IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 17, 2021, accepted August 30, 2021, date of publication October 1, 2021, date of current version October 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3117140

Depth Map Reconstruction and Enhancement With Local and Patch Manifold Regularized Deep Depth Priors

ABBAS K. ALI[1,2], PEYMAN ADIBI[1], AND MOHAMMAD-SAEED EHSANI[1]
[1]Department of Artificial Intelligence, Faculty of Computer Engineering, University of Isfahan, Isfahan 81746-73441, Iran
[2]Computer Center, University of Thi-Qar, Nasiriyah 64001, Iraq

Corresponding author: Peyman Adibi (adibi@eng.ui.ac.ir)

ABSTRACT The depth map captured by depth sensors (e.g., the time of flight (ToF) and Kinect) is often prone to low resolution, degradation, noise, and poor quality. This paper proposes a novel model for the robust depth estimation of RGB-D images through local and nonlocal manifold regularizations. The first stage called deep depth prior manifold (DDPM), is inspired partly by the deep depth prior (DDP) model, that is a deep convolutional neural network (CNN) integrated with a local manifold regularization term. The local neighboring relationships between depth pixels and color images are employed to promote smoothing in the results. The Laplacian Eigenmap technique used for local manifold modeling produces over-smooth depth map. To improve the quality of the reconstructed image, a nonlocal manifold modeling stage was suggested, where the similarity between the depth and the corresponding color image is determined by characterizing their matching aspects. These objectives are aggregated within an optimization problem. Moreover, to extract edges better considering visual nonlocal characteristics, the structured low-rank Hankel approximation was adopted to better eliminate depth degradations, and to extract highly promoted edges and sharp points. Three types of the degradations were handling in this work, containing undersampling, ToF-like, and Kinect-like degradations. Experimental results indicate that the proposed method outperformed the state-of-the-art restoration techniques on standard benchmark images, in terms of well-known criteria like PSNR.

INDEX TERMS Inverse problems, deep depth prior (DDP), manifold regularization, patch manifold, depth map, auto-encoder.

I. INTRODUCTION

In recent decades, many studies have been dedicated to estimate a depth map from a single image. The depth map plays an essential role in many image processing and computer vision applications such as the augmented reality (AR), 3DTV, 3D pose estimation, 3D scene analysis, 3D robot vision, and autonomous navigation [1]–[5].

The manifold modeling mechanism for depth estimation has not been analyzed thoroughly. According to the literature, the first study to use manifold models for depth estimation problems was conducted by Liu *et al*. in 2019 [1]. The goal of depth reconstruction and enhancement is to increase the usefulness of depth for more subjectively pleasing depth images for human viewing. There have been insufficient attempts at actual depth image estimation processes, the models of which are mostly ad hoc. Failing to increase the inherent contents of data, these processes include edge sharpening, noise reduction, filtering, interpolation and magnification, and pseudo coloring.

The depth information is the distance between a place where a camera is located and an object in the 3D scene. In these applications, the depth is captured with special cameras (e.g., Microsoft Kinect and ToF), which usually suffer from low resolution and high noise [4]. The previous studies sought to provide high-quality depth maps [6]. However, the real-world depth maps might be affected by different levels of degradation that can lead to low quality in comparison with color images due to the dissimilarity between a projector and a sensor [7]. This can cause random missing in flat

The associate editor coordinating the review of this manuscript and approving it for publication was Charalambos Poullis.

areas and the loss of depth information discontinuities. The depth map degradation may occur in different stages such as formation, transmission, and storage [8]. In other words, most methods of ToF depth map restoration can also be applied to the Kinect depth map restoration [3].

Although the estimation process aims to improve the quality of objects in the input depth map which is a degraded image by minimizing the artifacts, the input depth image is nearly corrupted by the missing information and noise texture needing additional information to achieve satisfactory performance. A serious challenge to this recovery process is the emergence of an ill-posed inverse problem. An important part of image restoration would be the ''prior'' which helps improve estimation and optimization in problems used even in deep learning or dimensionality reduction. The prior is also employed in image models. Many different models have been proposed to deal with the ill-posed optimization in computer vision problems by offering efficient solutions to many computer science tasks such as low-rank problems [2], independence of statistical elements [9], nonnegative matrix factorization [10], total variation (TV) [11], nonlocal similarity [12], [13], and video-like active recognition [14].

The proposed depth reconstruction model seeks to handle depth recovery on three typical depth map degradations, i.e. under-sampling degradation, Kinect-like degradation, and ToF-like degradation, by integrating deep learning and manifold modeling.

In fact, deep learning is used because it has yielded great performance [15], [16], especially in image restoration, and manifolds are used as proper models to represent depth maps.

Recently, deep convolutional neural networks (CNNs or ConvNets) have yielded satisfactory outputs, especially in super-resolution natural images such as the well-known analysis of SRCNN [17]. There is also a novel model representing an untrained deep image prior (DIP) [18], [19]. It has recently been introduced in computer vision and has proven to be the most interesting fully convolutional networks. This model can be employed to optimize untrained weight parameters (*i.e.* starting from a random weight $\boldsymbol{\theta}_0$) [19]. The reasons why a CNN can be used as a prior were given in [18]. Accordingly, the network resists ''bad'' solutions and descends much more quickly toward natural-looking images [20].

There are many methods that take advantage of low-dimensional manifolds in image recovery [21]. For instance, the paper reviewed by [22] developed a promising method called manifold modeling in embedded space (MMES). It is mainly based on the idea of interpreting the deep learning notions (*e.g.* in a deep image prior or a convolutional neural network) used in image restoration as patch manifold learning. The manifold can be used for regularization, whereas the local manifold is a smoothness regularization operator that utilizes the local relationships between the pixels of a depth image located nearby [1].

So far, the previous studies have mostly worked on depth estimation by using several image features offered no further systematic methods. For example, they did not exploit the image statistical models seriously. In other words, some approaches try to estimate the depth based on either depth images or color images. To solve this problem and eliminate the gaps in these methods, this study proposes a novel and powerful method for depth recovery from RGB-D images through deep learning frameworks with local and nonlocal manifolds. The proposed method exploits the relationships between a depth map and its corresponding color image. Therefore, the gaps between previous studies are addressed. This method was designed to be as simple as possible while having an essential ConvNet structure. The addition of a local manifold modeling term to the standard DIP framework with nonlocal manifold appears to be able to solve all the aforementioned issues. The depth estimation quality of the proposed method was analyzed successfully in comparison with the previous methods.

