Color-Guided Depth Recovery via Joint Local Structural and Nonlocal Low-Rank Regularization

Weisheng Dong, Guangming Shi, *Senior Member, IEEE*, Xin Li, Kefan Peng, Jinjian Wu, and Zhenhua Guo

***Abstract*—High-quality depth recovery from RGB-D data has received increasingly more attention in recent years due to their wide applications from depth-based image rendering to three-dimensional imaging and video. Sharp contrast between high-quality color images and low-quality depth maps presents severe challenges to the development of color-guided depth recovery techniques. Previous works have emphasized either *locally* varying characteristics of color-depth dependence or *nonlocal* similarities around the discontinuities of the scene geometry. Therefore, it is desirable to exploit both local and nonlocal structural constraints for optimizing the performance of color-guided depth recovery. In this work, we propose a unified variational approach via joint local and nonlocal regularization. The local regularization term consists of two complementary parts—one characterizing the color-depth dependence in the gradient domain and the other in the spatial domain; nonlocal regularization involves a low-rank constraint suitable for large-scale depth discontinuities. Extensive experimental results are reported to show that our approach outperforms several existing state-of-the-art depth recovery methods on both synthetic and real-world data sets.**

***Index Terms*—Color-guided depth recovery, dual autoregressive model, joint local/nonlocal regularization, low-rank method, weighted total-variation.**

Manuscript received February 24, 2016; revised July 13, 2016 and August 26, 2016; accepted September 19, 2016. Date of publication September 26, 2016; date of current version January 17, 2017. The work was supported in part by NSF Award CCF-1420174, in part by the Natural Science Foundation of China under Grant 61622210, Grant 61471281, Grant 61632019, Grant 61472301, Grant 61390512, and Grant 61372131, and in part by the Shenzhen Overseas High Talent Innovation Fund under Grant KQCX20140521161756231. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Balakrishnan Prabhakaran.

W. Dong is with the State Key Laboratory on Integrated Services Networks, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: wsdong@mail.xidian.edu.cn).

G. Shi is with Collaborative Innovation Center of Information Sensing and Understanding, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: gmshi@xidian.edu.cn).

X. Li is with the Lane Department of CSEE, West Virginia University, Morgantown, WV 26506-6109 USA (e-mail: xin.li@ieee.org).

K. Peng and J. Wu are with the School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: pkfxidian@163.com; jinjian.wu@mail.xidian.edu.cn).

Z. Guo is with Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China (e-mail: zhenhua.guo@sz.tsinghua.edu.cn).

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. This includes a video that shows the comparison of original noisy depth map sequence and the restored sequences by three competing methods (including the proposed). This material is 197 MB in size.

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/TMM.2016.2613824

I. Introduction

ACTIVE methods of acquiring depth information have advanced rapidly in recent years. Unlike passive methods, active sensing directly project lights to the scene and measure the depth information from echoed signals. Time-of-flight (ToF) based [1] and structured-light based (e.g., Kinect [2]) sensing techniques are two representative breakthroughs of achieving real-time depth capturing for dynamic scenes. Despite the progress in active depth sensing, the quality of acquired depth maps - when compared against their color image counterparts - is still low. Therefore it is highly desirable to develop efficient and effective depth recovery techniques (especially for improving the spatial resolution of depth cameras) to better support their uses in real-world applications such as depth-based image rendering (DIBR) [3], [4], human-computer interaction (HCI) [5], [6], 3D imaging/reconstruction [7] and 3D video [8].

Existing approaches toward enhancing the spatial resolution of depth maps can be classified into the following three categories: 1) treat a depth map like a grayscale image but develop specialized super-resolution (SR) technique (e.g., [9]); 2) multiple depth sensor fusion - similar to SR techniques for image sequence [10], multiple depth maps can be combined into one map for the purpose of achieving a higher resolution; 3) color-guided depth recovery - use a high-resolution color image to guide the resolution enhancement of low-resolution depth map. The last class of approaches have been widely studied in the literature because it is often practically convenient to acquire a pair of color/depth images by ToF or Kinect sensors. Moreover, public availability of middlebury benchmark data set facilitates the comparison among competing methods of color-guided depth recovery [11]–[19]. For those reasons, we opt to focus on color-guided depth recovery in this paper.

How can color information facilitate the recovery of depth information? On one hand, since color and depth sensors acquire the same physical scene (up to some small parallax), discontinuities in depth maps are likely to associate with those in color images (i.e., scene depth changes usually produce large-scale edges in both color images and depth maps). Therefore, a high-resolution color image does contain relevant clue to resolve the location uncertainty of discontinuities in depth maps. On the other hand, since homogeneous textured regions in color images correspond to flat surface areas in depth maps [17], it is wise to treat textures in color images as nuisance during the inference about high-resolution depth maps. Due to

1520-9210 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

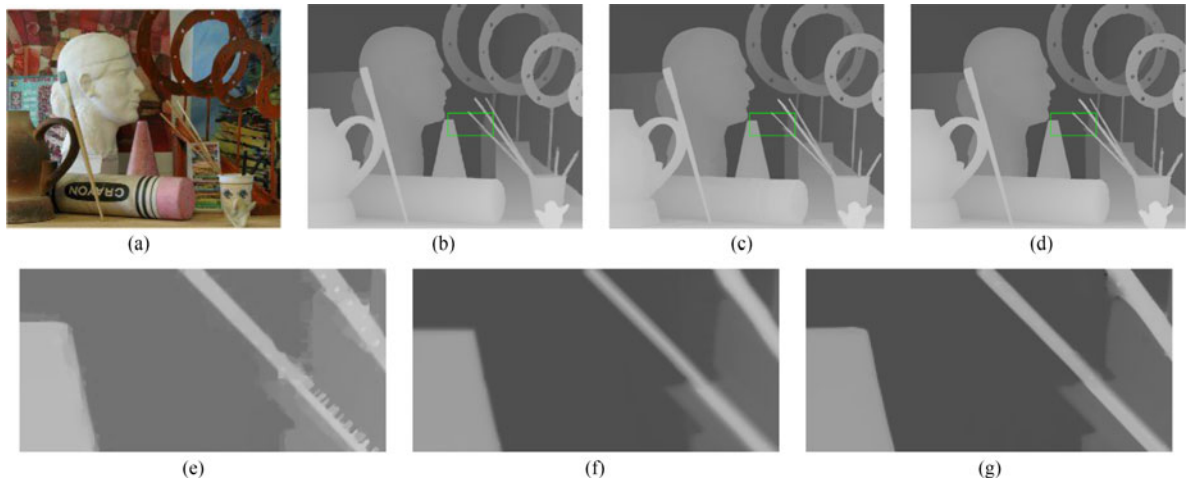


Fig. 1. Demonstration of the merit of joint local and nonlocal regularization on color-guided depth recovery of art at the upsampling rate of $\times$ 8. Left: local regularization only (PSNR $=$ 32.38); middle: nonlocal regularization only (PSNR $=$ 34.93); right: joint local and nonlocal regularization (PSNR $=$ 36.84). (a) Original color image. (b) Local only. (c) Nonlocal only. (d) Local $+$ nonlocal. (e) Zoomed portion of (b). (f) Zoomed portion of (c). (g) Zoomed portion of (d).

