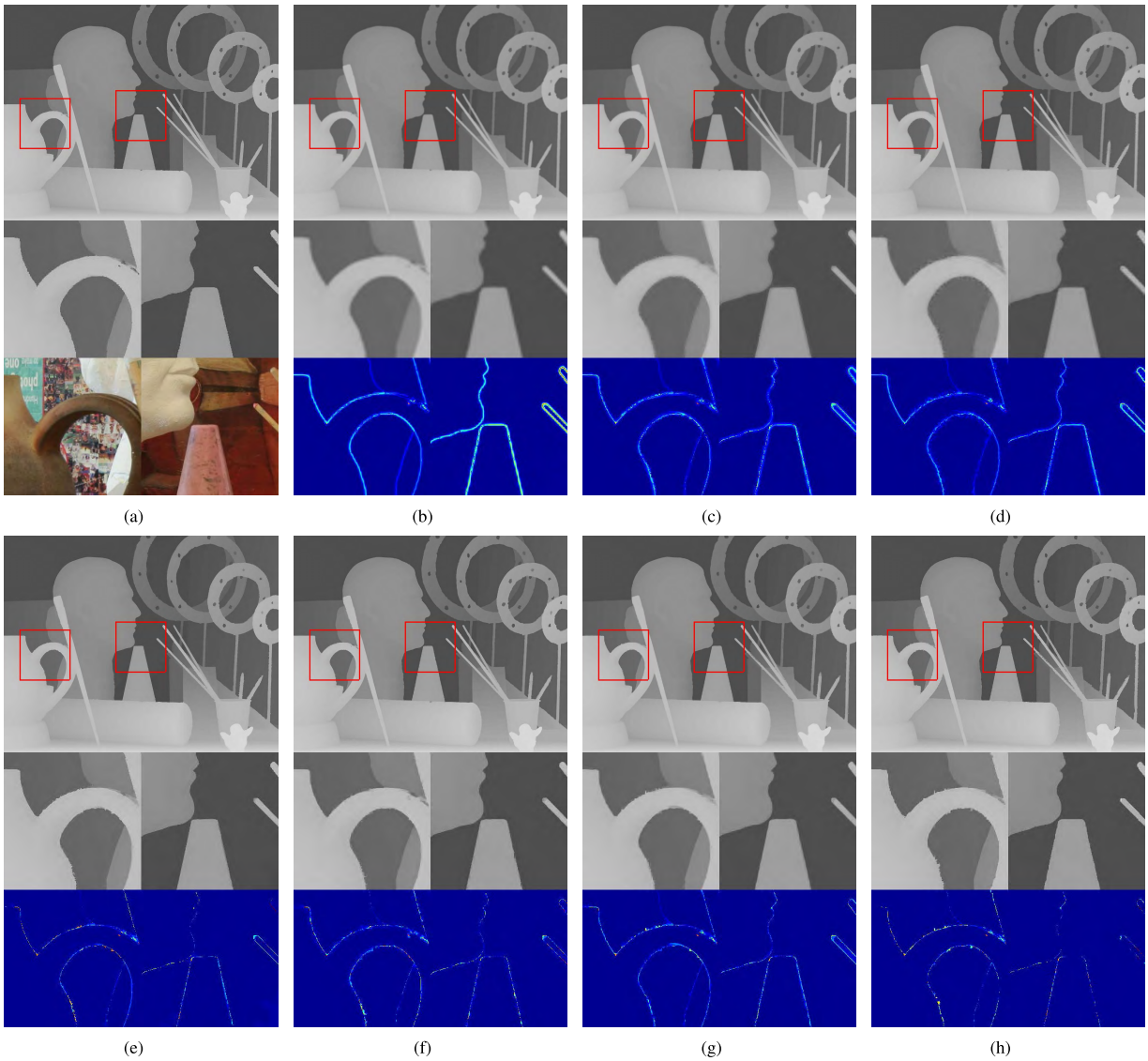


FIGURE 6. Visual comparison of 4× super resolution results on *Art*: (a) The groundtruth. Super resolution results of (b) Bilinear interpolation, (c) MRF [20], (d) GF [18], (e) JBFcv [13], (f) NLM-WLS [22], (g) TGV [24], and (h) the proposed method.

in smooth regions with homogeneous colors, *staircasing* artifacts still exist. Moreover, for those regions where color edges are inconsistent with depth discontinuities, *texture copying* artifacts appear in Fig. 3(d) and Fig. 5(d).

By introducing the fused edge map, texture edges in color images are eliminated thus *texture copying* artifacts can be effectively suppressed, as shown in Fig. 3(e). However, in Fig. 5(e), the *staircasing* artifacts still exist. In Fig. 3(f), due to the well balance between Tikhonov regularization and tensor total variation regularization in different regions, the discontinuities are well preserved and the recovered surface inside objects changes smoothly.