Introduced by Ulyanov *et al.* in [18], deep image prior (DIP) is among the most well-known unsupervised approaches. In their pioneering work, they empirically showed how the architecture of a deep CNN would be able to recover depth images more easily without the need for a fixed set of training examples. Therefore, this possibility was taken into account by adding manifold regularization terms. Furthermore, the manifold-based terms have been efficient in dealing with depth images [1]. The RGB and depth images were also employed to build weights by integrating DIP with manifold and modeling ideas of remedying various depth map degradations. Finally, the outputs of this deep depth prior manifold (DDPM) network are considered the inputs to the nonlocal manifold modeling module, a process which is described thoroughly in Section III.

The main contributions of this study are as follows:

1) The study provides a formulation to show the depth estimation task as a combination of deep learning and manifold learning to develop a powerful model leading to the best success of depth image reconstruction. Therefore, it is easy to apply the proposed model inn other image processing applications.

2) The study also develops a refined depth map with cleaner and sharper edges and preserves high-frequency parts due to using a CNN with a manifold and a patch manifold model. This model can exploit all the information of the prior from the depth image and the color image, a process which results in the robust depth recovery.

3) This study helps solve other problems in other studies such as jagged artifacts on edges.

Optimization techniques were employed to achieve good results considered the best of both areas. The architecture of the proposed model (DIP with manifold) for depth map restoration has not been completely explored yet; thus, this is the first study to use this combination for inverse-depth restoration. The rest of this paper consists of different sections. Section II reviews the literature and related works, whereas Section III presents the proposed method and describes the problem formulation. Experimental results

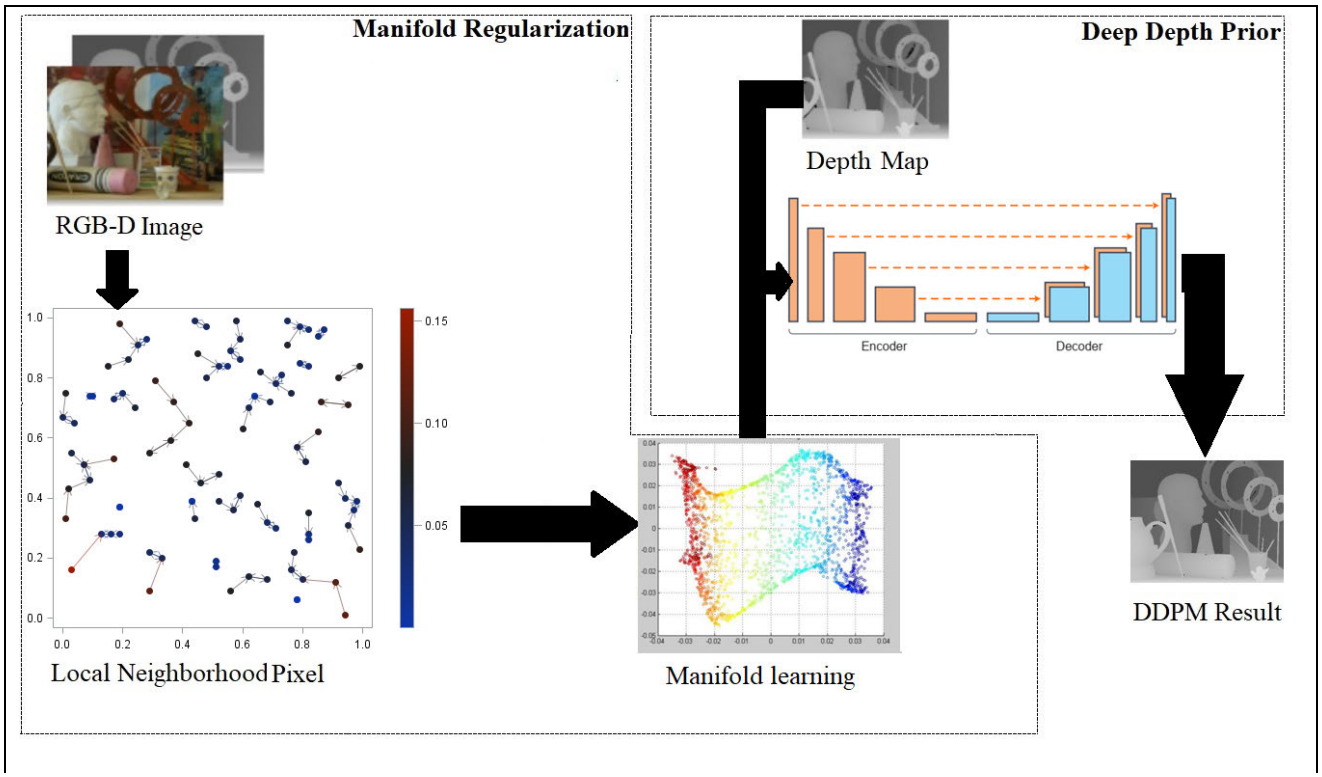


FIGURE 1. The deep depth prior local manifold (DDPM) modeling components.

are reported in Section IV. Finally, Section VI draws a conclusion.

II. RELATED WORKS

The depth map recovery is considered a popular and important field in computer vision tasks. Numerous methods and models of depth recovery have been proposed so far. Generally, the motivations and contributions of different studies can be summarized as follows: 1) Color and depth sensors are combined to develop a robust depth image. 2) The depth map can be restored by a single image. 3) Combination models are adopted simultaneously. This section reviews several related works containing the deep learning approaches and manifold learning ones.

A. DEPTH RECOVERY THROUGH DEEP LEARNING

The CNN method is employed for image restoration, especially while using a single image (in ill-posed conditions). It relies highly on the prior 3D shapes. Two recent CNN depth restoration models were proposed by Song and Kim [23] and Alhashim and Wonka [24]. In fact, Song and Kim [23] proposed a novel and simple model by utilizing two color-to-depth networks and employing the color image to develop a depth map. Afterwards, the latent space of the depth-to-depth network was utilized and compared with the ground truth depth to determine the loss. Alhashim and Wonka [24] proposed a similar network architecture (encoding-decoding) presenting a CNN for the depth map estimation, which obtains a high resolution image from a solid color image through transfer learning. Yielding better results, these methods are over-smooth as opposed to the other kinds of images such as ToF images. In [25] and [16], the contexts of global scenes were used for depth denoising. Li *et al.* [16] introduced a supervised deep convolutional network based on the filter structures and transferred the structural information from the prior guidance images to the noisy ones. Zhong *et al.* [25] proposed a deep neural network model to estimate a 3D face content through the single depth captured by a Kinect camera. They handled the lack of depth images by exploiting the bidirectional CycleGAN based on a generator for denoising and simulating noisy depths. The model was trained by synthetic depth images for real noise. However, supervised learning plays a key role in using ConvNet in image reconstruction functions [17], [26], [27].

