Color-Guided Depth Recovery From RGB-D Data Using an Adaptive Autoregressive Model

Jingyu Yang, Xinchen Ye, Kun Li, Chunping Hou, and Yao Wang, *Fellow, IEEE*

***Abstract*— This paper proposes an adaptive color-guided autoregressive (AR) model for high quality depth recovery from low quality measurements captured by depth cameras. We observe and verify that the AR model tightly fits depth maps of generic scenes. The depth recovery task is formulated into a minimization of AR prediction errors subject to measurement consistency. The AR predictor for each pixel is constructed according to both the local correlation in the initial depth map and the nonlocal similarity in the accompanied high quality color image. We analyze the stability of our method from a linear system point of view, and design a parameter adaptation scheme to achieve stable and accurate depth recovery. Quantitative and qualitative evaluation compared with ten state-of-the-art schemes show the effectiveness and superiority of our method. Being able to handle various types of depth degradations, the proposed method is versatile for mainstream depth sensors, time-of-flight camera, and Kinect, as demonstrated by experiments on real systems.**

***Index Terms*— Depth recovery (upsampling, inpainting, denoising), autoregressive model, RGB-D camera.**

I. Introduction

ACQUIRING depth information of real scenes is an essential task for many applications such as 3DTV, augmented reality, and 3D reconstruction. Generally, 3D information of a scene consists of texture information and position information, i.e., depth information in our context. While texture information can be readily captured by popular color cameras, depth information is not so easy to acquire. Until now, there are mainly two categories of methods to obtain depth information: passive methods and active methods.

In passive methods, depth information is computed from two-view images or multiview images via correspondence matching and triangulation. Being an active area for several decades, the accuracy of stereo matching has been significantly improved. However, there are still some inherent problems for practical application, e.g., the requirement of accurate image rectification and the inefficiency for textureless areas [1], [2].

The alternatives to acquire depth information are the active methods in which lights are intentionally projected to the scene and the depth information is measured from the echoed signals. Laser range scanner techniques are the earliest active methods and usually achieve high accuracy [3]. However, the slice-by-slice scanning of laser scanners makes them rather time-consuming and inapplicable to dynamic scenes. Time-of-flight (ToF) based technique is a recent advance in active depth sensing [4]. In ToF cameras, depth information is determined by measuring the phase difference between the emitted light and the reflected light. ToF cameras can capture depth information for dynamic scenes in real time, but are noisy and subject to low resolutions, e.g., 176×144 and 200×200, compared with popular color cameras. Structured-light based sensing technique is another breakthrough to achieve real-time depth capturing for dynamic scenes, and the Microsoft Kinect is a representative commodity device of this kind. In Kinect, an infrared light source projects a dot pattern on the scene and an offset infrared camera receives the pattern and estimates the depth information. The generated depth maps contain considerable holes due to the occlusion caused by the relative displacement of the projector and infrared camera.

While the new depth capturing techniques are promising, the use of depth cameras is limited by the low quality of produced depth maps, e.g., low resolution, noise, and depth missing in some areas. There have been some previous work on depth recovery for depth cameras. To compensate the undersampling of ToF cameras, an auxiliary color camera is equipped and the resolution of depth maps is enhanced by joint image filtering techniques from a low resolution depth map and a high resolution color image [4]–[7]. Some depth recovery methods for Kinect are adapted from image inpainting techniques [8], [9]. These methods achieve good quality for smooth regions, but may introduce artifacts, e.g. jagging, blurring, and ringing, around thin structures or sharp discontinuities. Both taking a low quality depth map and a high quality color image as input, depth recovery problems for ToF camera and Kinect are essentially the same, but are treated separately in literature before our preliminary results of this work [10].

This paper proposes an adaptive color-guided AR model to construct a unified depth recovery framework for both ToF and Kinect depth cameras. We first verify the fitness of AR

Manuscript received July 7, 2013; revised December 15, 2013 and March 28, 2014; accepted May 24, 2014. Date of publication June 9, 2014; date of current version July 1, 2014. This work was supported in part by the National Natural Science Foundation of China under Grant 61372084, Grant 61302059, Grant 61228104, and Grant 91320201, in part by the Program for New Century Excellent Talents under Grant NCET-11-0376, in part by the Ph.D. Programs Foundation under Grant 20110032110029 through the Ministry of Education of China, and in part by the Tianjin Research Program of Application Foundation and Advanced Technology under Grant 12JCYBJC10300 and Grant 13JCQNJC03900. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Chang-Su Kim.

J. Yang, X. Ye, K. Li, and C. Hou are with Tianjin University, Tianjin 300072, China (e-mail: yjy@tju.edu.cn; yexch@tju.edu.cn; lik@tju.edu.cn; hcp@tju.edu.cn).

Y. Wang is with the Polytechnic School of Engineering, New York University, Brooklyn, NY 11201 USA (e-mail: yao@poly.edu).

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

Digital Object Identifier 10.1109/TIP.2014.2329776

1057-7149 © 2014 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.

model for depth maps, and then design pixel-wise adaptive AR predictors based on the non-local similarity of both the depth map and the accompanied color image. The depth map is recovered by minimizing AR prediction errors subject to the observation consistency. The stability of the proposed method is analyzed from the linear system point of view. Inspired by the stability analysis, parameters are adaptively set according to the local characteristics of the depth-texture pair to achieve stable and reliable recovery. Experiments demonstrate that our method can handle various depth degradation modes and is applicable to both ToF and Kinect cameras. Without resorting to higher level tools such as segmentation used in [11], our proposed method achieves the best quality among several state-of-the-art depth recovery methods.

The contribution of our work is summarized into the following three aspects:

- **First Attempt at AR Modeling of Depth Maps**: We demonstrate that the AR model is able to tightly fit the depth signals if AR coefficients are carefully designed according to the signal characteristics. This accurate depth model brings great success in depth recovery, and also provides a promising tool for other depth-related processing.
- **A Unified Depth Recovery Framework With an Efficient Color-Guided AR Model**: We design high performance AR predictors by fully exploiting characteristics of RGB-D data: non-local correlations, non-stationary nature of depth maps, and structural correlations in RGB-D data. Several depth enhancement problems are unified into an elegant depth recovery framework with a versatile observation model that includes four commonly-existed degradation modes. The global formulation and optimization provide inherent closed-loop interactions between observed pixels and latent pixels. This prominent feature well complements the open-loop nature of filtering-based schemes such as JBF [12], guided filter [13], and their variants [14]–[16].
- **Systematic Stability Analysis and Effective Parameter Adaptation**: The stability behavior of the proposed method is systematically analyzed by the conditioning of linear systems. The influence of parameters on the stability and recovery quality is investigated. Based on the analysis, we propose an effective parameter adaptation scheme to achieve stable and accurate depth recovery.

II. Related Work

There are mainly two types of mainstream depth cameras: one is ToF cameras and the other is the structured-light based depth cameras, e.g., Kinect. The recovery of high quality depth information from measurements sensed by these devices is a crucial step for subsequent processing in many computer vision tasks, and many algorithms have been proposed in literature. This section briefly reviews the related work.

A. Depth Recovery for ToF Cameras

As shown in Fig. 1(a), the depth map captured by ToF camera has a much lower resolution than the color image.

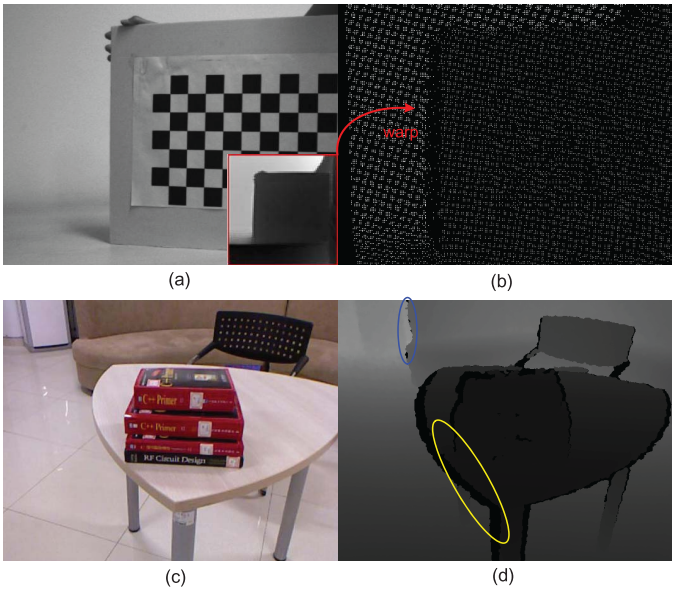

Fig. 1. Illustration of RGB-D pairs captured by ToF camera and Kinect: (a) noisy low-resolution depth map from ToF and high-resolution color image from the coupled color camera, (b) the depth samples warped from the ToF view to the color camera view, (c) Kinect color image, and (d) Kinect depth maps in which structural (random) missing is marked by yellow (blue) ellipse.

Such degradation would impede their practical applications. It seems impossible to recover high quality depth maps from severely undersampled versions due to the loss of salient information around notable discontinuities. However, the depth information and texture information are two descriptions of the same scene from different perspectives, and thus present strong structural correlations [17]. In particular, as shown in Fig. 1 (a) and (c), discontinuities often simultaneously present at the same locations in a depth map and the corresponding (registered) color image, and homogeneous regions in color image tend to have similar depth. Although the viewpoints between the depth sensor and image sensor are different, as shown in Fig. 1(a) and (b), the ToF depth map can be aligned with the color image via view warping with camera calibration parameters. Then, the strong structural correlation between the low-resolution depth map and high resolution color image can be conveniently exploited. Therefore, the common wisdom is to couple a color camera with a ToF camera and to recover high quality depth maps with the help of the accompanied color images [5], [6], [18]–[20].[1]

In an early work, Diebel and Thrun [20] proposed a two-layer MRF to model the correlation between range measurements and solve the MRF optimization with the conjugate gradient algorithm. This method is able to improve the quality of depth maps, but tends to produce oversmooth results. To reduce oversmoothing, Hannemann *et al.* [22] incorporated amplitude values generated by ToF into an MRF model to improve the quality of interpolation. The amplitude can be evaluated as a confidence measurement for the depth values. Lu *et al.* [23] further extended this work by designing a data term that fits to the characteristics of depth maps. Huhle *et al.* [24] added a third layer to the MRF

[1]We also note that some methods, e.g., [21], use depth maps only.

framework [20], where image gradients are encoded as nodes in the graph. Zhu *et al.* [25] extended the traditional spatial MRFs to dynamic MRFs so that both the spatial and the temporal relationship can be propagated in local neighbors, improving the accuracy and robustness of depth recovery for dynamic scenes. Aodha *et al.* [26] presented an algorithm to synthetically increase the resolution of a solitary depth image using only a generic database of local patches. They match against each low resolution input depth patch, and search database for a list of appropriate high resolution candidate patches. The training of data, matching, and fusion are quite computationally intensive.

Another category for the recovery of ToF depth maps is to use advanced filters such as bilateral filters and non-local means (NLM) filters [14], [15], [27]. Garro *et al.* [28] used an efficient graph-based segmentation method on color image to interpolate the missing depth information. Joint bilateral filtering [12] and its variations are readily available tools for depth recovery using high quality auxiliary color images [14]–[16]. Yang *et al.* [14] used the joint bilateral filtering method for range images super resolution. Yang *et al.* [29] also proposed a hierarchical joint bilateral filtering scheme for depth map upsampling. Chan *et al.* designed an adaptive multilateral upsampling filter to further address the noise in depth measurements [30]. Min *et al.* [31] proposed a weighted mode filtering method based on a joint histogram of depth video and color video. He *et al.* [13] investigated guided filtering to perform as an edge-preserving smoothing operator like the popular bilateral filter. Park *et al.* [11] used a non-local term to regularize depth maps and combined with a weighting scheme that involves edge, gradient, and segmentation information extracted from high quality color images, but jaggy artifacts occur in some boundaries. Lu *et al.* [32] formulated the filtering process as a local multipoint regression problem, consisting of multipoint estimation within a shape-adaptive local support, and aggregation of a number of multipoint estimates available for each point. It models a zero-order or linear relation between observed low resolution depth patch and color patch. However, the low-order model is not accurate for regions with complex color textures. Liu *et al.* [33] used a geodesic distance to compute the filtering coefficients based on the similarity between pixels. The algorithm is accelerated to a low computational complexity by dynamic programming. The geodesic upsampling method provides impressive recovered results for most areas of depth maps, but also introduces some annoying artifacts in regions where the associated color image has rich textures. Ferstl *et al.* [34] modeled the smooth term as a second order total generalized variation regularization, and guided the depth upsampling with an anisotropic diffusion tensor calculated from a high-resolution intensity image, providing high-quality upsampling results. It is noted that the moving least squares (MLS) method and its various variants are powerful in 3D surface fitting [35] and image recovery [36]. MLS schemes are also expected to have promising performance in depth recovery as evidenced in the comparison results in Section VI.

