This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

Explicit Edge Inconsistency Evaluation Model for Color-guided Depth Map Enhancement

Y. Zuo, Q. Wu, *Member, IEEE*, J. Zhang, *Member, IEEE*, P. An, *Member, IEEE*

***Abstract*—Color-guided depth enhancement is to refine depth maps according to the assumption that the depth edges and the color edges at the corresponding locations are consistent. In the methods on such low-level vision task, Markov Random Fields (MRF) including its variants is one of major approaches, which has dominated this area for several years. However, the assumption above is not always true. To tackle the problem, the state-of-the-art solutions are to adjust the weighting coefficient inside the smoothness term of MRF model. These methods are lack of explicit evaluation model to quantitatively measure the inconsistency between the depth edge map and the color edge map, so it cannot adaptively control the efforts of the guidance from the color image for depth enhancement leading to various defects such as texture-copy artifacts and blurring depth edges. In this paper, we propose a quantitative measurement on such inconsistency and explicitly embed it into the smoothness term. The proposed method demonstrates the promising experimental results when compared with benchmark and the state-of-the-art methods on Middlebury datasets, ToF-Mark datasets and NYU datasets.**

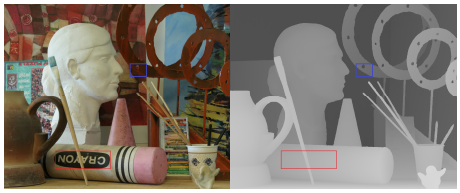
***Index Terms*—Depth Map Super-resolution (SR), Depth Map Completion, Depth Map Enhancement, Markov Random Field (MRF), RGB-D Camera**

I. Introduction

ACQUIRING high-quality depth maps is the key problem in the field of 3-D computer vision, which is required in many applications, e.g., interactive view interpolation, 3DTV, 3D object modeling, robot navigation, and 3D tracking. Generally speaking, methods on depth map acquisition consist of two categories: passive methods and active methods. Passive methods can generate a depth map from two-view or multi-view color images using stereo matching algorithms. In several decades, the performances of such methods are significantly improved. However, these methods still suffer from the inherent problems such as matching difficulties in texture-less areas and occlusion [1].

Active depth acquisition methods can obtain depth videos with the same frame rate as color cameras using depth sensors.

Y. Zuo and P. An are with School of Communication and Information Engineering, Shanghai University, Shanghai, 200072, China. (e-mail: Yifan.Zuo-1@student.uts.edu.au; anping@shu.edu.cn). Q. Wu and J. Zhang are with Faculty of Engineering and Information Technology, University of Technology, Sydney, 15 Broadway, Ultimo NSW 2007, Australia (e-mail: Qiang.Wu@uts.edu.au; Jian.Zhang@uts.edu.au). This work was supported in part by the National Natural Science Foundation of China, under Grants U1301257, 61571285, 61422111, and 61172096.



(a) Color image (b) Registered depth map

Fig. 1 An illustration of edge inconsistency (red window: edges occur on the color image but not on the depth map, blue window: edges occur on the depth map but not on the color image)

Compared with passive methods, depth acquisition through active methods is much more efficient. Particularly, in texture-less areas, active methods are able to achieve more robust performances than passive methods. So far, there are mainly two types of depth sensors, which are ToF sensors and structured-light sensors. In ToF sensors, depth maps are computed by measuring the phase difference between the emitted light and the reflected light [2]. The drawback is that the captured depth maps are noisy with low resolutions e.g., 176×144 or 200×200. In structured-light sensors, an infrared light source projects a dot pattern on the scene. Another offset infrared camera senses the pattern and estimates the depth map. Although structured-light sensors can obtain depth maps with higher resolutions, the quality of the depth maps obtained by such sensors is not satisfying. There are some holes (i.e. places without depth information sensed) appearing on the depth map. These holes may be caused by occlusion, weak reflection to the infrared light on some surfaces or even shadow reflection of the light patterns. Overall, objects in darker colors, specular surfaces, or fine-grained surfaces e.g., human hair are difficult to get depth sensing through depth sensors [3].

According to the analysis above, the main problems of depth maps obtained by depth sensors are low resolution, noisy depth values and holes. In this work, all these issues will be investigated under a uniform solution based on the improved MRF-based model. Given a low quality depth map and a companion high quality color image, a high quality depth map is produced.

This is so-called color-guided depth enhancement method. Such kind of methods always refine a depth map under the assumption that the depth edges and the color edges at the corresponding locations are consistent [6]. However, this assumption is not always true. The incorrect guidance from the companion color image will lead to texture-copy artifacts and blurring depth edges on the reconstructed depth map. Texture-copy artifacts derive from the situation that the smooth depth region corresponds to the color region with rich texture.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

By contrast, blurring depth edges result from the case that the smooth color region corresponds to the depth region with edges. Fig. 1 illustrates the edge inconsistency explained above. There have been several methods [4, 5, 12, 13, 18] trying to sort out these problems, the solutions (e.g. the methods based on MRF and its variants) are more-or-less to introduce a weighing scheme to balance the contribution from the original depth map and the companion color image. Very recently, W. Liu et al. [52] proposes a bandwidth adaption method for each patch on the depth map by checking its relative smoothness. Such method can be used in many existing methods. But all the methods shown above do not explicitly evaluate the edge inconsistency between the color image and the corresponding depth map. Therefore, they cannot adaptively control the efforts of the guidance from the color image when enhancing the depth map leading to texture-copy artifacts and blurring depth edges. In this paper, the main contributions are in three aspects. 1. Our method explicitly considers the inconsistency occurring between depth edges and the corresponding color edges and measures such inconsistency quantitatively. This quantitative measurement can provide a more precise definition of the inconsistency between the depth edge and the color edge in a numerical way. 2. Our method explicitly embeds the inconsistency measurement above into the smoothness term of the MRF energy function. Such model is able to mitigate texture-copy artifacts and preserve depth edges. 3. Our method solves depth map super-resolution (SR) and depth map completion via a uniform model. And the proposed method is evaluated on the performances of depth map SR and depth map completion on Middlebury datasets [43], ToF-Mark datasets [16] and NYU datasets [45], comparing with the state-of-the-art algorithms. Furthermore, the robustness of the proposed method is evaluated on extremely lower quality depth maps which have lower resolution and significant holes. All experimental results show the improved performance against the state-of-the-art depth enhancement methods.

The rest of this paper is organized as follows: In Sec. II, we review and analyze related work in depth map SR and depth map completion. Sec. III presents the proposed algorithm via Markov Random Fields (MRF) with edge inconsistency measurement. In sec. IV, the experimental results are presented. Sec. V concludes this paper.

II. Related Work

In recent years, there are a few proposed works on depth map SR and depth map completion, which are applied on depth maps captured by either ToF sensors or structured-light sensors.

A. Depth enhancement for ToF sensors

As mentioned above, depth maps captured by ToF sensors are noisy and subject to low resolutions. Therefore, the main tasks are to up-sample and de-noise the captured depth maps. Existing methods can be classified into two categories: non-color-guided methods [23-25] and color-guided methods [4-22].

In non-color-guided methods, paper [23] only requires a single depth map for up-sampling by using smoothing priors based on local self-similarities, but it either has difficulties in textured areas, or only works well for the case of small up-sampling factor. J. Xie et al. [25] proposes a single depth map SR method via a modified joint bilateral filter. Such bilateral filter is guided by a HR edge map which is constructed from the edges of the low-resolution (LR) depth map through an MRF optimization in a patch synthesis based manner. Another type of non-color-guided approach [24] is to fuse multiple displaced LR depth maps into a single HR depth map, which is not convenient for real applications because the geometrical relationship between all the depth sensors cannot be easily determined. Furthermore, such methods cannot obtain depth information where depth information of certain objects (e.g. human hair) cannot be sensed by any one of the sensors.

For the color-guided methods, it intends to improve the quality and the resolution of the captured depth map with the support of a registered HR color image. The fundamental assumption of such methods is that the depth edges and the color edges at the corresponding locations are consistent [6]. Such color-guided methods can be further classified into three categories that are local methods [7-13], global methods without machine learning [4-6, 14-19] and global methods with machine learning [20-22].

For local methods, according to the bilateral filtering techniques, a joint bilateral up-sampling (JBU) framework is proposed by J. Kopf et al. [7], which can be used for up-sampling LR depth maps. The edges of the HR depth map can be refined according to the edges of the registered HR color image. M. Liu et al. [8] proposes a variant of JBU. It computes weighting coefficients based on geodesic which is a joint space of color and distance instead of separating color space and distance space. Q. Yang et al. [9] also proposes a depth map SR method based on the joint bilateral filtering (JBF) techniques, which refines depth maps iteratively via a set of depth candidates. D. Min et al. [11] proposes a Weighted Mode Filtering method (WMF) based on joint histogram of depth candidates to better preserve the depth edge.

Compared with local methods, global methods are more robust to noise. MRF-based methods are major approaches in this category of methods. There are two terms in MRF, which are a data term and a smoothness term. The data term indicates the compatibility of the reconstructed depth values with the sensed ones and the smoothness term contributes to a piecewise smooth solution. J. Diebel et al. models depth map SR as solving a multi-labeling optimization problem via MRF [6]. J. Lu et al. [14] further extends this work by designing a data term which can better fits to the characteristics of depth maps. J. Zhu et al. [15] updates the traditional spatial MRF to dynamic MRF. Therefore, both spatial and temporal information can be introduced in an energy function, which improves the accuracy and the robustness for dynamic scenes. J. Park et al. [4] uses a non-local term to regularize depth maps and combines it with a weighting scheme which involves edge, gradient, and segmentation information extracted from HR color images. D. Ferstl et al. [16] models the smoothness term as a second order total generalized variation regularization, and guides the depth

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

map SR with an anisotropic diffusion tensor which is computed from the registered HR color image. K. Lo et al. [18] also presents a framework by solving an MRF labeling optimization problem. Such method shows the capability of preserving depth edges while suppressing the artifacts of texture copying caused by inconsistent color edges. The main constraints are in two aspects. 1. [18] only uses range information of local patches on the depth map to classify smooth regions and unsmooth regions. Based on the classification above, different weighting scheme are performed. However, they do not take both local edge structure and global edge structure into account. Therefore, the improvement is limited. 2. [18] only detects the edges on the depth map only, such method does not explicitly measure inconsistency between the color edge map and the depth edge map. Therefore, the guidance may not be correct in some situations. In addition to MRF-based methods, J. Yang et al. [5] achieves depth enhancement via the color-guided auto-regression model (AR). The depth recovery task is formulated as a problem for minimizing AR prediction errors. The AR predictor for each pixel is constructed according to both the local correlation on the initial depth map and the nonlocal similarity in the accompanied high quality color image.

Motivated by common RGB image SR methods, sparse representation techniques are introduced for depth map SR. Y. Li et al. [21] jointly trains dictionaries for registered patches of the LR depth maps, the HR depth maps and the color images. Then, each patch of the HR depth map is reconstructed through sparse representation of learned dictionaries independently. The final result is constructed by using such patches with averaging the overlapped regions. M. Kiechle et al. [22] exploits the co-sparsity of analysis operators and reconstructs the HR depth map through data fidelity and color-guided sparsity constraint. The results of such methods heavily rely on the selection of external datasets and always suffer from over-smoothed artifacts on either the overlapping regions of adjacent patches or the depth edges.