B. DEPTH RECOVERY USING DIMENSIONALITY REDUCTION METHODS

Many methods benefit from the fusion of multiple depths to integrate the multiple degraded depth map estimations into a monocular depth map with higher quality, usually based on data reduction or idea embedment [1]–[5], [28]. For instance, Gu *et al.* [28] proposed a method for weight-based depth map enhancement by using the relationship between intensity and depth images in a conventional manner in two ways driven by learning tasks and image guidance. There are many methods that follow the filtering path such as joint bilateral

filter [29] and edge guided filter [30]. Liu *et al.* [5] developed a combination method that utilized the internal smoothness prior and external gradient consistency constraints within a graph domain for the depth super-resolution. Yang *et al.* [4] proposed a model to handle big holes and huge noise with an optimization framework for the color-guided depth map restoration.

Liu *et al.* [1] introduced a novel model combining local and nonlocal manifolds. They used the local manifold for regularization and the non-local manifold for 3D thresholding on the manifold. They achieved good but over-smooth results. In the current study, this weakness is removed by integrating local manifolds with the deep image priors, a process which yields better results and works with all types of datasets. Dong *et al.* [2] proposed joining the weighted prior of the total variation to the guided color autoregressive (AR) method in [3] as well as the low-rank property to consider the similarity of both local and nonlocal manifolds for information guidance utilized to compute the patch-wise and pixel-wise similarities. Those are all vastly used in the natural image reconstruction. Both AR and TV methods were utilized in [6]. However, the AR method promotes smoothness, whereas the TV method improves the piecewise constant. In this paper, an unsupervised method was employed to develop a combined model for depth reconstruction in order to handle the degradation and noise problems by introducing a multi-shape consistency constraint. This is the distinct difference between the reviewed studies and the current one.

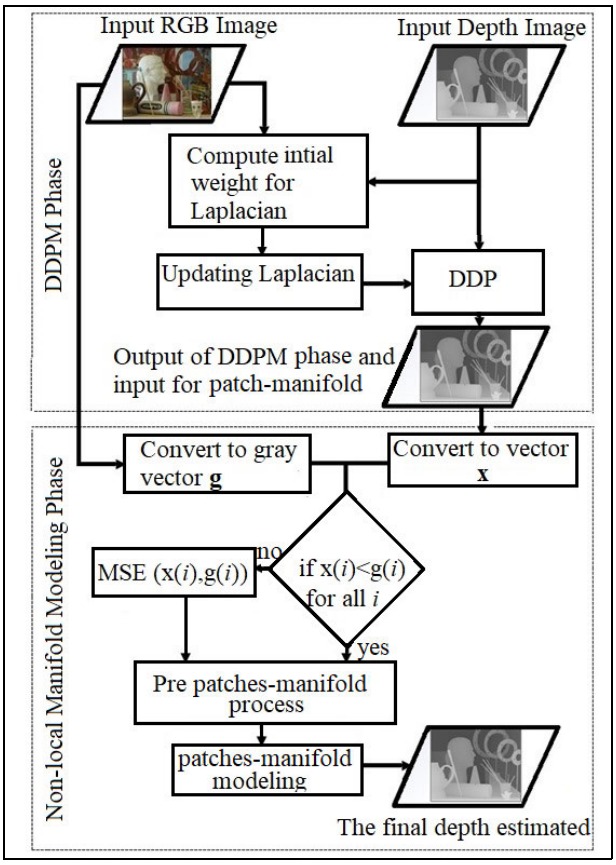


FIGURE 2. Flowchart of the proposed model.

III. PROPOSED METHOD

This study aims to exploit all image details(both RGB and depth) through different methods for the depth reconstruction to recover a proper, accurate, clean, and sharp image from a degraded and noisy observation.

For this purpose, an RGB image was first used with depth image for depth map reconstruction within an autoencoder neural network. Moreover, a CNN was adopted in the proposed approach, for it plays a key role in image recovery with manifold regularization, which yields accurate results, especially in image estimation. Regarding different types of manifolds, the local manifold used in this study was modeled on the Laplacian Eigenmap technique. All pixels of both RGB and depth images were utilized to build weights forming a Laplacian matrix for the local manifold. However, this approach produced an over-smooth depth map. Therefore, it still needs to be modified to handle the problem. For this purpose, a nonlocal manifold was suggested, for it could help obtain the depth with more accuracy and without noise. The input of this part is the output of the deep depth prior manifold (DDPM).

The DDPM is a fixed CNN with a mapping function $f_{\boldsymbol{\theta}}$, the weights of which are shown as $\boldsymbol{\theta}$. Its input $\mathbf{x}$ is a depth image vector. The CNN generator $f_{\boldsymbol{\theta}}$ is initialized with a set of random weights that are iteratively optimized by means of standard gradient-based algorithms. An approximation ($\mathbf{x}^*$) of the target solution ($\mathbf{x}$) is then computed as $f_{\theta*}(\mathbf{x})$, in which $\boldsymbol{\theta}^*$ is a solution obtained by applying the early stopping procedure to the involved iterative optimization scheme.

Before the depth image was treated, the input depth was made more reliable through the mean square error minimization on each pixel. After that, the input image of the nonlocal manifold was considered $s \times s$ patches. The edges obtained from the proposed method were sharper and cleaner than those of other approaches.

This study provided a formulation for depth map estimation which preserves the high-frequency parts, and therefore results in cleaner and sharper edges. Furthermore, a CNN was adopted for modeling the local manifold to reconstruct the depth map from RGB-D data, the results of which still had noise and blurring edges. Thus, furthermore a nonlocal manifold modeling module was proposed in order to provide more reliable results.

In a mathematical representation, a typical depth degradation method can be defined as below:

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

where $\mathbf{y} \in \mathbb{R}^v$ is the blurred image (represented as a column vector), and $\mathbf{H}$ is the matrix of a degradation operator on depth map images for missing pixels, downsampling, and blurring. Moreover, vector $\mathbf{x} \in \mathbb{R}^v$ contains the given depth information and is taken into account along with its mapping $\mathbf{x}$: $\boldsymbol{\Omega} \rightarrow \mathbb{R}$, in which Ω is the sampling grid of the spatial

TABLE 1. Depth map undersampling with differences between DDPM and DDPM with patch manifold in PSRN (dB).