inevitable errors with explicit texture segmentation, alternative approaches toward exploiting color-depth dependency (e.g., joint bilateral upsampling [20], guided image filtering [21] and autoregressive model based [18]) have been proposed in the literature.

Despite the above advances, several fundamental issues related to color-guided depth recovery remain unsettled. First, color-depth dependency is locally-varying and could change rapidly around large yet fine-detailed structures of a scene (e.g., long and thin paint-brushes in art as shown in Fig. 1). How to characterize such fast-evolving color-depth dependency in a principled fashion is still an open problem to the best of our knowledge. Consequently, it is often difficult for existing color-guided depth recovery methods to faithfully recover those delicate structures in depth maps. Second, due to lack of interference from surface albedo in depth sensing, depth maps often demonstrate stronger nonlocal similarity than their color image counterparts. How to effectively exploit such nonlocal similarity under the context of color-guided depth recovery is nontrivial especially in the scenario of SR (note that similar idea has been studied for depth map denoising in [22]). We also note that the issue of exploiting color-depth dependency is relevant to the problem of depth map coding which has received increasingly more attention from the multimedia community in recent years (e.g., [23]–[25]).

In this paper, we present a unified framework that can *jointly* exploit local and nonlocal color-depth dependencies and develop spatially adaptive techniques for color-guided depth recovery. The main ideas behind our approach are two-fold. The *local* regularization part consists of two complementary terms - one is weighted total-variation penalty similar to [17] defined by the local *gradients* of color/depth pair; the other is a pair of autoregressive-based models [18] characterizing the dependency between local *statistics* of color/depth pair. The *nonlocal* regularization part reflects the strategy of exploiting nonlocal similarity among large-scale depth discontinuities - we first search similar patches in the guided color image and then enforce low-rank constraints on the recovered depth maps [26]–[29]. It has been found that such joint exploitation of both locally varying color-depth dependency and nonlocal color-guided similarity of depth maps are beneficial to accurate depth recovery in the presence of sophisticated scene geometry. As can be verified from our experimental results in Section IV, the proposed approach is capable of outperforming several existing state-of-the-art methods for both synthetic and real-data sets (please refer to supplementary material).

II. Background and Motivation

A. Problem Formulation: Color-Guided Depth Recovery

The observation model in color-guided depth recovery can be described as follows. Let $\boldsymbol{y}, \boldsymbol{x}$ denote the low-resolution and high-resolution depth maps respectively; then the low-resolution representation is connected with the target of recovery (high-resolution depth map) via

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

where $\mathbf{H}$ represents the observation matrix and $\boldsymbol{n}$ is additive noise. Depending on the scenario of degradation (e.g., undersampling vs. missing data), the observation matrix can take different forms. For example, Kinect depth maps might suffer both degradations of structural missing and random missing data; while ToF depth maps often experience both undersampling and noise contamination [30]. Note that similar formulation has been widely studied in the literature of image interpolation and inpainting for color images; and a regularization/prior functional about the target of recovery $\boldsymbol{x}$ plays the key role in variational/Bayesian approaches toward image restoration. Color-guided depth recovery differs from such conventional formulation due to the existence of *side information* $\boldsymbol{z}$ - a high-resolution color image representation of the same scene. How to make the best use of this supplementary data (i.e., so-called "color-guided") to facilitate the recovery of depth map presents both new challenges and opportunities to the technical community.

Among early attempts, joint bilateral upsampling [20], [31] represents an elegant approach of leveraging existing bilateral filtering tool into color-guided depth map upsampling.

The key idea is to incorporate the high-resolution guide image into the definition of range filter (domain filter is still defined with respect to the low-resolution depth map). Since depth discontinuities correspond to large-scale object boundaries in color images, higher resolution of the guide image is useful to resolve the location uncertainty of depth discontinuities during upsampling. This idea has been further extended into joint geodesic upsampling in [15] where a novel geodesic distance sensitive to thin structures and fine scale changes is proposed to better preserve sharp changes in depth maps. Another line of related work is the development of guided image filtering (GIF) in [32] and [21] where local statistics of the guided image (high-resolution color image) are leveraged into edge-preserving filtering of the target image (depth map). Although depth map upsampling was not directly considered as a potential application of GIF, it has been adopted as a standard benchmark of color-guided depth upsampling in recent years (e.g., [11], [17], [18]).

In this paper, we attempt to sharpen our understanding of color-guided depth recovery from two complementary perspectives. On one hand, we propose to pay close attention to the *locally* varying characteristics of color-depth dependency and address the issue of model consistency between available and missing data. The former is particularly important to the recovery of large-scale but fine-detailed structures in depth maps (e.g., paint brushes in Art); the latter can be related to the idea of cross-validation in the literature of statistics [33]. On the other hand, we argue that it is equally important to develop nonlocal regularization tools for color-guided depth recovery due to lack of interference from albedo in depth sensing. In other words, positions of similar patches in color images could be highly correlated with those in depth maps (even though they are visually strikingly different). Along this line of reasoning, we propose to leverage the result of patch clustering in color images into nonlocally-regularized depth recovery (similar idea exists in [28]).