B. Depth enhancement for structured-light sensors

There are two major problems on the depth maps obtained by structured-light sensors: holes and noise. The operation, aiming to sort out these two problems, is called depth map completion [33]. The state-of-the-art methods can be grouped into two categories: depth fusion-based methods [26, 27] and color-guided methods which consist of image in-painting-based methods [28-34] and SR-based methods [4-11, 14, 16, 35].

For depth fusion methods, KinectFusion [26] integrates noisy depth maps captured at various viewpoints. In contrast to the single raw Kinect depth, the fused depth map has less holes and less noise. Multi-Kinect-Fusion [27] uses multiple low-cost depth sensors to obtain a fused depth map that can be arbitrary positioned between the input sensors. However, these methods either have problem in capturing a depth video or are lack of robustness due to the overlay of different infrared patterns on the scene. L. Wang et al. [28] proposes a stereoscopic in-painting algorithm which jointly completes the missing texture and depth via two pairs of RGB and depth cameras. Holes occluded by foreground are completed by minimizing a predefined energy function. Such system requires an additional pair of RGB and depth cameras. Y. Berdnikov et al. [29] combines the "deepest neighbor" method with the spatial interpolation method to address two different kinds of holes. One is caused by the edges of foreground objects, and the other is caused by shiny surfaces, certain special object characteristics or other uncertain factors. Although such method achieves real-time performance, the reconstructed depth maps are not always consistent with the corresponding color images, especially in the regions near the boundaries between the background and foreground. Very recently, S. Xiang et al. [30] proposes a method which validates the pair of edges on the depth map and the corresponding color image. Such validation is only based on local structure. A simple threshold is adopted to filter unreliable edge pairs. For depth map enhancement, [30] enhances the low quality pixels detected by the validation process above using JBF [7]. The constraint is that [30] does not enhance the rest of pixels or assumes noise free in the rest pixels. To tackle this constraint, the proposed method assumes that noise across the whole original depth map and enhances every depth pixel through an improved MRF optimization instead of JBU. In addition, many SR-based methods for depth map SR mentioned in subsection *A* can also be used for depth map completion task, such as methods [4-11, 14, 16, 35].

C. The main contributions of the proposed method when compared with the existing methods

The proposed method is an advanced color-guided depth enhancement method which is based on a modified MRF-based model. The method can be applied onto depth maps acquired by either ToF sensors or Structured-light sensors. Compared with the existing color-guided methods, the proposed method can best mitigate texture-copy artifacts and preserve depth edges.

In this paper, we propose a quantitative measurement on edge inconsistency between the registered color image and the depth map. Then such inconsistency measurement is explicitly embedded into the proposed MRF-based model which can provide high quality depth maps for the task of depth map SR and depth map completion.

Although some color-guided methods also formulate a uniform model for both types of depth sensors (i.e. ToF sensors and Structured-light sensors), they may have difficulties in addressing depth map SR and depth map completion simultaneously. By contrast, the proposed method successfully improves the performances on such problem. A preliminary work is published in [51]. The extensive work is carried out in this paper. It not only presents more details but also revises the previous work in order to deal with broader cases. The revised solution is able to enhance depth maps at smooth regions which appear on both the depth map and the corresponding color image. This case was not dealt with precisely in the previous work [51]. Furthermore, as mentioned above, this paper adopts the proposed solution to deal with both depth map SR and depth map completion. More comparison experiments are presented, which shows the improved performance of this work against

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

the previous work [51].

III. Proposed Method

Markov Random Field, also known as Markov network or undirected graphical model has been widely utilized for many image processing applications and tasks. MRF formulates depth map enhancement as solving an optimization problem. The input includes a high quality color image and a low quality depth map. According to the Hammersely-Clifford theorem [37], solving MRF inference problem is equivalent to optimizing the Gibbs energy function for which general formulation is defined as follows:

$$\mathbf{D}=\arg\min_{\mathbf{D}=\{d_p\}}\sum_{p\in\mathbf{O}}E_{data}\left(d_p,d_p^{\,0}\right)+\lambda\sum_{p}\sum_{q\in\mathbf{N_p}}E_{smooth}\left(d_p,d_q\right) \quad (1)$$

$$E_{data}\left(d_p,d_p^{\,0}\right)=\left(d_p-d_p^{\,0}\right)^2 \quad (2)$$

$$E_{smooth}\left(d_p,d_q\right)=\omega_{pq}\left(d_p-d_q\right)^2 \quad (3)$$

where $\mathbf{D}$ indicates the value set of the reconstructed depth map, d_p indicates the reconstructed depth value of the pixel p, $d_p^{\,0}$ is the observed depth value of p. $\mathbf{O}$ is the pixel set consisting of pixels with observed depth values. E_{data} is called the data term which maintains the consistency between the reconstructed depth value and the observed one. E_{smooth} is called the smoothness term which penalizes the difference of the reconstructed depth values between certain pixel and each neighboring pixel. The parameter λ is used to balance the data term and the smoothness term. $\mathbf{N_p}$ is the set of 8-connected neighboring pixels for p.

According to the MRF-based depth enhancement framework shown in Eq. (1), a common method models $E_{data}\left(d_p,d_p^{\,0}\right)$ and $E_{smooth}\left(d_p,d_q\right)$ as Eq. (2) and Eq. (3) respectively based on the assumption of Gaussian White Noise [44]. ω_{pq} links the color image to the depth map, which provides the guidance from the color image for depth map enhancement based on the assumed consistency between the color edge and the depth edge [6]. As mentioned above, this assumption is not always true. It is the fundamental problem leading to texture-copy artifacts and blurring depth edges. To tackle such problems, this paper proposes a novel weighting $\lambda_{smooth-pq}$ to replace ω_{pq} in Eq. (3) for the first time by introducing the quantitative measurement on the inconsistency between color edges and depth edges. In addition, it should be explained that the proposed MRF-based model is constructed in continuous domain rather than discrete labels. The reason is twofold: on the one hand, depth value is always recorded in mm unit as a continuous floating value, e.g. ToF-Mark datasets [16]; on the other hand, when the number of labels is large, e.g. 0-255 for the case of multi-label graph cut, it is difficult to obtain the global minimum via discrete inference even though the energy function is convex.

Subsection *A* presents the details of inconsistency measurement. Subsection *B* and *C* explicitly embed such measurement into MRF-based model. Subsection *D* provides the optimization method for our model. The complexity discussion is shown in subsection *E*.

A. Edge inconsistency measurement between the corresponding color image and the depth map

Motivated by [38], the inconsistency measurement between the color edge map and the depth edge map is formulated as a bi-directional edge map quality assessment. In order to introduce image quality assessment into the edge inconsistency measurement, a few specific points should be discussed.

1) To measure the inconsistency between the depth edge map and the color edge map, the resolutions of these two edge maps must be the same. In our case, the depth map in lower resolution or with holes is roughly interpolated through gridded or scattered interpolation methods before edge detection [46, 47].

2) Because the color image and the corresponding depth map have the structural similarity which is clearly observed on the relevant binary edge maps, this paper measures the inconsistency between the binary edge maps generated from the color image and the corresponding depth map respectively.

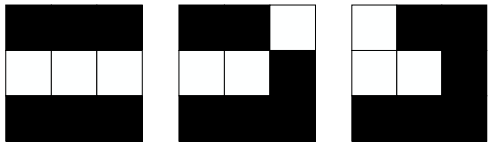
3) In [38], common edge map quality measurement is based on the position shift of each edge pixel against the position on the ground truth. However, the case to be investigated here is different. In the case of this paper, the matched edge pixels on the depth edge map and the color edge map which should have located in the same position always have displacement with each other. The reasons are some preprocessing such as coarse interpolation as explained above or noise in depth sensors. Thus, it is impossible to measure the inconsistency on the difference between the positions of each pair of matched edge pixels like the existing edge quality measurement methods [38, 48, 49]. Instead, the inconsistency measurement in this paper is based on the structure similarity of the edge maps which considers the structure presented by local neighboring regions as well as the global structure of the whole depth map.

Canny operator [39] is applied upon the intensity component of the color image and the coarsely interpolated depth map to generate the relevant edge maps. Due to low quality of the interpolated depth map, the positions of the corresponding edge pixels on the color edge map and the depth edge map are not consistent strictly. For convenience, the following explanation is based on the reference edge map and the target edge map whose meaning can be found in the end of this subsection.

For each edge pixel on the reference edge map, it will search the best consistency on the target edge map within a neighboring region around the corresponding position. This implies that if the color edge and the depth edge are consistent, the displacement of matched edge pixels should be constrained in a small range. Moreover, strength and orientation of the displacements of all matched edge pixels in a nearby region should be consistent. These two constraints are solved in an MRF optimization through its data term and smoothness term respectively. The data term implies local structure information and smoothness term implies global structure information. Therefore, the inconsistency measurement is robust to the errors in the original depth edge map.

$$\mathbf{L}=\arg\min_{\mathbf{L}=\{l_p\}}\sum_{p\in\mathbf{ref}}C\left(p,p+l_p\right)+\mu\cdot\sum_{p\in\mathbf{ref}}\sum_{q\in\mathbf{N_p}}V\left(l_p,l_q\right) \quad (4)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology



(a) Reference patch (b) Target patch 1 (c) Target patch 2

Fig. 2 An illustration on the advantage of Weighted Bipartite Matching

where $C(p, p+l_p)$ and $V(l_p, l_q)$ are the data term and the smoothness term in the MRF energy function respectively. μ is a balance factor between the data term and the smoothness term. It is set to 0.1 in this paper. p represents the position of an edge pixel on the reference edge map **ref** . $\mathbf{N_p}$ is the set of 8-connected neighboring pixels of p . l_p which is an element of $\mathbf{L}$ stands for the displacement for p . Therefore, $p+l_p$ represents the position of the edge pixel k on the target edge map. Since the sub-pixels created virtually by interpolation process may not be stable, each edge pixel in the reference edge map is mapped to the existing edge pixel detected in the target edge map. In other words, the proposed method assigns $\mathbf{L}$ in integer pixel precision. In our work, the size of the neighboring region is 5×5 for 2× SR, 7×7 for 4× SR, 9×9 for 8× SR, 11×11 for 16× SR and 7×7 for depth map completion without SR. The data term $C(p,k)$ matches the reference edge pixel p against the target edge pixel k . Given p , if certain target pixel k on the target edge map is not an edge pixel, it is regarded as definite inconsistency. In that case, $C(p,k)$ is assigned to the maximum inconsistency value (i.e. 1 in our work). Otherwise, this inconsistency is measured on two patches where the edge pixel p and the edge pixel k are the center positions respectively. In this paper, the size of the patch is 3×3. This measurement is sorted out through Minimum Weighted Bipartite Matching [40] which is more robust than Mean of Absolute Difference (MAD). In a weighted bipartite graph, each graph edge has an associated value. A Minimum Weighted Bipartite Matching is to find the best matching where the sum of the values of graph edges (graph edges are the set of arcs or lines in graph theory) linking matched vertices is a minimum. In our case, the quality of the bipartite matching is measured according to the difference between the locations of the matched edge pixels and the amounts of edge pixels in two patches. Fig. 2 provides an illustration on the advantage of Weighted Bipartite Matching compared with MAD. In the three patches shown in Fig. 2, white pixels and black pixels represent edge pixels and non-edge pixels respectively. When MAD is applied, both b) and c) (i.e. the target patches) are similar to the reference patch a). However, it can be observed that, in term of local structure, target patch b) is closer to a). Such fine-grained level similarity can be successfully picked up by Bipartite Graph Matching used in this paper. Based on the analysis above, the data term $C(p,k)$ is expressed as Eq. (5).