Image/ rate / Method	Art				Book				Dolls			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
DDPM	48.22	41.01	37.01	34.35	51.62	49.37	45.06	40.50	52.21	48.26	46.01	44.06
DDPM +Patch manifold	50.23	43.32	38.22	35.21	53.98	51.32	46.90	41.69	54.61	50.67	47.10	45.50
Image/ rate / Method	Moebius				Reindeer				Laundry			
	2×	4×	8×	16×	2×	4×	8×	16×	2×	4×	8×	16×
DDPM	53.59	49.50	45.74	40.06	45.90	43.15	39.52	37.88	48.18	45.00	43.78	36.79
DDPM +Patch manifold	55.90	51.89	47.85	42.41	48.13	44.06	41.61	38.02	49.93	45.83	44.09	39.01

domain reshaped to a v-dimensional column vector. In addition, $\mathbf{n}$ denotes the noise, which is typically an additive white Gaussian noise (AWGN) with a mean of zero and a standard deviation of σ [31]. Yielded by $\mathbf{g}$: $\Omega \rightarrow \mathbb{R}$ that is the intensity of the counterpart color image, the guidance image $\mathbf{g} \in \mathbb{R}^v$ is also considered. However, in the image restoration, many models consider Equation (1), which was proposed to supply a recovered clean image $\mathbf{x}$ of the required estimate. In the current study, this equation shows the depth map, which is an inverse and inherently challenging ill-posed problem in restoration of $\mathbf{x}$ from $\mathbf{y}$. Thus, an additional prior is required to make the problem tractable.

Inspired partly by the deep depth prior (DDP) with the local manifold model [1], deep learning was employed in this study. Afterwards, the results were integrated into those of the nonlocal manifold model. It is also a prior model, which provides low-dimensional parameterization. According to Figure 2, these two models were coupled together to regularize the ill-posed problem of the depth map. The reconstruction of the depth map was considered smooth on manifolds, *i.e.* they had smooth regions separated with sharp edges. The reconstructed image was represented as the observation. It is observed that the reconstruction performance was efficiently improved when a DDP was used in the local manifold regularization. Moreover, it is advisable that the designed regularization be general enough to be applicable for many recovery tasks [8].

The following subsections define a manifold with a DDP based on a DIP by adding a regularization operator to handle the overfitting problem in the depth image restoration tasks. A nonlocal manifold term is also defined based on the manifold model of the embedded space [22] to address denoising and sharpness refinements of RGB-D images.

A. MANIFOLD WITH DEEP DEPTH PRIOR (DDP)

This section focuses on the deep depth prior (DDP) framework employed for image restoration and utilized on the depth map called the deep depth prior (DDP) in a bid to make it more robust by adding the local manifold regularization operator improving the solution. The local parameterization was used for this purpose.

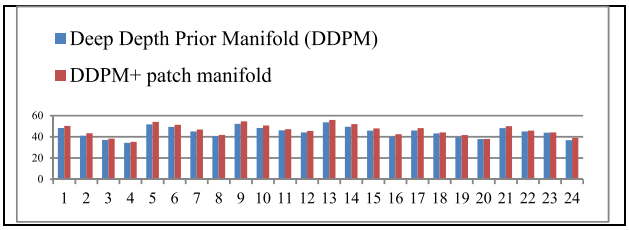


FIGURE 3. DDPM compared to DDPM with patch manifold in term of PSNR (dB) with 24-cases of rate.

Based on the nonlinear image processing approaches, the local manifold ($\mathfrak{M}$) can be employed to reflect the geometric properties of an image in which the geometry of the depth map $\mathbf{x}$ can favor some local relationships between local neighbors in Ω [6]. This model depends on the lifting function δ_f defined as below:

$$\delta_f(i) = (i, x(i)) \in \mathfrak{M} = \{i, x(i) | i \in \Omega\} \subset \Omega \times \mathbb{R} \subset \mathbb{R}^3 \tag{2}$$

The local lifting is similar to the bilateral filter [29] based on both the geometric feature and the value of each pixel, whereas d refers to an embedded surface of the depth map on a feature space with a dimension of $d = 3$.

To reflect the similarity between points on the local manifold, the weights of the data neighborhood graph are defined as the diffusion kernel as below:

$$w(i,j) = exp^{(\frac{\|i-j\|_2^2}{\sigma_p})} exp^{(\frac{\|x(i)-x(j)\|_2^2}{\sigma_d})} \times exp^{(\frac{\|r(i)-r(j)\|_2^2}{\sigma_r})} \quad \forall i \in \Omega,\ j \in \mathcal{N}(i) \tag{3}$$

where $\mathcal{N}(i)$ indicates the local neighborhood for pixel i. For every pixel z that is not in the neighborhood $\mathcal{N}(i)$, the corresponding weight is $w(i,z) = 0$. The geometric measurement is $\|i-j\|_2^2$ for the distance between pixels i and j. The photometric measurement representation to indicate the distance between pixel i and j in depth image is $\|x(i) - x(j)\|_2^2$. Moreover, $\|r(i) - r(j)\|_2^2$ is employed to determine the pixel value and compute the distance between points i and j in the color image. The hyper-parameters σ_p, σ_x, and σ_r are used to control the sensitivity of weights to these three

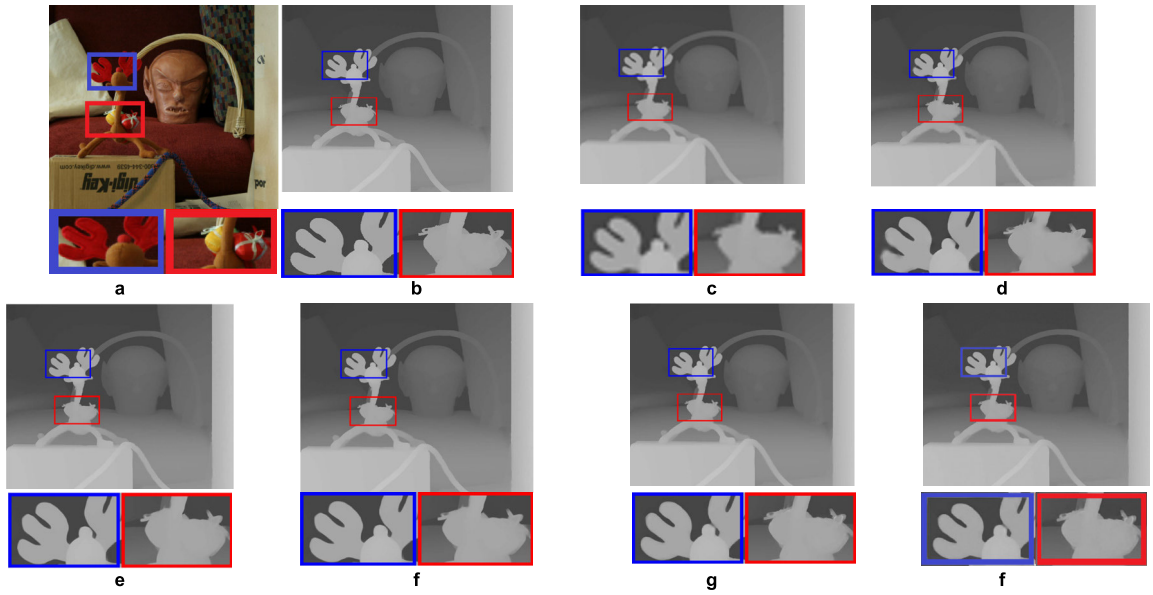


FIGURE 4. The comparison of estimation methods on reindeer image at upsampling degradation with rate 8×. (a) RGB, (b) the ground truth, (c) bicubic, (d) AR, (e) LN, (f) RCG, (g) manifold, and (h) the proposed method.