Generally, depth recovery schemes based on filtering techniques are competitive to MRF-based methods in recovery accuracy, but have lower asymptotic computational complexities.

B. Depth Recovery for Kinect

Depth maps provided by Kinect contain numerous structural missing around depth discontinuities due to occlusions between the build-in infrared projector and the infrared sensor. This is clearly shown in a typical depth map captured by Kinect in Fig. 1(d). Moreover, shiny objects or transparent surfaces can lead to loss of depth information. These defects correspond to the main degradations of structural depth missing for occlusion regions and random depth missing on the background.

For low delay and low complexity, many methods focus on the enhancement of Kinect depth within the spatial domain. Abdul Dakkak *et al.* [37] proposed an iterative diffusion method that incorporates RGB-D segmentation results to recover missing depth information. Andrew *et al.* [38] devised a fast modified two-pass median filter with dynamic window scales, which is effective in filling small holes, but cannot deal with large missing areas. Lai *et al.* [9] filled missing depth values by recursively applying a median filter in the construction of the RGB-D object dataset, but blurring occurs for large occlusions. Berdnikov *et al.* [39] used the "deepest neighbor" method and the simple spatial interpolation method to handle two different kinds of depth missing. This method achieves real-time processing, but the recovered depth maps are not always consistent with the accompanied color image, particularly around the boundaries between the background and foreground. We also note that there are some work departing from the inpainting approach. For example, Yu *et al.* [40] proposed refining noisy depth map in the framework of shape-from-shading.

Noting that depth maps along the temporal present strong correlations, another category is to use temporal information besides spatial information. S. Lee *et al.* [8] filled out the missing areas by an image inpainting algorithm [41], and extended the joint bilateral filter to the joint multilateral filter to improve depth quality and temporal consistency. Matyunin *et al.* [42] proposed a depth restoration method via a simple temporal filtering scheme. Camplani and Salgado [43] reduced the depth noise with a joint-bilateral filter on the spatial domain and repaired the depth value fluctuation on the temporal domain. These methods mentioned above use both the spatial and temporal information for depth recovery, but quality of the filtered depth maps are not satisfactory around depth discontinuities.

III. Degradation Modes and AR Model

A. Degradation Modes of Depth Maps

Current depth capturing systems are far from perfect. A captured depth map is a degraded version of the underlying groundtruth. Let $\boldsymbol{d}$ and $\boldsymbol{d}^0$ denote the vector form of the underlying perfect depth map and the captured one, respectively. The observation model for depth capturing is described as

$$\boldsymbol{d}^0 = \boldsymbol{P}\boldsymbol{d} + \boldsymbol{n}, \tag{1}$$

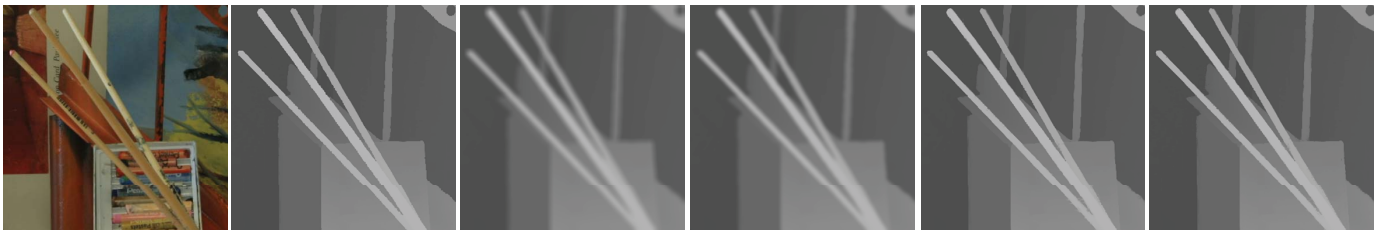

Fig. 2. Prediction efficiency for four AR predictors: (a) the associated color image, (b) the input depth map, and prediction results of AR predictors constructed by (c) average filter, (d) Gaussian filter, (e) bilateral filter, and (f) our proposed filter. The prediction error (MAD) between the predictions in (c), (d), (e) and (f) against the original depth maps are 3.992, 3.131, 0.129, and 0.051, respectively.

where $\boldsymbol{P}$ represents the observation matrix and $\boldsymbol{n}$ is the introduced additive noise.

There are mainly four types of degradations: undersampling, random depth missing, structural depth missing, and pollution with additive noise. For the former three ones, the observed depth map $\boldsymbol{d}^0$ has a smaller number of elements than $\boldsymbol{d}$ and $\boldsymbol{P}$ is a flat matrix to identify valid pixels with depth values. However, $\boldsymbol{P}$ has different structures for different degradation modes. For the degradation of undersampling, $\boldsymbol{P}$ is a sampling matrix; while in structural and random missing, $\boldsymbol{P}$ is constructed from an identity matrix by removing those rows associated with the depth-missing pixels. Depth capturing systems may suffer from different combinations of the degradation modes. As shown in Fig. 1(a), the depth map captured by ToF camera is undersampled (lower resolution than the accompanied color images), and polluted by noise. After viewpoint registration, the warped depth map contains disoccluded regions around object boundaries, and thus suffers from degradation of structural depth missing. As shown in Fig. 1 (d) (see also more examples in figures in Section VI-B on experiments), the Kinect depth map contains both random and structural missing degradations. Our method is to recover high quality depth maps from low quality observations, and all the four kinds of degradations are handled in the proposed unified depth recovery framework.

The observation model (1) considers degradation modes that commonly present in depth sensing technologies. It should be noted that there are other sources of degradation in real depth sensors. Related investigations, see [4], [44], show that depth cameras generally have some systematic errors for various reasons, e.g., anharmonic LED modulation, integration time offset, pixel offsets, intensity dependent response, and different Lambertian reflectance properties. It is possible to consider these degradations in the observation model. In applications using depth cameras, these complex systematic errors are usually calibrated and compensated as a preprocessing step before subsequence processing [44], [45].

B. AR Model of Depth Maps

The AR model has been applied in many image processing applications, such as detecting and interpolating missing areas in image sequences [46], super-resolution, forecasting of spatial-temporal data [47], [48], as well as backward adaptive video coding [49]. This demonstrates that the simple AR model is versatile for many applications as long as the AR predictors are properly designed.

As shown in Fig. 7, Fig. 11, and Fig. 13, depth maps for generic 3D scenes contain mainly smooth regions separated by curves. The AR model can well describe such type of 2D signals. The key insight is that a signal can be regenerated by the signal itself. Denote by $\boldsymbol{D}$ a depth map, and $\boldsymbol{D}_{\boldsymbol{x}}$ the depth value at location $\boldsymbol{x}$. The predicted depth map $\tilde{\boldsymbol{D}}$ by the AR model from the depth map $\boldsymbol{D}$ is expressed as

$$\tilde{\boldsymbol{D}}_{\boldsymbol{x}} = \sum_{\boldsymbol{y}\in\mathcal{N}(\boldsymbol{x})} a_{\boldsymbol{x},\boldsymbol{y}} \boldsymbol{D}_{\boldsymbol{y}}, \tag{2}$$

where $\mathcal{N}(\boldsymbol{x})$ is the neighborhood of pixel $\boldsymbol{x}$ and $a_{\boldsymbol{x},\boldsymbol{y}}$ denotes the AR coefficient for pixel $\boldsymbol{y}$ in the neighborhood $\mathcal{N}(\boldsymbol{x})$. The accuracy of the AR model can be measured by the difference between $\boldsymbol{D}$ and $\tilde{\boldsymbol{D}}$, e.g., mean absolute difference (MAD) or root mean squared error (RMSE).

To verify the fitness of the AR model for depth maps, we check the prediction errors between the predicted depth maps and the groundtruth for a set of test depth maps. Four AR predictors are tested: an average filter, a Gaussian filter, a bilateral filter and our proposed filter, all with a 11×11 neighborhood. As shown in Fig. 2, all the four filters have good prediction for smooth regions. However, when coming to discontinuities, we can see that the results of the average filter and gaussian filter are apparently of low quality and subject to oversmoothing around the edges. Since the proposed filter adapts the AR model to the nonlocal structures of signals, it almost regenerates the depth map: the average prediction error in MAD is only 0.051/pixel. These results demonstrate that the AR model is quite effective in modeling the depth maps, and thus encourage the application of this model to the recovery of depth maps.

Depth-color pairs have strong correlation in terms of geometrical structures, and are often acquired and used together [4], [20]. As shown in Fig. 2(a) and (b), edges in the depth map have their counterparts in color image. This suggests that the locations of edges in depth maps can be inferred from the accompanied color images, and motivates the proposed color-guided AR model for depth recovery from low resolution and incomplete observations.

IV. Color-Guided AR Model for Depth Recovery

A. Depth Recovery Based on AR Model

Denote by $\boldsymbol{D}^0$ the observed depth map and $\mathcal{O}$ the set of pixels with observed depth values. Given the observed depth map $\boldsymbol{D}^0$, we propose the following depth recovery model based on AR:

$$\min_{\boldsymbol{D}} E_{\text{data}}(\boldsymbol{D}, \boldsymbol{D}^0) + \lambda E_{\text{AR}}(\boldsymbol{D}), \tag{3}$$

where $E_{\text{data}}(\boldsymbol{D}, \boldsymbol{D}^0)$ is the data term to make the recovered depth consistent with the observation, $E_{\text{AR}}(\boldsymbol{D})$ is the AR term to impose AR model on the recovered depth map. The data term and the AR term are weighted by λ.

The data term is expressed as

$$E_{\text{data}}(\boldsymbol{D}, \boldsymbol{D}^0) \triangleq \sum_{\boldsymbol{x}\in\mathcal{O}} \left(\mathbf{D}_{\boldsymbol{x}} - \mathbf{D}^0_{\boldsymbol{x}}\right)^2, \tag{4}$$

and the AR model is incorporated into the depth recovery as the AR term

$$E_{\text{AR}}(\boldsymbol{D}) \triangleq \sum_{\boldsymbol{x}} \left(\boldsymbol{D}_{\boldsymbol{x}} - \sum_{\boldsymbol{y}\in\mathcal{N}(\boldsymbol{x})} a_{\boldsymbol{x},\boldsymbol{y}} \boldsymbol{D}_{\boldsymbol{y}} \right)^2, \tag{5}$$

where the AR coefficient $a_{\boldsymbol{x},\boldsymbol{y}}$ is defined according to both depth and color information in the following section. The proposed method has a similar form, but is a departure essentially, to the work in [11]. In [11], in addition to color information, segmentation and edge saliency are taken into account in confidence weights. Although such features can be readily incorporated in our recovery model, we found that the elegant AR model can well describe the characteristics of depth maps. Therefore we insist on the low level processing in depth recovery, and retain the simplicity of the model.

As shown in Section III-B, the AR model is powerful in describing depth maps only when the AR coefficients are properly designed. However, an accurate AR model is difficult to infer from only the degraded depth map $\boldsymbol{D}^0$. Since the depth-color pairs have strong structural correlations, the information loss due to depth degradation can be complemented by the accompanied color image. To achieve high quality depth recovery, we design pixel-wise adaptive AR predictors in Section IV-B using both the initial depth map and the auxiliary color image.

B. Color-Guided AR Model

As demonstrated by the simulation results in Section III-B, the AR model has very different performance with different AR predictors. The common way in AR-based image processing is to divide images into small units and each unit shares an AR predictor. However, we observe that the AR model cannot provide sufficient adaptivity when each unit contains considerable variations. Therefore, we design pixel-wise adaptive AR predictors: an AR predictor $\{a_{\boldsymbol{x},\boldsymbol{y}}\}$, $\boldsymbol{y} \in \mathcal{N}(\boldsymbol{x})$ is constructed for each pixel $\boldsymbol{x}$ by considering both the depth and color information.