$$C(p,k)=\begin{cases}1(\textit{definite inconsistency}), & \textit{If } k \notin \textit{edge pixels}\\ \textit{BMatching}(\mathbf{V_p},\mathbf{V_k},\mathbf{E},\mathbf{W}), & \textit{Otherwise}\end{cases} \quad (5)$$

where *BMatching* stands for Minimum Weighted Bipartite Matching [40]. The bipartite graph $\mathbf{G}(\mathbf{V_p},\mathbf{V_k},\mathbf{E},\mathbf{W})$ is defined. $\mathbf{V_p}$ and $\mathbf{V_k}$ are vertices, $\mathbf{E}$ represents graph edges between vertices and $\mathbf{W}$ is the vector which assigns weighting to each graph edge in $\mathbf{E}$. Specifically, $\mathbf{V_p}=\{ep_1, ep_2, ..., ep_M\}$ and $\mathbf{V_k}=\{ek_1, ek_2, ..., ek_N\}$ represent the sets of edge pixels in the two patches (excluding p and k which are the center edge pixels of these two patches). M and N are the amount of edge pixels inside these two sets. Thus, the inconsistency measurement between p and k is regarded as a matching problem between two data sets $\mathbf{V_p}$ and $\mathbf{V_k}$. In addition, the locations of an edge pixel and its true matched edge pixel are assumed to be close to each other. This assumption complies with the similarity of local structural information. Therefore, $\mathbf{W}$ is defined as $\phi(ep_i, ek_j)$ which is a monotonic function that returns a positive penalty for local structural matching.

$$\phi(ep_i, ek_j)=f\left(\left|ep_i^x-ek_j^x\right|+\left|ep_i^y-ek_j^y\right|\right) \quad (6)$$

where $f(0)=0, f(1)=1, f(2)=1.6$ and $f(x)=2$ when $x>2$. ep_i, ek_j are vertices in the bipartite graph, ep_i^x, ep_i^y are the coordinates of the edge pixel ep_i .

Minimum Weighted Bipartite Matching [40] is employed to enforce one-to-one matching between the edge pixel data sets above. That is, it assures any edge pixel in $\mathbf{V_p}$ / $\mathbf{V_k}$ matches only one edge pixel in $\mathbf{V_k}$ / $\mathbf{V_p}$ with $|M-N|$ unmatched pixels. Fig. 3 gives an illustration of Minimum Weighted Bipartite Matching with unmatched pixels marked. The amount of unmatched pixels also reflects the structure differences between the edge pixel sets $\mathbf{V_p}$ and $\mathbf{V_k}$. Furthermore, to effectively mitigate the errors of edge detection in noisy depth maps, the difference reflected by these unmatched pixels should be taken into account. It can be observed that when the amounts of edge pixels in both patches are similar, the additional matching cost is low. By contrast, when such amounts are very different, the proposed method considers that this edge may be caused by noise or this matching is not reliable by adding high additional matching cost. To consider these issues above, the edge inconsistency measurement term $\textit{BMatching}(\mathbf{V_p},\mathbf{V_k},\mathbf{E},\mathbf{W})$ in Eq. (5) is carefully adjusted and defined as Eq. (7).

$$\textit{BMatching}(\mathbf{V_p},\mathbf{V_k},\mathbf{E},\mathbf{W})=\left(\sum_{(mp_s,mk_s)\in\mathbf{V'_{pk}}}\phi(mp_s,mk_s)/2+|M-N|\right)/8 \quad (7)$$

where $\mathbf{V'_{pk}}=\{(mp_1,mk_1),(mp_2,mk_2),...(mp_r,mk_r)\}$ is the set of edge pixel pairs selected by Minimum Weighted Bipartite Matching [40]. $\phi(mp_s,mk_s)$ is the weight of the edge linking the edge pixel mp_s with the edge pixel mk_s and $s=\{1,2,3...r\}$. Therefore, $\sum_{(mp_s,mk_s)\in\mathbf{V'_{pk}}}\phi(mp_s,mk_s)$ is the matching cost of Minimum Weighted Bipartite Matching mentioned above. In order to constrain the data term $C(p,k)$ in the range of [0, 1],

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

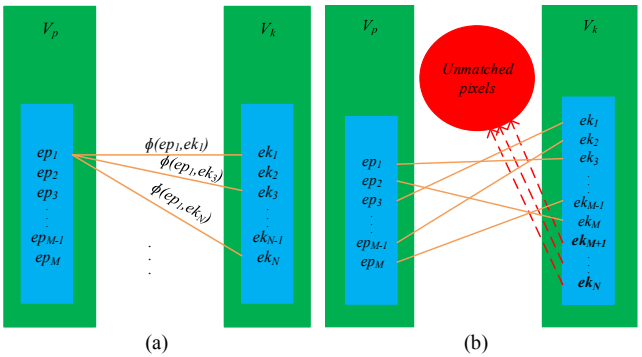


Fig. 3 An illustration of Minimum Weighted Bipartite Matching problem with (a) configuration of a bipartite graph and (b) a result of Minimum Weighted Bipartite Matching with unmatched pixels marked in bold.

the common normalization process is applied on Eq. (7).

$V(l_p,l_q)$ is the smoothness term in Eq. (4), which gives a penalty when adjacent edge pixels have different displacements as,

$$V_{q\in\mathbf{N_p}}(l_p,l_q)=\begin{cases}0, & l_p=l_q;\\ 1, & l_p\neq l_q;\end{cases} \tag{8}$$

where l_p is the label of the MRF inference problem in Eq. (4), representing the displacement vector for the edge pixel p. The relationship between labels and displacement vectors is one-to-one mapping.

Once the data term and the smoothness term in Eq. (4) are defined (see Eq. (5) and Eq. (8)), Graph cut [41] is adopted to resolve the discrete MRF inference problem. Then, the inconsistency measurement for edge pixels p represented by $C(p,k)$ can be computed by the optimized displacement l_p. In this paper, the inconsistency map $\mathbf{C}_{\mathbf{refer}}$ is defined as the set of inconsistency measurements for all edge pixels in the reference edge map. If there is no matching found for certain edge pixel, the displacement l of this edge pixel is meaningless. And its inconsistency value is assigned to maximum value (i.e. 1 in our work).

The inconsistency is measured based on the reference edge map against the target edge map. Thus, the measurement results will be different when swapping these two edge maps. In this work, the two edge maps are the color edge map and the corresponding depth edge map. When the color edge map is regarded as the reference edge map, it can be observed that the most inconsistent positions detected reflect the texture-copy happening areas. On the other hand, when the depth edge map is regarded as the reference edge map, it is observed that the most inconsistent positions detected reflect happening areas of blurring depth edges. Fig. 4 illustrates the bi-direction inconsistency measurement for Middlebury dataset "Art". The bi-direction inconsistency measurement is expressed in false color images. In Fig. 4 (c) and (d), the color along the edge pixels represents the strength of inconsistency of the edges between the reference edge map and the target edge map. According to the color scale coding in Fig. 4 (e), the color code on the leftmost side (i.e. dark blue) means the most consistent case. On the contrary, the color code on the rightmost side (i.e. dark red) means the most inconsistent case.

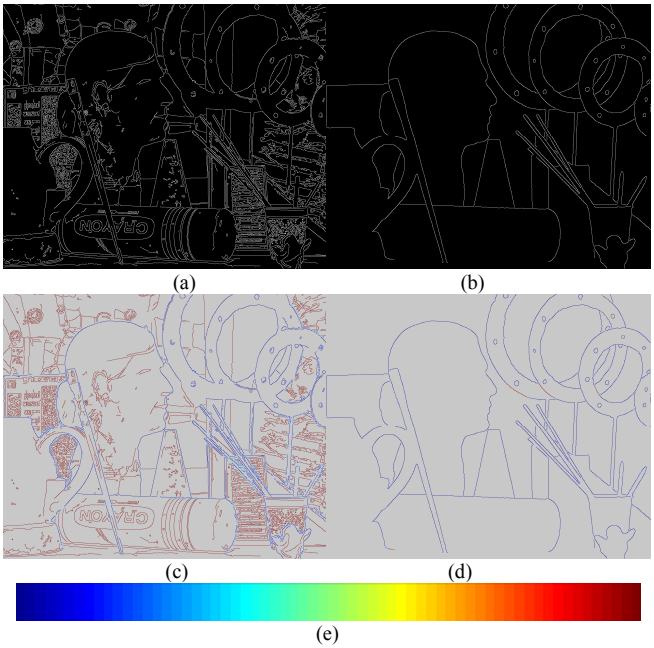


Fig. 4 The bi-direction inconsistency measurement for Middlebury dataset "Art", (a) the color edge map, (b) the depth edge map, (c) the inconsistency measurement in the case that the color edge map is the reference edge map, (d) the inconsistency measurement in the case that the depth edge map is the reference edge map, (e) the color scale coding, (The color code on the leftmost side (i.e. dark blue) means the most consistent case. The color code on the rightmost side (i.e. dark red) means the most inconsistent case). The inconsistency values of non-edge pixels in (b) and (d) are unavailable. They are shown in gray which is out of the color scale coding (e).

B. *Alignment of inconsistency maps*

After the bi-direction evaluation, there are two inconsistency maps $\mathbf{C}_{\mathbf{color}}$ (the color edge map is regarded as the reference edge map), $\mathbf{C}_{\mathbf{depth}}$ (the depth edge map is regarded as the reference edge map) as well as two sets of displacements $\mathbf{L}_{color}$, $\mathbf{L}_{depth}$ (defined in the same way as $\mathbf{C}_{\mathbf{color}}, \mathbf{C}_{\mathbf{depth}}$) for an image pair. Before embedding the inconsistency measurement values into the proposed MRF-based model, these two inconsistency maps must be consolidated to each other.

As mentioned before, the positions of edge pixels on the coarsely interpolated depth map are unreliable. On the contrary, the positions of edge pixels on the color edge map are more precise because of high quality of the color image. Through the solution of the MRF inference problem in Eq. (4) with the depth edge map as the reference edge map, the displacement between each depth edge pixel p and its matched color edge pixel k is $\mathbf{L}_{\mathbf{depth}}(p)$. Consequently, the true location of the observed depth edge pixel p supposes to be more close to $p+\mathbf{L}_{\mathbf{depth}}(p)$ when $\mathbf{C}_{\mathbf{depth}}(p)\neq 1$. For the case of definite inconsistency $\mathbf{C}_{\mathbf{depth}}(p)=1$, the position of the edge pixel p is unchanged because there has not any matched edge pixel in the color edge map. Moreover, due to the uncertainty of displacements, given a color edge pixel p', it may correspond to more than one depth edge pixels. Therefore, the adjusted $\mathbf{C}'_{\mathbf{depth}}$ is as,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

$$\begin{aligned}\mathbf{C}'_{\mathbf{depth}}(p') &= \min_{p\in\{p|p'=p+\mathbf{L}_{\mathbf{depth}}(p)\}} \mathbf{C}_{\mathbf{depth}}(p) \quad if\ \mathbf{C}_{\mathbf{depth}}(p)\neq 1 \\ \mathbf{C}'_{\mathbf{depth}}(p) &= \mathbf{C}_{\mathbf{depth}}(p) \qquad\qquad otherwise\end{aligned} \tag{9}$$

Eq. (9) defines that if there are more than one depth edge pixel p mapping to the same color edge pixel p', the best mapping with the lowest cost is adopted. And the proposed method maintains the positions and values of the rest mappings unchanged from $\mathbf{C}_{\mathbf{depth}}$ to $\mathbf{C}'_{\mathbf{depth}}$.