distances, respectively. Based on the matrix of diffusion kernels $\mathbf{W}$ with elements $w(i, j)$ from equation (3), the Laplacian matrix is computed as below:

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \tag{4}$$

where $\mathbf{D}$ represents a diagonal matrix of the node degrees with i'th diagonal element $\sum_j w(i, j)$. The depth recovery can be regarded as an optimization problem with the following objective function:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \gamma \mathbf{x}^{\mathbf{T}} \mathbf{L} \mathbf{x} \tag{5}$$

The second term plays the role of a manifold regularization term, and its minimization is equivalent to the smooth depth map on the manifold represented by Laplacian $\mathbf{L}$. The positive scalar γ balances the strength of the regularization. It is small only if the depth image $\mathbf{x}$ has close values on vertices i and j when the edge (i, j) has a large weight, or if the weight of the edge $w(i, j)$ is small. The minimization problem (5) has a closed form solution as:

$$\mathbf{x}^* = (\mathbf{H}^{\mathbf{T}}\mathbf{H} + \gamma \mathbf{L})^{-1} \mathbf{H}^{\mathbf{T}} \mathbf{y} \tag{6}$$

It is also possible to consider other metrics [32], [33] based on the Laplacian matrices, which are symmetric. First we build an orthogonal basis through the Eigen-Decomposition of Laplacian into a set of eigenvectors ($\mathbf{U} = \{\mathbf{u}_i\}$ $\mathrm{i} = 1, \ldots, n$), in which $\mathbf{U}$ refers to the eigenvector matrix with $\mathbf{u}_i$'s showing columns, and the real and non-negative eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ constituting the diagonal matrix $\mathbf{\Lambda}$ with λ_i's representing diagonal elements that indicate frequency in spectral graph theory:

$$\mathbf{L} = \mathbf{U}\Lambda\mathbf{U}^{\mathbf{T}} = \mathbf{U} \begin{bmatrix} \lambda_1 & 0\,0 & \ldots.0 \\ 0 & \lambda_2\,0 & \ldots.0 \\ \ldots & \ldots & .. \\ 0 & \cdots & \ldots\lambda_n \end{bmatrix} \mathbf{U}^{\mathbf{T}} \tag{7}$$

The manifold regularization term in (5) can be shown in the spectral domain as below:

$$\mathbf{x}^T \mathbf{L} \mathbf{x} = \boldsymbol{\alpha}^T \mathbf{\Lambda} \boldsymbol{\alpha} = \sum\nolimits_k \lambda_k \alpha_k^2 \tag{8}$$

where $\boldsymbol{\alpha} = \mathbf{U}^{\mathbf{T}}\mathbf{x}$ is the graph Fourier transform (GFT) coefficients, $\mathbf{U}$ and $\mathbf{\Lambda}$ are the spectral bases and eigenvalues derived from $\mathbf{L}$. Its mean that $\mathbf{x}^T\mathbf{L}\mathbf{x}$ is the sum of the squared graph Fourier transform coefficients α_k^2 scaled with frequencies (eigenvalues) λ_k. This study tried to estimate the solution to the ill-posed problem given in (5) by combining the manifold regularized restoration based on [1], [34].

Inspiring from various state-of-the-art deep image recovery networks [31], [35], the optimization problem can be defined as follows:

$$\theta^* = \arg\min_{\boldsymbol{\theta}} \mathrm{E}(\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}), \mathbf{x}_{\mathbf{0}}) \tag{9}$$

where $\mathbf{x}^* = \mathbf{f}_{\boldsymbol{\theta}^*}(\mathbf{x})$ is the recovered image, the image $\mathbf{x}$ is used to compute the task-dependent loss $\mathrm{E}(\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}), \mathbf{x}_{\mathbf{0}})$ which is selected as MSE loss for the used learning procedure, and $\mathbf{f}_{\boldsymbol{\theta}}$ represents the deep learning architecture with parameter $\boldsymbol{\theta}$. The empirical information is also available to the reconstruction process, which is the noisy image $\mathbf{x}_0$.

The method proposed in [18] uses not only the image space but also certain parameters of a parameterized space for image restoration. For the posterior used in image degradation modeling, the likelihood and the prior are also necessary. Without a prior, the resultant degradation depth map will not be affected by the domain knowledge. According to Figure 1, the local manifold regularization part can be added. The proposed combination of CNN-based DDP and local manifold modeling outperformed the method introduced in [19]. The clean depth image $\mathbf{x}^*$ was obtained from the degraded depth map $\mathbf{y}$ as $\mathbf{x}^* = \mathbf{f}_{\boldsymbol{\theta}^*}(\mathbf{x})$ in which $\mathbf{f}_{\boldsymbol{\theta}}$ represents a deep neural

network with weights $\boldsymbol{\theta}$ [11]. The minimization problem of DDP can be extended by using a manifold regularization $\mathbf{R}(\mathbf{x})$ term as below:

$$\boldsymbol{\theta}^* = \arg\min_{\theta} \|\mathbf{H}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}) - \mathbf{y}\|_2^2 + \gamma \mathbf{R}(\mathbf{x}) \tag{10}$$

which is rewritten through a Laplacian-based regularization term as below:

$$\boldsymbol{\theta}^* = \arg\min_{\theta} \|\mathbf{H}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}) - \mathbf{y}\|_2^2 + \gamma \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x})^T \mathbf{L}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}) \tag{11}$$