A depth map is reliably recovered with the optimal AR predictors, which can be derived only when the depth map is available. To break this chicken-egg dilemma, we design AR predictors using the available depth map and the accompanied color image. Note that the observed depth map $\boldsymbol{D}^0$ is not directly applicable due to degradations such as the undersampling or depth missing. Denote by $\hat{\boldsymbol{D}}$ the rough estimated depth map obtained by bicubic interpolation from $\boldsymbol{D}^0$. Represent the accompanied color image with $\boldsymbol{I} = \{\boldsymbol{I}^i, i \in \mathcal{C}\}$, where $\boldsymbol{I}^i$ is the intensity of the color channel with index i and $\mathcal{C}$ is the index set of color channels in a certain color space. We had investigated three color spaces (RGB, YUV, and Lab). All three color spaces yield similar results, and we choose the YUV color space due to its slightly better performance, i.e., $\mathcal{C} = \{Y, U, V\}$ in our implementation. The AR coefficient $a_{\boldsymbol{x},\boldsymbol{y}}$ consists of two terms:

$$a_{\boldsymbol{x},\boldsymbol{y}} = \frac{1}{S_{\boldsymbol{x}}} a^{\hat{\boldsymbol{D}}}_{\boldsymbol{x},\boldsymbol{y}} a^{\boldsymbol{I}}_{\boldsymbol{x},\boldsymbol{y}}, \tag{6}$$

where $S_{\boldsymbol{x}}$ is the normalization factor, $a^{\hat{\boldsymbol{D}}}_{\boldsymbol{x},\boldsymbol{y}}$ and $a^{\boldsymbol{I}}_{\boldsymbol{x},\boldsymbol{y}}$ are the depth term and color term, respectively.

The depth term $a^{\hat{\boldsymbol{D}}}_{\boldsymbol{x},\boldsymbol{y}}$ is defined on the initial estimated depth map $\hat{\boldsymbol{D}}$ by a range filter:

$$a^{\hat{\boldsymbol{D}}}_{\boldsymbol{x},\boldsymbol{y}} = \exp\left(- \frac{\left(\hat{\boldsymbol{D}}_{\boldsymbol{x}} - \hat{\boldsymbol{D}}_{\boldsymbol{y}}\right)^2}{2\sigma_1^2} \right), \tag{7}$$

where σ_1 is the decay rate of the range filter. Qualitatively, $a^{\hat{\boldsymbol{D}}}_{\boldsymbol{x},\boldsymbol{y}}$ has a large value if $\hat{\boldsymbol{D}}_{\boldsymbol{x}}$ is close to $\hat{\boldsymbol{D}}_{\boldsymbol{y}}$. This term is also designed to avoid incorrect depth prediction due to depth-color inconsistency: pixels of the same depth layers may have very different colors; pixels of similar colors may belong to different depth layers.

The color term $a^{\boldsymbol{I}}_{\boldsymbol{x},\boldsymbol{y}}$ is designed to take benefit of the correlations in the depth-color pair. Edges in a depth map co-occur with their counterparts in the accompanied color image. The color term $a^{\boldsymbol{I}}_{\boldsymbol{x},\boldsymbol{y}}$ should be able to prevent the AR model from predicting across depth discontinuities. Based on the non-local principle, we propose the following color terms :

$$a^{\boldsymbol{I}}_{\boldsymbol{x},\boldsymbol{y}} = \exp\left(- \frac{\sum_{i\in\mathcal{C}} ||\boldsymbol{B}_{\boldsymbol{x}} \circ \left(\mathcal{P}^i_{\boldsymbol{x}} - \mathcal{P}^i_{\boldsymbol{y}}\right)||_2^2}{2 \times 3 \times \sigma_2^2} \right), \tag{8}$$

where σ_2 controls the decay rate of the exponential function, $\mathcal{P}^i_{\boldsymbol{x}}$ denotes an operator that extracts a $w \times w$ patch centered at $\boldsymbol{x}$ in color channel i, "$\circ$" represents the element-wise

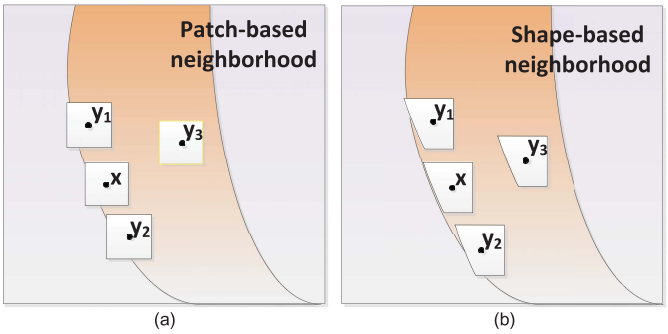

Fig. 3. Illustrations for contrast between the traditional predictors and our proposed AR predictors: (a) patch-based neighborhood (b) shape-based neighborhood.

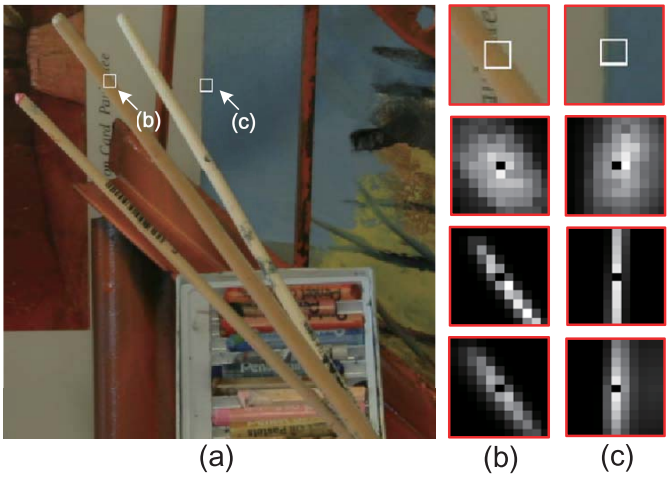

Fig. 4. Illustrations for the color term of AR predictors: (a) two pixels with their neighborhoods, (b) and (c) present the enlarged versions (top row) and AR predictors for the two pixels constructed from bilateral filter (2nd row), standard NLM filter (3rd row), and the proposed filter (bottom row).

multiplication. The bilateral filter kernel $\boldsymbol{B}_{\boldsymbol{x}}$ is defined in the extracted $w \times w$ patch:

$$\boldsymbol{B}_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{||\boldsymbol{x}-\boldsymbol{y}||_2^2}{2\sigma_3^2}\right)\exp\left(-\frac{\sum_{i\in\mathcal{C}}(\boldsymbol{I}_{\boldsymbol{x}}^i - \boldsymbol{I}_{\boldsymbol{y}}^i)^2}{2\times 3\times \sigma_4^2}\right), \quad (9)$$

where σ_3 and σ_4 are parameters of the bilateral kernel to adjust the importance of the spatial distance and intensity difference, respectively.

The difference between the proposed filter in the color term and the standard NLM filter is that the proposed one uses a bilateral kernel to weight the distance of local patches while the standard one uses a Gaussian kernel. The bilateral kernel $\boldsymbol{B}_{\boldsymbol{x}}$ has a strong response for pixels of similar intensities to $\boldsymbol{x}$, and hence carries the shape information of local image structures. This extends the NLM filter from patch-based to shape-based in measuring the resemblance of local structures, and has a significant impact on the structure of AR predictors for pixels around edges. As shown in Fig. 3, two homogeneous regions are separated by smooth curves and $\boldsymbol{x}$ is close to a curve. To construct the AR predictor for $\boldsymbol{x}$, the similarity between $\boldsymbol{x}$ and each pixel $\boldsymbol{y}$ in the neighborhood $\mathcal{N}_{\boldsymbol{x}}$ is evaluated. Constrained by the patch structure, the standard NLM filter produces large coefficients only for pixels that are parallel to the edge, e.g. $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, and produces small coefficients for other pixels, such as $\boldsymbol{y}_3$, even though they have the same intensity as $\boldsymbol{x}$. On the contrary, our bilateral-weighted NLM filter has a shape-adaptive neighborhood, and increases opportunities to exploit more correlations for pixels around discontinuities. This is illustrated in Fig. 4. With the shape-adaptive neighborhood, the proposed filter produces an equally large coefficient for $\boldsymbol{y}_3$ as for $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$. As verified later in Section V, small supports of AR predictors would underdetermine the recovery system and can lead to fail recovery for related pixels; while the proposed AR predictors of larger supports form a more well-determined system, and achieve stable recovery.

V. Stability Analysis and Parameter Adaptation

The depth recovery system often works under perturbations: 1) observation perturbation: sensed depth measurements may contain some noise; and 2) system perturbation: AR predictors can be affected by highly-textured regions of the color image, which can make the system ill-conditioned. So, it is quite necessary to analyze the behavior of the depth recovery method under perturbations for stable and high-quality recovery.

In this section, we formulate the quadratic minimization (3) into a quadratic programming and analyze the stability of the recovery model by the conditioning of linear systems (Section V-A). Then we investigate how the parameters affect the system stability, and design a parameter adaptation scheme to achieve stable and high quality depth recovery (Section V-B).

A. Stability of the Depth Recovery System

The depth recovery model (3) are quadratic with respect to $\boldsymbol{D}$. Therefore, it can be reformulated as an unconstrained quadratic programming and analyzed with the conditioning of linear systems. Let $\boldsymbol{d}$ and $\boldsymbol{d}^0$ be the vector form of $\boldsymbol{D}$ and $\boldsymbol{D}^0$, respectively. Then, the depth recovery model is equivalent to the following minimization with respect to $\boldsymbol{d}$:

$$\min_{\boldsymbol{d}} \|\boldsymbol{d}^0 - \boldsymbol{P}\boldsymbol{d}\|_2^2 + \lambda\|\boldsymbol{d} - \boldsymbol{Q}\boldsymbol{d}\|_2^2, \quad (10)$$

where $\boldsymbol{P}$ is the observation matrix and $\boldsymbol{Q}$ is the prediction matrix corresponding to AR predictors $\{a_{\boldsymbol{x},\boldsymbol{y}}\}$. In (10), the first term is the data term and the second term is the AR term.

The unconstrained quadratic programming (10) is convex, and its global minima can be obtained by solving the first-order conditions:

$$\underbrace{\left(\boldsymbol{P}^\top\boldsymbol{P} + \lambda(\boldsymbol{I}-\boldsymbol{Q})^\top(\boldsymbol{I}-\boldsymbol{Q})\right)}_{\boldsymbol{H}}\boldsymbol{d} = \underbrace{\boldsymbol{P}^\top\boldsymbol{d}^0}_{\boldsymbol{c}}, \quad (11)$$

where $\boldsymbol{H}$ is a squared matrix. Therefore, the stability of the depth recovery model can be analyzed via the conditioning of linear systems. Denote the condition number of $\boldsymbol{H}$ by $\kappa := \sigma_{max}/\sigma_{min}$, where σ_{max} and σ_{min} are the maximal and minimal singular values of $\boldsymbol{H}$, respectively. Denote by $\delta\boldsymbol{c}$ the noise in $\boldsymbol{c}$, $\delta\boldsymbol{H}$ the perturbation in $\boldsymbol{H}$, and $\delta\boldsymbol{d}$ the resulting error in $\boldsymbol{d}$. The sensitivity of the linear system can be obtained by considering the perturbed system: $(\boldsymbol{H}+\delta\boldsymbol{H})(\boldsymbol{d}+\delta\boldsymbol{d}) = \boldsymbol{c}+\delta\boldsymbol{c}$ [50]. The sensitivity of the linear system is described by:

$$\frac{\|\delta\boldsymbol{d}\|}{\|\boldsymbol{d}\|} \le \kappa\left(\frac{\|\delta\boldsymbol{H}\|}{\|\boldsymbol{H}\|} + \frac{\|\delta\boldsymbol{c}\|}{\|\boldsymbol{c}\|}\right), \quad (12)$$

which shows that the relative error of the recovery depth map is proportional to the relative noise in $\boldsymbol{H}$ and $\boldsymbol{c}$ up to a magnification of the condition number κ. When κ is large, a small relative change in either $\boldsymbol{H}$ or $\boldsymbol{c}$ can cause a large change in $\boldsymbol{d}$, which would severely degrade the performance of depth recovery. Therefore, the depth recovery model should be carefully designed so that the matrix of the resulting linear system has a low condition number. In the linear equations of first-order conditions, the coefficient matrix is the combination of the sampling matrix $\boldsymbol{P}$, the