B. Modeling Locally Varying Color-Depth Dependency

In order to model color-depth dependency in a principled fashion, one has to carefully pick the appropriate mathematical tools. Previous work [17] has introduced an anisotropic diffusion tensor for the purpose of penalizing large depth discontinuities at homogeneous regions but allowing sharp depth edges around texture ones. We have found this anisotropic diffusion tensor is conceptually similar to the existing idea of anisotropic weighted total-variation model (e.g., [34]) - i.e.,

$$E_{\mathrm{TV}}(\boldsymbol{x}) = \sum_{i,j} |\nabla \boldsymbol{x}_{i,j} \circ \mathbf{W}| \tag{2}$$

where $\nabla \boldsymbol{x}_{i,j}$ denotes the gradient of depth map $\boldsymbol{x}$ at a pixel location (i,j), $\circ$ is a point-wise multiplication operator and $\mathbf{W}$ is defined by anisotropic weighted total-variation model as follows:

$$\mathbf{W} = \exp\left(-\sum_{c} \frac{(\nabla \mathbf{I}^c)^\alpha}{\theta}\right) \tag{3}$$

where $\mathbf{I}^c$ represents the intensity of color channel ($c = R, G, B$ or $c = Y, U, V$), ∇ is the gradient operator and α, θ two weighting constants.

More recently, autoregressive (AR) model has been proposed in [18], where both ideas of joint bilateral and nonlocal-mean filtering are combined to better characterize the spatially varying color-depth dependencies from region to region. More specifically, the objective function considered in the AR model can be written as

$$E_{\mathrm{AR}}(\boldsymbol{x}) = \sum_{i,j} [\boldsymbol{x}_{i,j} - \sum_{(k,l)\in\mathcal{N}(i,j)} a_{k,l} \boldsymbol{y}_{k,l}]^2 \tag{4}$$

where (i,j) refers to the spatial location of a pixel, (k,l) is the pixel in the neighborhood of (i,j) and $a_{k,l}$ denotes AR coefficients. It has been suggested in [18] that careful design of pixel-adaptive AR coefficients $a_{k,l}$'s is critical to the accuracy of color-guided depth recovery. For instance, the coefficient $a_{k,l}$ in [18] consists of two balancing terms

$$a_{k,l} = \frac{1}{\mathbf{N}_{k,l}} a^{\boldsymbol{y}}_{k,l} a^{\boldsymbol{z}}_{k,l} \tag{5}$$

where $\mathbf{N}_{k,l}$ is a normalization factor. The first depth term $a^{\boldsymbol{y}}_{k,l}$ takes the form of a range filter (σ_1 is decaying rate)

$$a^{\boldsymbol{y}}_{k,l} = \exp\left\{-\frac{(\boldsymbol{y}_{i,j} - \boldsymbol{y}_{k,l})^2}{\sigma_1^2}\right\} \tag{6}$$

and the second color term $a^{\boldsymbol{z}}_{k,l}$ is built upon bilateral and nonlocal-mean filtering (σ_2, σ_3 are decaying rates)

$$a^{\boldsymbol{z}}_{k,l} = \exp\left\{-\frac{(i-k)^2 + (j-l)^2)}{\sigma_2^2}\right\} \exp\left\{-\frac{(\boldsymbol{z}_{i,j} - \boldsymbol{z}_{k,l})^2}{\sigma_3^2}\right\} \tag{7}$$

where $\boldsymbol{z}_{i,j}, \boldsymbol{z}_{k,l}$ denote the intensity values of color images at locations (i,j) and (k,l).

In this paper, we propose to further refine above model of color-depth dependency by introducing the dual minimization problem to (4) - that is

$$E^{\mathrm{new}}_{\mathrm{AR}}(\boldsymbol{x}) = \sum_{i,j} \left[\boldsymbol{x}_{i,j} - \sum_{(k,l)\in\mathcal{N}(i,j)} a_{k,l} \boldsymbol{y}_{k,l}\right]^2 + \sum_{k,l} \left[\boldsymbol{y}_{k,l} - \sum_{(i,j)\in\mathcal{N}(k,l)} \tilde{a}_{i,j} \boldsymbol{x}_{i,j}\right]^2 \tag{8}$$

where $\tilde{a}$ (different from a but defined similarly) denotes AR coefficients associated with the dual formulation. In other words, we require that the underlying dual AR models well fit not only the missing depth values interpolated from their surrounding neighborhood by a's but also the true depth values predicted from the surrounding interpolated ones by $\tilde{a}$'s. For notational simplicity, we rewrite above equation into a more compact matrix form

$$E^{\mathrm{new}}_{\mathrm{AR}}(\boldsymbol{x}) = ||\boldsymbol{x} - \mathbf{A}\boldsymbol{y}||^2 + ||\boldsymbol{y} - \tilde{\mathbf{A}}\boldsymbol{x}||^2. \tag{9}$$

Such idea of spatially enforcing model consistency is inspired by the previous work on AR-based image interpolation [35]; however, unlike [35], the model whose consistency gets enforced here is the one characterizing color-depth dependency. In theory, this idea is related to the classical tool of cross-validation used in statistical analysis [33] in that the second term of (8) can interpreted as the validation on the available depth values. In

practice, we have found that the combination of weighted-TV and dual-AR model is most effective for the recovery of fine-detailed structures in the scene (e.g., long paint brushes in *art*) as can be seen from Fig. 1.

C. Color-Guided Nonlocal Regularization Model for Depth Maps

In addition to spatially varying local correlation, nonlocal similarity contributes to the global characterization of color-depth dependency. For instance, although textured regions of a color image often correspond to flat surfaces of a depth map (since albedo and shape of an object are independent from each other), their boundaries are aligned - in other words, depth discontinuities leave their traces in both color images and depth maps. Along this line of reasoning, we conclude that it is also beneficial to develop color-guided nonlocal regularization tools for depth maps. In the literature, several previous works (e.g., nonlocal-mean filtering based regularization [11], patch-based synthesis [13], low-rank based depth enhancement [28]) have been developed for *nonlocal-regularized* depth recovery based on similar observation. One drawback of those methods is the increased computational burden by globally searching similar patches.

We propose to strike an improved tradeoff between the performance and the cost by exploiting nonlocal regularization *only* around edges in the guided color image (similar idea exists in the literature such as [36]). In our implementation, we opt to search similar patches only for those exemplar ones whose variance is above a certain threshold; then the results of patch grouping were passed onto the depth map during the formation of data matrices by similar patches. Note that unlike [28] where the definition of patch distance involves both RGB and depth, we opt to count on color image only in the process of searching similar patches. Our strategy can be justified from the perspective of minimizing the impact of aliasing from low-resolution depth maps especially when the upsampling ratio is high (e.g., $\times$ 8 or $\times$ 16 in our experiments). More specifically, the proposed color-guided nonlocal regularization functional is written as

$$E_{\mathrm{NL}}(\mathbf{L}_n) = \sum_{n\in\mathbf{E}} (||\mathbf{R}_n - \mathbf{L}_n||_F^2 + \tau \mathrm{rank}(\mathbf{L}_n)) \tag{10}$$