Once two inconsistency maps $\mathbf{C}'_{\mathbf{depth}}$ and $\mathbf{C}_{\mathbf{color}}$ are aligned, a confidence map $\boldsymbol{\alpha}$ is defined as Eq. (10) which has considered the inconsistency measurement in the bi-direction calculation. It describes the final inconsistency status between the color edge map and the depth edge map.

$$\boldsymbol{\alpha} = \max\left(\mathbf{C}'_{\mathbf{depth}}, \mathbf{C}_{\mathbf{color}}\right) \tag{10}$$

In the next subsection, this measurement is applied into MRF to fine tune the efforts of the guidance from the color image in order to improve the performance of color-guided depth enhancement.

C. Improved MRF with the inconsistency measurement

To simplify the explanation in the following, Eq. (3) is updated as Eq. (11),

$$E_{smooth}\left(d_p, d_q\right) = \lambda_{smooth-pq}\left(d_p - d_q\right)^2 \tag{11}$$

where $\lambda_{smooth-pq}$ is to replace ω_{pq} in Eq. (3). Generally speaking, guidance information for depth enhancement task can be derived from two sources. One is from the registered color image, and the other is from the original depth map itself. Based on the confidence map $\boldsymbol{\alpha}$ computed in Eq. (10), this paper combines these two kinds of information systematically to generate a new guidance for computing the weighting coefficient $\lambda_{smooth-pq}$.

$$\lambda_{smooth-pq} = e^{-\frac{\left(\left|\nabla_{color}{}^{pq}\right|\cdot(1-\alpha_{pq})+\left|\nabla_{depth}{}^{pq}\right|\cdot\alpha_{pq}\right)^2}{2\delta^2}} \tag{12}$$

where $\nabla_{color}{}^{pq}$ and $\nabla_{depth}{}^{pq}$ represent color difference and depth difference between the position p and its neighboring pixel q in the guided color image and the coarsely interpolated depth map respectively. δ controls decay rate of the exponential function. In addition, as mentioned above, only edge pixels have available confidence values, "max" operation is adopted when integrating $\boldsymbol{\alpha}(p)$ and $\boldsymbol{\alpha}(q)$ together to mitigate texture-copy artifacts and preserve depth edges better because of the single pixel width edges detected by Canny operator. α_{pq} is expressed as $\alpha_{pq} = \max\left(\boldsymbol{\alpha}(p), \boldsymbol{\alpha}(q)\right)$. More specifically, when neighboring pixel pair p,q are located across the edges in the color image as well as the depth map, α_{pq} is more close to 0 and $\nabla_{color}{}^{pq}$ plays a more important role in computing the weighting coefficient $\lambda_{smooth-pq}$. In such situation, the guidance from the registered color image helps recover sharp edges in the reconstructed depth map. By contrast, when neighboring pixel pair p,q only across the edge either in the color image or on the depth map, but not both, α_{pq} is more close to 1 and $\nabla_{depth}{}^{pq}$ provides main guidance. In these two situations, depth enhancement is through the approach of single depth map enhancement method. Indeed, some single depth enhancement method can be adopted to provide more accuracy depth map instead of simple interpolated depth map. However, the improvement is not significant when up-sampling factor is small. On the other hand, it is difficult to obtain accurate depth edges for large up-sampling factor by using single depth enhancement methods. Therefore, by considering the complexity and equity, we use coarse interpolated depth map as the guidance source for all cases. The benefit is twofold. On the one hand, it mitigates texture-copy artifacts. On the other hand, the guidance from the interpolated depth map is more reasonable than the incorrect guidance from the color image.

The scenario discussed above is on the regions around edge pixels. For pixels located on smooth regions where there is no edge pixel on neither the color image nor the coarsely interpolated depth map, Eq. (12) cannot satisfy such case because it is impossible to calculate the edge inconsistency in a local region where there is no edge pixel at all. In this paper, it is updated as Eq. (13) for this special case, where the guidance information for depth enhancement is from the coarsely interpolated depth map only to better mitigate texture-copy artifacts.

$$\lambda_{smooth-pq} = e^{-\frac{\left(\nabla_{depth}{}^{pq}\right)^2}{2\delta^2}} \tag{13}$$

Based on the analysis above, the proposed method can preserve depth edges and mitigate texture-copy artifacts efficiently by adaptively controlling the efforts of the guidance from the color image.

In addition, in regions near depth edges, δ should be small to preserve depth edges. By contrast, δ should be large to suppress noise in smooth regions. This paper uses fixed values for smooth regions and non-smooth regions respectively which are determined by the depth edge map. More specifically, if there is no edge pixel in the local windows centered at p and its neighboring pixel q respectively, the pixel pair of p,q is located at a smooth region. Otherwise, such pixel pair is located at a non-smooth region. In this paper, δ is set to 2 and 4 for non-smooth regions (Eq. (12)) and smooth regions (Eq. (13)) respectively.

D. Convex optimization for our MRF inference problem

It can be noticed that the energy function in Eq. (1) is a quadratic function which is a convex function. Therefore, there is a unique minimum for such energy function, standing for the global minimum. According to the Karush–Kuhn–Tucker conditions (KKT constraints) [50], we take the derivative of the energy function in Eq. (1) with respect to $\mathbf{D}$ and let it equal to zero. Specifically, for each $d_p \in \mathbf{D}$, such normal equation is shown in Eq. (14) with symbols defined in the same way as Eq. (1). After a simple merging of similar items, Eq. (14) can be converted to Eq. (15). Then the optimization problem is equivalent to solving a linear system $\mathbf{AD} = \mathbf{B}$, where $\mathbf{A}$ is a $n \times n$ symmetric matrix (n is the number of pixels), $\mathbf{D}$ is a $n \times 1$ matrix which consists of recovery depth values, $\mathbf{B}$ is a $n \times 1$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

matrix, made of observed depth values. This linear system is formulated as below. i, j stand for row i, column j in matrices. And they represent indexes of pixels in vectorized depth map under row order as well. $\mathbf{O}$ is the pixel set consisting of pixels with observed depth values. $\mathbf{N_i}$ is the set of pixels' indexes in the vectorized depth map which are 8-connected neighboring pixels of the pixel i in the two-dimension depth map.

$$\begin{cases} \dfrac{\partial E}{\partial d_p} = 2\times\left(d_p - d_p{}^0\right) + 4\lambda \sum\limits_{q\in\mathbf{N_p}} \lambda_{smooth-pq}\left(d_p - d_q\right) = 0, & p \in \mathbf{O} \\ \dfrac{\partial E}{\partial d_p} = 4\lambda \sum\limits_{q\in\mathbf{N_p}} \lambda_{smooth-pq}\left(d_p - d_q\right) = 0, & p \notin \mathbf{O} \end{cases} \tag{14}$$

$$\begin{cases} (1 + 2\lambda \sum\limits_{q\in\mathbf{N_p}} \lambda_{smooth-pq}) d_p - 2\lambda \sum\limits_{q\in\mathbf{N_p}} \lambda_{smooth-pq} d_q = d_p{}^0, & p \in \mathbf{O} \\ 2\lambda \sum\limits_{q\in\mathbf{N_p}} \lambda_{smooth-pq} d_p - 2\lambda \sum\limits_{q\in\mathbf{N_p}} \lambda_{smooth-pq} d_q = 0, & p \notin \mathbf{O} \end{cases} \tag{15}$$

$$\begin{cases} \mathbf{A}_{ii} = 1 + 2\lambda \sum\limits_{j\in\mathbf{N_i}} \lambda_{smooth-ij}, & i \in \mathbf{O} \\ \mathbf{A}_{ii} = 2\lambda \sum\limits_{j\in\mathbf{N_i}} \lambda_{smooth-ij}, & i \notin \mathbf{O} \\ \mathbf{A}_{ij} = -2\lambda\lambda_{smooth-ij}, & j \in \mathbf{N_i} \\ \mathbf{A}_{ij} = 0, & j \notin \mathbf{N_i} \end{cases} \tag{16}$$

$$\begin{cases} \mathbf{B}_i = d_i{}^0, & i \in \mathbf{O} \\ \mathbf{B}_i = 0, & i \notin \mathbf{O} \end{cases} \tag{17}$$

In this paper, we solve this linear system via Preconditioned Conjugate Gradients method (PCG) [42].

E. Algorithm complexity discussion

In the edge inconsistency measurement stage, the multi-label Graph Cut problem is solved by several binary-label Graph Cut sub-problems through $\alpha-$ expansion method [41]. The complexity of binary-label Graph Cut is up to $O(mn^2|C|)$, where m is the number of graph edges, n is the number of nodes in the graph (i.e. the number of edge pixels detected in the reference edge map) and $|C|$ is the cost of the minimum cut which is the smallest total weight of the edges which if removed would disconnect the source from the sink [53]. Therefore, the complexity of multi-label Graph Cut is up to $O(Lmn^2|C|)$, where L is the number of labels [41]. In addition, the complexity of the Hungarian algorithm [40] in our Weighted Bipartite Graph Matching is $O(V^2E)$, where V and E represents the number of vertices (i.e. the number of edge pixels in the two patches) and graph edges respectively.

IV. Experimental Results

The platform to carry out the experiments is a PC with Intel i7 2.60 GHz, 12G RAM. The plain Matlab implementation of our method (Graph Cut is implemented in C code) takes 115.39s on average to up-sample the low quality depth map up to the resolution of 1376×1088 in the case of 16×. The running time of each step is listed in Table I.

Our experiments consist of three parts. The first part (i.e. subsection *B*) is to evaluate the proposed method's performance on Middlebury datasets [43] in which the synthetic depth maps are degraded manually in various ways.

TABLE I
Average Running Time of Our Method (16×)

Running Time on Average	*Bi-inconsistency Measurement*	*Solve Linear Equation*	*Total*
Unit: Second	69.72	45.67	115.39

The comparison performance between the proposed method and several existing methods are shown. The second part (i.e. subsection *C*) is to apply our method on real datasets (ToF-Mark datasets [16] and NYU datasets [45]) to obtain high quality depth maps in order to show the robustness of the proposed method. The third part (i.e. subsection *D*) is to demonstrate the performance of the proposed method on depth map enhancement which is to tackle a difficult situation when the complex degradation occurs. It involves both lower resolution and significant holes.

A. Parameters setting

1) Canny thresholds

All the edge maps are computed through Canny operator. The proposed method intentionally sets the threshold of the detector low such that more edges especially main edges could be extracted. For color edges detection, the dual thresholds are 0.04 and 0.12. For the depth edge map calculation, in the case of depth map SR including subsection *D*, the dual thresholds of Canny operator are set in two ranges. Such parameters are determined in an empirical way through cross-validation process. In addition, thanks to the MRF optimization, the proposed edge inconsistency measurement is robust to the quality of depth edge map in some degree.

$$\begin{aligned} Th_L &= [(\log_2 factor)\times 0.01, (\log_2 factor)\times 0.02] \\ Th_h &= [(\log_2 factor)\times 0.03, (\log_2 factor)\times 0.04] \end{aligned} \tag{18}$$

where *factor* is the corresponding up-sampling factor. In the case of depth map completion, the dual thresholds are 0.03 and 0.07.