This optimization problem is equivalent to:

$$\begin{aligned} &\arg\min_{\boldsymbol{\theta},\mathbf{a}} \|\mathbf{H}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}) - \mathbf{y}\|_2^2 + \gamma \|\mathbf{a}\|_2^2 \\ &\text{subject to } \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}) = \mathbf{a} \end{aligned} \tag{12}$$

where constraint $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}) = \mathbf{a}$, denotes the image vector with a_i being the ith element of $\mathbf{a}$. If the constraint is assumed to be always satisfied, a_i is a derivative computed on the ith pixel for $i = 1, \ldots, v$, whereas v denotes the whole image size. To solve this problem, the nonconvex optimization framework for depth restoration was utilized in [11], [19]. The standard derivation for the minimization of the problem (12) read [36]:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}, \mathbf{a}, \boldsymbol{\gamma}_a) = &\|\mathbf{H}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}) - \mathbf{y}\|_2^2 + \gamma \|\mathbf{a}\|_2^2 \\ &+ \frac{\beta_a}{2} \|\mathbf{R}(\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x})) - \mathbf{a}\|_2^2 \\ &+ \boldsymbol{\gamma}_a^{\mathrm{T}} (\mathbf{R}(\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x})) - \mathbf{a}) \end{aligned} \tag{13}$$

where βa is a positive scalar, called penalty parameter, and $\boldsymbol{\gamma}_a$ is the Lagrangian-Coefficient associated with the constraint $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}) = \mathbf{a}$. According to the ADMM framework [36], itssaddle point can be determined by minimizing the primal variables $\boldsymbol{\theta}$ and $\mathbf{a}$, auxiliary variable, and $\boldsymbol{\gamma}_a$ which is the dual variable when the variables involved are properly initialized. Thus, the t-th iteration of the optimization algorithm [19] is defined as below:

$$\begin{aligned} \boldsymbol{\theta}^{t+1} = &\arg\min_{\boldsymbol{\theta}} \|\mathbf{H}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}) - \mathbf{y}\|_2^2 \\ &+ \frac{\beta_a}{2} \left\| \mathrm{R}(\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x})) - \mathbf{a}^t + \frac{\boldsymbol{\gamma}_a^t}{\beta_a} \right\|_2^2 \end{aligned} \tag{14}$$

$$\begin{aligned} \mathbf{a}^{t+1} = &\arg\min_{\boldsymbol{a}} \gamma \|\mathbf{a}\|_2^2 \\ &+ \frac{\beta_a}{2} \left\| \mathbf{a}^t - \mathrm{R}\left(\mathbf{f}_{\boldsymbol{\theta}^{t+1}}(\mathbf{x})\right) + \frac{\boldsymbol{\gamma}_a^t}{\beta_a} \right\|_2^2 \end{aligned} \tag{15}$$

$$\boldsymbol{\gamma}_{\mathbf{a}}^{\mathbf{t+1}} = \boldsymbol{\gamma}_{\mathbf{a}}^{\mathbf{t}} + \boldsymbol{\beta}_{\mathbf{a}}(\mathrm{R}\left(\mathbf{f}_{\boldsymbol{\theta}^{\mathbf{t+1}}}(\mathbf{x})\right)) - \mathbf{a}^{\mathbf{t+1}} \tag{16}$$

Problem (14) is solved inexactly by applying a prefixed number of iterations of a gradient-based method. In particular, the ADMM framework was adopted [36]. The numerical gradient is determined by employing the automatic differentiation provided by Pytorch with respect to the variable $\boldsymbol{\theta}$ [37]. Evidently, the optimization problem is similar to the one solved in the classical DIP framework (9). In this case, $\mathrm{R}\left(\mathbf{f}_{\boldsymbol{\theta}^{t+1}}(\mathbf{x})\right)$ is forced to be closed to $\mathbf{a}^{\mathrm{t}} - \frac{\boldsymbol{\gamma}_{\mathrm{a}}^t}{\beta_{\mathrm{a}}}$.

Problem (15) is separable and can be solved in a closed form by using the 2D L_2-norm, which is a proximity operator to the v components of $\mathrm{R}\left(\mathbf{f}_{\boldsymbol{\theta}^{t+1}}(\mathbf{x})\right) + \frac{\boldsymbol{\gamma}_{\mathrm{a}}^t}{\beta_{\mathrm{a}}}$. In the implementation process, the local manifold was selected as a regularization parameter.

Adding the manifold regularization term to the deep depth prior framework can help improve the quality of the resultant restored depth images by taking the underlying data manifold and the data of intrinsic dimensionality into account as well as dealing with the curse of dimensionality in the problem. Algorithm 1 shows the steps of the proposed method called the deep depth prior manifold (DDPM).

Algorithm 1 Deep Depth Prior Manifold (DDPM)

Input: $\mathbf{g}, \mathbf{y} \in \mathbb{R}^n$ (RGB image and corrupted depth image); $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x})$ (CNN); and t_{max} (the maximum number of iterations).

Output: $\boldsymbol{\theta}^*$(optimal learned network parameters); and $\mathbf{x}^*$ (the final depth map)

Initialize γ (regularization coefficient) $\boldsymbol{\theta}^{(0)}$ (random initial network weights), $\mathbf{x}^{(0)} \leftarrow$ initial depth image, $t \leftarrow 1$, and $\sigma_p, \sigma_x, \sigma_r$

Compute the matrices $\mathbf{W}$ and $\mathbf{L}$ according to (3) and (4)

While not converged and $t \leq t_{max}$

1. Update $\boldsymbol{\theta}^{(t+1)}$ based on $\mathbf{x}^{(t-1)}$ according to (14)
2. Update $\boldsymbol{\gamma}^t$ according to (16)
3. $t \leftarrow t + 1$

End while

$\boldsymbol{\theta}^* = \boldsymbol{\theta}^{(t-1)}$, $\mathbf{x}^* = \mathbf{x}^{(t-1)}$