B. MIDDLEBURY DATASETS

We also test our model on Middlebury datasets [44]. Three datasets *Art*, *Books* and *Moebius* are used for evaluation. We compare the proposed methods with six state-of-the-art methods: Bilinear interpolation, MRF-based method [20], guided image filtering (GF) [18], joint bilateral filtering on cost volumes (JBFcv) [13], the nonlocal means regularized weighted least square (NLM-WLS) [22], total generalized variation (TGV) [24]. Since the original resolutions of depth image and color image are different, we crop and cut those input images to the same resolution. The low resolution depth images are downsampled from groundtruth depth image directly and three upsampling factors are tested in this paper.

Table 1 shows the quantitative results of the three datasets in terms of Mean Absolute Difference (MAD). From this table we find that our method gives lower MAD, especially at small upsampling factors. Furthermore, another widely used metric to evaluate the quality of depth image in stereo matching [45] named the Percentage of Bad Pixels (PBP) is

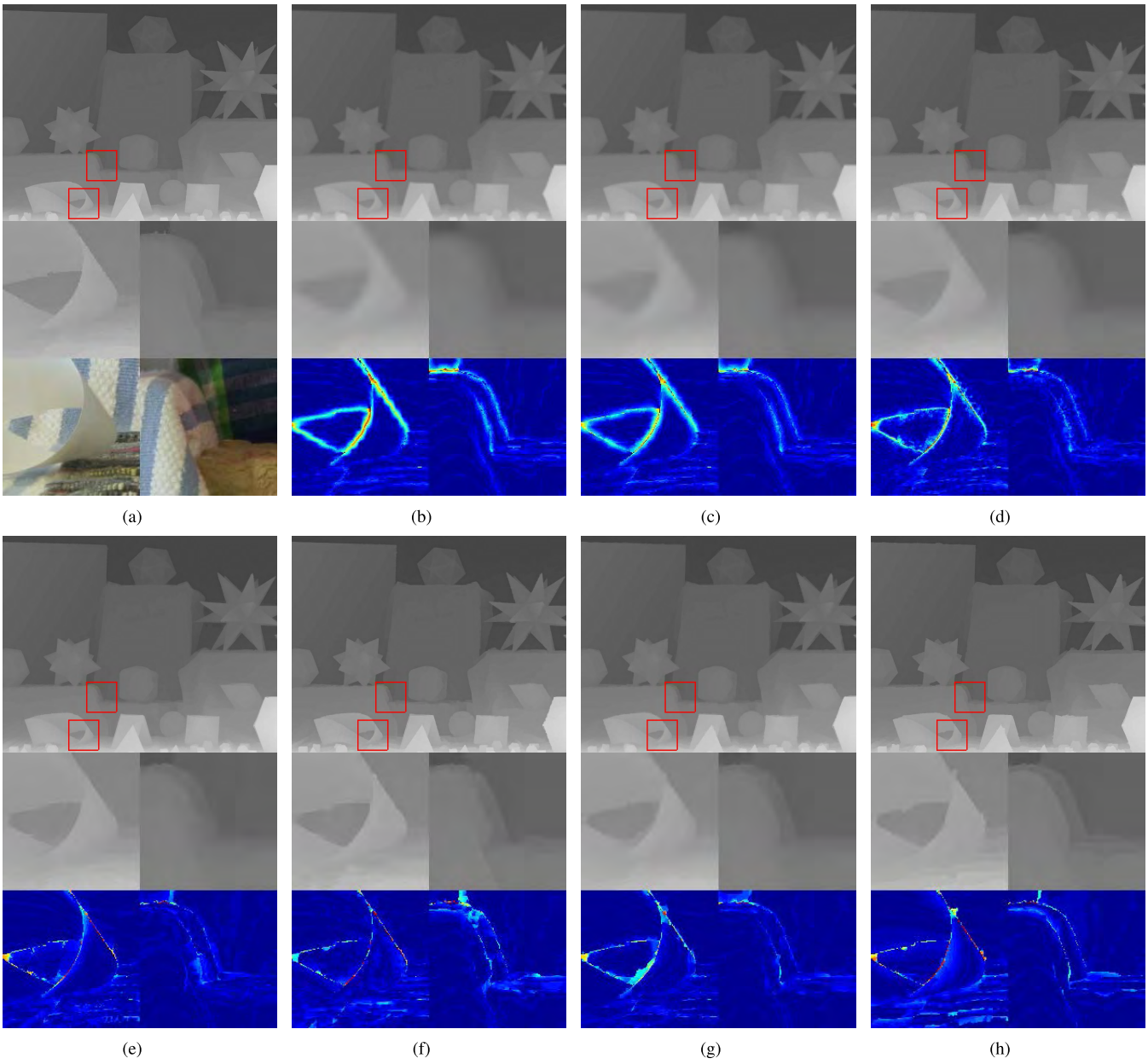


FIGURE 7. Visual comparison of 8× super resolution results on *Moebius*: (a) The groundtruth. Super resolution results of (b) Bilinear interpolation. (c) MRF [20]. (d) GF [18]. (e) JBFcv [13]. (f) NLM-WLS [22]. (g) TGV [24]. (h) The proposed method.