2) Balance factor λ in Eq. (1)

Regarding λ in Eq. (1), it can be theoretically analyzed through two aspects which are based on the up-sampling factor and the noise situation on the LR depth map. On the one hand, λ should decrease as the up-sampling factor increases. A larger up-sampling factor will cause sparsity in $\mathbf{d^0}$ dataset (i.e. less observed data in the set $\mathbf{O}$ in Eq. (1)) so the contribution of the data term in MRF (see Eq. (1)) is light. To balance the contributions of the data term and the smoothness term, it is necessary to reduce λ thus the contribution of the data term will be lift up relatively even the observed depth data in $\mathbf{d^0}$ is sparse. On the other hand, increasing λ for the case of stronger noise is able to provide the MRF model more robustness to noise by enhancing the efforts of the smoothness term.

In our work, it is also observed that when noise on the depth map is weak, the up-sampling factor has less impact to λ. That is, for different up-sampling factors, the optimal λ may have the close values as long as noise on the depth map is not significant. Through cross-validation process, we fix λ to 0.01 for all up-sampling experiments in which the LR depth maps are from Middlebury dataset without adding noise. For the case of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

adding noise, the relation between λ and the up-sampling factor is defined as,

$$\lambda = \frac{\kappa}{factor}, \quad factor > 1 \tag{19}$$

where κ is a constant, it is set to 3.2 in all depth map SR experiments. *factor* is the up-sampling factor. In addition, λ is fixed to 5 for all the depth map completion experiments.

B. Experiments on Datasets with Synthetic Degradations

In this subsection, six datasets including "Art", "Book", "Moebius", "Reindeer", "Laundry", and "Dolls" from the Middlebury's benchmark [43] are used for the evaluation. Three kinds of degradations are considered in experiments which are 1) down-sampling without adding noise, 2) down-sampling with noise and 3) structural error and random missing.

1) Degradation by down-sampling without adding noise

We run our tests on filled ground truth data downscaled by nearest neighbor interpolation. The proposed method is compared with 12 benchmark and the state-of-the-art methods: Bicubic interpolation, MRF-based method (MRF) [6], Joint bilateral up-sampling [7] (JBU), Improved JBU with our guidance weights $\lambda_{smooth-pq}$ (IJBU), Spatial-depth super resolution for range images (JBUV) [9], guided image filtering (Guided) [10], edge-weighted NLM-regularization (NLMR) [4], joint geodesic filtering (JGF) [8], total generalized variation (TGV) [16], moving least squares filter (MLS) [35], auto-regression model (AR) [5] and our previous work (PRE) [51]. Moreover, it is realized that the existing papers MRF [6] and JBUV [9] did not report the experimental results on the datasets of "reindeer", "laundry" and "doll".

Table II shows the up-sampling results under four different up-sampling factors with optimal and suboptimal results marked in bold and underlined respectively. It is noticed that the proposed method obtains the lowest MAD for most cases. In the case of 16× SR, the coarsely up-sampled depth map introduces significant errors, which affects the quality of the depth edge map. However, the performance of the proposed method in such case achieves the best ones in 2 out of 6 cases, sub-optimal ones in 2 out of 6 cases. In the rest cases, our method achieves the performance on top rank 3. It is shown that our method is robust to the quality of the depth edge map. In addition, improved JBU with $\lambda_{smooth-pq}$ (IJBU) can improve the performances of JBU a little. However, the overall performances are worse than global methods, such as AR [5] and TGV [16].

Fig. 5 shows the experimental results of 8× up-sampled depth maps (where the specific details can be seen by zooming in the image) for "Dolls" dataset compared with 5 state-of-the-art methods: NLMR [4], MLS [35], JGF [8], AR [5] and TGV [16]. From the highlighted regions, it is shown that NLMR [4], MLS [35], JGF [8] and TGV [16] severely suffer from blurring depth edges and texture-copy artifacts. AR [5] provides comparable results to ours, but it does not well deal with texture-copy artifacts and blurring depth edges either. Compared with the existing methods, our method generates the best depth map SR results.

2) Degradation by down-sampling with noise added

In real situation, depth maps captured by sensors are accompanied by unavoidable noise. To simulate such cases, we run our tests on the datasets provided by AR [5] which firstly introduce Gaussian noise with a variance of 25 to the original datasets, and then down-sample these datasets at four up-sampling factors by nearest neighbor interpolation. Table III gives the depth enhancement results of our method as well as 7 benchmark and the state-of-the-art methods with optimal and suboptimal results marked in bold and underlined respectively. From Table III, it is shown that our method obtains the lowest or the second lowest MAD for all cases. The de-noising ability of JGF [8] is very poor. The performances of NLMR [4], MLS [35] and Guided [10] are similar. TGV [16] provides comparable results to ours in 2× and 4× cases, but it is lack of robustness in 16× case. Overall, AR [5] provides comparable results to ours and obtains better performances on "Reindeer" and "Laundry" datasets. To compare results visually, Fig. 6 illustrates the results of depth map SR with noise up-sampled by the state-of-the-art methods: NLMR [4], Guided [10], MLS [35], TGV [16] and our method. It is shown that there is strong noise left in the results of Guided [10] and MLS [35]. The TGV [16] provides cleaner depth maps, but fails to preserve tiny structures such as sticks in the cup. Overall, our method can suppress noise and protect most details. However, there are some blurry artifacts near a small part of edge in our result on Moebius dataset. The main reason is that this weak edge in the color image cannot be detected by Canny detector with the predefined thresholds. However, the corresponding region on depth map has strong depth discontinuity. According to the proposed method, it is a case of definite inconsistency that color image is not adopted as the guidance for up-sampling. Therefore, it may lead to blurry artifacts in the case of higher up-sampling factors, e.g. 8x, 16x. On the contrary, NLMR [4] may performs better in such region due to the higher level cues, e.g. segmentation and/or edge saliency map. But these high level cues are not stable. It is shown that its results have clear noise left and are also seriously polluted by blurring depth edges and texture-copy artifacts.

3) Degradation by structural errors and random missing

To quantitatively test the effectiveness of depth map completion, the proposed method uses the datasets created by AR [5] which manually adds some holes in the ground truth of Middlebury datasets. The holes consist of structural errors and random missing which are generated near depth edges and in smooth regions respectively. The experimental results are listed in Table IV compared with 5 benchmark and the state-of-the-art methods. As shown in the Table IV, the proposed method obtains the lowest MAD in four datasets and the sub-optimal results in the rest two datasets, which proves the effectiveness of the proposed methods. The visual results of ours and Guided [10], JBF [7] and AR [5] are shown in Fig. 7. Although all the methods obtain good results in depth map completion, our results can provide more accurate depth edges which is shown

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

TABLE II
QUANTITATIVE UP-SAMPLING RESULTS (IN MAD) ON MIDDLEBURY DATASETS AT FOUR UP-SAMPLING FACTORS

Methods	*Art*				*Book*				*Moebius*				*Reindeer*				*Laundry*				*Doll*			
	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x
Bicubic	0.48	0.97	1.85	3.59	0.13	0.29	0.59	1.15	0.13	0.30	0.59	1.13	0.30	0.55	0.99	1.88	0.28	0.54	1.04	1.95	0.20	0.36	0.66	1.18
MRF[6]	0.59	0.96	1.89	3.78	0.21	0.33	0.61	1.20	0.24	0.36	0.65	1.25	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
JBUV[9]	0.55	0.68	1.44	3.52	0.29	0.44	0.62	1.45	0.38	0.46	0.67	1.10	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
JBU[7]	0.45	0.85	1.68	3.35	0.17	0.36	0.74	1.56	0.18	0.37	0.76	1.46	0.27	0.50	1.00	1.89	0.26	0.49	0.94	1.95	0.20	0.38	0.74	1.46
IJBU	0.43	0.83	1.62	3.26	0.16	0.34	0.72	1.47	0.17	0.36	0.74	1.39	0.27	0.49	0.98	1.87	0.25	0.48	0.92	1.94	0.20	0.37	0.73	1.44
Guided[10]	0.63	1.01	1.70	3.46	0.22	0.35	0.58	1.14	0.23	0.37	0.59	1.16	0.42	0.53	0.88	1.80	0.38	0.52	0.95	1.90	0.28	0.35	0.56	1.13
NLMR[4]	0.41	0.65	1.03	2.11	0.17	0.30	0.56	1.03	0.18	0.29	0.51	1.10	0.20	0.37	0.63	1.28	0.17	0.32	0.54	1.14	0.16	0.31	0.56	1.05
JGF[8]	0.29	0.47	0.78	**1.54**	0.15	0.24	0.43	0.81	0.15	0.25	0.46	0.80	0.23	0.38	0.64	1.09	0.21	0.36	0.64	1.20	0.19	0.33	0.59	1.06
TGV[16]	0.45	0.65	1.17	2.30	0.18	0.27	0.42	0.82	0.18	0.29	0.49	0.90	0.32	0.49	1.03	3.05	0.31	0.55	1.22	3.37	0.21	0.33	0.70	2.20
MLS[35]	0.27	0.68	1.04	2.20	0.16	0.26	0.48	1.16	0.15	0.25	0.49	0.93	0.32	0.64	0.74	1.43	0.23	0.39	0.81	1.53	0.24	0.36	0.61	0.98
AR[5]	**0.18**	0.49	**0.64**	2.01	0.12	0.22	**0.37**	0.77	**0.10**	**0.20**	0.40	**0.79**	0.22	0.40	0.58	**1.00**	0.20	0.34	**0.53**	1.12	0.21	0.34	0.50	**0.82**
PRE[51]	0.25	0.47	0.76	1.96	0.11	0.22	0.39	0.76	0.11	0.24	0.45	0.90	0.17	0.34	0.61	1.30	0.15	0.32	0.59	1.28	0.14	0.28	0.51	1.05
Ours	**0.18**	**0.45**	0.71	1.97	**0.10**	**0.20**	**0.37**	**0.74**	**0.10**	**0.20**	**0.39**	0.80	**0.14**	**0.31**	**0.56**	1.10	**0.14**	**0.30**	**0.53**	**1.10**	**0.12**	**0.26**	**0.49**	0.83

TABLE III
QUANTITATIVE UP-SAMPLING WITH NOISE RESULTS (IN MAD) ON MIDDLEBURY DATASETS AT FOUR UP-SAMPLING FACTORS

Methods	*Art*				*Book*				*Moebius*				*Reindeer*				*Laundry*				*Doll*			
	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x
Bicubic	3.52	3.84	4.47	5.72	3.30	3.37	3.51	3.82	3.28	3.36	3.50	3.80	3.39	3.52	3.82	4.45	3.35	3.49	3.77	4.35	3.28	3.34	3.47	3.72
MLS[35]	1.43	1.95	3.37	4.67	0.81	1.39	2.68	3.21	0.87	1.40	2.65	3.16	0.92	1.49	2.86	3.53	0.94	1.53	2.83	3.58	0.81	1.34	2.57	3.09
Guided[10]	1.49	1.97	3.00	4.91	0.80	1.22	1.95	3.04	1.18	1.90	2.77	3.55	1.29	1.99	2.99	4.14	1.28	2.05	3.04	4.10	1.19	1.94	2.80	3.50
NLMR[4]	1.69	2.40	3.60	5.75	1.12	1.44	1.81	2.59	1.13	1.45	1.95	2.91	1.20	1.60	2.40	3.97	1.28	1.63	2.20	3.34	1.14	1.54	2.07	3.02
JGF[8]	2.36	2.74	3.64	5.46	2.12	2.25	2.49	3.25	2.09	2.24	2.56	3.28	2.18	2.40	2.89	3.94	2.16	2.37	2.85	3.90	2.09	2.22	2.49	3.25
TGV[16]	0.82	1.26	2.76	6.87	0.50	0.74	1.49	2.74	0.56	0.89	1.72	3.99	0.59	0.84	1.75	4.40	0.61	1.59	1.89	4.16	0.66	1.63	1.75	3.71
AR[5]	0.76	**1.01**	**1.70**	3.05	0.47	0.70	1.15	1.81	0.46	0.72	**1.15**	1.92	**0.48**	**0.80**	**1.29**	**2.02**	**0.51**	**0.85**	1.30	**2.24**	0.59	0.91	1.32	2.08
Ours	**0.74**	1.02	1.72	**3.01**	**0.45**	**0.66**	**1.07**	**1.80**	**0.45**	**0.68**	1.18	**1.85**	0.53	0.82	1.31	2.14	0.54	0.89	**1.24**	2.33	**0.52**	**0.84**	**1.25**	**1.92**

