

Robust Color Guided Depth Map Restoration

Wei Liu, Xiaogang Chen, Jie Yang, and Qiang Wu

Abstract—One of the most challenging issues in color guided depth map restoration is the inconsistency between color edges in guidance color images and depth discontinuities on depth maps. This makes the restored depth map suffer from texture copy artifacts and blurring depth discontinuities. To handle this problem, most state-of-the-art methods design complex guidance weight based on guidance color images and heuristically make use of the bicubic interpolation of the input depth map. In this paper, we show that using bicubic interpolated depth map can blur depth discontinuities when the upsampling factor is large and the input depth map contains large holes and heavy noise. In contrast, we propose a robust optimization framework for color guided depth map restoration. By adopting a robust penalty function to model the smoothness term of our model, we show that the proposed method is robust against the inconsistency between color edges and depth discontinuities even when we use simple guidance weight. To the best of our knowledge, we are the first to solve this problem with a principled mathematical formulation rather than previous heuristic weighting schemes. The proposed robust method performs well in suppressing texture copy artifacts. Moreover, it can better preserve sharp depth discontinuities than previous heuristic weighting schemes. Through comprehensive experiments on both simulated data and real data, we show promising performance of the proposed method.

Index Terms—Color guided depth map restoration, preserve depth discontinuity, suppress texture copy artifacts, ToF, Kinect.

I. INTRODUCTION

DEPTH maps captured by modern depth cameras such as Kinect and Time-of-Flight (ToF) are usually of low resolution. They are also contaminated by missing depth values and noise. However, many applications require accurate and high resolution depth maps such as object reconstruction, robot navigation and automotive driver assistance. To facilitate the use of depth data, tremendous efforts have been spent on the restoration of depth maps captured by modern depth cameras. As stated by Yang *et al.* [38], for ToF depth maps,

Manuscript received March 30, 2016; revised July 8, 2016 and September 10, 2016; accepted September 11, 2016. Date of publication September 22, 2016; date of current version November 18, 2016. This work was supported in part by 973 Plan, China, under Grant 2015CB856004 and in part by NSFC, China, under Grant 61572315 and Grant 61503250. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jing-Ming Guo. (Corresponding author: Jie Yang.)

W. Liu and J. Yang are with the Key Laboratory of Ministry of Education for System Control and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: liuwei.1989@sjtu.edu.cn; jieyang@sjtu.edu.cn).

X. Chen is with the College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai 200093, China (e-mail: xg.chen@live.com).

Q. Wu is with the School of Computing and Communications, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: Qiang.Wu@uts.edu.au).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2016.2612826

their problems are of low resolution and noise. The corresponding restoration mainly focuses on enlarging the spatial resolution and smoothing the noise. For Kinect depth maps, their problems are the missing depth values and noise. The corresponding restoration mainly focuses on filling the “hole” formed with missing depth values and smoothing the noise. Yang *et al.* [38] also pointed out that the degradation of these two kinds of depth maps could be formulated with the same formulation. Although the restoration tasks of these two kinds of depth maps are different, their analysis [38] shows that they can be formulated as the same problem. This implies that most methods for ToF depth map restoration are also applicable to Kinect depth map restoration. For example, the work of Yang *et al.* [38] can be used for both ToF and Kinect depth map restoration. The work of Park *et al.* [26] which was proposed for ToF depth map restoration has also been proved to be applicable to Kinect depth map restoration in their extended work [27].

Inspired by the work of texture synthesis [6], Mac Aodha *et al.* [22] first presented an algorithm to synthetically increase the resolution of a single depth map by using a generic database of high resolution local patches. They casted the task of selecting the right candidate from the database at each location on the target depth map as a Markov Random Fields (MRF) labeling problem. Their work was further extended by Li *et al.* [17] with more regularization terms such as label coherency term and self-similarity term. Hornacek *et al.* [11] proposed to upsample a single depth map without any external database. They upsampled the depth map by identifying and merging patch correspondences within the input depth map itself. Their work exploited patch-wise scene self-similarity across depth with a new 3D variant of patch match. These methods tend to fail to cope with large upsampling factors and scenes with complex structures.

Another research direction of depth map restoration is to fuse multiple frames of depth measurements into one depth map. Schuon *et al.* [31] combined multiple low resolution depth maps captured with different camera centers with an optimization framework that incorporated ToF sensor characteristics. Hahne and Alexa [8] combined depth maps taken with different exposure time in order to produce high quality depth maps. By modeling the original HR range as an MRF, Rajagopalan *et al.* [30] proposed a Bayesian framework to fuse several low resolution depth images. These methods are based on the assumption that the scene is static. They are more practical for static scenes rather than for dynamic scenes.

There are also research interests in developing guided restoration schemes which restore the depth map with the

guidance of the registered (aligned) color image. Joint Bilateral Upsampling (JBU) [14] extended the Bilateral Filter (BF) [36] for depth upsampling where the bilateral weights are based on the guidance color image. Camplani *et al.* [1] proposed to restore Kinect depth maps with a variant of JBU. To achieve sharper depth discontinuities, Liu *et al.* [18] evaluated pixel dissimilarities based on the geodesic distance instead of the Euclidean distance in JBU [14]. Min *et al.* [24] proposed to enhance the quality of depth map with a novel weighted mode filter. Very recently, a new edge preserving filter named guided image filter [10] was introduced and used to perform guided upsampling. Their work was further improved by Lu *et al.* [20] and Tan *et al.* [35] to preserve sharp depth discontinuities and spatial variation. Diebel and Thrun [2] performed the restoration using Markov Random Fields (MRF) with a pairwise appearance consistency data term and an image guided smoothness term. Park *et al.* [26] proposed an optimization framework which regularized the depth map with a nonlocal structure regularization term in order to maintain fine details and structures. Ham *et al.* [9] cast color guided depth upsampling as a graph-based transduction problem. Yang *et al.* [37], [38] performed color guided depth map restoration using a color guided Auto-Regressive (AR) model. Sparse representation models [7], [13], [16] were also proposed to learn the statistical dependency between color and depth in a scene. When compared with the first two categories of methods, this category of methods can produce promising restoration quality for larger upsampling factors and are also not restricted to static scenes.

The fundamental assumption of color guided depth map restoration is that there exists joint occurrences between depth discontinuities and the corresponding color image edges. However, when depth discontinuities are inconsistent with the color edges, these methods may have the following two problems: (I) texture copy artifacts in smooth depth regions when the corresponding color image regions are highly textured, (II) blurring depth discontinuities when the corresponding color edges are weak (or even no corresponding color edges). To handle these two issues, most recent methods proposed to design complex guidance weight based on guidance color images and heuristically took the bicubic interpolation of the input depth map into account, for example, the definition of the AR coefficient in the color guided AR model proposed by Yang *et al.* [38], the weight of the smoothness term in the weighted least squares model proposed by Park *et al.* [26] and the RGB-D structure similarity in the sparse model proposed by Kwon *et al.* [16]. However, complex guidance weight does not always help to improve the upsampling quality while its computational cost is quite heavy. Also, the bicubic interpolation of the input depth map becomes unreliable especially when the upsampling factor is large (e.g., 8 \times) and the input depth map contains large holes and heavy noise. This is the case as the resolution of modern ToF depth cameras (e.g., SwissRanger SR4000) is rather low and their depth maps usually contain heavy noise. Kinect depth maps can also contain large holes when the object contains large black areas. In such cases, their methods often blur depth discontinuities on the restored depth map.

The method proposed in this paper belongs to the color guided depth map restoration scheme. We attempt to alleviate the above two issues. We propose a robust optimization framework for color guided depth map restoration. By adopting a robust penalty function to model the smoothness term of our model, we show that the proposed method is robust against the inconsistency between color edges and depth discontinuities even when we use simple guidance weight. To the best of our knowledge, it is the first time to solve this problem with a principled mathematical formulation rather than the previous heuristic weighting schemes. The proposed method performs well in suppressing texture copy artifacts. Moreover, it can better preserve sharp depth discontinuities than previous heuristic weighting schemes such as the AR model [38] and the model proposed by Park *et al.* [26] especially for real data.

The rest of this paper is organized as follows: in Sec. II, we briefly present the Weighted Least Square (WLS) model [23] and other choices of guidance weighting schemes. Then we compare and analyze these weighting schemes in terms of handling the inconsistency between color edges and depth discontinuities. We propose our robust model and its parameter adaptation based on a robust smoothness measurement in Sec. III. Experimental results and comparison with other state-of-the-art methods are shown in Sec. IV. We draw the conclusion of this paper in Sec. V.

II. RELATED WORK AND MOTIVATION

A. The Weighted Least Squares Model

Given a degraded depth map and the aligned high resolution color image, the degraded depth map can be restored through the Weighted Least Squares (WLS). The WLS is a basic optimization framework that can be applied to many image processing tasks [3], [25]. Recently, Min *et al.* [23] showed that the WLS could also be applied to color guided depth map restoration. The WLS is formulated as:

$$D^H = \arg \min_D \left\{ \sum_{i \in \Omega^0} (D_i - D_i^0)^2 + \alpha \sum_{i \in \Omega} \sum_{j \in N(i)} \omega_{i,j}^c (D_i - D_j)^2 \right\} \quad (1)$$

where D^0 represents the valid measurements that have been mapped to the high resolution coordinate Ω . Ω^0 represents the corresponding coordinate of D^0 . $N(i)$ is the neighborhood of i inside the square patch centered at i .¹ We define $\omega_{i,j}^c$ as:

$$\omega_{i,j}^c = \exp \left(-\frac{|i-j|^2}{2\sigma_s^2} \right) \exp \left(-\frac{\sum_{k \in C} |I_i^k - I_j^k|^2}{3 \times 2\sigma_c^2} \right) \quad (2)$$

where $C = \{R, G, B\}$ or $C = \{Y, U, V\}$ represents different channels of the guidance color image I . σ_c and σ_s are constants defined by the user.

This WLS model cannot handle the case where color edges are inconsistent with depth discontinuities. It has the following two problems: (I) When the homogeneous depth region corresponds to a highly textured color area, this region

¹We have slightly adjusted the 4-connected/8-connected neighborhood system in the original model to a larger square support $N(i)$ of radius r . However, the mechanism remains the same.