B. LOW-DIMENSIONAL PATCH MANIFOLD PRIOR

The nonlocal computations introduced in [6] for the structure of the depth map image exploited through the previously obtained formulation can improve the quality of the restored image. They can also be used for image denoising [38]. This subsection uses a nonlocal manifold inspired partly by the MMES [22] in the previous study as the last stage of the proposed algorithm shown in Figure 2. The Hankel structured framework [39] was used with an MMES for the image restoration applications such as in-painting and super-resolution. The optimization problem related to Equation (1) can be formulated for the patch manifold through the following problem:

$$\begin{aligned} &\min_{\mathbf{x}} \left\| \mathbf{Y} - \mathcal{G}(\mathbf{X}) \right\|_2^2 \\ &s.t.\, \mathcal{H}(\mathbf{X}) = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_P] = \mathbf{H} \\ &\quad \mathbf{f}_p \in \mathcal{M}_{\mathbf{k}} \text{ for } p = 1, 2, \ldots, \mathrm{P} \end{aligned} \tag{17}$$

where $\mathbf{Y} \in \mathbb{R}^{\mathrm{J}\times\mathrm{Q}}$ is the observed corrupted image (J and Q are the dimensions of image $\mathbf{Y}$), and $\mathbf{X} \in \mathbb{R}^{\mathrm{I}\times\mathrm{Q}}$ is the estimated image (I and Q are the dimensions of image $\mathbf{X}$), whereas $\mathcal{G}$: $\mathbb{R}^{\mathrm{I}\times\mathrm{Q}} \rightarrow \mathbb{R}^{\mathrm{J}\times\mathrm{Q}}$ is a linear operator representing the observation system and the degradation operator on depth map images such as missing pixels or blurring. Furthermore, $\mathcal{H}$: $\mathbb{R}^{\mathrm{I}\times\mathrm{Q}} \rightarrow \mathbb{R}^{\mathrm{D}\times\mathrm{P}}$ is padding for a Hankelization operator [22] with a sliding window of size $(\mathrm{S}_1, \mathrm{S}_2, \ldots, \mathrm{S}_{\mathrm{Q}})$.

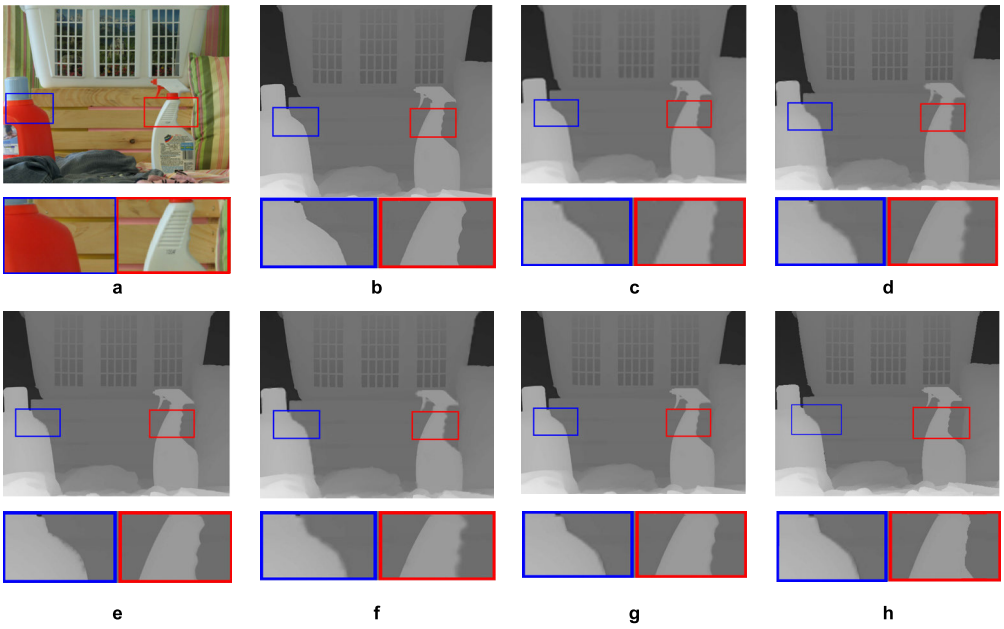


FIGURE 5. The comparison of estimation methods on laundry image at upsampling degradation with rate 8×. (a) RGB, (b) the ground truth, (c) bicubic, (d) AR, (e) LN, (f) RCG, (g) manifold, and (h) the proposed method.

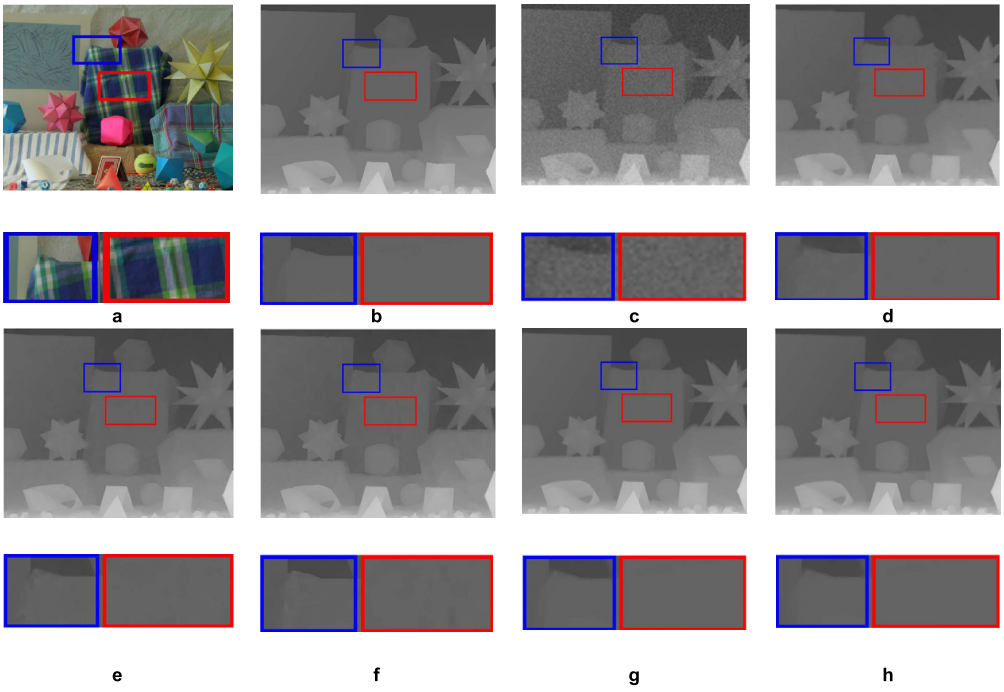


FIGURE 6. Comparison between methods of ToF-like degradation on the dataset Moebius (a) the RGB, (b) the ground truth, (c) bicubic, (d) AR, (e) LN, (f) RCG, (g) manifold, and (h) the proposed method.

Each vector $\mathbf{f}_p$ was obtained from S_q, $\mathrm{q} = 1, 2, \ldots, \mathrm{Q}$, on image $\mathbf{X}$ [22] in the proposed method ($Q = P$) as shown in Figure 2. In addition, each vector $\mathbf{f}_p$ is a point in a k-dimensional manifold $\mathcal{M}_k$ embedded in the D-dimensional Euclidean space with $k \leq D$. The linear operator $\mathcal{G}$ can vary in different tasks. In this study, $\mathcal{G}$ was considered an AWGN noise (as mentioned in Section III) with three types of degradation called undersampling, ToF-like, and Kinect-like degradations.