where $\mathbf{E}$ denotes edge regions in the color image, $\mathbf{R}_n = [\boldsymbol{x}_{n_0}, \boldsymbol{x}_{n_1}, \ldots, \boldsymbol{x}_{n_{k-1}}]$ is a matrix formed by k similar patches $\boldsymbol{x}$'s extracted from the depth map $\boldsymbol{x}$ (again note that the indexes of similar patches $\{n_0, \ldots, n_{k-1}\}$ are leveraged from color image $\boldsymbol{z}$), and $rank(\mathbf{L}_n)$ is the rank of recovered data matrix $\mathbf{L}_n$. In [26], it has been shown that the above rank minimization problem can be approximated by a more tractable problem of minimizing the nuclear norm of a matrix. That is, we can rewrite (10) into

$$E_{\mathrm{NL}}(\mathbf{L}_n) = \sum_{n\in\mathbf{E}} (||\mathbf{R}_n - \mathbf{L}_n||_F^2 + \tau |\mathbf{L}_n|_*) \tag{11}$$

where $|\mathbf{L}_n|_*$ is the nuclear norm of matrix $\mathbf{L}_n$ (i.e., the summation of its singular values) [37]. In the literature (e.g., [38]), the above optimization problem of (11) lends itself to the class of singular-value thresholding techniques. Putting things together, here is the joint local and nonlocal regularization model for color-guided depth recovery

$$\begin{aligned}(\boldsymbol{x}, \mathbf{L}_n) = \ & \underset{\boldsymbol{x},\mathbf{L}_n}{\operatorname{argmin}} ||\boldsymbol{y} - \mathbf{H}\boldsymbol{x}||_2^2 + \eta \sum_{i,j} |\nabla \boldsymbol{x} \circ \mathbf{W}| \\ & + \beta ||\boldsymbol{x} - \mathbf{A}\boldsymbol{y}||^2 + \delta ||\boldsymbol{y} - \tilde{\mathbf{A}}\boldsymbol{x}||^2 \\ & + \lambda \sum_{n\in\mathbf{E}} (||\mathbf{R}_n - \mathbf{L}_n||_F^2 + \tau |\mathbf{L}_n|_*)\end{aligned} \tag{12}$$

where the first term is data term, the second to the fourth corresponding to local terms defined in $E_{\mathrm{TV}}, E_{\mathrm{AR}}^{\mathrm{new}}$, the last two terms are nonlocal regularization functional and $\eta, \beta, \delta, \lambda, \tau$ are regularization parameters. the above optimization problem can be solved by the method of alternating minimization, as we will elaborate next.

TABLE I
UPSAMPLING RESULTS [IN PSNR (dB)] FROM UNDERSAMPLED DEPTH MAPS ON MIDDLEBURY DATASETS AT THREE SUBSAMPLING RATES

method	rate	Art	Book	Dolls	Laundry	Moebius	Reindeer
Bicubic	4	36.38	44.05	45.62	40.50	45.63	39.16
Edge [41]	4	37.47	44.48	45.71	41.56	45.61	40.50
Guided [32]	4	36.84	44.24	45.85	40.86	45.19	39.53
JGF [15]	4	36.25	42.95	44.61	39.79	44.07	39.29
AR [18]	4	**39.03**	44.44	45.28	41.07	**47.86**	40.42
Proposed	4	38.54	**45.88**	**46.95**	**42.63**	46.83	**40.79**
Bicubic	8	33.39	40.79	42.64	37.37	42.23	36.12
Edge [41]	8	36.02	42.22	43.15	39.28	43.09	38.63
Guided [32]	8	34.46	41.75	43.33	37.98	42.90	36.82
JGF [15]	8	33.74	40.56	42.16	36.59	40.62	37.30
AR [18]	8	36.78	42.84	43.63	39.04	**44.32**	**38.74**
Proposed	8	**36.84**	**43.55**	**45.02**	**40.63**	44.11	38.70
Bicubic	16	29.93	37.70	39.68	33.99	39.03	32.83
Edge [41]	16	32.46	39.32	40.40	35.75	40.71	35.47
Guided [32]	16	30.52	38.07	40.04	34.27	39.32	33.19
JGF [15]	16	30.95	37.12	39.33	33.17	36.95	35.17
AR [18]	16	32.70	38.83	40.73	34.77	39.92	35.90
Proposed	16	**33.77**	**40.20**	**41.70**	**37.23**	**40.39**	**36.85**

III. Optimization Algorithm for Color-Guided Depth Recovery

In this section, we show how to solve the optimization problem in (12) by alternatively solving the following two subproblems. Note that the method of alternative minimization has been developed for total-variation image restoration in [39]; however, since the regularization model has changed, we have to re-derive the optimization algorithm as follows.

A. Subproblem of Minimizing $\boldsymbol{x}$ for fixed $\mathbf{L}_n$

The first subproblem consists of all penalty terms except that the last one of nuclear term in (12).

$$\begin{aligned}\boldsymbol{x} = \ & \underset{\boldsymbol{x}}{\operatorname{argmin}} ||\boldsymbol{y} - \mathbf{H}\boldsymbol{x}||_2^2 + \eta \sum_{i,j} |\nabla \boldsymbol{x} \circ \mathbf{W}| \\ & + \beta ||\boldsymbol{x} - \mathbf{A}\boldsymbol{y}||^2 + \delta ||\boldsymbol{y} - \tilde{\mathbf{A}}\boldsymbol{x}||^2 \\ & + \lambda \sum_{n\in\mathbf{E}} ||\mathbf{R}_n - \mathbf{L}_n||_F^2\end{aligned} \tag{13}$$

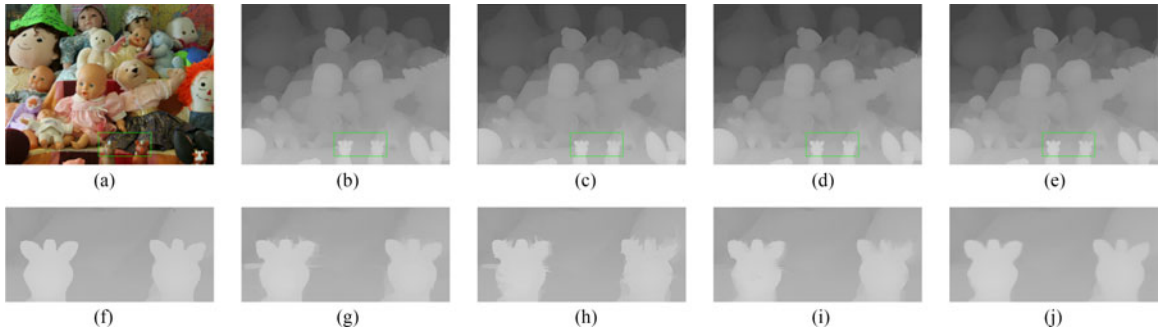


Fig. 2. Visual comparison of recovered depth maps at the upsampling rate $\times$ 8 for *dolls*. (a) Original color. (b) Edge[41]. (c) JGF[15]. (d) AR[18]. (e) This work. (f) Ground truth depth (zoomed). (g) Edge [41] (zoomed). (h) JGF [15] (zoomed). (i) AR [18] (zoomed). (j) This work (zoomed).