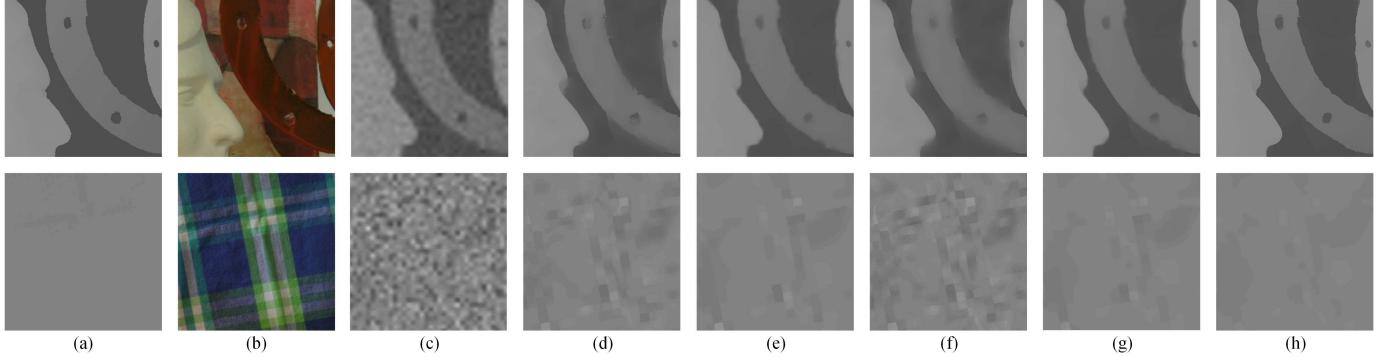


Fig. 1. Comparison on $8 \times$ upsampling. (a) The groundtruth depth maps. (b) The color images. (c) The bicubic interpolation of the noisy low resolution depth maps. Results obtained by (d) the WLS defined in Eq. (1), (e) the EWLS defined in Sec. II, (f) Experiment I and (g) Experiment II defined in Sec. II, (h) the proposed method.

suffers from texture copy artifacts on the restored depth map. (II) When the depth discontinuity corresponds to a weak color edge (or no color edge), it will be blurred on the restored depth map. Taking depth map upsampling for example, Fig. 1(d) illustrates these two problems on $8 \times$ upsampling.

To handle the above two issues, most state-of-the-art methods adopt two strategies: adopting more complicated color guidance weight and making use of the bicubic interpolation of the input depth map. For example, the Auto-Regressive (AR) coefficient in the color guided AR model [38] consists of the combination of the “shape-based” guidance weight based on the guidance color image and the bilateral weight based on the bicubic interpolated input depth map. The guidance weight in the model proposed by Park *et al.* [26] consists of the segmentation, color information and edge saliency of the guidance color image as well as the bicubic interpolated input depth map. The RGB-D structure similarity in the sparse model proposed by Kwon *et al.* [16] also adopted the bicubic interpolated input depth map. Promising results were reported in their papers in handling the above two problems. To show the effectiveness of these strategies, we extend the WLS model in Eq. (1) by replacing its guidance weight in Eq. (2) with the AR coefficient in the AR model [38]. The AR coefficient is defined as:

$$\begin{aligned} a_{i,j} &= a_{i,j}^{\hat{D}} a_{i,j}^I, \\ a_{i,j}^{\hat{D}} &= \exp\left(-\frac{|\hat{D}_i - \hat{D}_j|^2}{2\sigma_1^2}\right) \\ a_{i,j}^I &= \exp\left(-\frac{\sum_{k \in C} \|B_i \circ (P_i^k - P_j^k)\|^2}{3 \times 2\sigma_2^2}\right) \end{aligned} \quad (3)$$

where $a_{i,j}^{\hat{D}}$ is the depth term defined on \hat{D} which is the bicubic interpolation of the input depth map. $a_{i,j}^I$ is the color term defined on the guidance color image I . σ_1 and σ_2 are user defined constants. P_i^k denotes an operator that extracts a $\omega \times \omega$ patch centered at i in color channel k , “ \circ ” represents the element-wise multiplication. The bilateral filter kernel B_i is defined as:

$$B_i(i, j) = \exp\left(-\frac{|i-j|^2}{2\sigma_3^2}\right) \exp\left(-\frac{\sum_{k \in C} |I_i^k - I_j^k|^2}{2\sigma_4^2}\right) \quad (4)$$

where σ_3 and σ_4 are user defined constants.

By replacing $\omega_{i,j}^c$ in Eq. (1) with $a_{i,j}$ in Eq. (3), we obtain another WLS model and we denote it as Extended WLS (EWLS) in this paper. The EWLS can have much better performance than the WLS on handling the inconsistency between color edges and depth discontinuities, i.e. suppressing texture copy artifacts and preserving depth discontinuities. Fig. 1(e) shows illustrations for $8 \times$ upsampling.

B. Comparison and Analysis of Different Weighting Schemes

The weight $a_{i,j}$ in Eq. (3) consists of two terms: the depth term $a_{i,j}^{\hat{D}}$ and the color term $a_{i,j}^I$. Now we show which term contributes to the performance improvement. To this end, we perform another two experiments: (I) Only use the color term $a_{i,j}^I$ in Eq. (3) to replace the guidance weight $\omega_{i,j}^c$ in Eq. (1). We denote this experiment as Experiment I. (II) Replace the guidance weight $\omega_{i,j}^c$ in Eq. (1) with $a_{i,j} = \omega_{i,j}^c a_{i,j}^{\hat{D}}$. $\omega_{i,j}^c$ is defined in Eq. (2). $a_{i,j}^{\hat{D}}$ is defined in Eq. (3). We denote this experiment as Experiment II. Then we perform four comparisons: (I) comparison between Experiment I and the WLS defined in Eq. (1), (II) comparison between Experiment II and the EWLS defined in Sec. II-A, (III) comparison between Experiment I and the EWLS, (IV) comparison between Experiment II and the WLS. The first two comparisons show whether $a_{i,j}^I$ defined in Eq. (3) has better performance than $\omega_{i,j}^c$ defined in Eq. (2). The last two comparisons show whether guidance weight with additional $a_{i,j}^{\hat{D}}$ defined in Eq. (3) contributes to the performance improvement.

Fig. 1(f) shows the experimental results of Experiment I and Fig. 1(g) shows the experimental results of Experiment II. For Experiment I, depth discontinuities are severely blurred and texture copy artifacts are quite obvious. For Experiment II, depth discontinuities are properly preserved and texture copy artifacts are also well suppressed. Then we perform the four experimental comparisons mentioned above. Based on the visual comparison, we have the following two observations: (I) The first two comparisons show that the guidance weight $a_{i,j}^I$ defined in Eq. (3) makes little performance improvement over the bilateral guidance weight $\omega_{i,j}^c$ defined in Eq. (2). (II) The last two comparisons show that when the guidance weight makes use of the bicubic interpolation of the input depth map, i.e., containing $a_{i,j}^{\hat{D}}$, there is large

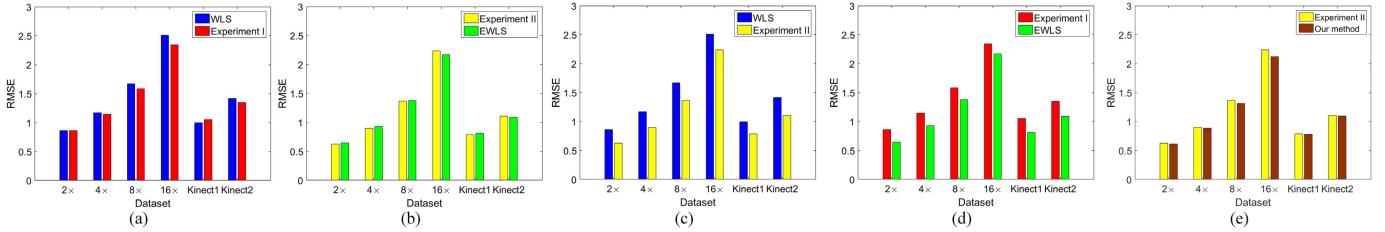


Fig. 2. Quantitative comparisons on simulated ToF dataset and Kinect datasets in terms of average MAE. Comparison between (a) the WLS and Experiment I, (b) the EWLS and Experiment II, (c) the WLS and Experiment II, (d) Experiment I and the EWLS. (e) Experiment II and our method. Experiment I and Experiment II are defined in Sec. II. Comparisons in (a) and (b) show whether $a_{i,j}^I$ defined in Eq. (3) has better performance than $\omega_{i,j}^c$ defined in Eq. (2). Comparisons in (c) and (d) show whether $a_{i,j}^{\hat{D}}$ defined in Eq. (3) contributes to the performance improvement.

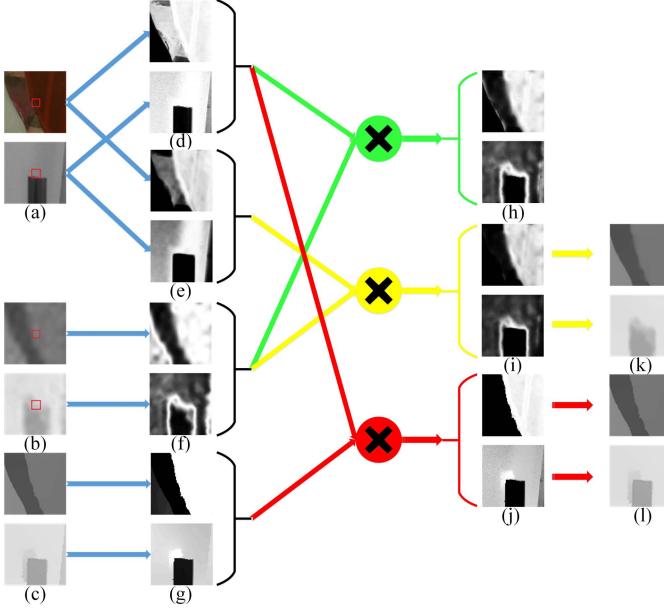


Fig. 3. Kernel comparison. (a) The color images. (b) The bicubic interpolation of the noisy low resolution depth maps. (c) The newly updated depth map in the proposed method. Kernels with respect to the central pixel labeled with the red box. (d) The kernel computed on the color image based on the $\omega_{i,j}^c$ defined in Eq. (2). (e) The kernel computed on the color image based on the $a_{i,j}^I$ defined in Eq. (3). (f) The kernel computed on the bicubic interpolation in (c) based on $a_{i,j}^{\hat{D}}$ defined in Eq. (3). (g) The kernel computed on the updated depth map in the proposed method. (h) Combination of (d) and (f) which is used in Experiment II defined in Sec. II. (i) Combination of (e) and (f) used in the EWLS, i.e. $a_{i,j}$ in Eq. (3). (j) Combination of (d) and (g) used in the proposed method. The upsampling results of (k) the EWLS and (l) the proposed method. Depth maps in the first row are simulated data. Depth maps in the second row are real data.

performance improvement when compared with the results of the guidance weight that is only based on the color image. When the guidance weight contains $a_{i,j}^{\hat{D}}$, texture copy artifacts are well suppressed and depth discontinuities are also better preserved. To further validate our observations, we show quantitative comparisons on the simulated ToF dataset and Kinect datasets used in this paper in terms of average Mean Absolute Error (MAE) in Fig. 2. As shown in the figure, quantitative comparisons also validate the above two observations.