adopted to evaluate different methods in this paper. Quantitative results in PBP of the proposed method and other methods are reported in Table 2. Occasionally, the method in [24] yields lower PBP than the propose method. The reason lies in the fact that TGV favors piecewise affine solutions, which is quite appropriate for *Books* dataset. In contrast, the proposed method is more general and achieves better performance in most cases.

Fig. 6 shows illustration of 4× upsampling results on *Art*. In order to show a more clearly comparison, the enlarged depth image and error maps of the given regions are also shown. In Fig. 6, bilinear interpolation gives smooth results at boundaries. By introducing the auxiliary color image, sharper edges are recovered. Comparing those error maps in Fig. 6, we find that the proposed method obtains results with more accurate edges due to the fused edge map and the edge preserving property of total variation.

We also give 8× upsampling results on *Moebius* in Fig. 7. Compared with other methods, the proposed method recovers the sharpest strip and tablecloth.

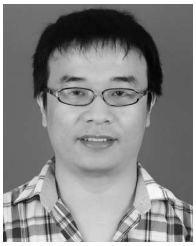
V. CONCLUSION

In this paper, a spatially adaptive tensor total variation-Tikhonov regularization model is proposed to recover

piecewise smooth depth image with sharp edges. In order to solve the convex but not differentiable model, a first order primal dual algorithm is adopted. Quantitative results show that the proposed method obtain high resolution depth image with lower percentage of bad pixels and mean absolute difference. Furthermore, the visual comparison demonstrates that the proposed method obtains sharper edges in the recovered depth images.