At the first look, the above equation is long and tedious. But note that all terms in the first subproblem of optimization are L_2-based; therefore, it is possible to solve in closed form. More specifically, we define

$$\mathbf{B} = \mathbf{H}^T\mathbf{H} + \lambda \sum_{n\in\mathbf{E}} \mathbf{R}_n^T\mathbf{R}_n + \beta\mathbf{A}^T\mathbf{A} + \delta\mathbf{I} + \eta\nabla^T\mathbf{W}^2\nabla \tag{14}$$

where $\mathbf{I}$ is an identity matrix and

$$\boldsymbol{b} = \mathbf{H}^T\boldsymbol{y} + \lambda \sum_{n\in\mathbf{E}} \mathbf{R}_n^T\mathbf{L}_n + \beta\mathbf{A}^T\boldsymbol{y} + \delta\tilde{\mathbf{A}}\boldsymbol{y}. \tag{15}$$

Then, the Least-Square solution to (13) is given by

$$\boldsymbol{x} = (\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{B}\boldsymbol{b}). \tag{16}$$

B. Subproblem of Minimizing $\mathbf{L}_n$ for Fixed $\boldsymbol{x}$

The second subproblem

$$\mathbf{L}_n = \underset{\mathbf{L}_n}{\operatorname{argmin}} \sum_{n\in\mathbf{E}} (||\mathbf{R}_n - \mathbf{L}_n||_F^2 + \tau|\mathbf{L}_n|_*). \tag{17}$$

This problem has been extensively studied in the literature (e.g., [26], [38]) and can be solved by the method of singular value thresholding

$$(\mathbf{U}, \Sigma, \mathbf{V}) = \operatorname{svd}(\mathbf{R}_n) \tag{18}$$

and

$$\mathbf{L}_n = US_\tau(\Sigma)V^T \tag{19}$$

where S_τ denotes the soft thresholding operator $S_\tau(\sigma_n) = sign(\sigma_n)(|\sigma_n| - \tau)_+$.

Putting things together, we have the following iterative depth recovery algorithm.

C. Implementation Details

Algorithm 1 has been implemented under Matlab with the following parameter settings: in the case of noisy observation, we have used $\tau = 1.3$, $\lambda = 0.03$, $\beta = 3 - 5.6$, $\eta = 3$, $\delta = 0.2 - 0.8$; otherwise (noise-free observation data), we use $\tau = 1.3$, $\lambda = 0.015 - 0.03$, $\beta = 3 - 6$, $\eta = 2 - 3$, $\delta = 0.15 - 0.2$. It usually takes around 10-20 iterations for Algorithm 1 to converge; and the average processing time for each iteration is 8-10 seconds on a machine with 3.4G processor and 12G memory. For example, the total running time for upsampling depth maps by a factor of eight to 1376 $\times$ 1088 is around 220 seconds (25 iterations in total). Such computational complexity is comparable to that of previous work such as [18] which has reported an average processing time of two minutes for upsampling depth maps by a factor of four. It should be noted that local-only method (such as [17]) usually has lower computational complexity but sacrifices on the accuracy of reconstruction (as shown in Fig. 1).

TABLE II
DEPTH RECOVERY RESULTS FROM TOF-LIKE DEGRADATIONS (UNDERSAMPLING WITH NOISE) AT THREE SUBSAMPLING RATES

method	rate	Art	Book	Dolls	Laundry	Moebius	Reindeer
Bicubic	4	33.26	35.29	35.45	34.61	35.43	34.23
Edge [41]	4	32.59	39.45	39.78	36.92	40.17	36.39
Guided [32]	4	35.59	41.25	39.68	37.96	39.73	37.41
JGF [15]	4	33.46	37.48	38.31	36.22	37.97	35.57
AR [18]	4	38.14	43.49	43.63	41.44	44.04	**40.70**
Proposed	4	**38.30**	**43.91**	**44.01**	**41.46**	**44.27**	40.16
Bicubic	8	31.55	34.73	35.04	33.57	34.99	33.01
Edge [41]	8	30.06	37.79	37.69	35.04	37.83	33.60
Guided [32]	8	33.57	38.58	36.75	35.02	36.82	34.51
JGF [15]	8	30.77	35.89	36.99	33.97	36.18	33.21
AR [18]	8	35.11	39.93	41.13	38.67	40.93	37.37
Proposed	8	**36.23**	**41.56**	**41.99**	**39.77**	**41.80**	**38.13**
Bicubic	16	29.01	33.65	34.33	31.83	34.11	31.08
Edge [41]	16	27.30	34.30	35.17	32.08	34.86	30.57
Guided [32]	16	29.98	35.24	34.84	32.26	34.67	31.56
JGF [15]	16	27.61	33.34	34.52	31.01	33.52	30.66
AR [18]	16	30.67	37.02	38.00	33.78	37.23	34.72
Proposed	16	**33.31**	**39.05**	**39.24**	**36.15**	**39.00**	**36.36**

Algorithm 1: Color-Guided Depth Recovery via Joint Local and Nonlocal Regularization.

Input: degraded depth map $\boldsymbol{y}$, color image $\boldsymbol{z}$
Output: recovered depth map $\boldsymbol{x}$

- (Initialization) obtain initial estimate $\boldsymbol{x}^{(0)}$ via linear interpolation
- Compute matrices $\mathbf{H}, \mathbf{A}, \tilde{\mathbf{A}}, \mathbf{W}$;
- Set regularization parameters $\eta, \beta, \delta, \lambda, \tau$;
- Main loop: for $t = 1, 2, \ldots, iter$
  - Update $\mathbf{L}_n^{(t+1)}$ with $\boldsymbol{x}^{(t)}$ using Eqs. (18) and (19);
  - Update $\boldsymbol{x}^{(t+1)}$ with $\mathbf{L}_n^{(t+1)}$ using Eqs. (14)-(16).