C. The Drawback of Making Use of the bicubic Interpolated Input Depth Map

In fact, to make use of the bicubic interpolation of the input depth is to make use of the information from the depth

map itself. However, simply making use of the information from the bicubic interpolation is far from good enough. Fig. 3 shows the corresponding kernel of (I) $\omega_{i,j}^c$ used in the original WLS, (II) $a_{i,j}^I$ defined in Eq. (3), (III) $a_{i,j}^{\hat{D}}$ defined in Eq. (3), (IV) $a_{i,j} = \omega_{i,j}^c a_{i,j}^I$ used in Experiment II, (V) $a_{i,j} = a_{i,j}^{\hat{D}} a_{i,j}^I$ used in the EWLS. It can be observed that depth discontinuities on the bicubic interpolation of the noisy low resolution input are quite blurring as shown in Fig. 3(b). Kernel $a_{i,j}^{\hat{D}}$ based on the bicubic interpolation is also quite blurring as shown in Fig. 3(f). When combined with this kernel, kernel $a_{i,j} = \omega_{i,j}^c a_{i,j}^{\hat{D}}$ in Fig. 3(h) and kernel $a_{i,j} = a_{i,j}^{\hat{D}} a_{i,j}^I$ in Fig. 3(i) are also blurred. As a result, depth discontinuities on the final restored depth map are also blurred as shown in Fig. 3(k). In fact, as shown in the second row in Fig. 3, the quality of real data is much worse than simulated data. Blurring depth discontinuities are not the only problem of the bicubic interpolated depth map. As shown in Fig. 3(b), pixels belong to the background are assigned to depth values of the foreground. As a result, kernels in Fig. 3(f), (h) and (i) are not only blurred but also incorrect. This results in the bad restoration shown in Fig. 3(k). From results of both the simulated data and real data, it shows that $a_{i,j}^{\hat{D}}$ based on the bicubic interpolation is not good enough to preserve sharp depth discontinuities.

III. THE PROPOSED METHOD

A. The Robust Model

The smoothness term of the WLS in Eq. (1) is modeled with L_2 norm. The L_2 norm is known to be not robust against outliers in relevant research areas such as image denoising [4], [28]. In the previous sections, we also show it is not robust against the inconsistency between color edges and depth discontinuities in color guided depth map restoration.

In this section, we propose to penalize the smoothness term with a robust penalty function instead of L_2 norm. The proposed optimization framework is defined as:

$$D^H = \arg \min_D \left\{ \sum_{i \in \Omega^0} (D_i - D_i^0)^2 + \alpha \sum_{i \in \Omega^0} \sum_{j \in N(i)} \omega_{i,j}^c \psi(|D_i - D_j|^2) \right\} \quad (5)$$

where $\psi(\cdot)$ is defined as:

$$\psi(x^2) = 2\mu^2 \left(1 - \exp \left(-\frac{x^2}{2\mu^2} \right) \right) \quad (6)$$

where μ is a user given parameter. We will present a data driven approach in Sec. III-D to properly adapt μ according to the smoothness of depth maps.

As our analysis shown in Sec. II, the color guidance weight $a_{i,j}^I$ in Eq. (3) makes little contribution to the performance improvement. Note that $a_{i,j}^I$ is based on the patch to patch difference. It is much more time-consuming than $\omega_{i,j}^c$ in Eq. (2) which is based on the pixel to pixel difference. We thus adopt $\omega_{i,j}^c$ as the color guidance weight in our model.

In fact, Eq. (6) has long been used as a robust penalty function in relevant areas such as image denoising [29] and optical flow [39] to preserve sharp edges. In this paper, we introduce it to the color guided depth map restoration to handle the inconsistency between color edges and depth discontinuities.

B. Analysis of the Model

Now we analyze why the proposed model is robust against the inconsistency between color edges and depth discontinuities. First, we present the normal equation of our model

$$(D_i - D_i^0) + 2\alpha \sum_{j \in N(i)} \omega_{i,j}^c d_{i,j} (D_i - D_j) = 0, \quad i \in \Omega \quad (7)$$

where we define:

$$d_{i,j} = \psi'(|D_i - D_j|^2) = \exp\left(-\frac{|D_i - D_j|^2}{2\mu^2}\right) \quad (8)$$

The closed-form solution to Eq. (7) is not available. Thus it should be solved iteratively. If we keep $d_{i,j}$ as constant in each iteration where $d_{i,j}^n = \psi'(|D_i^n - D_j^n|^2)$ for iteration $n+1$, then Eq. (7) becomes the normal equation of the following iteratively re-weighted least squares optimization framework:

$$D_{n+1} = \arg \min_D \left\{ \sum_{i \in \Omega^0} (D_i - D_i^0)^2 + \alpha \sum_{i \in \Omega} \sum_{j \in N(i)} \omega_{i,j} (D_i - D_j)^2 \right\} \quad (9)$$

where $\omega_{i,j} = \omega_{i,j}^c d_{i,j}^n$.

Now we compare the guidance weight $\omega_{i,j}$ in Eq. (9) with the guidance weight $\omega_{i,j}^c$ of WLS in Eq. (2) and the guidance weight $a_{i,j}$ of the EWLS in Eq. (3). The $\omega_{i,j}^c$ only based on the guidance color image results in blurring depth discontinuities and texture copy artifacts. The $a_{i,j}$ based on the guidance color image and the bicubic interpolation of the input can suppress texture copy artifacts but depth discontinuities are still blurred. The $\omega_{i,j}$ is based on not only the guidance color image but also the newly updated depth map in each iteration. The quality of the newly updated depth map is much better than that of the bicubic interpolated depth map. This helps our model to not only suppress texture copy artifacts but also preserve sharper depth discontinuities than the EWLS. This is also validated in Fig. 1(h), Fig. 3(j) and Fig. 3(l) where the proposed method results in both the sharp kernel and sharp depth discontinuities on the restored depth maps. Note that the proposed method is especially effective for real data as shown in Fig. 3. Fig. 2(e) also shows the comparison between our method and Experiment II (using bicubic interpolation of the input depth map) on different datasets in terms of

average MAE. The improvement of the average MAE also validates the effectiveness of our method.

C. The Numerical Solution

We start from the normal equation of our model in Eq. (7). If we keep $d_{i,j}$ as constant in each iteration where $d_{i,j}^n = \psi'(|D_i^n - D_j^n|^2)$ for iteration $n+1$, then we can solve it through its fixed point equation. It is defined as:

$$D_i^{n+1} = \frac{D_i^0 + 2\alpha \sum_{j \in N(i)} \omega_{i,j}^c d_{i,j}^n D_j^n}{1 + 2\alpha \sum_{j \in N(i)} \omega_{i,j}^c d_{i,j}^n}, \quad i \in \Omega \quad (10)$$

We initialize Eq. (10) with the bicubic interpolation of the input depth map. As the bicubic interpolation has already been close to the “groundtruth”, we thus assume that it satisfies the condition of the fixed point equation. We iteratively update the depth map through Eq. (10) to get the final output. According to our experimental results, it works well for satisfying results.

In fact, we can solve the proposed model in Eq. (5) by iteratively solving the normal equation in Eq. (7) in each iteration. Eq. (7) is a linear system which can be solved by modern solvers such as Preconditioned Conjugate Gradient (PCG) [15]. However, one of the most challenging problems is that solving this linear system needs a large amount of memory to store the affinity matrix when the size of the neighborhood system $N(i)$ becomes large (e.g., 19×19). It cannot be accomplished without a computer of large memory. On the contrary, the memory cost of Eq. (10) is quite little even for the 19×19 neighborhood system. It can be accomplished even with a laptop of 4GB memory. According to our experimental results, both these two methods can get similar results. We thus solve our model through Eq. (10).

D. Parameter Adaptation Based on Local Relative Smoothness

The parameter μ in Eq. (6) is an important parameter in our model. We denote μ as the *bandwidth* of our model in this paper. Small bandwidth can well preserve depth discontinuities but performs worse in smoothing noise. On the contrary, large bandwidth performs better in smoothing noise but can blur depth discontinuities. A better choice of bandwidth based on the local smoothness of the depth map can result in both better noise smoothing and depth discontinuity preserving. To this end, we first adopt our previous work [19] to measure the local smoothness of the depth map as follows:

$$\rho_i = \frac{\sum_{s \in N_\rho(x_{min})} x_s}{\sum_{t \in N_\rho(x_{max})} x_t}, \quad i \in \Omega \quad (11)$$

where x_{min} and x_{max} are the minimal and maximal depth values inside $N(i)$. $N(i)$ is the neighborhood of pixel i . $N_\rho(x_{min})/N_\rho(x_{max})$ denotes a small patch of radius $r_\rho \times r_\rho$ centered at the pixel whose value is x_{min}/x_{max} . This makes ρ_i robust against the noise. We set $r_\rho = r$ where r is the radius of $N(i)$. ρ_i is denoted as the *relative smoothness* of a given depth patch in this paper. In our

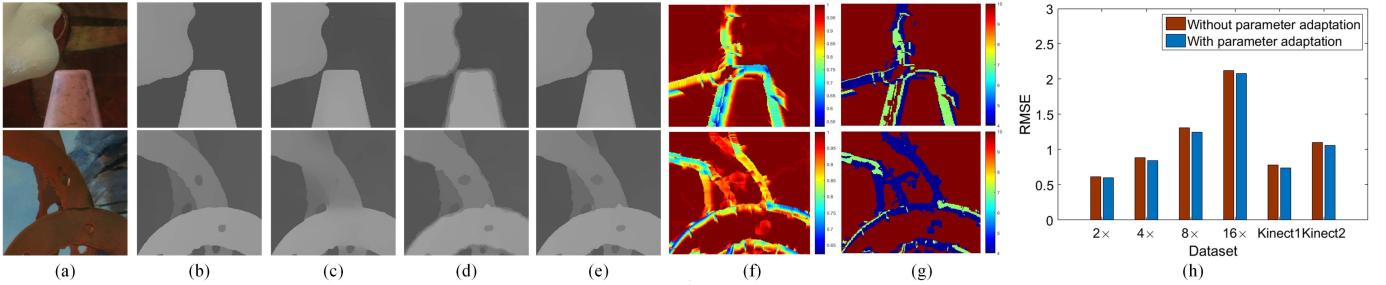


Fig. 4. (a) The color images. (b) The groundtruth depth maps. Depth maps restored by our method using (c) large bandwidth $\mu = 10$, (d) small bandwidth $\mu = 4$ and (e) adapted bandwidth based on Eq. (12). (f) Illustration of the relative smoothness and (g) the corresponding adapted bandwidth based on Eq. (12). (h) Quantitative comparison between the results restored with bandwidth adaptation and without adaptation in terms of average MAE.