REFERENCES

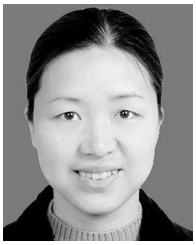
[1] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," *Proc. SPIE*, vol. 5291, pp. 93–104, May 2004.
[2] S. Schwarz, R. Olsson, and M. Sjöström, "Depth sensing for 3DTV: A survey," *IEEE Multimedia Mag.*, vol. 20, no. 4, pp. 10–17, Oct. 2013.
[3] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Jun. 2006, pp. 519–528.
[4] K. Nasrollahi and T. B. Moeslund, "Super-resolution: A comprehensive survey," *Mach. Vis. Appl.*, vol. 25, no. 6, pp. 1423–1468, 2014.
[5] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.
[6] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "Lidarboost: Depth superresolution for ToF 3D shape scanning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 343–350.
[7] G. Zhong, L. Yu, and P. Zhou, "Edge-preserving single depth image interpolation," in *Proc. Vis. Commun. Image Process. (VCIP)*, Sep. 2013, pp. 1–6.
[8] J. Xie, R. Feris, S.-S. Yu, and M.-T. Sun, "Joint super resolution and denoising from a single depth image," *IEEE Trans. Multimedia*, vol. 17, no. 9, pp. 1525–1537, Sep. 2015.
[9] 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.
[10] O. Mac Aodha, N. Campbell, A. Nair, and G. Brostow, "Patch based synthesis for single depth image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, pp. 71–84, 2012.
[11] C.-C. Su, L. K. Cormack, and A. C. Bovik, "Color and depth priors in natural images," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2259–2274, Jun. 2013.
[12] I. Eichhardt, D. Chetverikov, and Z. Jankó, "Image-guided ToF depth upsampling: A survey," *Mach. Vis. Appl.*, vol. 28, no. 3, pp. 267–282, 2017.
[13] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2007, pp. 1–8.
[14] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, p. 96, Jul. 2007.
[15] D. Min, J. Lu, and M. N. Do, "Depth video enhancement based on weighted mode filtering," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1176–1190, Mar. 2012.
[16] K. Sung-Yeol and H. Yo-Sung, "Fast edge-preserving depth image upsampler," *IEEE Trans. Consum. Electron.*, vol. 58, no. 3, pp. 971–977, Mar. 2012.
[17] M. Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 169–176.
[18] 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.
[19] K. L. Hua, K. H. Lo, and Y. C. F. F. Wang, "Extended guided filtering for depth map upsampling," *IEEE Multimedia Mag.*, vol. 23, no. 2, pp. 72–83, Feb. 2016.
[20] J. Diebel and S. Thrun, "An application of Markov random fields to range sensing," in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, pp. 291–298.
[21] J. I. Jung and Y. S. Ho, "Depth image interpolation using confidence-based Markov random field," *IEEE Trans. Consum. Electron.*, vol. 58, no. 4, pp. 1399–1402, Nov. 2012.
[22] J. Park, H. Kim, Y. Tai, M. S. Brown, and I. S. Kweon, "High-quality depth map upsampling and completion for RGB-D cameras," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5559–5572, Dec. 2014.
[23] J. Park, H. Kim, T. Yu-Wing, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3d-ToF cameras," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 1623–1630.
[24] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sep. 2013, pp. 993–1000.
[25] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3443–3458, Aug. 2014.
[26] W. Liu, X. Chen, J. Yang, and Q. Wu, "Robust color guided depth map restoration," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 315–327, Jan. 2017.
[27] 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.
[28] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.
[29] J. Jiang, X. Ma, Z. Cai, and R. Hu, "Sparse support regression for image super-resolution," *IEEE Photon. J.*, vol. 7, no. 5, pp. 1–11, Oct. 2015.
[30] J. Jiang, X. Ma, C. Chen, T. Lu, Z. Wang, and J. Ma, "Single image super-resolution via locally regularized anchored neighborhood regression and nonlocal means," *IEEE Trans. Multimedia*, vol. 19, no. 1, pp. 15–26, Jan. 2017.
[31] A. N. Tikhonov, V. I. Arsenin, and F. John, *Solutions of Ill-Posed Problems*, vol. 14. Washington, DC, USA: Winston, 1977.
[32] V. B. S. Prasath, D. Vorotnikov, R. Pelapur, S. Jose, G. Seetharaman, and K. Palaniappan, "Multiscale tikhonov-total variation image restoration using spatially varying edge coherence exponent," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5220–5235, Dec. 2015.
[33] W. Li, Q. Du, and M. Xiong, "Kernel collaborative representation with tikhonov regularization for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 48–52, Jan. 2015.
[34] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.
[35] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," *Math. Models Comput. Vis.*, vol. 17, no. 2, 2005.
[36] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 492–526, 2010.
[37] T. Zhen, J. Xun, Y. Kehong, P. Tinsu, and B. J. Steve, "Low-dose CT reconstruction via edge-preserving total variation regularization," *Phys. Med. Biol.*, vol. 56, no. 18, p. 5949, 2011.
[38] R. Ranftl, K. Bredies, and T. Pock, "Non-local total generalized variation for optical flow estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 439–454.
[39] D. Chan, H. Buisman, C. Theobalt, and S. Thrun, "A noise-aware filter for real-time depth upsampling," in *Proc. Workshop Multi-Camera Multi-Modal Sensor Fusion, Algorithms Appl.*, 2008.
[40] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.
[41] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1, pp. 89–97, 2004.
[42] C. Wu and X.-C. Tai, "Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 300–339, 2010.
[43] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.
[44] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. IEEE Int. Conf. CVPR*, Jun. 2007, pp. 1–8.
[45] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.



GANG ZHONG received the B.S. degree in communication engineering from Huazhong University of Science and Technology, Wuhan, China, in 2008. His research interests are video coding and image processing, focusing on depth image super resolution.

SEN XIANG (S'11) received the B.S. and Ph.D. degrees from Huazhong University of Science and Technology, Wuhan, China in 2010 and 2016, respectively. He was a Visiting Scholar with the University at Buffalo, State University of New York, USA, from 2013 to 2014. He is currently a Lecturer with the School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan. His research interests include depth image processing, 3-D video, and structured light depth acquisition and related areas.

PENG ZHOU received the B.S. degree in communication engineering from Taiyuan University of Technology in 2004, and the M.S. degree in communication and information systems from Yangtze University, Jingzhou, China, in 2009. He is currently pursuing the Ph.D. degree with Huazhong University of Science and Technology. His research interests are plenoptic sampling and spectral analysis, focusing on sampling theory in 3-D.



LI YU received the B.S. degree in electronics and information engineering, the M.S. degree in communication and information system, and the Ph.D. degree in electronics and information engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 1995, 1997, and 1999, respectively. In 2000, she joined the Electronics and Information Engineering Department, HUST, where she has been a Professor since 2005. She was a Co-Sponsor of the China AVS Standard Special Working Group, where she was involved as the Key Member. Her team has applied more than ten related patents and submitted 79 proposals to AVS standard organization. Her research interests include multimedia communication and processing, computer network, and wireless communication.