TABLE IV
QUANTITATIVE DEPTH MAP COMPLETION RESULTS (IN MAD) ON MIDDLEBURY DATASETS WITH STRUCTURAL ERRORS AND RADOM MISSING

	Art	*Book*	*Moebius*	*Reindeer*	*Laundry*	*Doll*
Bicubic	0.90	0.61	0.66	0.95	0.91	0.76
MLS[35]	0.91	0.58	0.72	**0.68**	0.72	0.82
JBF[7]	0.84	0.63	0.69	0.92	0.88	0.76
Guided[10]	1.20	0.63	0.67	0.96	0.94	0.76
AR[5]	**0.58**	0.53	0.60	**0.68**	0.75	0.69
Ours	0.60	**0.52**	**0.56**	0.70	**0.71**	**0.68**

in the highlighted regions.

C. Depth enhancement experiments using Real Datasets

The proposed method is also tested on ToF-Mark datasets [16] and NYU datasets [45] corresponding to two types of depth sensors respectively (i.e. ToF depth sensor and Structured-light depth sensor). The experiments prove that the proposed method can reconstructed high quality depth maps from low quality depth maps captured by different type of sensors.

1) Experiments on ToF-Mark datasets

The proposed method is assessed on ToF-Mark datasets [16] consisting of three RGB-D datasets, "Books", "Shark" and "Devil", with ground-truth depth maps. The resolution of the original depth maps is 120×160, and the corresponding intensity images are the size of 610×810. The suggested up-sampling factor is approximately 6.25× [16]. Table V illustrates quantitative comparison results with optimal and suboptimal results marked in bold and underlined respectively. The up-sampling errors are computed by MAD in mm unit. The proposed method obtains the lowest MAD error for all the three datasets compared with other 10 benchmark and the state-of-the-art methods. Fig. 8 shows the visual depth enhancement results of the proposed method against the 4 state-of-the-art methods (MLS [35], JGF [8], TGV [16] and AR [5]). It is observed that the results of MLS [35] and JGF [8] still contain considerable amount of noise due to the limited de-noising ability, while the depth enhanced by TGV [16], AR [5] and the proposed method are much better. However, the results of TGV [16] and AR [5] introduce texture-copy artifacts in some smooth regions, e.g., the eye of the fish in "Shark" dataset highlighted by red square. Results of the proposed method do not have such texture-copy artifacts. In addition, the edge of the rectangular box in "Shark" dataset highlighted by red square is more accurate in our result than that of others, which proves that the proposed method can efficiently preserve depth edges.

2) Experiments on NYU datasets

By using NYU datasets [45] in which the depth maps are captured by structured-light depth sensors, the proposed method is evaluated for depth map completion, compared with 4 state-of-the-art methods: AR [5], MLS [35], JBU [7] and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

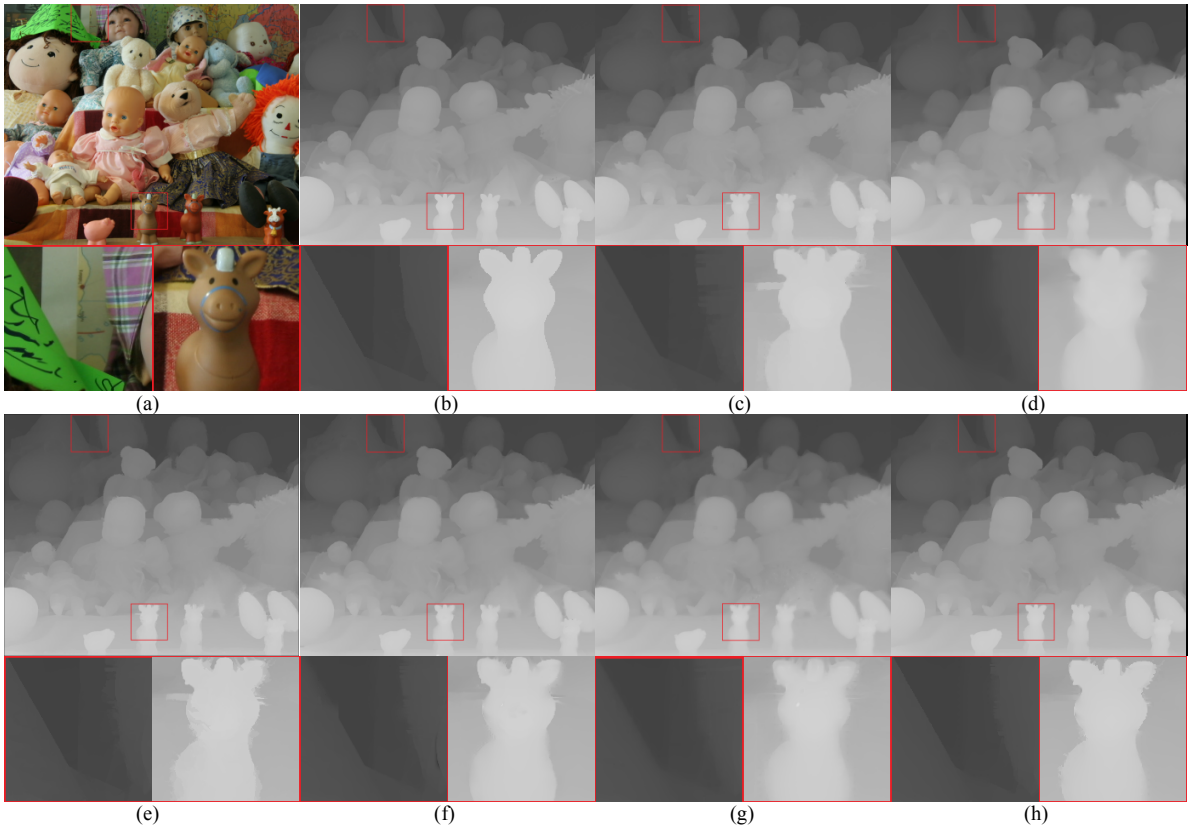


Fig. 5 The visual quality comparison for depth map SR on "Dolls" dataset: (a) color image, (b) depth ground truth, depth maps are up-sampled (8×) by (c) NLMR [4], (d) MLS [35], (e) JGF [8], (f) AR [5], (g) TGV [16], (h) ours.

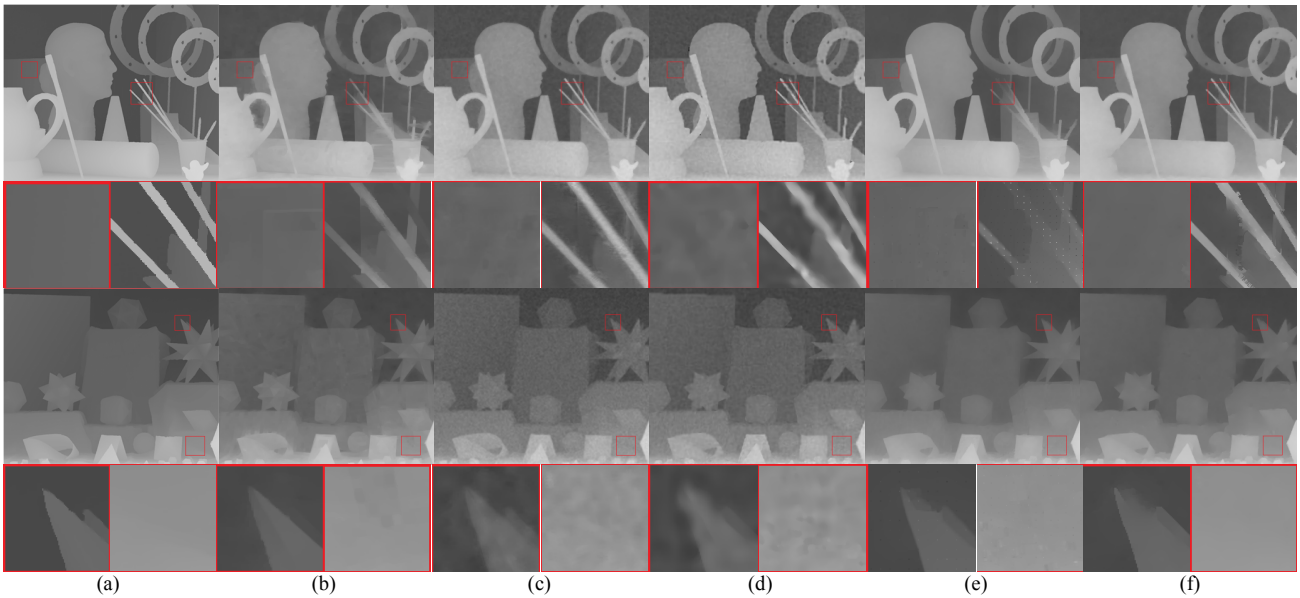


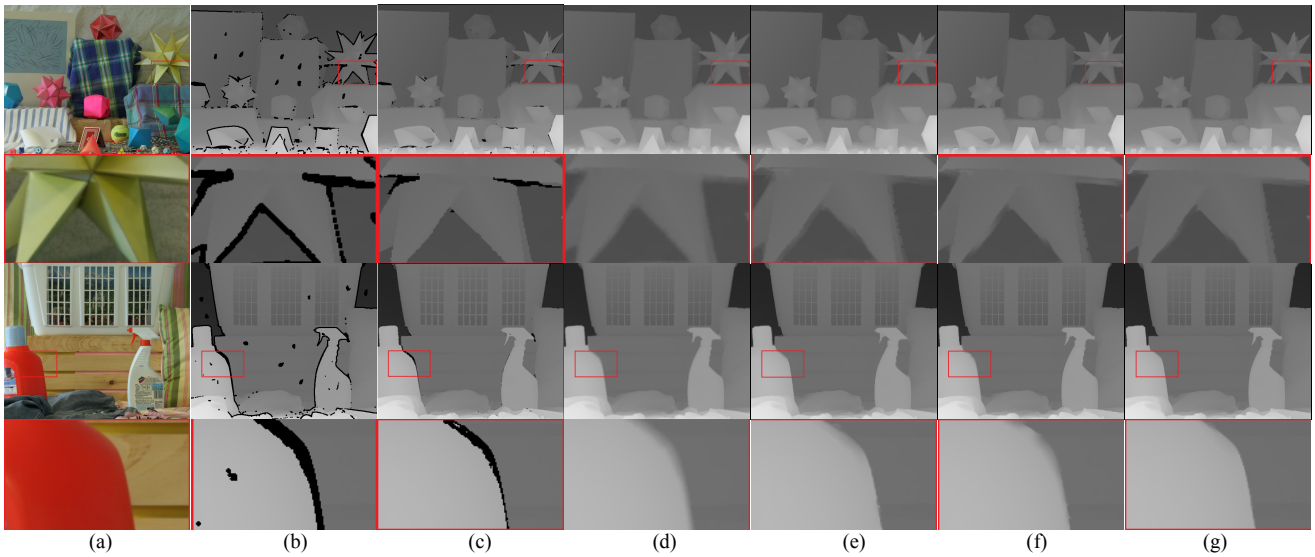
Fig. 6 The visual quality comparison for depth map SR with noise on "Art" and "Moebius" datasets. (a) ground truth, depth maps are up-sampled (8×) by (b) NLMR [4], (c) Guided [10], (d) MLS [35], (e) TGV [16], (f) ours.