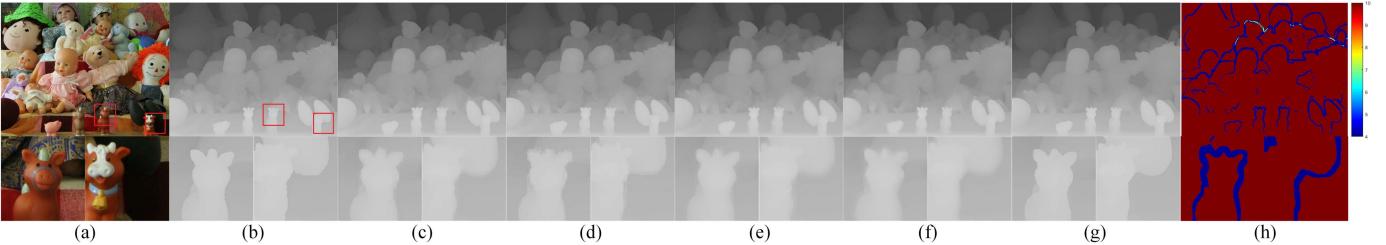


Fig. 5. 8x upsampling of simulated ToF data without noise. (a) Color images. (b) The groundtruth. Upsampling Results of (c) the NLM-WLS [26], (d) the TGV [5], (e) the WLS [23], (f) the AR model [38] and (g) the proposed method. (h) Bandwidth maps of the proposed method. Regions in red boxes are highlighted.

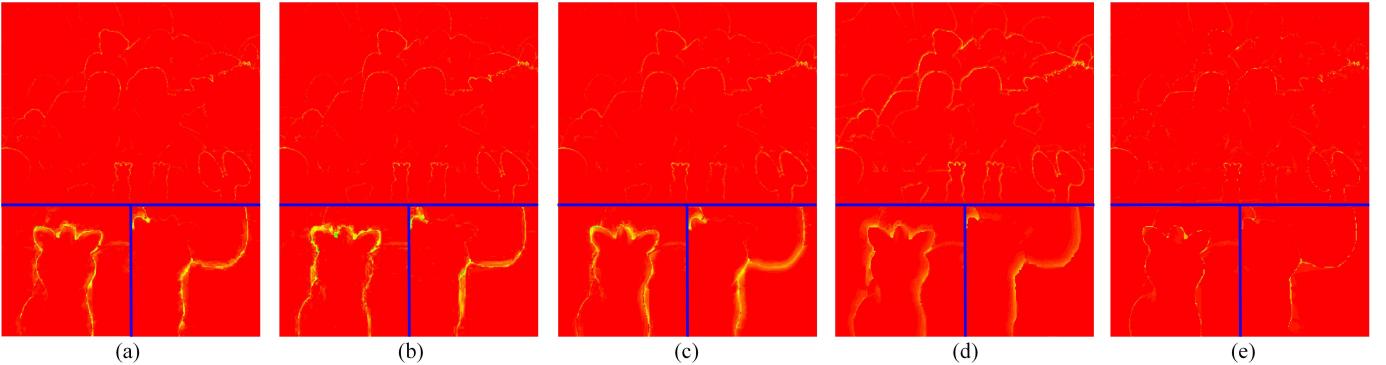


Fig. 6. Corresponding error maps of the results in Fig. 5. (a) the NLM-WLS [26], (b) the TGV [5], (c) the WLS [23], (d) the AR model [38] and (e) the proposed method.

experiments, we iteratively update the relative smoothness based on the newly updated depth map in Eq. (10) in each iteration.

After we get the local relative smoothness of the depth map, we can properly adapt the bandwidth for each pixel according to $\rho_i (i \in \Omega)$. However, unlike our previous work [19] which only divided a depth map into homogenous regions and depth discontinuities, we find that depth discontinuities should be explored on more fine-grained levels. Our experiments show that not all depth discontinuities favor small bandwidth μ . Weak depth discontinuities favor smaller bandwidth while strong depth discontinuities favor larger bandwidth. If we use small bandwidth for strong depth discontinuities, there will be multiple layers around the discontinuities. We illustrate this in Fig. 4 with 8x upsampling examples. Fig. 4(c) shows the results restored with our method using $\mu = 10$ (large bandwidth) where fine details and weak discontinuities in the second row are

heavily blurred. Fig. 4(d) shows the results restored with our method using $\mu = 4$ (small bandwidth) where fine details and weak discontinuities are properly preserved. However, the strong discontinuities in the first row result in multiple layers which are ‘‘blurred’’. Thus, depth discontinuities of different scales on the depth map should be treated differently: weak depth discontinuities need smaller bandwidth. On the contrary, strong depth discontinuities need larger bandwidth. Luckily, the value of the proposed relative smoothness in Eq. (11) can properly reflect the strength of depth discontinuities: strong depth discontinuities have smaller μ than weak depth discontinuities.

After we have made all this clear, we can properly adapt the bandwidth μ for each pixel $i \in \Omega$ according to the corresponding relative smoothness ρ_i . Here we show one possible parameter setting that can yield satisfying results in our experiments. Assuming all the depth values on the depth map are in interval $[0, 255]$, then we adopt the bandwidth by

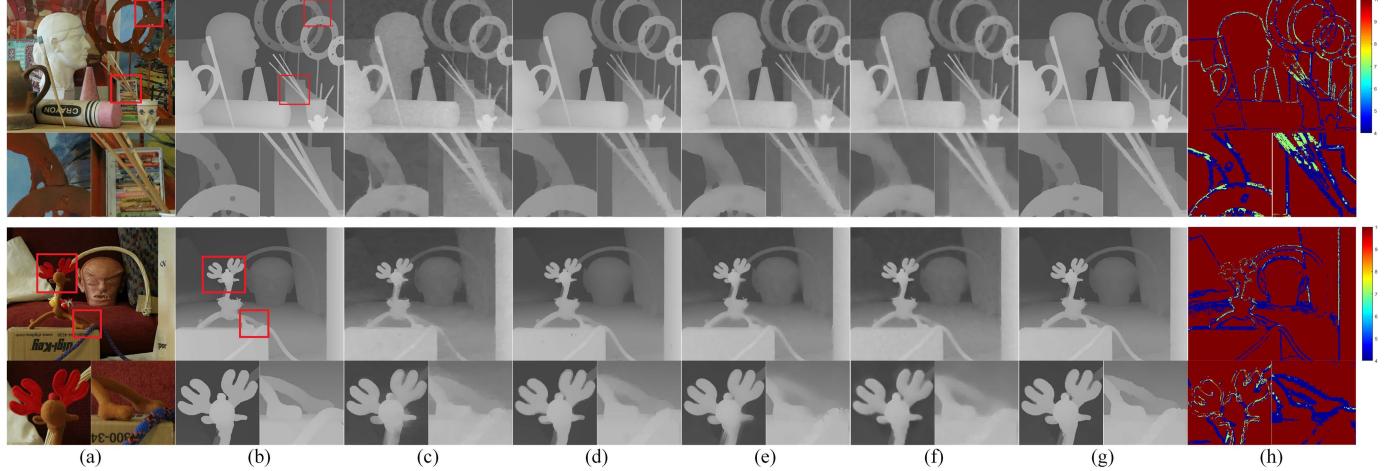


Fig. 7. 8 \times upsampling of noisy simulated ToF data. (a) Color images. (b) The groundtruth. Upsampling Results of (c) the NLM-WLS [26], (d) the TGV [5], (e) the WLS [23], (f) the AR model [38] and (g) the proposed method. (h) Bandwidth maps of the proposed method. Regions in red boxes are highlighted.

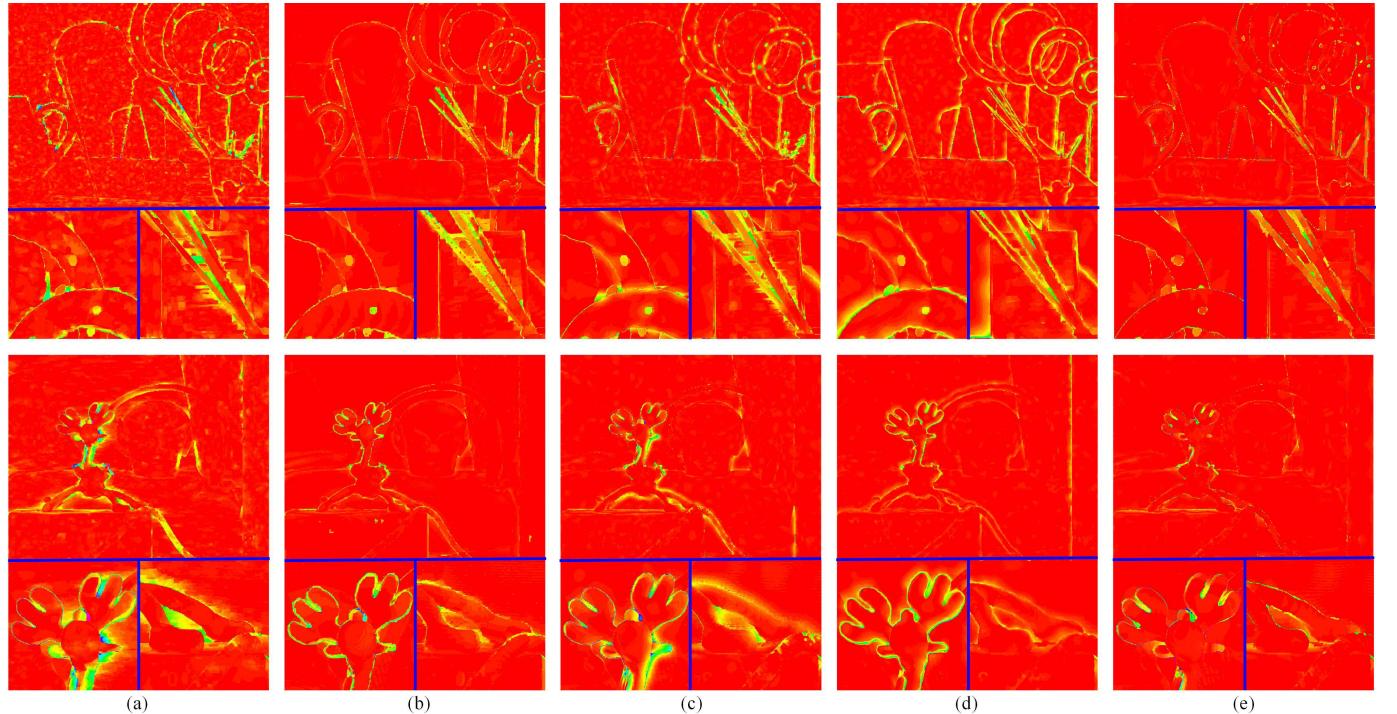


Fig. 8. Corresponding error maps of the results in Fig. 7. (a) the NLM-WLS [26], (b) the TGV [5], (c) the WLS [23], (d) the AR model [38] and (e) the proposed method.