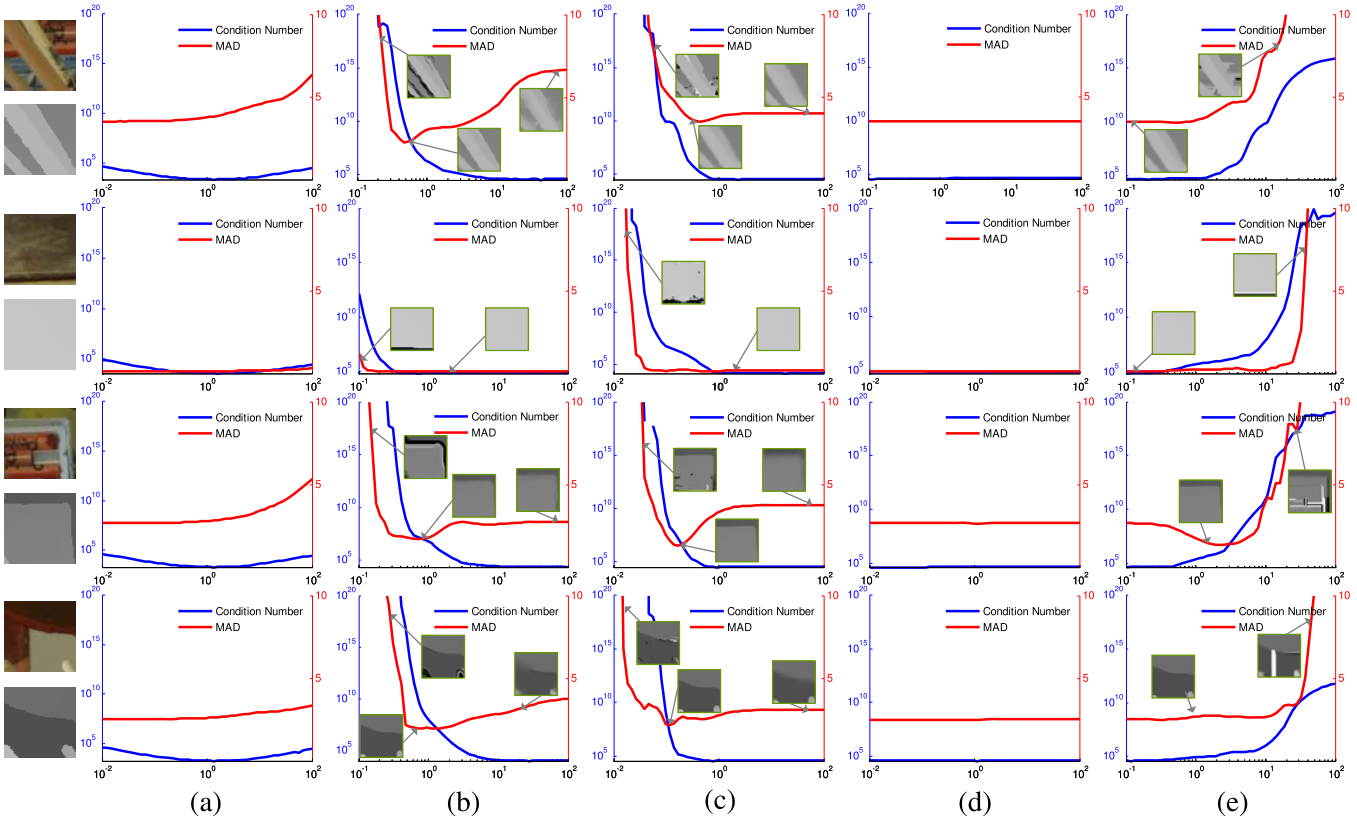

Fig. 5. Condition number κ of the recovery system (left vertical axes in blue) and the recovery quality in MAD (right vertical axes in red) with respect to the parameters: (a) λ, (b) σ_1, (c) σ_2, (d) σ_3, and (e) σ_4. For σ_1, σ_2, and σ_4 that significantly affect the recovery quality, some representative recovered depth maps are presented for comparison and analysis.

prediction matrix $\boldsymbol{Q}$, and their transposes. Note that $\boldsymbol{P}^\top \boldsymbol{P}$ is a highly deficient diagonal matrix, e.g., the rank is only the 1/64 of the full rank for 8×8 super-resolution. Therefore, the invertibility and stability of the linear system (11) are determined by the prediction matrix $\boldsymbol{Q}$ governed by five parameters: λ and $\{\sigma_i\}_{i=1}^4$. The specific influences of the five parameters on the stability as well as the parameter adaptation are detailed in the following section (Section V-B).

B. Parameter Adaptation

To test the influence of the parameters on the stability and the recovery quality, we randomly extract a large number of patches from degraded depth maps, and perform 8×8 super-resolution (other upsampling rates also yield the same conclusions). Instead of traversing the whole parameter space, we perform depth recovery by varying each parameter while setting other parameters at fixed reasonable values: $\lambda = 0.01$, $\sigma_1 = 2$, $\sigma_2 = 9$, $\sigma_3 = 5$, and $\sigma_4 = 2$. The varying range for each parameter is $[0.01, 100]$ which covers the interest fraction of the parameter space. For each test point, we evaluate the condition number of the resulting matrix $\boldsymbol{H}$ and the quality of recovered depth measured in MAD. Results for four test depth-texture pairs are presented in Fig. 5. We analyze the sensitivity of each parameter as well as its adaptation as follows.

1) λ: This parameter adjusts the importance of the data term and the AR term. Note that both $\boldsymbol{P}^\top \boldsymbol{P}$ and $(\boldsymbol{I} - \boldsymbol{Q})^\top (\boldsymbol{I} - \boldsymbol{Q})$ are rank-deficient matrices. Either a very small or a very large λ will produce $\boldsymbol{H}$ of a large condition number. Fig. 5 shows that $\lambda \in [0.01, 100]$ yield condition numbers lower than 10^5, which is stable for the recovery system. We observe that the MAD of the recovered depth is monotonically increasing with respect to λ. Therefore, we set $\lambda = 0.01$ in our implementation.

2) σ_1: In the depth term, the weights for candidates are assigned according to the closeness of their values to the reference $\hat{\boldsymbol{D}}_{\boldsymbol{x}}$. σ_1 controls the tolerance for two different depth values to be considered close enough to assign a significant weight. As shown in Fig. 5, the condition number would dramatically increase when σ_1 is small (e.g., below 0.5), and depth around edges cannot be recovered due to the instability of the recovery system. We also observe that too large a σ_1 tends to produce oversmooth results. Therefore, we design an adaptive scheme: σ_1 is assigned to a large value for smooth depth regions to include more depth values for stable and accurate prediction, and is assigned to a small value for regions around depth discontinuities to avoid prediction across depth edges. To this end, σ_1 is determined by the local smoothness:

$$\sigma_1(\boldsymbol{x}) = a_1 + b_1 \exp\big(-c_1 \|\nabla \hat{\boldsymbol{D}}\|_{[\boldsymbol{x}]}\big), \tag{13}$$

where $\|\nabla \hat{\boldsymbol{D}}\|_{[\boldsymbol{x}]}$ denotes the gradient magnitude of $\hat{\boldsymbol{D}}$ at $\boldsymbol{x}$; a_1 and b_1 determine the lower and upper bounds of σ_1, and are set at 0.5 and 2.5, respectively; c_1 controls the decay rate of the exponential mapping, and is set at 10.

3) σ_2: The most important role of the color term is to provide clues of depth edges lost due to undersampling based on the assumption of strong structural correlation between color images and the associated depth maps. Being similar to σ_1 in the depth term, σ_2 in the color term is the tolerance for two patches to be considered similar enough to assign a significant weight. The behavior of the stability and recovery performance with respect to σ_2 is quite similar to those with respect to σ_1. The recovery system become instable when σ_2 is very small, and the depth recovery fails for pixels around depth discontinuities. Also, σ_2 should be large in flat depth regions to have stable prediction and small around depth edges to avoid prediction across discontinuities. Therefore, σ_2 is adapted according to the same strategy as for σ_1:

$$\sigma_2(\boldsymbol{x}) = a_2 + b_2 \exp\big(-c_2 \|\nabla \hat{\boldsymbol{D}}\|_{[\boldsymbol{x}]}\big), \tag{14}$$

where a_2, b_2, and c_2 are parameters. Note that color images are much more spatially-variant than depth maps. The lower and upper bounds of σ_1 are adapted to local characteristics of the color image with the following piecewise function:

$$\begin{cases} a_2 = 1.9, & b_2 = 3.8, & \|\nabla \boldsymbol{I}\|_{[\boldsymbol{x}]} < 5, \\ a_2 = 4.8, & b_2 = 2.9, & 5 \le \|\nabla \boldsymbol{I}\|_{[\boldsymbol{x}]} < 8, \\ a_2 = 6.7, & b_2 = 2.9, & \|\nabla \boldsymbol{I}\|_{[\boldsymbol{x}]} \ge 8, \end{cases} \tag{15}$$

where $\|\nabla \boldsymbol{I}\|_{[\boldsymbol{x}]}$ is the gradient magnitude of the Y-channel image at $\boldsymbol{x}$. The values of a_2 and b_2 in (15) are obtained by numerically fitting the two parameters and the gradient magnitude to have the best recovery performance.

4) σ_3: The two parameters σ_3 and σ_4 of the bilateral kernel in Formula (9) control the shape and the size of the non-local patches. As shown in Fig. 5, the conditional number and recovery accuracy with respect to σ_3 are quite stable. For example, the fluctuation of MAD is usually within 0.01 for smooth regions and is within 0.15 for depth regions around edges. Therefore, σ_3 is fixed at 5 in our implementation.

5) σ_4: In the bilateral kernel (9), σ_4 controls the support of the bilateral kernel. As σ_4 increases, the bilateral kernel $\boldsymbol{B}_{\boldsymbol{x}}$ to define the shape of the patch tends to have equal weights within the $w \times w$ window. This will reduce the shape-adaptive patch to a squared patch, and thus the proposed non-local kernel degenerates into a conventional non-local mean filter. As shown in Fig. 3 and Fig. 4, for pixels around edges, the squared patch produces AR predictors of small supports. This would lead to the instability of the recovery system. As verified by the results shown in Fig. 5, the conditional number increases rapidly when σ_4 is larger than 5; Accordingly, the recovery performance is stable when $\sigma_4 < 10$, but will severely drop beyond this range. When σ_4 has large values, depth values cannot be reliably recovered due to the ill-conditioning of the system. Therefore, we set $\sigma_4 = 3$ in our implementation.

6) Neighborhood Size in AR Predictors: We investigate the influence of neighborhood size $\mathcal{N}_{\boldsymbol{x}}$ on the recovery quality and computational complexity. To this end, low resolution depth patches are recovered with various neighborhood size, i.e., $3 \times 3, 5 \times 5, \ldots, 17 \times 17$. As shown in Fig. 6, as the neighborhood size becomes larger, more samples are included into AR prediction, yielding more stable recovery. However, the recovery accuracy does not significantly increase as the neighborhood size beyond the size of 11×11. Moreover, increasing the support size will also increase complexity. The computation is approximately linear with respect to the neighborhood size. Therefore, the neighborhood size of 11×11 is chosen in our implementation.

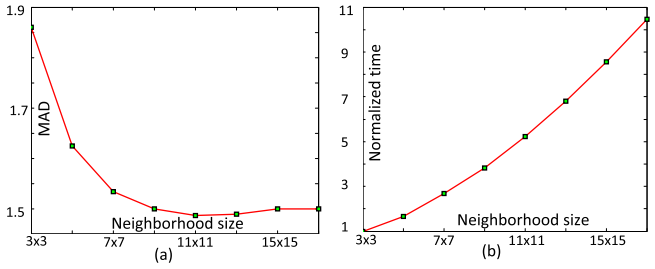

Fig. 6. Influence of neighborhood size on (a) recover quality (MAD) and (b) computational complexity in normalized time (relative to the 3×3 case normalized to one).

VI. Experiments and Results

Our method is first evaluated on Middlebury datasets with various synthetic degradations and compared with several existing methods. Then, our method is applied on two real depth sensing systems to obtain high quality depth maps. All the datasets, results, and recovered depth maps are available in the project website.[2] We direct interested readers to the website for more results on real datasets.