Colorization [36]. Fig. 9 shows the depth map completion results of two datasets. From these highlighted regions, it is shown that the existing methods [5, 35] suffer from texture-copy artifacts (e.g. highlighted in the second row). By contrast, there are no such artifacts in our results. In term of preserving depth edges, AR [5] performs best in these existing methods. Through comparison, it is shown that the proposed method demonstrates the best performances on preserving depth edges (e.g. highlighted in the fourth row). Therefore, the proposed method can provide more robust results of depth map completion than the state-of-the-art methods.

D. Experiments on tackling both depth map SR and depth map completion

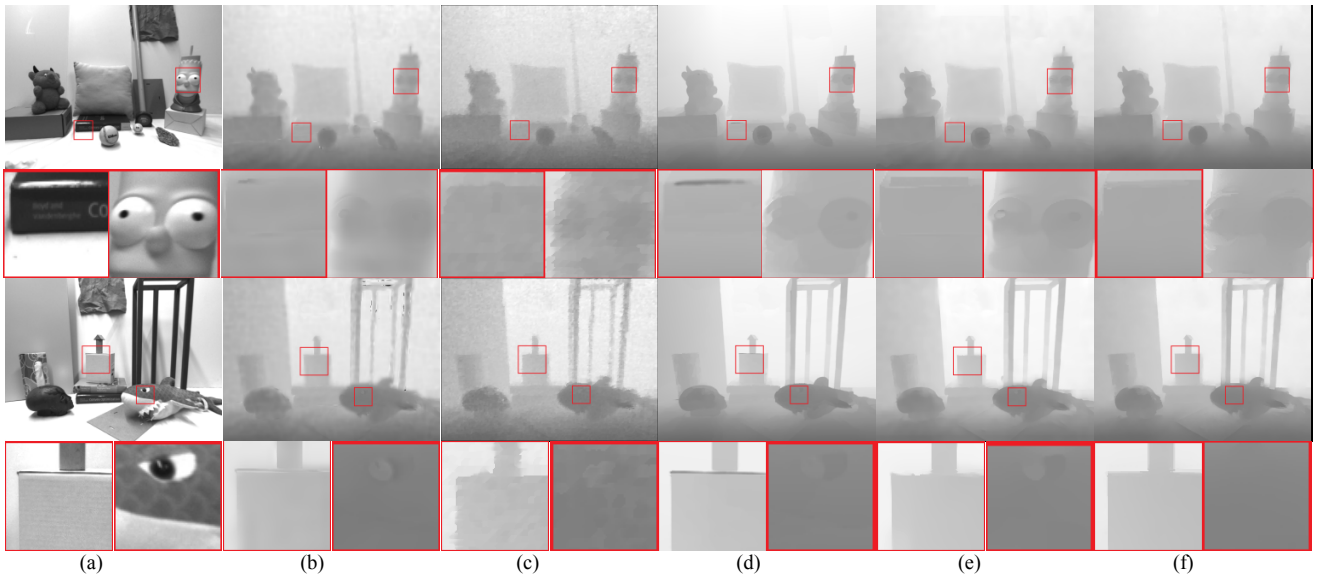
In the previous experiments, we testify the performances of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology



(a) (b) (c) (d) (e) (f) (g)

Fig. 7 The visual quality comparison for depth map completion on "Moebius" and "Laundry" datasets with structural errors and random missing: (a) color images, (b) degraded depth maps, (c) ground truth, depth completed by (d) Guided [10], (e) JBF [7], (f) AR [5] and (g) ours.



(a) (b) (c) (d) (e) (f)

Fig. 8 The visual quality comparison for depth map SR on "Devil" and "Shark" datasets: (a) color images, depth maps are up-sampled by (b) MLS [35], (C) JGF [8], (d) TGV [16], (e) AR [5] and (f) ours.

TABLE V
QUANTITATIVE DEPTH UP-SAMPLING RESULTS (IN MAD) ON TOF-MARK DATASETS

	Bicubic	*MRF[6]*	*Guided[10]*	*MLS[35]*	*JBU[7]*	*JGF[8]*	*NLMR[4]*	*TGV[16]*	*AR[5]*	*PRE[51]*	*Ours*
Books	16.23	13.87	15.74	14.50	16.03	17.39	14.31	12.36	12.25	12.39	**12.23**
Shark	17.78	16.07	18.21	16.26	18.79	18.17	15.88	15.29	14.71	14.23	**14.14**
Devil	16.66	15.36	27.04	14.97	27.57	19.02	15.36	14.68	13.83	13.86	**13.71**

the proposed method on depth map SR and depth map completion independently. In order to further verify the robustness of the proposed method, the experiments in this subsection are to tackle an extremely difficult case in which the proposed method is going to up-sample the LR depth map (i.e. 4x up-sampling factor) and complete holes simultaneously. NYU datasets [45] are adopted in the experiments. Fig. 10 shows the results of our method, compared with Colorization [36], JBU [7] and TGV [16]. From the highlighted regions, it is shown that our method provides the best performances in holes filling, mitigating texture-copy artifacts and preserving depth edges. By contrast, Colorization [36] shows texture-copy artifacts (e.g. highlighted in the second row) and blurring depth edges (e.g. highlighted in the fourth row) in such results. JBU [7] and TGV [16] cannot give satisfying results with holes left uncompleted.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

Fig. 9 The visual quality comparison for depth map completion on NYU datasets: (a) color images, (b) Registered raw depth maps from Kinect v1, completed by (c) AR [5], (d) MLS [35], (e) JBU [7], (f) Colorization [36] and (g) ours.

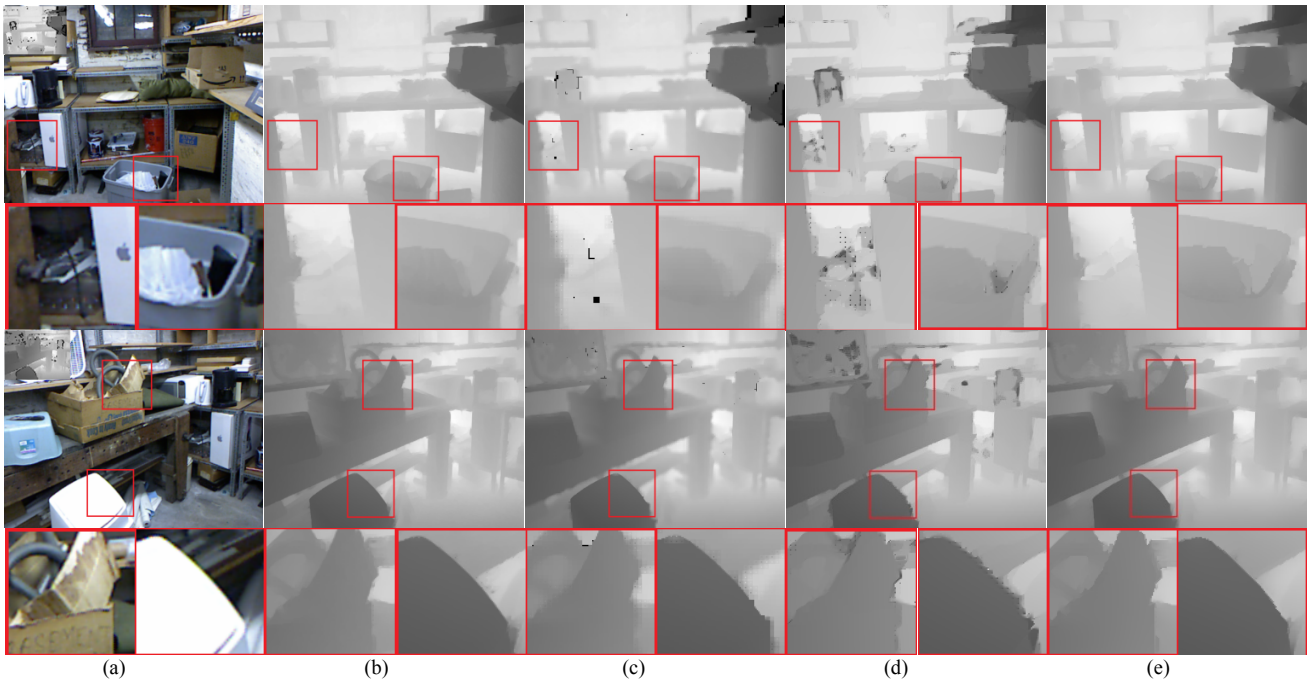


Fig. 10 The visual quality comparison for depth map enhancement with the complex degradation (down-sampling and depth values missing) on NYU datasets: (a) color images with the corresponding LR depth maps shown on the upper left corner, depth maps are enhanced by (b) Colorization [36], (c) JBU [7], (d) TGV [16] and (e) ours.

V. Conclusion and Future Work

This paper proposes a novel color-guided depth map enhancement method via MRF optimization. The key contribution is to explicitly measure the inconsistency between the color edge map and the corresponding depth edge map. In the following step, such quantitative measurement is embedded into MRF-based model. It controls the efforts of the guidance from the color image. Therefore, the proposed method can mitigate texture-copy artifacts and preserve depth edges. This kind of solution has not been seen in any existing methods of the same category. To verify the proposed method, enough experiments on Middlebury datasets, ToF-Mark datasets and NYU datasets for depth map SR and depth map completion tasks are conducted. Furthermore, the proposed method is able to handle both depth map SR and depth map completion simultaneously. All the experimental results prove the improved performances of the proposed method when compared with the state-of-the-art methods.

Although the promising depth enhancement results can be obtained by using the proposed method, there are some failure cases due to missing the weak edges in binary edge detection. A corresponding failure case is shown in the result on Moebius dataset in Fig. 6. In future, the inconsistency measurement can be extended on intensity gradient images rather than binary edge map to obtain more robust structure inconsistency measurements.

REFERENCES

[1] R. Szeliski et al.. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(6):1068–1080, Jun. 2008.

[2] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight cameras in computer graphics. *Comput. Graph. Forum*, 29(1):141–159, 2010.

[3] J. Cho, S. Kim, Y. Ho, and K. Lee. Dynamic 3d human actor generation method using a time-of-flight depth camera. *IEEE Trans. Consumer Electronics*, 54:1514-1521, 2008.

[4] J. Park, H. Kim, Y. Tai, M. Brown, I. Kweon. High-Quality Depth Map Upsampling and Completion for RGB-D Cameras. *IEEE Trans. Image Process.*, 23(12):5559-5572, 2014.

[5] J. Yang, X. Ye, K. Li, C. Hou, Y. Wang. Color-Guided Depth Recovery From RGB-D Data Using an Adaptive Autoregressive Model. *IEEE Trans. Image Process.*, 23(8):3443-3458, 2014.