directly thresholding the relative smoothness as follows:

$$\mu_i = \begin{cases} 10, & \text{if } \rho_i \geq 0.96, \\ 4, & \text{if } 0.8 \leq \rho_i < 0.96, \\ 7, & \text{if } 0.7 \leq \rho_i < 0.8, \\ 10, & \text{if } \rho_i < 0.7, \end{cases} \quad i \in \Omega \quad (12)$$

The first inequality in Eq. (12) means that the depth map $N(i)$ locates in homogeneous region which needs large bandwidth to smooth the noise. The second to the fourth inequalities mean that $N(i)$ contains a depth discontinuity. As the relative smoothness becomes smaller, the strength of the discontinuity becomes stronger and the bandwidth should

become larger based on our analysis above. Our experimental results show that this simple thresholding is enough for satisfying results. Note that the value of μ_i in Eq. (12) is based on the assumption that depth values are in interval $[0, 255]$. If depth values on the depth map are not in $[0, 255]$, it should be normalized into $[0, 255]$. However, the condition of ρ_i on the right side of Eq. (12) does not depend on either the range of depth values or the noise on the depth map since we always have $\rho_i \in (0, 1]$. If the noise on the depth map becomes larger, to guarantee the accuracy of Eq. (12), the radius r_ρ of Eq. (11) should become larger to make ρ_i more robust against the noise.

Fig. 4(e) shows the results by our method with the proposed parameter adaptation. It shows that both the strong

TABLE I

QUANTITATIVE COMPARISON ON THE NOISE FREE SIMULATED TOF DATA. RESULTS ARE EVALUATED IN MAE AND THE BEST RESULTS ARE IN BOLD

	Art				Book				Dolls				Laundry				Moebius				Reindeer			
	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x
JB [14]	0.67	0.97	1.64	3.39	0.3	0.4	0.59	1.12	0.37	0.45	0.62	1.08	0.4	0.55	0.88	1.76	0.31	0.4	0.6	1.11	0.44	0.59	0.88	1.65
JGU [18]	0.66	1	1.76	3.59	0.3	0.4	0.61	1.18	0.37	0.45	0.64	1.16	0.4	0.57	0.96	1.87	0.3	0.4	0.62	1.17	0.44	0.61	0.97	1.84
NLM-WLS [26]	0.67	0.81	1.44	3.33	0.27	0.32	0.53	1.1	0.35	0.37	0.54	1.04	0.36	0.44	0.8	1.76	0.29	0.33	0.54	1.09	0.38	0.46	0.77	1.66
TGV [5]	0.51	0.74	1.29	2.8	0.19	0.29	0.45	0.83	0.22	0.33	0.57	0.96	0.29	0.4	0.76	1.55	0.2	0.31	0.51	1.03	0.3	0.41	0.71	1.45
WLS [23]	0.53	0.94	1.69	3.28	0.2	0.38	0.62	1.12	0.23	0.4	0.63	1.06	0.29	0.51	0.9	1.71	0.21	0.39	0.66	1.15	0.32	0.54	0.89	1.57
AR [38]	0.56	1.12	2.21	4.62	0.27	0.47	0.89	1.92	0.29	0.55	1.02	1.91	0.37	0.62	1.23	2.45	0.25	0.51	0.99	1.98	0.75	0.69	1.33	2.67
Ours	0.45	0.73	1.3	2.48	0.19	0.31	0.51	0.93	0.22	0.3	0.51	1	0.27	0.4	0.68	1.46	0.19	0.33	0.5	1	0.3	0.43	0.68	1.33

TABLE II

QUANTITATIVE COMPARISON ON THE NOISY SIMULATED TOF DATA. RESULTS ARE EVALUATED IN MAE AND THE BEST RESULTS ARE IN BOLD

	Art				Book				Dolls				Laundry				Moebius				Reindeer			
	2x	4x	8x	16x	2x	4x	8x	16x																
GF [10]	1.91	2.23	3.08	4.87	0.84	1.12	1.73	2.82	0.84	1.11	1.69	2.71	1.01	1.31	2.0	3.33	0.92	1.19	1.78	2.84	1.06	1.32	1.98	3.31
JB [14]	1.59	2.06	3.11	5.08	0.82	1.24	2.04	3.19	0.81	1.12	1.98	3.05	0.94	1.38	2.25	3.67	0.89	1.28	2.05	3.18	0.95	1.36	2.24	3.67
JGU [18]	1.33	1.81	2.9	4.92	0.79	1.24	2.05	3.17	0.8	1.23	2.01	3.05	0.88	1.36	2.23	3.65	0.82	1.25	2.03	3.15	0.91	1.37	2.26	3.68
NLM-WLS [26]	1.66	2.47	3.44	5.55	1.19	1.47	2.06	3.1	1.19	1.56	2.15	3.04	1.34	1.73	2.41	3.85	1.2	1.5	2.13	2.95	1.26	1.65	2.46	3.66
JID [13]	1.69	2.98	3.68	5.99	1.53	2.71	3.04	4.39	1.54	2.71	2.94	3.90	1.45	2.72	3.16	4.63	1.55	2.72	2.94	4.34	1.65	2.80	3.13	4.63
TGV [5]	0.8	1.21	2.01	4.59	0.61	0.88	1.21	2.19	0.66	0.96	1.38	2.88	0.61	0.87	1.36	3.06	0.57	0.77	1.23	2.74	0.61	0.85	1.3	3.41
AR [38]	1.17	1.7	2.93	5.32	0.98	1.22	1.74	2.89	0.97	1.21	1.71	2.74	1	1.31	1.97	3.43	0.95	1.2	1.79	2.82	1.07	1.3	2.03	3.34
WLS [23]	1.25	1.73	2.59	4.01	0.74	1.1	1.45	2.13	0.85	1.21	1.68	2.13	0.83	1.17	1.65	2.62	0.8	1.18	1.67	2.27	0.84	1.15	1.58	2.43
Ours	0.71	1.06	1.72	3.13	0.57	0.78	1.13	1.68	0.64	0.87	1.21	1.73	0.54	0.77	1.12	1.98	0.55	0.76	1.15	1.71	0.57	0.8	1.14	1.88

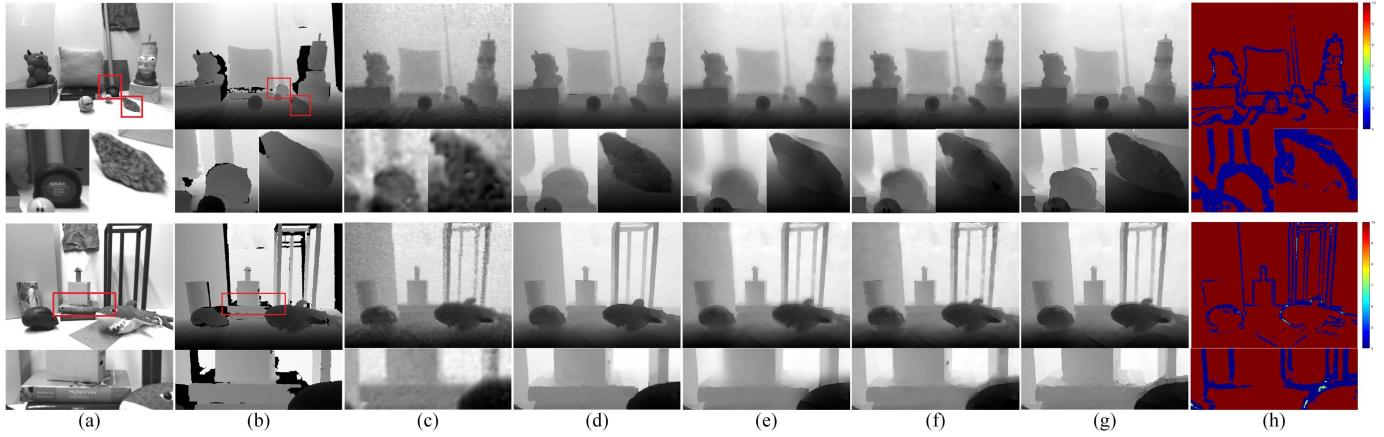


Fig. 9. Experimental results of real ToF data. (a) Intensity images. (b) The groundtruth. (c) The bicubic interpolation of the input depth maps. Upsampling Results of (d) the TGV [5], (e) the WLS [23], (f) the AR model [38] and (g) the proposed method. (h) Bandwidth maps of the proposed method. Regions in red boxes are highlighted.

depth discontinuities in the first row and the weak depth discontinuities in the second row are well preserved. To further validate the effectiveness of the proposed parameter adaptation, we perform experiments on both ToF and Kinect datasets. Quantitative comparison in terms of average MAE between our method with parameter adaptation and without parameter adaptation is shown in Fig. 4(h). The parameter adaptation also helps to achieve lower average MAE.

IV. EXPERIMENTS

In this section, we test the proposed method on both simulated and real ToF and Kinect datasets. Parameters are set as follows: α in Eq. (5) is set as 0.95. r of the neighborhood $N(i)(i \in \Omega)$ is set as 9. The bandwidth of the robust penalty function in Eq. (6) is set as Eq. (12). Depth values of all the depth maps are firstly normalized into $[0, 255]$ when performing the restoration and then normalized back to their own depth range for quantitative comparison. All the parameters remain the same for all the test datasets.

We also show visual and quantitative comparison between the proposed method and other state-of-the-art methods including Joint Bilateral Upsampling (JBU) [14], Joint Geodesic Upsampling (JGU) [18], the nonlocal means regularized weighted least squares [26] that we denote as NLM-WLS, the Joint Intensity and Depth (JID) co-sparse analysis model [13], the Total Generalized Variation (TGV) guided depth upsampling [5], the Weighted Least Squares (WLS) [23] and the color guided Auto-Regressive (AR) model [38].²

A. Experiments on ToF Depth Maps Upsampling

We first test our method on ToF depth map upsampling with the simulated ToF dataset. Two kinds of degradation are simulated: downsampling and downsampling with noise.