A. Experiments on Datasets With Synthetic Degradations

Six datasets, *Art*, *Book*, *Moebius*, *Reindeer, Laundry,* and *Dolls* from the Middlebury's benchmark [51] are used for evaluation. Three kinds of typical degradations are simulated: undersampling, ToF-like degradation (undersampling with noise), Kinect-like degradation (structural missing along depth discontinuities and random missing in flat regions). Our method is compared with ten state-of-the-art methods (if applicable): Bicubic interpolation, MRF-based method (MRF) [20], improved MLS (IMLS) [36], joint bilateral filtering on cost volumes (JBFcv) [14], guided image filtering (Guided) [13], edge-weighted NLM-regularization (Edge) [11], patch-based synthesis (PS) [26], cross-based local multipoint filtering (CLMF) [32], joint geodesic filtering (JGF) [33], total generalized variation (TGV) [34]. Upsampling results (Table I) on *Art*, *Book*, and *Moebius* for MRF, JBFcv, Guided, Edge, and TGV are quoted from [11] and [34]. The results for MRFs [20] and JBFcv [14] on other three RGB-D pairs were not available; For the Patch-based synthesis (PS) method [26], the released patch database trained for $4\times$ upsampling is poor for other upsampling rates. Therefore, we only present its results at $4\times$ upsampling. The rest results in Table I, and results in Table II and Table III are generated by the provided codes (if have) or our implementations. We improve the MLS scheme [36] by extending the Gussian weighting to the cross bilateral weighting [7] to avoid MLS fitting across depth discontinuities, hence named improved MLS (IMLS). The CLMF method [32]

[2] http://cs.tju.edu.cn/faculty/likun/projects/depth_recovery/index.htm

TABLE I
QUANTITATIVE UPSAMPLING RESULTS (IN MAD) FROM UNDERSAMPLED DEPTH MAPS ON MIDDLEBURY DATASETS AT FOUR SUBSAMPLING RATES

	Art				*Book*				*Moebius*				*Reindeer*				*Laundry*				*Dolls*			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
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 [20]	0.59	0.96	1.89	3.78	0.21	0.33	0.61	1.20	0.24	0.36	0.65	1.25	-	-	-	-	-	-	-	-	-	-	-	-
IMLS[36]	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
JBFcv[14]	0.55	0.68	1.44	3.52	0.29	0.44	0.62	1.45	0.38	0.46	0.67	1.10	-	-	-	-	-	-	-	-	-	-	-	-
Guided[13]	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
Edge [11]	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
PS[26]	-	0.93	-	-	-	**0.16**	-	-	-	**0.17**	-	-	-	0.56	-	-	-	1.13	-	-	-	0.83	-	-
CLMF0[32]	0.43	0.74	1.37	2.95	0.14	0.28	0.51	1.06	0.15	0.29	0.52	1.01	0.32	0.51	0.84	1.51	0.30	0.50	0.82	1.66	0.24	0.34	0.66	1.02
CLMF1[32]	0.44	0.76	1.44	2.87	0.14	0.28	0.51	1.02	0.15	0.29	0.51	0.97	0.32	0.51	0.84	1.55	0.30	0.50	0.80	1.67	0.23	0.34	0.60	1.01
JGF[33]	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[34]	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
Ours_FP	0.18	0.49	0.66	2.15	**0.12**	0.25	0.48	0.80	0.11	0.25	0.42	0.90	0.22	0.40	0.64	1.21	0.20	0.35	0.59	1.20	0.21	0.36	0.56	1.09
Ours_AP	**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**

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

	Art				*Book*				*Moebius*				*Reindeer*				*Laundry*				*Dolls*			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
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
IMLS[36]	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[13]	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
Edge[11]	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
PS[26]	-	1.46	-	-	-	1.09	-	-	-	1.17	-	-	-	1.21	-	-	-	1.53	-	-	-	1.33	-	-
CLMF0[32]	1.19	1.77	2.95	4.91	0.90	1.48	2.38	3.36	0.87	1.44	2.32	3.30	0.96	1.56	2.54	3.85	0.94	1.55	2.50	3.81	0.96	1.54	2.37	3.25
JGF[33]	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[34]	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
Ours_AP	**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**

has two versions: CLMF0 and CLMF1 for zero- and first-order polynomial model, respectively. In visual comparisons (Fig. 7, Fig. 8, and Fig. 10), regions highlighted by rectangles are enlarged, and the error maps are shown by subtracting between recovered depth and ground truth, for easy visual inspection.

1) Undersampling Degradation: Depth recovery results (in MAD) at four upsampling rates for each RGB-D pair are reported in Table I. For our method, we present the results with two configurations in Table I: 1) Ours_FP uses fixed parameters that are manually tuned to avoid instable recovery and also to obtain the best recovery performance, and 2) Ours_AP adopts the parameter adaptation schemes in Section V-B. As shown in Table I, our method nearly obtains the lowest MAD for most cases (especially for high upsampling rates of 8× and 16× upsampling), which demonstrates its effectiveness. The Edge method [11] provides slightly better results for low upsampling rates on *Reindeer, Laundry,* and *Dolls*. We also observe that the PS method [26] achieves slightly better results for *Book* and *Moebius* at 4× upsampling. However, it needs to train patch database at each upsampling rate, and its computational complexity is quite high: it takes about 40 minutes to super-resolve a depth map.

TABLE III
QUANTITATIVE DEPTH RECOVERY RESULTS FROM KINECT-LIKE DEGRADATIONS (STRUCTURAL MISSING AND RANDOM MISSING)

	Art	*Book*	*Moebius*	*Reindeer*	*Laundry*	*Dolls*
Bicubic	0.90	0.61	0.66	0.95	0.91	0.76
IMLS[36]	0.91	0.58	0.72	0.68	0.72	0.82
JBF[12]	0.84	0.63	0.69	0.92	0.88	0.76
Guided[13]	1.20	0.63	0.67	0.96	0.94	0.76
CLMF0[32]	1.01	0.60	0.64	0.94	0.89	0.74
Ours_AP	**0.58**	**0.53**	**0.60**	**0.68**	**0.75**	**0.69**

Fig. 7 shows 8× upsampled depth maps for *Dolls* and *Art*. Upsampled depth maps by three state-of-the-art methods, IMLS [36], Edge [11] and JGF [33], are also shown for comparison. The Edge method generates comparable results to ours for *Moebius*, but introduces some jaggy artifacts along edges. The JGF method provides promising quality for most areas of depth maps, but introduces annoying artifacts in regions where the associated color image has rich textures, e.g., the crayon of *Art*. The visual comparison show that our method not only achieves low average recovery errors, but also provides visually consistent results.

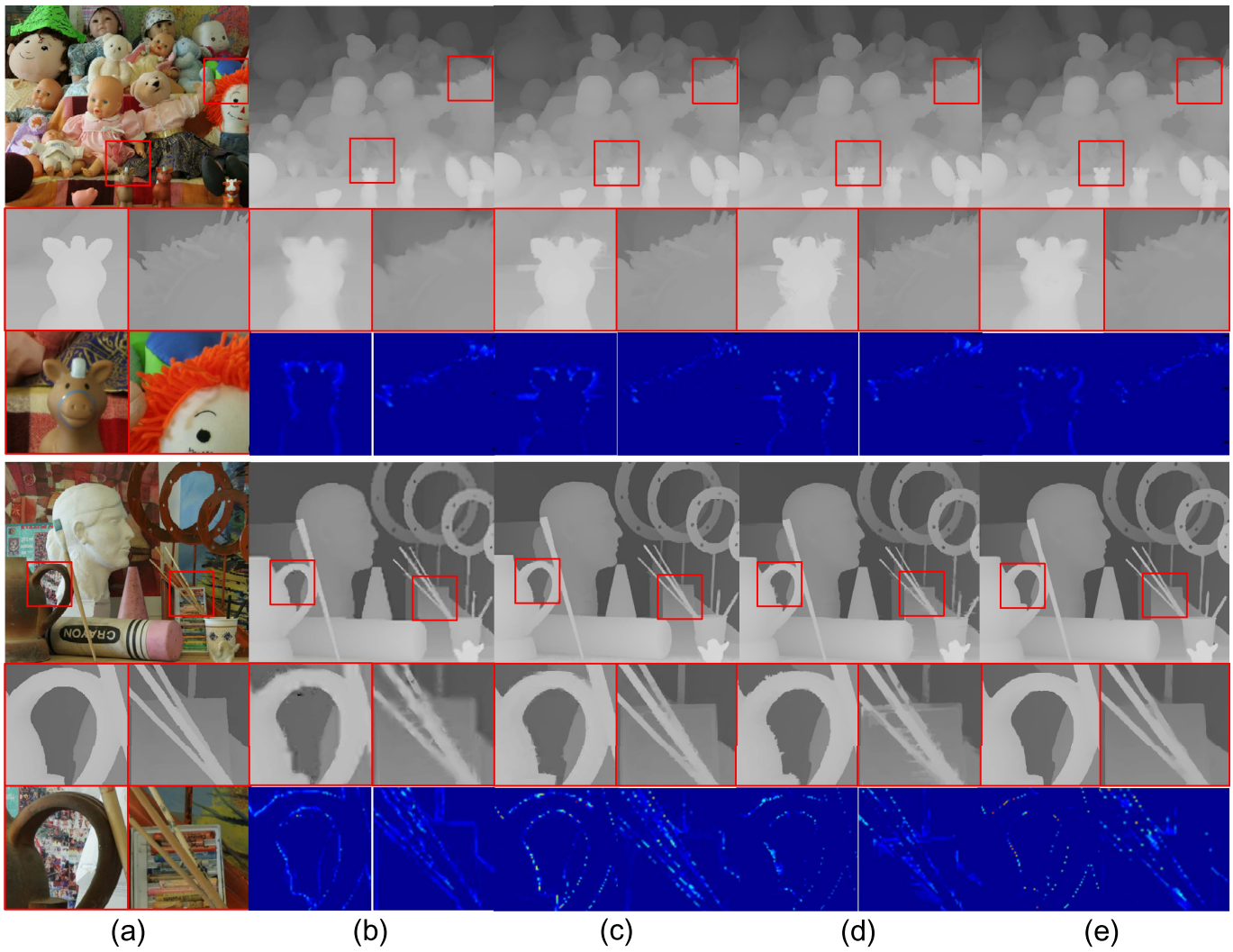

Fig. 7. Visual quality comparison for depth upsampling on two Middlebury RGB-D pairs: (a) color image and depth ground truth, depth maps upsampled (8×) by (b) IMLS [36] (MAD: 0.61; 1.04), (c) Edge [11] (MAD: 0.56; 1.03), (d) JGF [33] (MAD: 0.59; 0.78), and (e) our method (MAD: 0.50; 0.64). The first and second MADs for each method are for *Dolls* and *Art*, respectively.

2) ToF-Like Degradation: To simulate the ToF-like depth degradation, we first add Gaussian noise with a variance of 25 to the original datasets, and then downsample the polluted datasets at the four upsampling rates. Quantitative depth recovery results of our method and other eight methods are summarised in Table II. Our method obtains the lowest MAD for all cases. The JGF method does not perform as well as in pure upsampling due to the lack of denoising capability. The IMLS, Guided, Edge, PS, and CLMF0 methods provide comparative results thanks to their inherent denoising capabilities. The minimization of total variation in TGV is powerful in suppressing noise, and therefore yields promising results. To compare visual results, Fig. 8 presents depth maps on *Book* and *Reindeer* recovered by Bicubic, IMLS, Edge, TGV, and our method. The depth maps recovered by two fitting methods, Bicubic and IMLS, still contain significant redidual noise. The Edge method tends to over-smooth out more useful signal components. The TGV method provides cleaner depth maps, but fails to preserve tiny structures such as the ears of the reindeer. Our method is able to effectively remove noise in upsampling while avoiding contaminating depth content.

3) Kinect-Like Degradation: To simulate Kinect-like degradation, structural missing is created along depth discontinuities, and random missing is generated in flat areas. Depth maps with Kinect-like degradation are presented in Fig. 9. Recovery results from Kinect-like depth degradation are reported in Table III. Five methods applicable for hole filling are compared: Bicubic, IMLS [36], joint bilateral filtering (JBF) [12], Guided [13], and CLMF0 [32]. As shown in Table III, our method obtains the lowest MAD for all cases, which shows its effectiveness in handling Kinect-like degradation. For visual comparison, depth maps recovered by IMLS, JBF, CLMF0, and our method are presented in Fig. 10. All methods provide good recovery performance for random missing in flat regions. However, most methods have difficulties in correctly recovering sharp discontinuities within missing areas. Our method is able to recover better geometrical structures as suggested by the error maps.