[6] J. Diebel and S. Thrun. An application of Markov random fields to range sensing. *Advances in Neural Information Processing Systems*, 18:291, Cambridge, MA, USA: MIT Press, 2005.

[7] J. Kopf, M. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. *ACM Trans. Graph.*, 26(3):96, 2007.

[8] M. Liu, O. Tuzel, and Y. Taguchi. Joint geodesic upsampling of depth images. *In Proc. of IEEE Comput. Vis. Pattern Recognit. (CVPR)*, 2013, pp. 169-176.

[9] Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-depth super resolution for range images. *In Proc. of IEEE Comput. Vis. Pattern Recognit. (CVPR)*, 2007, pp. 1-8.

[10] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(6):1397 - 1409, 2012.

[11] D. Min, J. Lu, M. Do. Depth video enhancement based on weighted mode filtering. *IEEE Trans. Image Process.*, 21(3):1176-1190, 2012.

[12] O. Choi, S. Jung. A Consensus-Driven Approach for Structure and Texture Aware Depth Map Upsampling. *IEEE Trans. Image Process.*, 23(8):3321-3335, 2014.

[13] K. Hua, K. Lo, Y. Wang. Extended Guided Filtering for Depth Map Upsampling. *IEEE MultiMedia*, 23(2):72-83, 2016.

[14] J. Lu, D. Min, R. Pahwa, M. Do. A revisit to MRF-based depth map superresolution and enhancement. *In Proc. of IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2011, pp. 985-988.

[15] J. Zhu, L. Wang, J. Gao, R. Yang. Spatial-temporal fusion for high accuracy depth maps using dynamic MRFs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(5):899-909, 2010.

[16] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof. Image guided depth upsampling using anisotropic total generalized variation. *In Proc. of IEEE Int. Conf. Comput. Vis. (ICCV)*, 2013, pp. 993-1000.

[17] D. Kim, K. Yoon. High-quality depth map up-sampling robust to edge noise of range sensors. *In Proc. of IEEE Int. Conf. Image Process.(ICIP)*, 2012, pp. 553-556.

[18] K. Lo, K. Hua, Y. Wang. Depth map super-resolution via Markov Random Fields without texture-copying artifacts. *In Proc. of IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2013, pp. 1414-1418.

[19] J. Yang, X. Ye, K. Li, C. Hou. Depth Recovery Using an Adaptive Color-Guided Auto-Regressive Model. In Proc. of European Conf. Comput. Vis. (ECCV), 2012, pp. 158-171.

[20] S. Gudmundsson and J. Sveinsson. ToF-CCD Image Fusion using Complex Wavelets. *In Proc. of IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2011, pp. 1557-1560.

[21] Y. Li, T. Xue, L. Sun, and J. Liu. Joint Example-Based Depth Map Super-Resolution. *In Proc. of IEEE Int. Conf. Multimedia and Expo (ICME)*, 2012, pp. 152-157.

[22] M. Kiechle, S. Hawe, M. Kleinsteuber. A Joint Intensity and Depth Co-Sparse Analysis Model for Depth Map Super-Resolution. *In Proc. of IEEE Int. Conf. Comput. Vis. (ICCV)*, 2013, pp. 1545-1552.

[23] G. Freedman and R. Fattal. Image and Video Upscaling from Local Self-Examples. *ACM Trans. Graph.*, 30(2):1-11, 2011.

[24] S. Schuon, C. Theobalt, J. Davis, and S. Thrun. Lidar-Boost: Depth Superresolution for ToF 3D Shape Scanning. *In Proc. of IEEE Comput. Vis. Pattern Recognit. (CVPR)*, 2009, pp. 343-350.

[25] J. Xie, R. Feris and M. Sun. Edge-Guided Single Depth Image Super Resolution. *IEEE Trans. Image Process.*, 25(1):428-438, 2016.

[26] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. *24th annual ACM symposium on User interface software and technology, ser. UIST '11*, 2011, pp. 559-568.

[27] M. Schmeing, E. Krauskopf, J. Xiaoyi. Real-time depth fusion using a low-cost depth sensor array. *In Proc. of 3DTV Conf.*, 2014, pp. 1-4.

[28] L. Wang, H. Jin, R. Yang, and M. Gong. Stereoscopic inpainting: Joint color and depth completion from stereo images. *In Proc. of IEEE Comput. Vis. Pattern Recognit. (CVPR)*, 2008, pp. 1-8.

[29] Y. Berdnikov and D. Vatolin. Real-time depth map occlusion filling and scene background restoration for projected-pattern based depth cameras. *In Proc. of IETP Graph. Conf.*, 2011, pp. 121–126.

[30] S. Xiang, L. Yu, C. Chen. No-Reference Depth Assessment Based on Edge Misalignment Errors for T + D Images. *IEEE Trans. Image Process.,* 25(3):1479-1494, Mar. 2016.

[31] D. Miao, J. Fu, Y. Lu, S. Li, C. Chen. Texture-assisted Kinect depth inpainting. *In Proc. of IEEE Symposium on Circuits and Systems (ISCAS),* 2012, pp. 604-607.

[32] Y. Zhao, C. Zhu, Z. Chen, D. Tian, L. Yu. Boundary Artifact Reduction in View Synthesis of 3D Video: From Perspective of Texture-Depth Alignment. IEEE Trans. Broadcasting, 57(2):510-522, 2011.

[33] J. Shen, S. Cheung. Layer Depth Denoising and Completion for Structured-Light RGB-D Cameras. *In Proc. of IEEE Comput. Vis. Pattern Recognit. (CVPR)*, 2013, pp. 1187-1194.

[34] L. Chen, H. Lin, S. Li. Depth image enhancement for Kinect using region growing and bilateral filter. *In Proc. of IEEE Int. Conf. Pattern Recognition*, 2012, pp. 3070-3073.

[35] N. Bose and N. Ahuja. Superresolution and noise filtering using moving least squares. *IEEE Trans. Image Process.*, 15(8):2239–2248, Aug. 2006.

[36] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. *ACM Trans. Graph.*, 23(3):689–694, 2004.

[37] J. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. 1971.

[38] W. Jang, C. Kim. SEQM: Edge quality assessment based on structural pixel matching. *In Proc. of Vis. Comm. Image Process. (VCIP)*, 2012, pp. 1-6.

[39] J. Canny. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986.

[40] H. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1):83-97, 1955.

[41] Y. Boykov, O. Veksler, and R. Zabih. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222-1239, 2001.

[42] R. Barrett et al., "Preconditioners," in *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 2nd ed., Society for Industrial and Applied Mathematics, 1994, pp. 35-50.

[43] Middlebury Datasets [Online], Available: http://vision.middlebury.edu/stereo/data/

[44] C. Bishop et al., "Linear Model for Regression" in *Pattern recognition and machine learning*, 6th ed., New York: Springer Science and Business media, 2006, pp. 137-172.

[45] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. *In Proc. of European Conf. on Computer Vision (ECCV)*, 2012.

[46] R. Franke and G. Nielson, "Scattered Data Interpolation and Applications: A Tutorial and Survey," in *Geometric Modeling*, Berlin: Springer-Verlag, 1991, pp. 131-160.

[47] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE Trans. on Acoustics, Speech, and Signal Process.*, 29(6):1153-1160, 1981.

[48] M. Prieto and A. Allen. A similarity metric for edge images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(10):1265–1273, Oct. 2003.

[49] W. Pratt, "Edge Detection," in *Digital Image Processing*, 4th ed., New York: Wiley Interscience, 2007, pp. 465-529.

[50] H. Kuhn, A. Tucker. Nonlinear programming. *In Proc. of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1950, pp. 481–492.

[51] Y. Zuo, Q. Wu, J. Zhang, P. An. Explicit modeling on depth-color inconsistency for color-guided depth up-sampling. *In Proc. of IEEE Int. Conf. Multimedia and Expo (ICME)*, 2016.

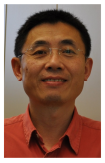
[52] W. Liu, X. Chen, J. Yang, Q. Wu. Variable Bandwidth Weighting for Texture Copy Artifacts Suppression in Guided Depth Upsampling. *IEEE Trans. Circuits and Systems for Video Technology*, PP(99):1-1, 2016.

[53] Y. Boykov, V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(9):1124-1137, 2004.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2609438, IEEE Transactions on Circuits and Systems for Video Technology

Yifan Zuo received the bachelor's degree in electronic information engineering from Nanchang University, Nanchang, China, in 2008, and the master's degree in signal processing from Shanghai University, Shanghai, China, in 2012, where he is currently pursuing the Ph.D. degree. From 2015 to 2017, he participates in a dual-doctoral program in University of Technology, Sydney. His research interests include depth map refinement and depth map estimation.

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 an associate professor and a core member of Global Big Data Technologies Centre at University of Technology Sydney. Dr. Qiang Wu's 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 ECCV, CVPR, ICIP, and ICPR and the major international journals such as IEEE TIP, IEEE TSMC-B, IEEE TCSVT, IEEE TIFS, PR, PRL, Signal Processing, Signal Processing Letter. Dr. Qiang Wu also serves several journals and conferences as reviewer including TPAMI, PR, TIP, TCSVT, TSMC-B, CVIU, IVC, PRL, Neurocomputing and EURASIP Journal on Image and Video Processing.



Jian Zhang (SM'04) received the B.Sc. degree from East China Normal University, Shanghai, China, in 1982; the M.Sc. degree in computer science from Flinders University, Adelaide, Australia, in 1994; and the Ph.D. degree in electrical engineering from the University of New South Wales (UNSW), Sydney, Australia, in 1999. From 1997 to 2003, he was with the Visual Information Processing Laboratory, Motorola Labs, Sydney, as a Senior Research Engineer, and later became a Principal Research Engineer and a Foundation Manager with the Visual Communications Research Team. From 2004 to July 2011, he was a Principal Researcher and a Project Leader with National ICT Australia, Sydney, and a Conjoint Associate Professor with the School of Computer Science and Engineering, UNSW. He is currently an Associate Professor with the Global Big Data Technologies Centre & School of Computing and Communication, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney. He is the author or co-author of more than100 paper publications, book chapters, and six issued patents filed in the U.S. and China. His current research interests include multimedia processing and communications, image and video processing, machine learning, pattern recognition, media and social media visual information retrieval and mining, human-computer interaction and intelligent video surveillance systems.

Dr. Zhang was the General Co-Chair of the International Conference on Multimedia and Expo in 2012 and Technical Program Co-Chair of IEEE Visual Communications and Image Processing 2014. He is Associated Editors for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2006-2015) and the EURASIP Journal on Image and Video Processing (2016-now).

Ping An is a professor of the video processing group at School of Communication and Information Engineering, Shanghai University, China. She received the BA and MS degrees from Hefei University of Technology in 1990, 1993, and PhD from Shanghai University in 2002. In 1993, she joined Shanghai University. Between 2011 and 2012, she joined the Communication Systems Group at Technische University Berlin, Germany, as a visiting professor. Her research interest is image and video processing, especially focuses on 3D video processing recent years. She has finished more than 10 projects supported by the National Natural Science Foundation of China, National Science and Technology Ministry, and Science & Technology Commission of Shanghai Municipality, etc. She awarded the Second Prize of Shanghai Municipal Science & Technology Progress Award in 2011 and the Second Prize in Natural Sciences of the Ministry of Education in 2016.