²We re-implement the AR model with C language based on the author provided MATLAB source code. The author provided source code is available here: http://cs.tju.edu.cn/faculty/liikun/projects/depth_recovery/index.htm. Our implementation is available here: <https://github.com/wliusjtu/Adaptive-Auto-Regressive-Model-for-Guided-Depth-Map-Restoration>. Our implementation is much faster and can handle much larger depth maps.

TABLE III
QUANTITATIVE COMPARISON ON REAL TOF DATASET. THE ERROR IS CALCULATED AS MAE TO THE MEASURED GROUNDTRUTH IN mm. THE BEST RESULTS ARE IN BOLD

	Bicubic	GF [10]	JGU [18]	JBK [14]	NLM-WLS [26]	TGV [5]	AR [38]	WLS [23]	Ours
<i>Books</i>	16.23mm	15.55mm	15.76mm	15.86mm	14.31mm	12.8mm	14.37mm	13.87mm	12.82mm
<i>Devil</i>	17.78mm	16.1mm	16.28mm	16.32mm	15.36mm	14.97mm	15.41mm	15.36mm	14.5mm
<i>Shark</i>	16.66mm	17.1mm	17.25mm	17.32mm	15.88mm	15.53mm	16.27mm	15.88mm	14.92mm

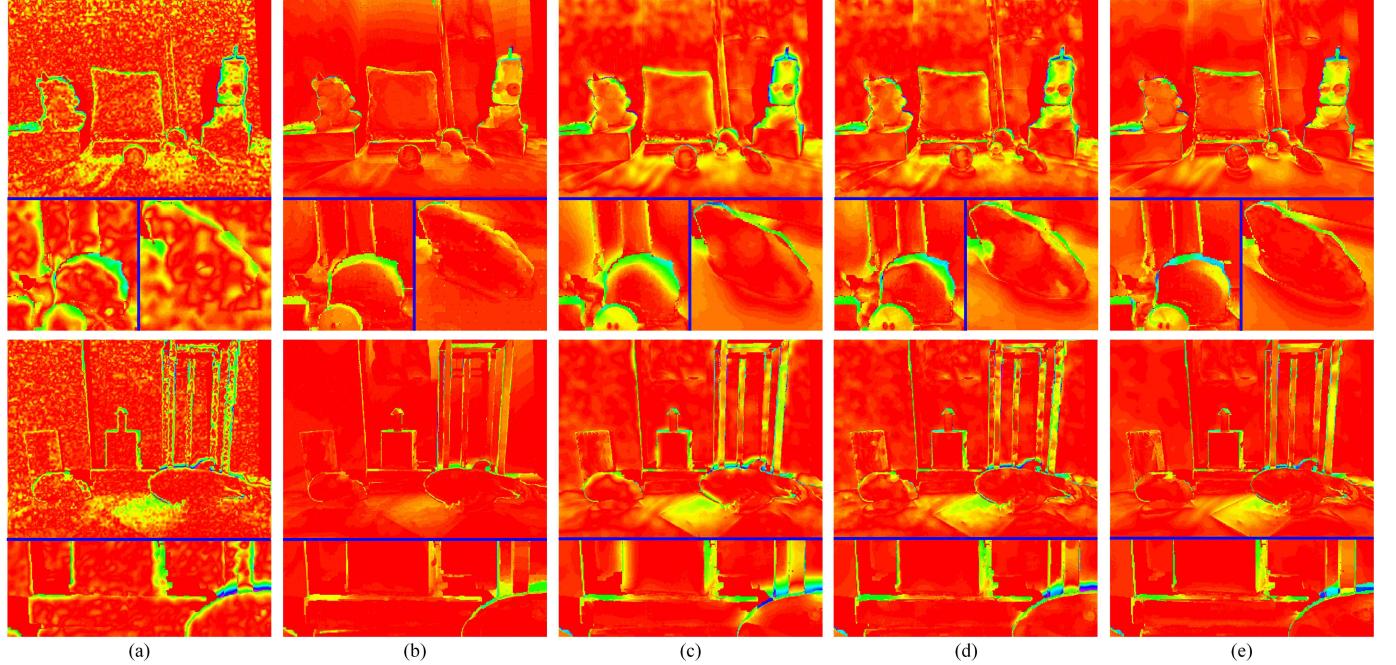


Fig. 10. Corresponding error maps of the results in Fig. 9 (a) The bicubic interpolation of the input depth maps. (b) The TGV [5], (c) the WLS [23], (d) the AR model [38] and (e) the proposed method.

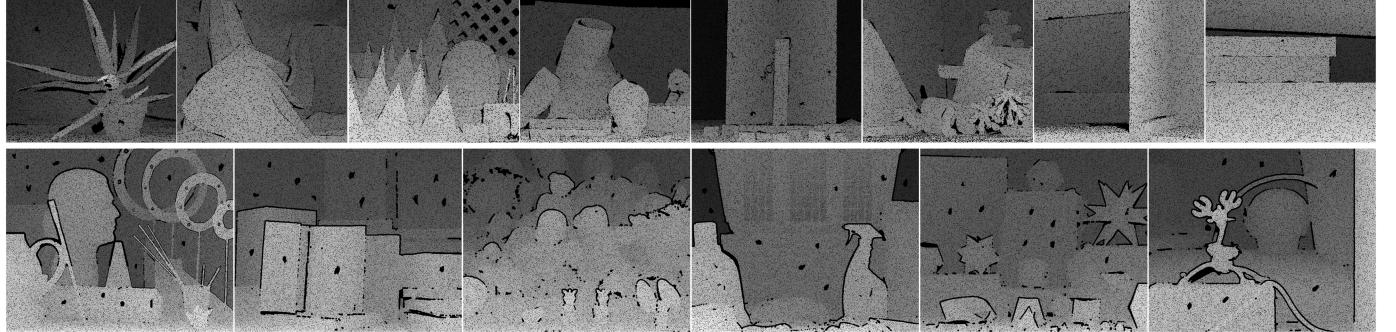


Fig. 11. Simulated Kinect datasets used in this paper. *Kinect dataset 1* in the first row are from [21]. *Kinect dataset 2* in the second row are slightly changed by adding noise and random missing depth values to the original data in [38].

The noisy downsampled depth maps are from the dataset provided by Yang *et al.* [38]. Four upsampling factors are tested in this paper: $2 \times / 4 \times / 8 \times / 16 \times$.

Table I shows the quantitative comparison of upsampling results of noise free depth maps in terms of MAE. Both our method and the TGV [5] outperform other methods in terms of lower MAE. Fig. 5 shows illustration of $8 \times$ upsampling. The corresponding error maps are shown in Fig. 6 where we show our methods performs better in preserving sharp depth discontinuities than other methods.

Table II shows the quantitative comparison of noisy simulated ToF upsampling results. Our method outperforms other compared state-of-the-art methods with lowest MAE. Fig. 7 further shows visual comparison of $8 \times$ upsampling results. As shown in the highlighted regions, when the edges in the guidance color image are weak, the corresponding depth discontinuities on the upsampled depth map by WLS [23] are severely blurred. This is clearly illustrated on the corresponding error maps in Fig. 8(c). Errors are quite large around the depth discontinuities. Note that there are also noticeable errors

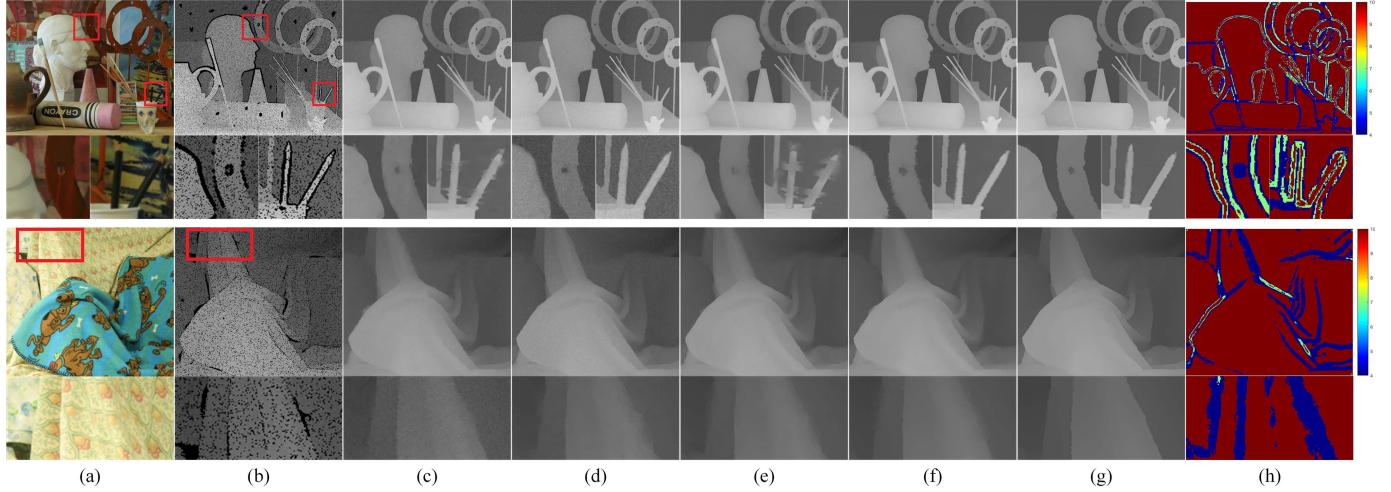


Fig. 12. Experimental of simulated Kinect data. (a) Color images. (b) The degraded depth maps. Restored results of (c) the NLM-WLS [26], (d) the JID [13], (e) the WLS [23], (f) the AR model [38] and (g) the proposed method. (h) Bandwidth maps of the proposed method. Regions in red boxes are highlighted. The depth map in the first row is from Kinect dataset 2. The depth map in the second row is from Kinect dataset 1.