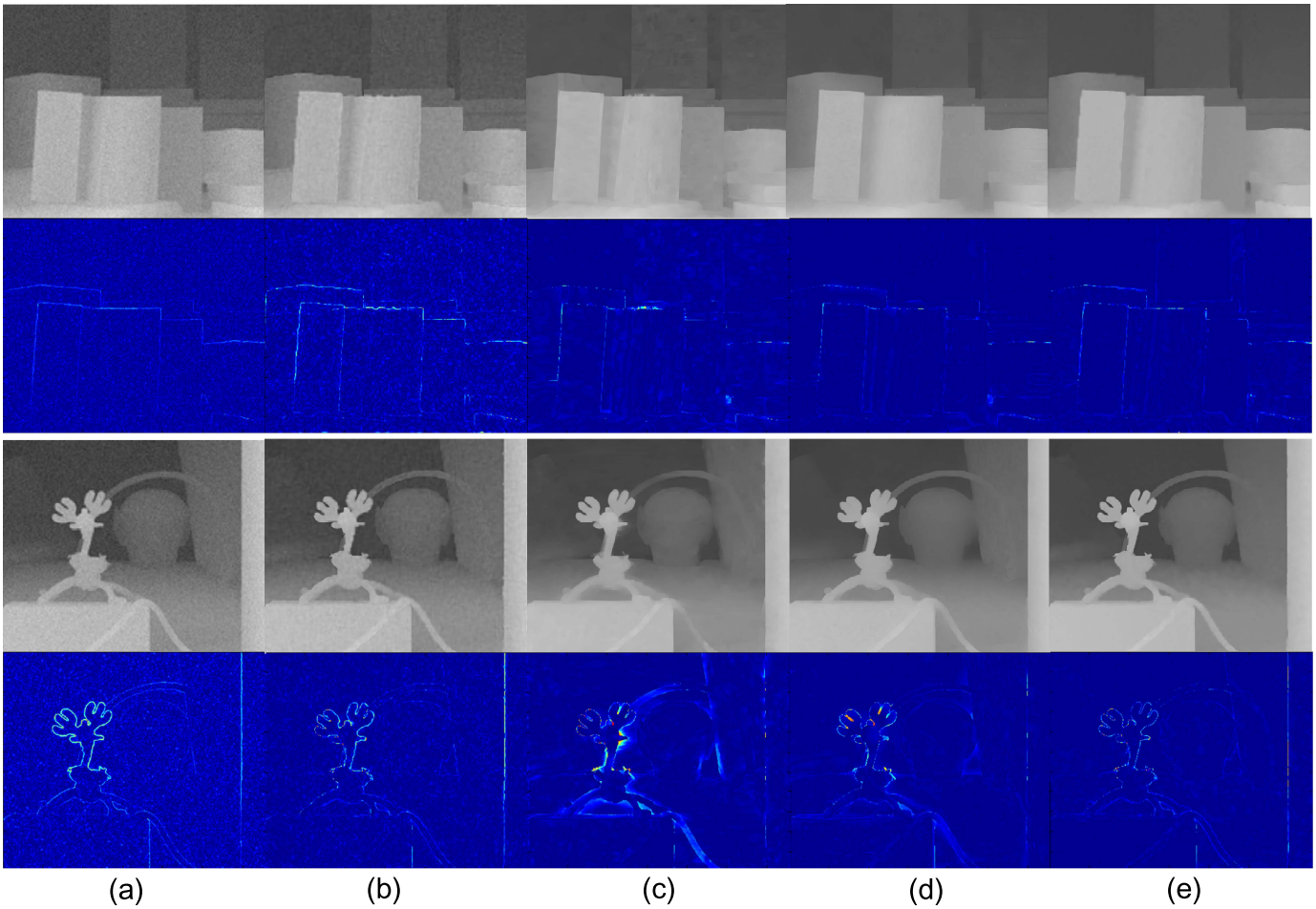

Fig. 8. Visual quality comparison for recovered depth maps from ToF-like degradation (8× upsampling with intense Gaussian noise): depth maps are recovered by (a) Bicubic (MAD: 3.51; 3.82), (b) IMLS [36] (MAD: 2.68; 2.86), (c) Edge [11] (MAD: 1.81; 2.40), (d) TGV [34] (MAD: 1.49; 1.75), and (e) our method (MAD: 1.15; 1.29). The two MADs for each method are for *Book* (first) and *Reindeer* (second), respectively.

Fig. 9. Depth maps contaminated by simulated Kinect-like degradations.

B. Experiments on Real Datasets

We apply our method on two types of depth sensors to achieve high quality depth recovery from the low quality sensor measurements.

1) ToF Depth Maps: We evaluate our method on datasets captured by two ToF-based RGB-D sensing systems: a) one is our depth-color camera rig and b) the other rig is constructed by Ferstl *et al.* [34].

a) Our RGB-D sensing system: Our depth camera rig is constructed by mounting a high resolution Point Grey Flea2 color camera on a PMD[vision] CamCube3 ToF depth camera. The ToF camera has a resolution of 200 × 200, and the resolution of color camera is set at 640 × 480 to obtain nearly the same field of view as the ToF camera. To compensate misalignment of different viewpoints, depth maps are warped to the viewpoint of the color camera using intrinsic parameters and extrinsic parameters for both cameras computed by the camera calibration module in the OpenCV library [52]. We first reject outliers using the associated amplitude images as confidence levels, and then rectify the intensity-dependent error with a pre-measured look-up table, similar to the approach in [45].

Two RGB-D pairs captured by our depth-color rig are shown in Fig. 11(a). We compare our method on these datasets with three representative methods: CLMF0 [32], JGF [33], and IMLS [36]. As the recovered depth maps shown in Fig. 11 (b)∼(d), CLMF0 and JGF tend to generate jaggy artifacts due to rich color textures and the discontinuity mismatch between the color images and depth maps, while IMLS brings a little bit over-smoothing around edges although we introduce a cross bilateral weighting for fair comparison.

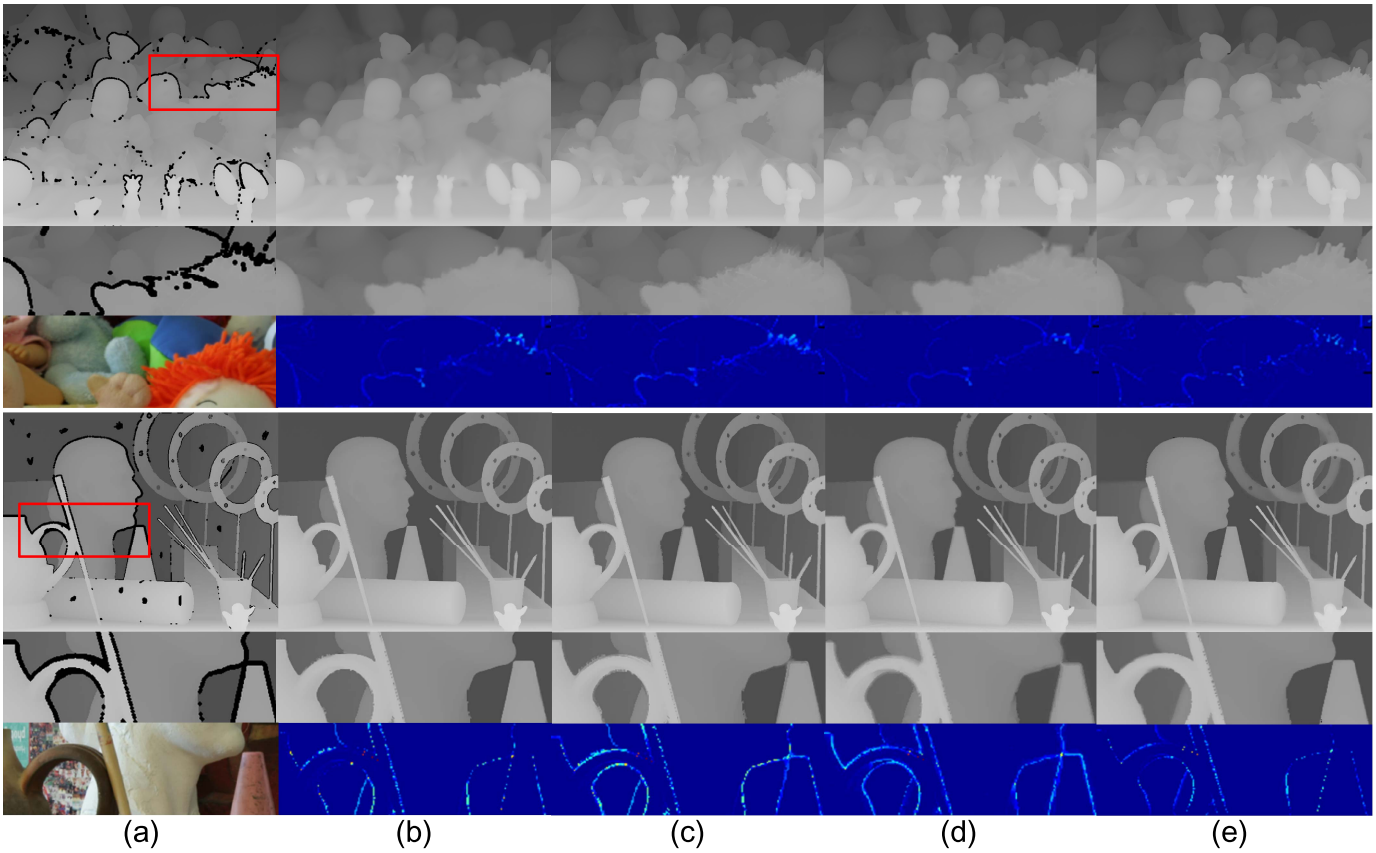

Fig. 10. Visual quality comparison for recovered depth maps from Kinect-like degradations: (a) degraded depth maps, depth maps recovered by (b) IMLS (MAD: 0.76; 0.90), (c) JBF [12] (MAD: 0.76; 0.84), (d) CLMF0 [32] (MAD: 0.74; 1.01), and (e) our method (MAD: 0.69; 0.58). The two MADs for each method are for *Dolls* (first) and *Art* (second), respectively. For visual inspection, regions highlighted by rectangles are enlarged, and the error maps are shown by subtracting between recovered depth and ground truth.

By inspecting the color-depth accordance, our method achieves quite promising recovery quality particularly around depth discontinuities.

b) ToFMark RGB-D sensing system: We also test on the *ToFMark* datasets [53] consisting of three RGB-D pairs, *Books, Shark, Devil*, with ground-truth depth maps. The depth maps are of size 120×160, and the intensity images are of size 610×810, suggesting approximately $6.25\times$ upsampling. Table IV presents quantitative results. The recovery error is measured by MAD in mm. Our method also obtains the lowest recovery error for all the three test cases compared with other seven classic or state-of-the-art methods. In Fig. 12, it is observed that depth maps recovered by Bicubic, IMLS, and CLMF0 still contain considerable amount of noise, while those recovered by TGV and the proposed method are much more clear. The TGV method in some cases introduces annoying artifacts in regions where the associated intensity image has rich textures, e.g., the bottom of cup and the edges of the book in *Books*. By closer inspection, TGV is superior in recovering slant planar surfaces that can be well characterized by the total variation minimization, while our method shows advantages in recovery high-order surfaces thanks to the powerful color-guided AR model.

2) Kinect Depth Maps: Microsoft Kinect is an integrated sensor array for natural user interaction, consisting of a depth camera and a color camera. The captured depth maps and color images are of size 640×480, and registered to the same viewpoint. We suppressed fake-color artifacts in color images by re-demosaicing the color images with an advanced method [54]. This experiment uses five RGB-D pairs, two of which are captured in our lab while the other three are from the NYU RGB-D dataset [55].