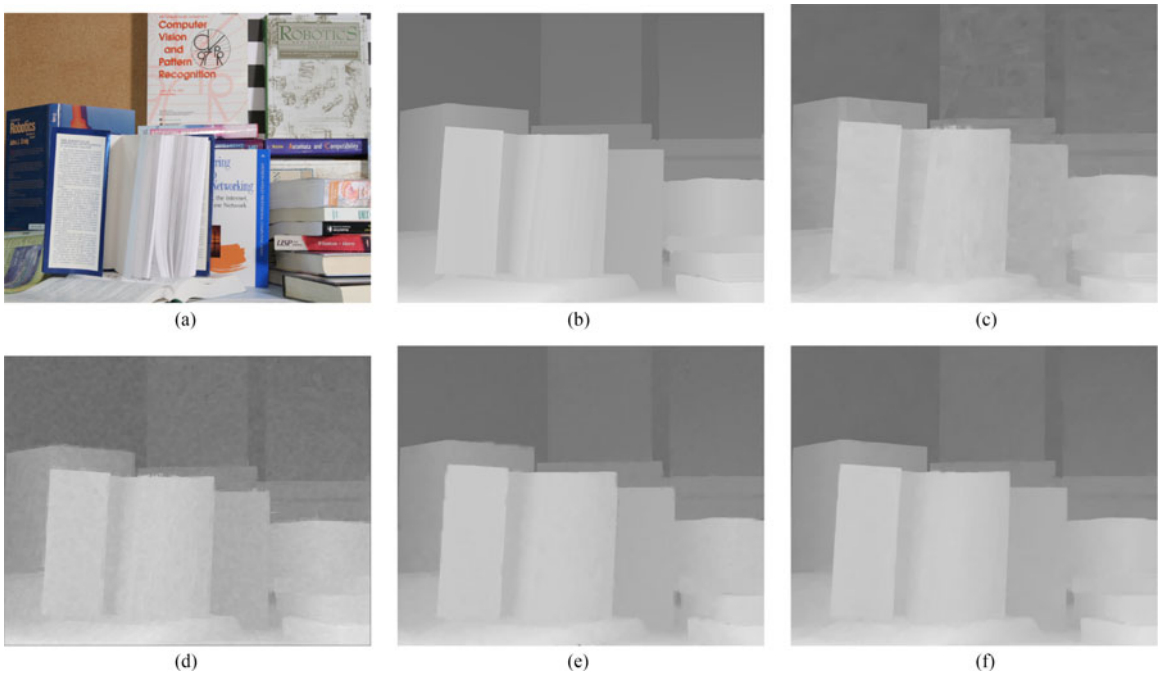


Fig. 3. Visual comparison of recovered depth maps among competing methods for ToF-like degradation. (a) Original color image. (b) Ground truth depth map. (c) Edge [41]. (d) JGF [15]. (e) AR [18]. (f) This work.

IV. Experimental Results

In this section, extensive experimental results with both synthetic and real-world data sets are reported to verify the effectiveness of the proposed color-guided depth recover method. We will compare the performance of Algorithm 1 against several other competing approaches in the latest literature. Supplementary material contains a video demo showing the reduction of flickering artifacts for real-world Kinect depth map sequences.

A. Comparisons Against Other Competing Methods on Synthetic Datasets

Six data sets in the Middlebury benchmark database - namely *Art, Book, Moebius, Reindeer, Laundry* and *Dolls* [40] are used in our study. We have experimented with three types of degradation: undersampling, ToF-like degradation (undersampling with noise) and Kinect-like degradation (structural missing along discontinuities and random missing in flat regions). This work is compared against the following six competing methods: Bicubic interpolation, guided image filtering (guided) [32], edge-weighted NLM-regularization (edge) [41], joint geodesic filtering(JGF) [15], adaptive autoregressive(AR) [18]. The last three represent the current state-of-the-art to the best of our knowledge. Experimental results of upsampling on *Art*, *Book*, and *Moebius* for [32] and [41] are quoted from [41] and [17]; others are quoted from [18].

1) Undersampling: The comparison of PSNR results at three upsampling rates ($\times$ 4, $\times$ 8, $\times$ 16) are shown in Table I. As can be seen from Table I, our method achieves the highest PSNR for most cases (especially at high upsampling rates). The only exceptions are $Moebius$ and $Reindeer$ at the upsampling rate of four - AR [18] are slightly better (the difference is only a few tenth decibels). Fig. 2 contains a comparison of original $Dolls$ color/depth and the recovered depth maps among four competing methods. It is apparent that ours achieves the most accurate recovery - almost no artifact around the depth discontinuities (boundary of dolls). Indeed for this specific setting, the PSNR gain of our method over others is in the range of 1.5-3 dB.

TABLE III
Quantitative Depth Recovery Results From Kinect-Like Degradations (Structural Missing and Random Missing)

	Art	Book	Dolls	Laund.	Moeb.	Reind.
Bicub.	35.00	40.46	39.70	37.00	40.81	36.03
Guided [32]	35.27	41.18	40.84	37.80	41.68	36.60
JBF [20]	33.75	39.06	40.11	36.32	39.65	34.70
AR [18]	36.14	40.41	41.02	37.77	40.34	37.20
Ours	**39.31**	**43.61**	**42.07**	**40.30**	**42.43**	**40.00**

2) ToF-like Degradation: In this experiment, we simulate ToF like degradation by adding Gaussian noise with a standard deviation of 5 and then downsampling depth maps at three different rates ($\times$ 4, $\times$ 8, $\times$ 16). The comparison results in Table II show that our method obtains the highest PSNR performance for all cases except two (ours falls behind AR [18] by less than 1 dB). The Guided [32], Edge [41] and AR [18] methods provide decent results due to their intrinsic denoising abilities. However, the JGF [15] method does not as well as in upsampling without noise because of its lack of denoising ability. Fig. 3 shows the subjective quality comparison among the ground-truth and recovered depth maps for $Book$ by four competing