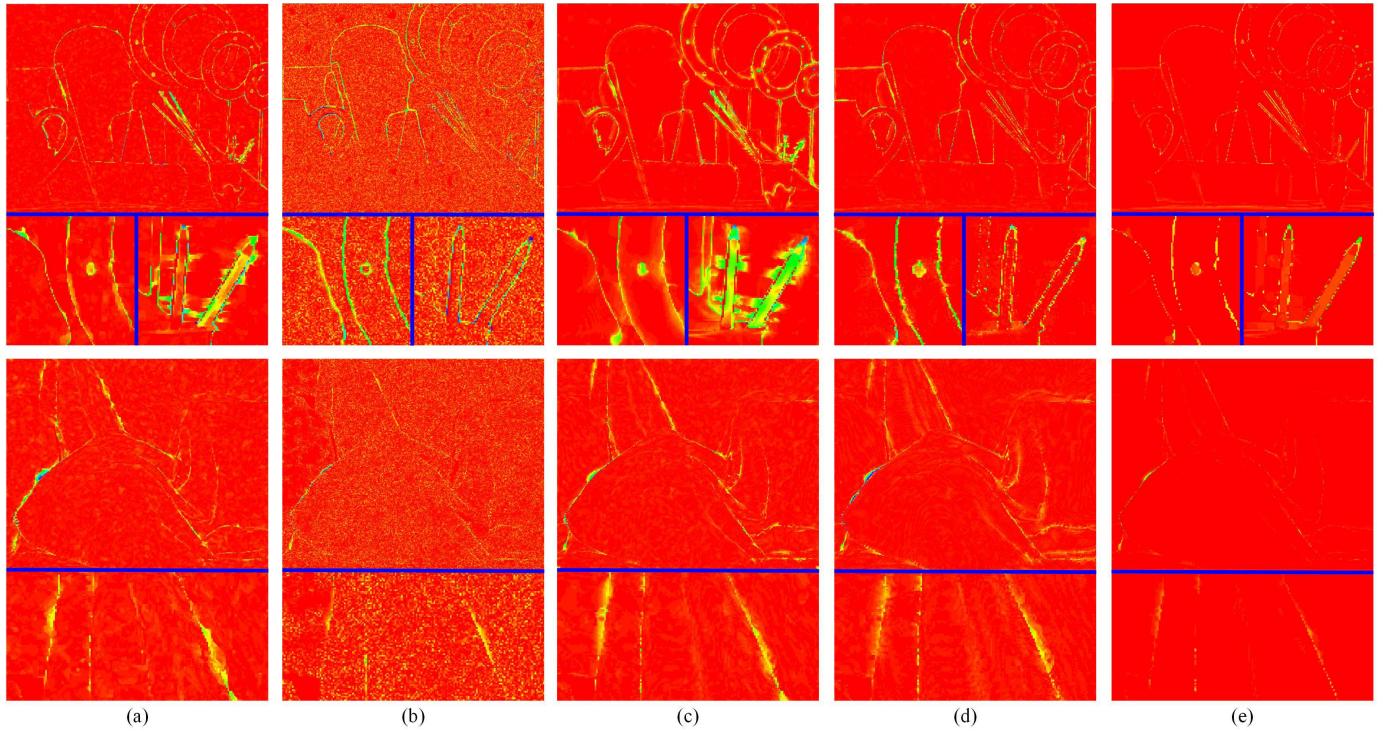


Fig. 13. Corresponding error maps of the results in Fig. 12. (a) the NLM-WLS [26], (b) the JID [13], (c) the WLS [23], (d) the AR model [38] and (e) the proposed method.

in the homogeneous regions in Fig. 8(c) which reflect texture copy artifacts on the upsampled depth map. Results of the NLM-WLS [26] in Fig. 7(c) and Fig. 8(a) are similar with that of the WLS [23]. The AR model [38] can properly suppress texture copy artifacts. However, depth discontinuities are still blurred. This is because the depth term in its AR coefficient is only based on the bicubic interpolation of the noisy low resolution input. On the contrary, our method can properly preserve sharp depth discontinuities. Our results have much smaller errors around the depth discontinuities in Fig. 8(e)

than that of the AR model in Fig. 8(d). Moreover, our method can preserve fine details, for example, the “small holes” in the rings as highlighted in the first row of Fig. 7(g). Bandwidth maps of our method shown in Fig. 7(h) also well correspond to the property of the corresponding depth maps. Weak depth discontinuities are assigned to small bandwidth value (the ones in blue) while strong depth discontinuities are assigned to large bandwidth value (the ones in red). This helps our method to properly preserve depth discontinuities of different scales and fine details as shown in the highlighted regions.

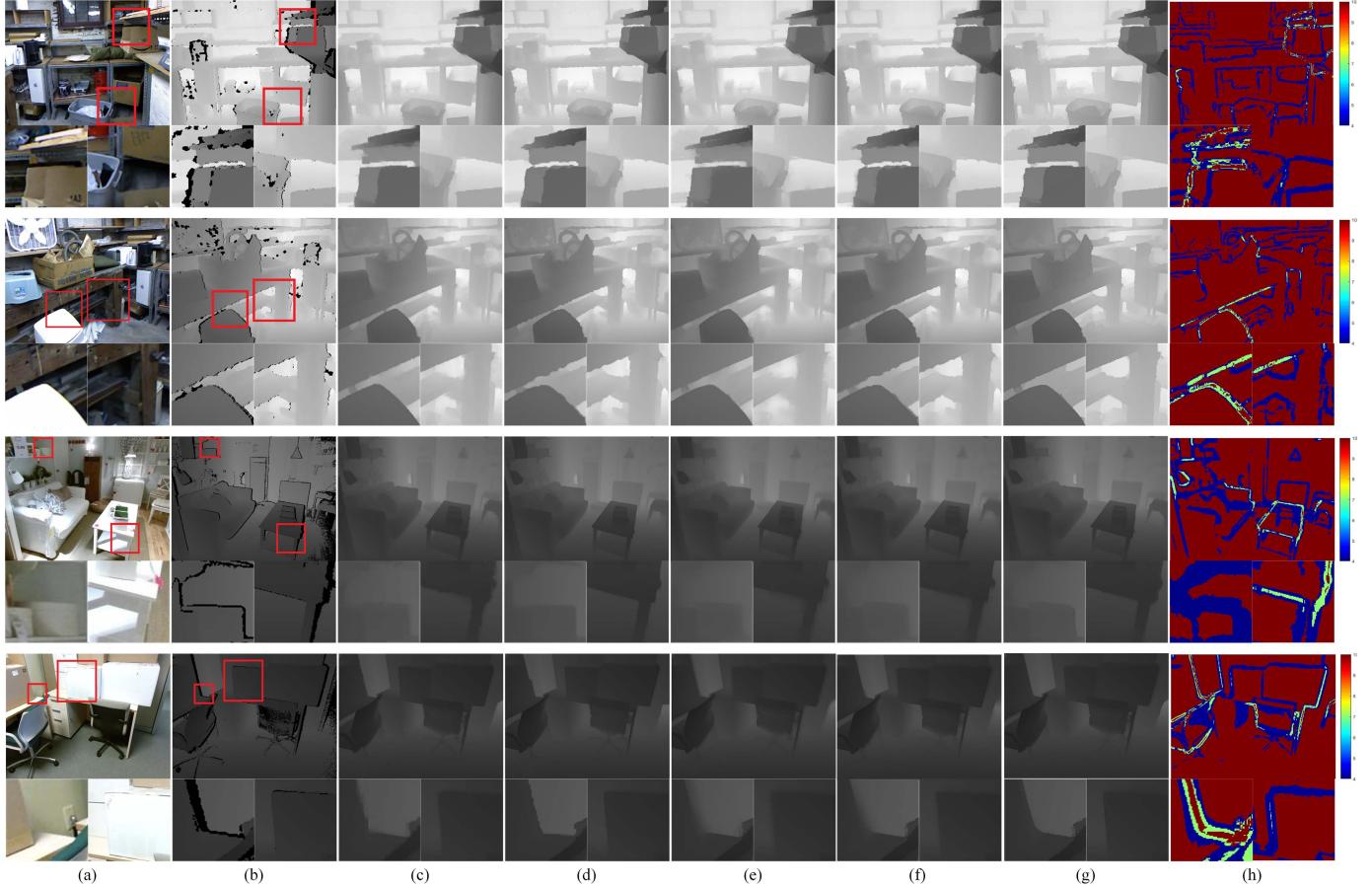


Fig. 14. Experimental results of real Kinect data. (a) Color images. (b) The degraded depth maps. Restored results of (c) the NLM-WLS [26], (d) the JID [13], (e) the WLS [23], (f) the AR model [38] and (g) the proposed method. (h) Bandwidth maps of the proposed method. Regions in red boxes are highlighted. The depth maps in the first two rows are captured by Kinect v1 from the NYU Kinect dataset [33]. The depth maps in the last two rows are captured by Kinect v2 from the SUN RGBD dataset [34].

We also test our method on the real ToF dataset [5] which contains three depth maps: *books*, *devil* and *shark*. The noisy low resolution depth maps are captured by a PMD Nano ToF camera delivering a 120×160 dense depth map and IR amplitude image. The intensity image is captured by a CMOS camera with the sensor size of 810×610 pixels. The upsampling factor is about $6.25\times$. Also, this dataset provides groundtruth measurements which are generated by using a structured light scanner with high accuracy. Table III shows the quantitative comparison of our method with other methods in terms of MAE. The MAE is measured in mm between the upsampled results and the corresponding groundtruth. Fig. 9 shows visual comparison. We also show the bicubic interpolation of the input in Fig. 9(c). As shown in the figure, the bicubic interpolation is quite noisy and also contains jaggy depth discontinuities, for example, the highlighted “rock” in the first row. As a result, depth discontinuities in the results of AR model [38] are not only blurred but also jaggy. This problem is well handled by our method as shown in Fig. 9(g). The corresponding error maps are shown in Fig. 10 for better comparison. Unlike the simulated results in Fig. 8(d), the AR model cannot well smooth the noise in homogeneous regions as shown in Fig. 10(d). This is also due to the heavy noise in the bicubic interpolation. Bandwidth maps in Fig. 9(h) also

show that our parameter adaptation scheme is robust against noise and can well handle real data.

B. Experiments on Kinect Depth Maps Restoration

We also test our method on Kinect depth map restoration. Two simulated Kinect datasets are used in this paper: we reuse the dataset in [21] that we denote as *Kinect Dataset 1*. There are eight depth maps in Dataset 1 which are shown in the first row of Fig. 11. We also reuse the simulated Kinect dataset in [38]. We denote this dataset as *Kinect Dataset 2*. This dataset contains six depth maps. However, depth maps in this dataset are noise free and only contain structure missing depth values. According to the analysis by Khoshelham and Elberink [12], the noise level on Kinect depth maps is proportional to the square of depth values. So we add additive Gaussian noise to the depth maps where the variance of noise at each pixel is proportional to the square of noise free depth value. This kind of simulation was also adopted by Park *et al.* [26]. According to the analysis by Shen and Cheung [32], Kinect depth maps should also contain random missing depth values other than structure missing depth values, so we also add random missing values to the depth maps. This kind of simulation is also adopted by Yang *et al.* [37]. The final

TABLE IV

QUANTITATIVE COMPARISON ON SIMULATED KINECT DATASETS. RESULTS ARE EVALUATED IN TERMS OF AVERAGE MAE OF ALL THE DATA.
THE BEST RESULTS ARE IN BOLD.