Fig. 13 shows depth recovery results for two RGB-D pairs: one is captured in our lab while the other is from the NYU RGB-D dataset [55]. It is observed that the depth maps contain lots of holes around depth discontinuities due to occlusions. This corresponds to the case of synthetic datasets with structural missing areas in Section VI-A. Note that methods designed for depth upsampling are not directly applicable to

TABLE IV
QUANTITATIVE DEPTH UPSAMPLING RESULTS FOR *ToFMark* DATASETS

	Bicubic	IMLS[36]	JBF[12]	Guided[13]	CLMF0[32]	JGF[33]	TGV[34]	Ours
Books	16.23	14.50	16.03	15.74	13.89	17.39	12.36	**12.25**
Shark	17.78	16.26	18.79	18.21	15.10	18.17	15.29	**14.71**
Devil	16.66	14.97	27.57	27.04	14.55	19.02	14.68	**13.83**

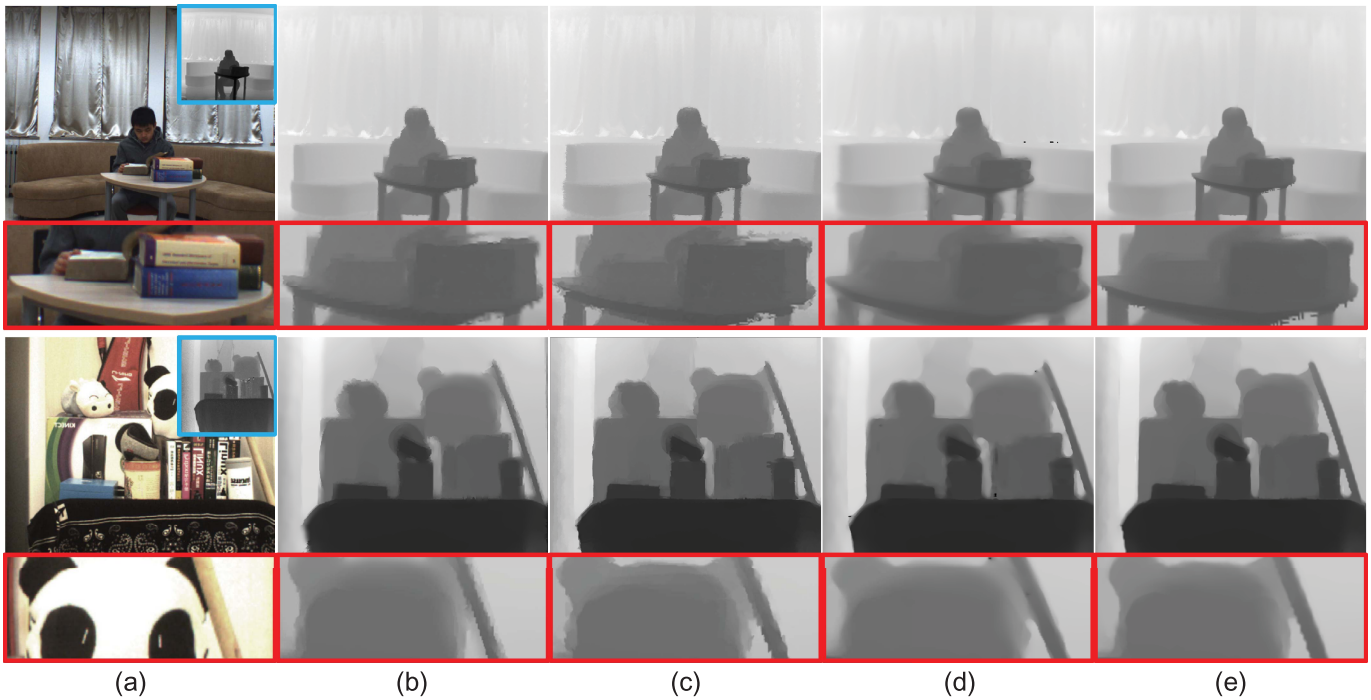

Fig. 11. Depth recovery results for our depth-color camera rig: (a) RDB-D pairs, recovered depth maps by (b) CLMF0 [32], (c) JGF [33], (d) IMLS [36], and (e) our method. Captured depth maps are overlaid on the color images to save space.

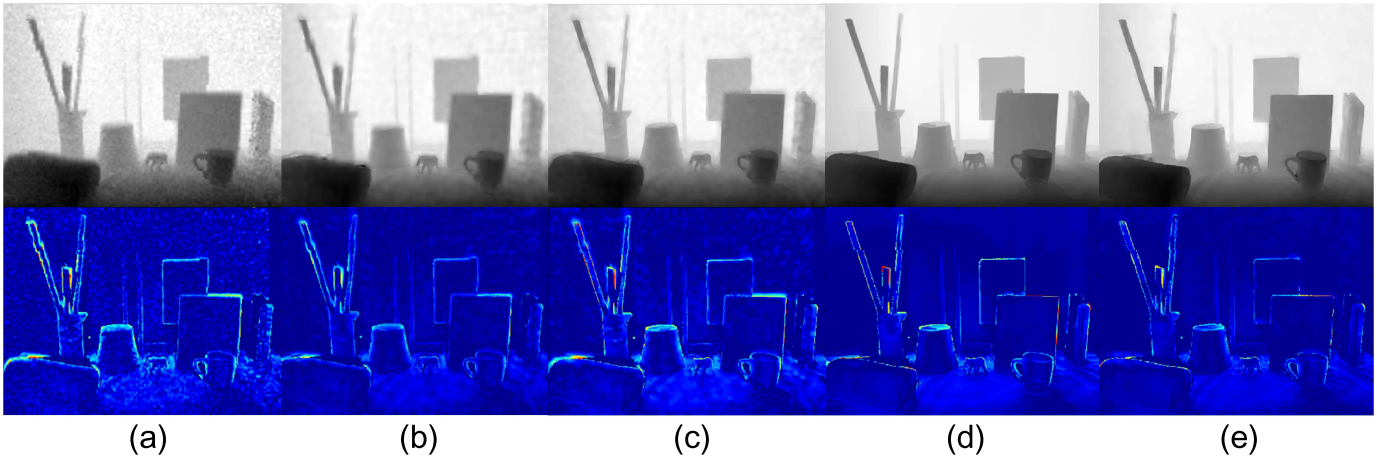

Fig. 12. Visual quality comparison on depth recovery for *Books* from *ToFMark* datasets: (a) Bicubic, (b) IMLS [36], (c) CLMF0 [32], (d) TGV [34], and (e) our method.

the recovery of Kinect degradations. Therefore, we compare with the IMLS [36] and JBF [12]. JBF produces annoying jaggy artifacts around depth discontinuities, while IMLS tends to smooth out sharp depth edges as in previous experiments. Our method outperforms the two methods, particularly in preserving prominent geometrical structures in depth maps.

The results in this section demonstrate that the proposed method is versatile for both the two types of mainstream depth cameras, and is applicable in various applications involved depth sensing.

C. Discussions and Future Work

1) Taxonomy-Based Discussions: Analogous to the taxonomy for stereo matching [1], most depth recovery methods can be classified into two categories: the global methods and local methods. Representatives of global methods include MRF-based methods [20], [23], [25], the edge-based NLM regularization [11], and our method. Most local methods use joint filtering schemes [13]–[16], [27], [29], [32], [33].

The global methods have very different behaviors from the local ones regarding the interactions between observed pixels and the latent ones. In global methods, there are closed-loop interactions between observed pixels and latent pixels. For example, in solving the MRF energy function, messages iteratively exchange between neighboring pixels. Our AR-based method also has a similar mechanism as all pixels including the observed ones should conform to the autoregression. We call this type of iterations as closed-loop prediction in global methods. On the contrary, in local filtering

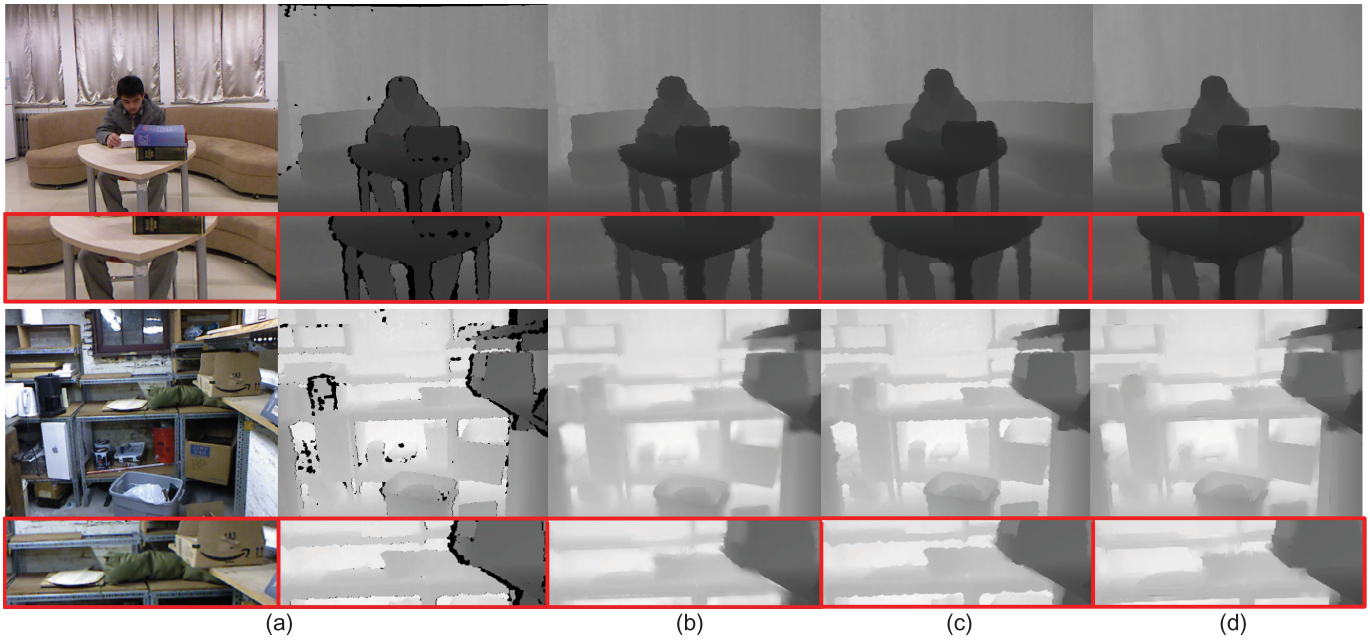

Fig. 13. Depth recovery results for Kinect datasets: (a) RGB-D pairs, recovered depth maps by (b) IMLS [36], (c) JBF [12], and (d) our method. The RGB-D pair in the top row is captured in our lab, while the other is from the NYU RGB-D dataset [55].

methods, latent pixels the one-hit prediction of neighboring observed ones, which is open-loop prediction. The closed-loop prediction generally achieves better performance than open-loop prediction with the same prediction scheme. However, in literature, local filtering schemes usually outperform global methods. The reason may be that prediction schemes in previous global methods are not so flexible and adaptive as those in local methods. Our work is a good example to show the superiority of global methods if the inherent predictors are properly designed.

However, the potential superiority of (closed-loop) global methods comes at the price of higher computational complexity than the (open-loop) methods, as their names imply. Generally, the running times of global methods are at the scale of $10^0 \sim 10^1$ minutes. For example, the MRF optmization in [26] needs 12.5 minutes for $4\times$ upsampling of a 200×200 depth map. The quadratic optimization based method in [11] takes nearly half minutes for $8\times$ upsampling to size of 1376×1088. There are two time-consuming parts in our methods: nonlocal filtering in the construction of AR predictors, and quadratic optimization as in [11]. Each has the similar computational complexity. The plain Matlab implementation of our method takes two minutes on average to super-resolve a low-resolution depth map to the resolution of 1376×1088, being independent of upsampling factors. A preliminary GPU version takes 2.8 seconds on average in a desktop (i5 3GHz CPU and 4GB RAM) with an NVIDIA Tesla 2050 GPU card. For the local filtering methods, the running time is at the scale of $10^0 \sim 10^1$ seconds. For example, the guided filter [13] (C++ implementation) takes about 0.48s to filtering a magepixel color image; and the CLMF method [32] takes about 0.50 seconds in matching a stereo pair of size of 384×288. Our purpose here is not to give a precise comparison, but to grasp the scale of required computation for these two types of methods. Clearly, we observe that the local filtering methods use far less computation than global methods. An interesting point is to develop approximate algorithms of global methods to enjoy both the high accuracy of global methods and low complexity of local filtering methods.