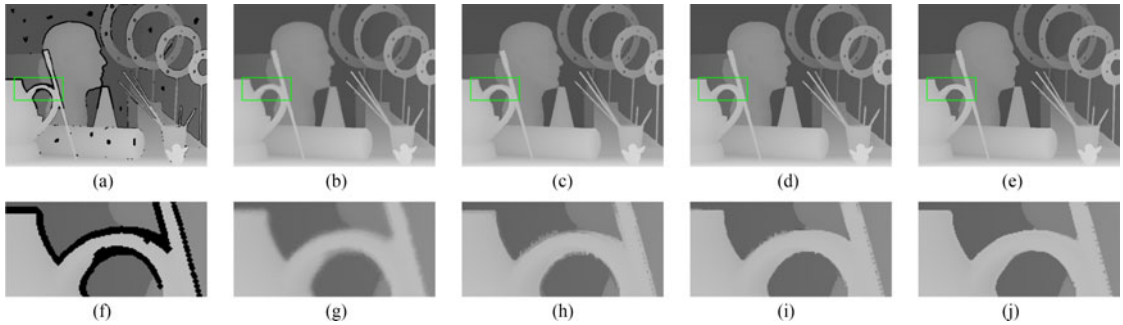


Fig. 4. Visual comparison of recovered depth maps among competing methods for Kinect-like degradation. (a) Degraded depth. (b) Guided [32]. (c) JBF [20]. (d) AR [18]. (e) This work. (f) Zoomed version of (a). (g) Zoomed version of (b). (h) Zoomed version of (c). (i) Zoomed version of (d). (j) Zoomed version of (e).

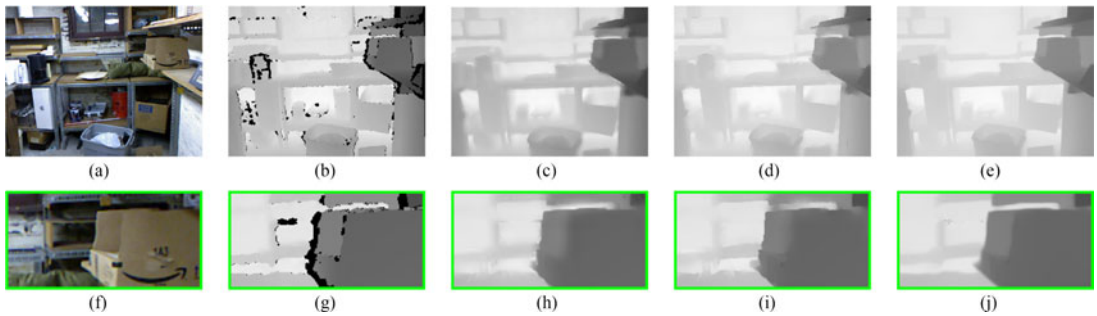


Fig. 5. Visual comparison of recovered depth maps among competing methods for real-world depth data (note that the way of encoding depth information into intensity values in this data set is different from those in previous figures). (a) Color image. (b) Degraded depth map. (c) MLS [42]. (d) AR [18]. (e) Ours. (f) Zoomed version of (a). (g) Zoomed version of (b). (h) Zoomed version of (c). (i) Zoomed version of (d). (j) Zoomed version of (e).

methods. Our method significantly outperforms in terms of sharpness around depth discontinuities and freedom from artifacts.

3) Kinect-like Degradation: To simulate Kinect-like degradation, we create both structural and random missing data (holes) around depth discontinuities and in flat regions. Table III includes the PSNR comparison results for this kind of degradation. Five competing methods are compared: Bicubic, Guided [32], joint bilateral filtering (JBF) [20], Adaptive Autoregressive (AR) [18] and ours. In fact, one can be more easily verify the benefit of joint local and nonlocal regularization in the situation of missing data than upsampling due to absence of frequency aliasing. Our method convincingly outperforms others by 1-3 dB for all six images. Fig. 4 includes the visual comparison of different recovery methods and it is easy to verify the superiority of this work.

B. Experimental Results on Real-world Datasets

We have also applied the proposed recovery method to depth maps captured from a Kinect sensor in the real world. The data set used in our experiment consists of five RGB-D pairs - three from [18] and the other two from the NYU RGB-D dataset.[1] Fig. 5 shows the comparison of depth recovery results among three competing methods (Moving Least-Square [42], AR [18] and ours) for two RGB-D pairs. Moving Least-Square (MLS) [42] is a method of noise filtering and superresolution for grayscale images but adapted here for depth map recovery. Since this is not ground truth for comparison, we can only count on subjective inspection of recovered depth maps in Fig. 5 - this work arguably delivers more visually pleasant recovery results than the other two. More convincing recovery results can be found in the supplementary material accompanying this paper where dramatic reduction of flickering artifacts in a real-world depth map sequence can be observed.

[1][Online]. Available: http://cs.nyu.edu/∼silberman/datasets/

V. Conclusion

In this paper, we have proposed a joint local and nonlocal regularization strategy for high-quality color-guided depth recovery. The local regularization part involves a combination of weighted TV and dual AR-based terms aiming at better exploiting spatially-varying color-depth dependencies especially around large yet fine-detailed structures. The nonlocal regularization part is based on the result of searching similar patches from the color image and designed to better characterize the self-repeating patterns in depth maps. An optimization algorithm via alternative minimization is developed and tested on both synthetics and real-world data sets. For both data sets, our experimental results have shown that the proposed method is superior to the existing state-of-the-art in the open literature in terms of both objective accuracy and subjective quality of recovered depth maps.

REFERENCES

[1] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE J. Quantum Electron.*, vol. 37, no. 3, pp. 390–397, Mar. 2001.

[2] 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.

[3] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Proc. Electron. Imaging*, 2004, pp. 93–104.

[4] P. J. Lee and Effendi, "Nongeometric distortion smoothing approach for depth map preprocessing," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 246–254, Apr. 2011.

[5] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1110–1120, Aug. 2013.

[6] H. Liang, J. Yuan, and D. Thalmann, "Parsing the hand in depth images," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1241–1253, Aug. 2014.

[7] D. S. Alexiadis, D. Zarpalas, and P. Daras, "Real-time, full 3-D reconstruction of moving foreground objects from multiple consumer depth cameras," *IEEE Trans. Multimedia*, vol. 15, no. 2, pp. 339–358, Feb. 2013.

[8] P. Ndjiki-Nya *et al.*, "Depth image-based rendering with advanced texture synthesis for 3-D video," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 453–465, Jun. 2011.

[9] J. Xie, R. S. Feris, S. S. Yu, and M. T. Sun, "Joint super resolution and denoising from a single depth image," *IEEE Trans. Multimedia*, vol. 17, no. 9, pp. 1525–1537, Sep. 2015.

[10] S. Park, M. Park, and M. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.