	GF [10]	JGU [18]	JBW [14]	NLM-WLS [26]	JID [13]	AR [38]	WLS [23]	Ours
Dataset 1	2.69	2.71	2.84	1.27	2.15	0.9	1	0.68
Dataset 2	2.72	2.87	2.86	1.59	2.52	1.25	1.42	0.96

TABLE V

COMPUTATIONAL TIME (IN SECONDS) COMPARISON OF DIFFERENT METHODS ON DIFFERENT DATASETS

	GF [10]	JBW [14]	JGU [18]	NLM-WLS [26]	JID [13]	TGV [5]	WLS [23]	AR [38]	Ours
ToF simulated	0.47	27.64	254.34	147.98	1046.12	3860.91	74.57	243.92	210.05
ToF real	0.16	8.79	83.13	50.72	-	1337.9	24.47	116.65	69.96
Kinect simulated	0.084	4.78	45.21	28.07	409.38	-	12.96	29.77	27.87
Kinect real	0.1	5.54	52.37	33.06	469.98	-	13.98	35.77	32.43

simulated Kinect depth maps in Kinect dataset 2 are shown in the second row of Fig. 11.

Table IV shows the quantitative comparison of simulated Kinect restoration results in terms of average MAE. The proposed method also yields the lowest average MAEs on the two datasets. Fig. 12 shows visual comparison. Our method performs well in preserving sharp depth discontinuities. The corresponding error maps are shown in Fig. 13 for better comparison. As shown in the error maps, our method clearly outperforms other methods with much smaller errors.

Real Kinect data are also used to test the proposed method. As the first generation Kinect (Kinect v1) is different from the second generation Kinect (Kinect v2), the real Kinect data used in this paper include depth maps captured by Kinect v1 and Kinect v2. We reuse the Kinect v1 data in [38] which are from the NYU dataset [33] and are shown in the first two rows of Fig. 14. The Kinect v2 data are from the SUN RGBD dataset [34] which are shown in the last two rows of Fig. 14. Both results of WLS [23], NLM-WLS [26] and AR model [38] suffer from blurring depth discontinuities as shown in highlighted regions. The inpainting based method JID [13] can preserve sharp depth discontinuities but they are quite jaggy. Our method can properly rectify the jaggy depth discontinuities. Depth discontinuities in our results are also quite sharp.

Table V shows the computational time of different methods on different datasets. Our method is slower than WLS [23] and NLM-WLS [26]. However, it is faster than the AR model [38] and much faster than TGV [5] and JID [13].

V. CONCLUSION

In this paper, we propose a robust framework for color guided depth map restoration. The proposed method is robust against the inconsistency between color edges and depth discontinuities. Through both mathematical analysis and experimental results, we show that the proposed method can both suppress texture copy artifacts and preserve sharp depth discontinuities. Experimental results on both simulated and real ToF and Kinect data have shown the promising performance of the proposed method.

REFERENCES

- [1] M. Camplani, T. Mantecon, and L. Salgado, "Depth-color fusion strategy for 3-D scene modeling with Kinect," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1560–1571, Dec. 2013.
- [2] J. Diebel and S. Thrun, "An application of Markov random fields to range sensing," in *Proc. NIPS*, 2005, pp. 291–298.
- [3] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," in *Proc. ACM Trans. Graph. (TOG)*, 2008, p. 67.
- [4] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.
- [5] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *Proc. ICCV*, 2013, pp. 993–1000.
- [6] W. Freeman and C. Liu, "Markov random fields for super-resolution and texture synthesis," in *Proc. Adv. Markov Random Fields Vis. Image Process.*, 2011, pp. 155–165.
- [7] X. Gong, J. Ren, B. Lai, C. Yan, and H. A. Qian, "Guided depth upsampling via a cosparse analysis model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2014, pp. 738–745.
- [8] U. Hahne and M. Alexa, "Exposure fusion for time-of-flight imaging," in *Proc. Comput. Graph. Forum*, 2011, pp. 1887–1894.
- [9] B. Ham, D. Min, and K. Sohn, "Depth superresolution by transduction," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1524–1535, May 2015.
- [10] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Sep. 2013.
- [11] M. Hornacek, C. Rhemann, M. Gelautz, and C. D. Rother, "Depth super resolution by rigid body self-similarity in 3D," in *Proc. CVPR*, 2013, pp. 1123–1130.
- [12] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of Kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.
- [13] M. Kiechle, S. Hawe, and M. Kleinsteuber, "A joint intensity and depth co-sparse analysis model for depth map super-resolution," in *Proc. ICcv*, 2013, pp. 1545–1552.
- [14] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," in *Proc. ACM Trans. Graph.*, 2007, p. 96.
- [15] D. Krishnan, R. Fattal, and R. Szeliski, "Efficient preconditioning of laplacian matrices for computer graphics," in *Proc. ACM Trans. Graph. (TOG)*, 2013, p. 142.
- [16] H. Kwon, Y.-W. Tai, and S. Lin, "Data-driven depth map refinement via multi-scale sparse representation," in *Proc. CVPR*, 2015, pp. 159–167.
- [17] J. Li, Z. Lu, G. Zeng, R. Gan, and H. Zha, "Similarity-aware patchwork assembly for depth image super-resolution," in *Proc. CVPR*, Jun. 2014, pp. 3374–3381.
- [18] M.-Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc. CVPR*, 2013, pp. 169–176.
- [19] W. Liu, X. Chen, J. Yang, and Q. Wu, "Variable bandwidth weighting for texture copy artifacts suppression in guided depth upsampling," *IEEE Trans. Circuits Syst. Video Technol.*, no. 99, p. 1, 2016.
- [20] J. Lu, K. Shi, D. Min, L. Lin, and M. N. Do, "Cross-based local multipoint filtering," in *Proc. CVPR*, 2012, pp. 430–437.
- [21] S. Lu, X. Ren, and F. Liu, "Depth enhancement via low-rank matrix completion," in *Proc. CVPR*, 2014, pp. 3390–3397.
- [22] O. M. Aodha, N. D. Campbell, A. Nair, and G. J. Brostow, "Patch based synthesis for single depth image super-resolution," in *Proc. ECCV*, 2012, pp. 71–84.
- [23] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, "Fast global image smoothing based on weighted least squares," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5638–5653, Dec. 2014.

- [24] 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.
- [25] D. Min and K. Sohn, "Cost aggregation and occlusion handling with WLS in stereo matching," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1431–1442, Aug. 2008.
- [26] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3D-TOF cameras," in *Proc. ICCV*, Nov. 2011, pp. 1623–1630.
- [27] J. Park, H. Kim, Y.-W. 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.
- [28] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [29] L. Pizarro, P. Mrázek, S. Didas, S. Grewenig, and J. Weickert, "Generalised nonlocal image smoothing," *Int. J. Comput. Vis.*, vol. 90, no. 1, pp. 62–87, 2010.
- [30] A. Rajagopalan, A. Bhavsar, F. Wallhoff, and G. Rigoll, "Resolution enhancement of PMD range maps," in *Proc. Pattern Recognit.*, 2008, pp. 304–313.
- [31] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "Lidarboost: Depth superresolution for ToF 3D shape scanning," in *Proc. CVPR*, 2009, pp. 343–350.
- [32] J. Shen and S.-C. S. Cheung, "Layer depth denoising and completion for structured-light RGB-D cameras," in *Proc. CVPR*, Jun. 2013, pp. 1187–1194.
- [33] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGB-D images," in *Proc. ECCV*, 2012, pp. 746–760.
- [34] S. Song, S. P. Lichtenberg, and J. A. Xiao, "RGB-D: A RGB-D scene understanding benchmark suite," in *Proc. CVPR*, 2015, pp. 567–576.
- [35] X. Tan, C. Sun, and T. D. Pham, "Multipoint filtering with local polynomial approximation and range guidance," in *Proc. CVPR*, 2014, pp. 2941–2948.
- [36] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. ICCV*, Jan. 1998, pp. 839–846.
- [37] J. Yang, X. Ye, K. Li, and C. Hou, "Depth recovery using an adaptive color-guided auto-regressive model," in *Proc. ECCV*, Oct. 2012, pp. 158–171.
- [38] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive auto-regressive model," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3443–3458, Aug. 2014.
- [39] H. Zimmer, A. Bruhn, and J. Weickert, "Optic flow in harmony," *Int. J. Comput. Vis.*, vol. 93, no. 3, pp. 368–388, 2011.



Wei Liu received the B.Eng. degree from the Department of Automation, Xi'an Jiaotong University, in 2012. He is currently pursuing the Ph.D. degree with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, under the supervision of Prof. J. Yang. His current research interests mainly include image/video denoise, image deblur, and depth image restoration.



Xiaogang Chen received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2013. Then he joined the University of Shanghai for Science and Technology, Shanghai, as an Assistant Professor. His current research interests mainly focus on image and video processing.



Jie Yang received the Ph.D. degree from the Department of Computer Science, Hamburg University, Germany, in 1994. He is currently a Professor with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China. He has led many research projects (e.g., National Science Foundation, 863 National High Tech. Plan), had one book published in Germany, and authored over 200 journal papers. His major research interests are object detection and recognition, data fusion and data mining, and medical image processing.



Qiang Wu received the B.Eng. and M.Eng. degrees from the Harbin Institute of Technology, Harbin, China, in 1996 and 1998, respectively, and the Ph.D. degree from the University of Technology Sydney, Australia, in 2004. He is currently an Associate Professor and a Core Member of the Global Big Data Technologies Centre, University of Technology Sydney. His research interests include computer vision, image processing, pattern recognition, machine learning, and multimedia processing. His research outcomes have been published in many premier international conferences including the ECCV, the CVPR, the ICIP, and the ICPR and the major international journals such as the IEEE TIP, the IEEE TSMC-B, the IEEE TCSVT, the IEEE TIFS, the PR, the PRL, the Signal Processing, and the IEEE SIGNAL PROCESSING LETTER. He also serves several journals and conferences as a Reviewer, including the TPAMI, the PR, the TIP, the TCSVT, the TSMC-B, the CVIU, the IVC, the PRL, the Neurocomputing, and the EURASIP Journal on Image and Video Processing.