2) Future Work: Depth recovery for a single frame is relatively well-investigated over the past few years. The remaining challenges includes: 1) accurate recovery of shining and transparent regions; 2) good complexity-quality tradeoff (as discussed above), and 3) temporally-coherent depth video recovery (e.g., the flicking issue). The first two have not been seriously addressed in the literature. Regarding depth video recovery, there have been some work to consider spatial-temporal recovery of depth squences [25], [31], [56].

VII. Conclusion

This paper proposes a novel framework to recover depth maps from low quality measurements with various types of degradations. We show that depth maps are well described by the AR model if the AR predictors can adapt to the characteristics of depth maps. Based on this observation, we design pixel-wise adaptive AR predictors using both the depth map and the accompanied color image. The depth map is recovered by minimizing AR prediction errors subject to the observation consistency. We show that the proposed depth recovery method is equivalent to a linear system, and its stability is analyzed by the conditioning of the linear system. To achieve stable and accurate depth recovery, the parameters are adaptively set according to the local structures of the depth maps and the accompanied color images. Experiments demonstrate that our method achieves high quality depth recovery from low quality

versions with various degradation. Experiments on two real systems demonstrate that our method is versatile for various depth capturing systems such as ToF cameras and Kinect.

ACKNOWLEDGMENT

The authors would like to thank J. Park for providing recovered depth maps in [11] for comparison, and thank anonymous reviewers for their comments which help to significantly improve the paper.

REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, no. 1, pp. 7–42, 2002.

[2] 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.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.

[3] P. Thanusutiyabhorn, P. Kanongchaiyos, and W. Mohammed, "Image-based 3D laser scanner," in *Proc. Int. Conf. Elect. Eng./Electron., Comput., Telecommun., Inform. Technol.*, 2011, pp. 975–978.

[4] A. Kolb, E. Barth, R. Koch, and R. Larsen, "Time-of-flight cameras in computer graphics," *Comput. Graph. Forum*, vol. 29, no. 1, pp. 141–159, 2010.

[5] M. Lindner, A. Kolb, and K. Hartmann, "Data-fusion of PMD-based distance-information and high-resolution RGB-images," in *Proc. Int. Symp. Signals, Circuits Syst.*, vol. 1. Jul. 2007, pp. 1–4.

[6] S. A. Guomundsson, R. Larsen, H. Aanæs, M. Pardas, and J. R. Casas, "TOF imaging in smart room environments towards improved people tracking," in *Proc. IEEE Comput. Vis. Pattern Recognit. Workshops, (CVPRW)*, Jun. 2008, pp. 1–6.

[7] A. Riemens, O. Gangwal, B. Barenbrug, and R. Berretty, "Multi-step joint bilateral depth upsampling," in *Proc. Vis. Commun. Image Process.*, 2009, pp. 1–12.

[8] S. Lee and Y. Ho, "Joint multilateral filtering for stereo image generation using depth camera," in *The Era of Interactive Media*. New York, NY, USA: Springer-Verlag, 2013, pp. 373–383.

[9] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view RGB-D object dataset," in *Proc. Int. Conf. Robot. Autom.*, 2011, pp. 1817–1824.

[10] J. Yang, X. Ye, K. Li, and C. Hou, "Depth recovery using an adaptive color-guided auto-regressive model," in *Proc. 12th Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 158–171.

[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. (ICCV)*, Nov. 2011, pp. 1623–1630.

[12] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, p. 96, 2007.

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

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

[15] J. Dolson, J. Baek, C. Plagemann, and S. Thrun, "Upsampling range data in dynamic environments," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2010, pp. 1141–1148.

[16] F. Li, J. Yu, and J. Chai, "A hybrid camera for motion deblurring and depth map super-resolution," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[17] A. Zomet and S. Peleg, "Multi-sensor super-resolution," in *Proc. 6th IEEE Workshop Appl. Comput. Vis. (WACV)*, pp. 27–31, Dec. 2002.

[18] T. Prasad, K. Hartmann, W. Weihs, S. Ghobadi, and A. Sluiter, "First steps in enhancing 3D vision technique using 2D/3D sensors," in *Proc. Comput. Vis. Winter Workshop*, 2006, pp. 82–86.

[19] A. Linarth, J. Penne, B. Liu, O. Jesorsky, and R. Kompe, "Fast fusion of range and video sensor data," in *Advanced Microsystems for Automotive Applications*. Berlin, Germany: Springer-Verlag, 2007, pp. 119–134.

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

[21] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "High-quality scanning using time-of-flight depth superresolution," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPR)*, Jun. 2008, pp. 1–7.

[22] W. Hannemann, A. Linarth, B. Liu, and G. Kokai, "Increasing depth lateral resolution based on sensor fusion," *Int. J. Intell. Syst. Technol. Appl.*, vol. 5, no. 3, pp. 393–401, 2008.

[23] J. Lu, D. Min, R. Pahwa, and M. Do, "A revisit to MRF-based depth map super-resolution and enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 985–988.

[24] B. Huhle, S. Fleck, and A. Schilling, "Integrating 3D time-of-flight camera data and high resolution images for 3DTV applications," in *Proc. 3DTV Conf.*, May 2007, pp. 1–4.

[25] J. Zhu, L. Wang, J. Gao, and R. Yang, "Spatial-temporal fusion for high accuracy depth maps using dynamic MRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 899–909, May 2010.

[26] 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. (ECCV)*, 2012, pp. 71–84.

[27] B. Huhle, T. Schairer, P. Jenke, and W. Straßer, "Fusion of range and color images for denoising and resolution enhancement with a non-local filter," *Comput. Vis. Image Understand.*, vol. 114, no. 12, pp. 1336–1345, 2010.

[28] V. Garro, P. Zanuttigh, and G. Cortelazzo, "A new super resolution technique for range data," in *Proc. Associazione Gruppo Telecomunicazioni e Tecnologie dell Informazione*, 2009.

[29] Q. Yang, K. Tan, B. Culbertson, and J. Apostolopoulos, "Fusion of active and passive sensors for fast 3D capture," in *Proc. IEEE Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2010, pp. 69–74.

[30] D. Chan, H. Buisman, C. Theobalt, and T. Sebastian, "A noise-aware filter for real-time depth upsampling," in *Proc. Workshop Multi-Camera Multi-Modal Sensor Fusion Algorithms Appl. (M2SFA2)*, 2008.

[31] D. Min, J. Lu, and M. Do, "Depth video enhancement based on joint global mode filtering," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1176–1190, Mar. 2012.

[32] J. Lu, K. Shi, D. Min, L. Lin, and M. N. Do, "Cross-based local multipoint filtering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 430–437.

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

[34] 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. (ICCV)*, Dec. 2013, pp. 993–1000.

[35] S. Fleishman, D. Cohen-Or, and C. T. Silva, "Robust moving least-squares fitting with sharp features," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 544–552, 2005.

[36] 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.

[37] A. Dakkak and A. Husain, "Recovering missing depth information from Microsoft's Kinect," in *Proc. Embedded Vis. Alliance*, Boston, MA, USA, 2012.

[38] A. Maimone and H. Fuchs, "Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras," in *Proc. 10th IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Oct. 2011, pp. 137–146.

[39] Y. Berdnikov and D. Vatolin, "Real-time depth map occlusion filling and scene background restoration for projected-pattern based depth cameras," in *Proc. IETP Graph. Conf.*, vol. 2. 2011, pp. 121–126.

[40] 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 Recognit. (CVPR)*, Jun. 2013, pp. 1415–1422.

[41] A. Telea, "An image inpainting technique based on the fast marching method," *J. Graph. Tools*, vol. 9, no. 1, pp. 23–34, 2004.

[42] S. Matyunin, D. Vatolin, Y. Berdnikov, and M. Smirnov, "Temporal filtering for depth maps generated by Kinect depth camera," in *Proc. 3DTV Conf.*, May 2011, pp. 1–4.

[43] M. Camplani and L. Salgado, "Adaptive spatio-temporal filter for low-cost camera depth maps," in *Proc. Int. Conf. Emerg. Signal Process. Appl.*, Jan. 2012, pp. 33–36.

[44] M. Frank, M. Plaue, H. Rapp, U. Köthe, B. Jähne, and F. A. Hamprecht, "Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras," *Opt. Eng.*, vol. 48, no. 1, p. 013602, 2009.

[45] J. Zhu, L. Wang, R. Yang, J. E. Davis, and Z. Pan, "Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1400–1414, Jun. 2011.
[46] 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.
[47] Y. Zhang, D. Zhao, X. Ji, R. Wang, and W. Gao, "A spatio-temporal auto regressive model for frame rate upconversion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 9, pp. 1289–1301, Sep. 2009.
[48] J. Owen, B. Eccles, B. Choo, and M. Woodings, "The application of auto–regressive time series modelling for the time–frequency analysis of civil engineering structures," *Eng. Struct.*, vol. 23, no. 5, pp. 521–536, 2001.
[49] D. Liu, E. Sara, and W. Sun, "Nested auto-regressive processes for MPEG-encoded video traffic modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 2, pp. 169–183, Feb. 2001.
[50] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.
[51] (2013, Apr. 4). *Middlebury Datasets* [Online]. Available: http://vision.middlebury.edu/stereo/data/
[52] (2012, Dec. 1). *Open Source Computer Vision (OpenCV)* [Online]. Available: http://opencv.org
[53] (2013, Dec. 20). *ToFmark Datasets* [Online]. Available: http://rvlab.icg.tugraz.at/tofmark/
[54] D. Paliy, A. Foi, R. Bilcu, and V. Katkovnik, "Denoising and interpolation of noisy Bayer data with adaptive cross-color filters," *Proc. SPIE*, vol. 6822, p. 68221K, Jan. 2008.
[55] (2013, Dec. 22). *NYU Datasets* [Online]. Available: http://cs.nyu.edu/ silberman/ datasets/
[56] M. Camplani and L. Salgado, "Efficient spatio-temporal hole filling strategy for Kinect depth maps," *Proc. SPIE*, vol. 8290, p. 82900E, Feb. 2012.

Jingyu Yang (M'10) received the B.E. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2003, and the Ph.D. (Hons.) degree from Tsinghua University, Beijing, in 2009.

He has been with the faculty of Tianjin University, Tianjin, China, since 2009, where he is currently an Associate Professor with the School of Electronic Information Engineering. His research interests mainly include image/video processing, 3D imaging, and computer vision. He was selected into the program for New Century Excellent Talents in University from the Ministry of Education of China in 2011, and the Elite Scholar Program of Tianjin University in 2012.

Xinchen Ye received the B.S. degree from Tianjin University, Tianjin, China, in 2006, where he is currently pursuing the Ph.D. degree with the School of Electronic Information Engineering. His research interests include depth recovery and 3D imaging.

Kun Li (M'12) received the B.E. degree in communication engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2006, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, in 2011.

She has been with the faculty of Tianjin Uinversity, Tianjin, China, since 2011, where she is currently an Assistant Professor with the School of Computer Science and Technology. Her research interests include image/video processing, image-based modeling, dynamic scene 3D reconstruction, and multicamera imaging. She was selected into the Elite Scholar Program of Tianjin University in 2012.

Chunping Hou received the M.Eng. and Ph.D. degrees in electronic engineering from Tianjin University, Tianjin, China, in 1986 and 1998, respectively.

She was a Post-Doctoral Researcher with the Beijing University of Posts and Telecommunications, Beijing, China, from 1999 to 2001. Since 1986, she has been with the faculty of the School of Electronic Information Engineering, Tianjin University, where she is currently a Full Professor and the Director of the Broadband Wireless Communications and 3D Imaging Institute. Her current research interests include wireless communication, 3D image processing, and the design and applications of communication systems.

Yao Wang (M'90–SM'98–F'04) received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1983 and 1985, respectively, and the Ph.D. degree in electrical and computer engineering from the University of California at Santa Barbara, CA, USA, in 1990.

She has been with the Faculty of Electrical and Computer Engineering, Polytechnic School of Engineering, New York University, Brooklyn, NY, USA, since 1990. She has authored a textbook titled *Video Processing and Communications* (Prentice Hall, 2001). Her current research interests include video coding and networked video applications, medical imaging, and pattern recognition. She has served as an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. She was a recipient of the New York City Mayor's Award for Excellence in Science and Technology in the Young Investigator Category in 2000, the IEEE Communications Society Leonard G. Abraham Prize Paper Award in the Field of Communications Systems in 2004, the Oversea Outstanding Young Investigator Award from the National Natural Science Foundation of China in 2005, and the Yangtze River Scholar Award by the Ministry of Education of China in 2008.