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

[12] H. Deng, L. Yu, J. Qiu, and J. Zhang, "A joint texture/depth edge-directed up-sampling algorithm for depth map coding," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2012, pp. 646–650.

[13] O. Mac Aodha, N. D. Campbell, A. Nair, and G. J. Brostow, "Patch based synthesis for single depth image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 71–84.

[14] J. Li, G. Zeng, R. Gan, H. Zha, and L. Wang, "A Bayesian approach to uncertainty-based depth map super resolution," in *Proc. ACCV*, 2013, pp. 205–216.

[15] M.-Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 169–176.

[16] M. Kiechle, S. Hawe, and M. Kleinsteuber, "A joint intensity and depth co-sparse analysis model for depth map super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1545–1552.

[17] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 993–1000.

[18] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3443–3458, Jun. 2014.

[19] 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.

[20] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Grap.*, vol. 26, p. 96, 2007.

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

[22] W. Hu, X. Li, G. Cheung, and O. Au, "Depth map denoising using graph-based transform and group sparsity," in *Proc. IEEE 15th Int. Workshop Multimedia Signal Process.*, Sep. 2013, pp. 001–006.

[23] S. Liu, P. Lai, D. Tian, and C. W. Chen, "New depth coding techniques with utilization of corresponding video," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 551–561, Jun. 2011.

[24] B. Yan and J. Zhou, "Efficient frame concealment for depth image-based 3-D video transmission," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 936–941, Jun. 2012.

[25] G. Petrazzuoli, T. Maugey, M. Cagnazzo, and B. Pesquet-Popescu, "Depth-based multiview distributed video coding," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 1834–1848, Nov. 2014.

[26] E. Candes and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.

[27] H. Ji, C. Liu, Z. Shen, and Y. Xu, "Robust video denoising using low rank matrix completion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2010, pp. 1791–1798.

[28] S. Lu, X. Ren, and F. Liu, "Depth enhancement via low-rank matrix completion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2014, pp. 3390–3397.

[29] W. Dong, G. Shi, Y. Ma, and X. Li, "Image restoration via simultaneous sparse coding: Where structured sparsity meets Gaussian scale mixture," *Int. J. Comput. Vis.*, vol. 114, pp. 217–232, 2015.

[30] S. Schwarz, M. Sjostrom, and R. Olsson, "A weighted optimization approach to time-of-flight sensor fusion," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 214–225, Jan. 2014.

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

[32] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 1–14.

[33] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell.*, 1995, pp. 1137–1143.

[34] M. Unger, T. Mauthner, T. Pock, and H. Bischof, "Tracking as segmentation of spatial-temporal volumes by anisotropic weighted tv," in *Proc. Energy Minimization Methods Comput. Vis. Pattern Recog.*, 2009, pp. 193–206.

[35] X. Zhang and X. Wu, "Image interpolation by adaptive 2-D autoregressive modeling and soft-decision estimation," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 887–896, Jun. 2008.

[36] X. Li and M. Orchard, "Edge directed prediction for lossless compression of natural images," *IEEE Trans. Image Process.*, vol. 10, no. 6, pp. 871–875, Jun. 2001.

[37] J. Cai, E. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, pp. 1956–1982, 2010.

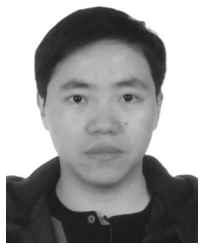
[38] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: a low-rank approach," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 700–711, Feb. 2013.

[39] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imaging Sci.*, vol. 1, no. 3, pp. 248–272, 2008.

[40] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2007, pp. 1–8.

[41] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin, "Shading-based shape refinement of RGB-D images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 1415–1422.

[42] N. K. Bose and N. A. Ahuja, "Superresolution and noise filtering using moving least squares," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2239–2248, Aug. 2006.



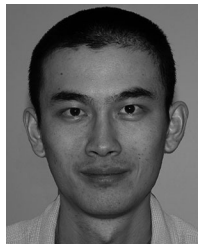
Weisheng Dong received the B.S. degree in electronic engineering from Huazhong University of Science and Technology, Wuhan, China, in 2004, and the Ph.D degree in circuits and system from the Xidian University, Xi'an, China, in 2010.

From September to December, 2006, he was a visiting student at Microsoft Research Asia, Bejing, China. From January 2009 to June 2010, he was a Research Assistant with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China. In September 2010, he joined the School of Electronic Engineering, Xidian University, as a Lecturer, and has been an Associate Professor at Xidian University since June 2012. His research interests include inverse problems in image processing, sparse signal representation, and image compression.

Dr. Dong was the recipient of the Best Paper Award at SPIE Visual Communication and Image Processing 2010.

Guangming Shi (M'07–SM'10) received the B.S. degree in automatic control, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, Xi'an, China, in 1985, 1988, and 2002, respectively.

He joined the School of Electronic Engineering, Xidian University, in 1988. From 1994 to 1996, as a Research Assistant, he cooperated with the Department of Electronic Engineering, University of Hong Kong, Hong Kong, China. Since 2003, he has been a Professor with the School of Electronic Engineering, Xidian University, and in 2004 the Head of National Instruction Base of Electrician and Electronic. From June to December 2004, he studied with the Department of Electronic Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, USA. Presently, he is the Deputy Director of the School of Electronic Engineering, Xidian University, and the academic leader in the subject of circuits and systems. He has authored or coauthored more than 60 research papers. His research interests include compressed sensing, theory and design of multirate filter banks, image denoising, low-bit-rate image/video coding, and implementation of algorithms for intelligent signal processing (using DSP&FPGA).



Xin Li received the B.S. degree with highest honors in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 1996, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2000.

He was a Member of Technical Staff with Sharp Laboratories of America, Camas, WA, USA, from August 2000 to December 2002. Since January 2003, he has been a faculty member with the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA. His research interests include image/video coding and processing.

Dr. Li is currently serving as a member of Image, Video and Multidimensional Signal Processing Technical Committee and an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING. He was the recipient of the Best Student Paper Award at the Conference of Visual Communications and Image Processing as the junior author in 2001, the Runner-Up Prize of the Best Student Paper Award at the IEEE Asilomar Conference on Signals, Systems, and Computers as the senior author in 2006, and the Best Paper Award at the Conference of Visual Communications and Itemage Processing as the single author in 2010.

Kefan Peng, photograph and biography not available at the time of publication.

Jinjian Wu, photograph and biography not available at the time of publication.

Zhenhua Guo, photograph and biography not available at the time of publication.