In the practical experiments, a pixel value in the guidance image was observed to be greater than its value in the depth image. Therefore, when the opposite was found, the pixel was infected as usual. According to Figure 2, it is possible to modify this issue through thresholding before the patch

TABLE 2. Quantitative comparison of the performance [in PSNR (dB)] (up-sampling without noise).

Method	rate	Art	Book	Dolls	Laundry	Mobius	Reindeer	Average
Bicubic	2 ×	39.92	47.75	48.94	44.05	48.98	42.49	45.36
AR	2 ×	40.49	46.33	49.05	45.61	49.16	43.67	45.72
RCG	2 ×	40.79	46.39	46.59	44.84	47.11	43.35	44.85
LN	2 ×	41.22	49.36	49.36	45.61	50.29	43.80	46.61
Manifold	2 ×	47.56	**56.61**	53.19	50.84	54.28	49.09	51.93
Ours	2 ×	**50.23**	53.98	**54.61**	**50.93**	**55.90**	**49.13**	**52.46**
gain	2 ×	2.67	-2.63	1.42	0.09	1.62	0.04	0.53
Bicubic	4 ×	36.38	44.05	45.62	40.50	45.63	39.16	41.89
AR	4 ×	39.03	44.44	45.28	41.07	47.86	40.42	43.02
RCG	4 ×	37.54	43.91	45.15	41.93	44.54	40.24	42.22
LN	4 ×	38.57	45.92	47.00	42.63	46.83	40.79	43.62
Manifold	4 ×	40.81	48.26	49.23	44.11	49.32	42.76	45.75
Ours	4 ×	**43.32**	**51.32**	**50.67**	**45.83**	**51.89**	**44.06**	**47.85**
gain	4 ×	2.51	3.06	1.44	1.72	2.57	1.30	2.10
Bicubic	8 ×	33.39	40.79	42.64	37.37	42.23	36.12	38.76
AR	8 ×	36.78	42.84	43.63	39.04	44.32	38.74	40.89
RCG	8 ×	35.09	41.84	43.81	39.22	42.20	38.26	40.07
LN	8 ×	36.57	43.61	44.89	40.63	43.84	38.66	41.37
Manifold	8 ×	37.77	44.32	45.97	42.19	44.91	41.07	42.71
Ours	8 ×	**38.22**	**46.90**	**47.10**	**44.09**	**47.85**	**41.61**	**44.30**
gain	8 ×	0.45	2.58	1.13	1.90	2.94	0.54	1.59
Bicubic	16 ×	29.93	37.70	39.68	33.99	39.03	32.83	35.53
AR	16 ×	32.70	38.83	40.73	34.77	39.92	35.90	37.14
RCG	16 ×	30.64	38.69	41.59	35.15	40.15	34.14	36.73
LN	16 ×	33.81	40.16	41.55	37.09	40.39	36.66	38.28
Manifold	16 ×	33.14	40.44	43.10	37.53	41.24	36.82	38.71
Ours	16 ×	**35.21**	**41.69**	**45.50**	**39.01**	**42.41**	**38.02**	**40.31**
gain	16 ×	2.07	1.25	2.40	1.48	1.17	1.20	1.60

Note: the best performance is boldfaced on each column.

manifold is considered. This thresholding process is formulated as below:

$$x(i) = \begin{cases} x(i) & \text{if } x\,(i) < g(i) \\ \dfrac{1}{2}(x(i) - g(i))^2 & \text{otherwise} \end{cases} \tag{18}$$

when the problem occurs, the mean square error is obtained for both images (depth image **x** and guidance image **g**).

From the multilinear algebra standpoint [40], a multilinear image provides a powerful theoretical-mathematical framework for analyzing the multifactor formation of image ensembles and handling the difficult problem of disentangling the constituent elements or modes [41].

The proposed multilinear modeling technique employs an image extension of the conventional matrix, and the multilinear was used through the duplication of multiple matrices and image reshapes.

However, it can also be utilized in this section by adding a padding operation as introduced in [22]. Based on the k-dimensional manifold $\mathcal{M}_k$ in the D-dimensional Euclidean

TABLE 3. Quantitative comparison of the performance [in PSNR (dB)] (up-sampling without noise).

Method	rate	Art	Book	Dolls	Laundry	Mobius	Reindeer	Average
Bicubic	2 ×	34.51	35.72	35.78	35.36	35.78	35.11	35.38
AR	2 ×	41.56	47.08	46.87	45.09	47.45	43.89	45.32
RCG	2 ×	40.85	46.59	46.66	44.49	46.71	42.92	44.70
LN	2 ×	40.15	44.52	44.18	43.05	45.13	41.01	43.01
Manifold	2 ×	45.68	53.51	52.83	**49.51**	53.68	46.71	50.32
Ours	2 ×	**48.13**	**54.02**	**53.21**	48.83	**53.81**	**46.83**	**50.81**
gain	2 ×	2.45	0.51	0.38	-0.68	0.13	0.12	0.49
Bicubic	4 ×	33.26	35.29	35.45	34.61	35.43	34.23	34.71
AR	4 ×	38.14	43.49	43.63	41.44	44.04	40.70	40.70
RCG	4 ×	37.69	43.44	44.07	41.37	43.72	40.09	41.73
LN	4 ×	38.30	43.91	44.01	41.46	44.27	40.16	42.02
Manifold	4 ×	41.22	47.84	47.73	44.77	47.58	44.24	45.56
Ours	4 ×	**42.06**	**49.62**	**48.52**	**45.05**	**49.66**	**45.15**	**46.68**
gain	4 ×	0.84	1.78	0.79	0.28	2.08	0.91	1.11
Bicubic	8 ×	31.55	34.73	35.04	33.57	34.99	33.01	33.82
AR	8 ×	35.11	39.93	41.13	38.67	40.93	37.37	38.86
RCG	8 ×	34.55	40.63	41.42	38.37	40.34	37.70	38.84
LN	8 ×	36.23	41.56	41.99	39.77	41.80	38.13	39.91
Manifold	8 ×	36.61	43.46	**45.44**	41.51	43.98	40.44	41.91
Ours	8 ×	**37.17**	**45.70**	44.95	**42.14**	**45.41**	**40.78**	**42.69**
gain	8 ×	0.56	2.24	-0.49	0.63	1.43	0.34	0.79
Bicubic	16×	29.01	33.65	34.33	31.83	34.11	31.08	32.34
AR	16×	30.67	37.02	38.00	33.78	37.23	34.72	35.24
RCG	16×	30.47	37.49	39.10	34.69	38.10	33.78	35.61
LN	16×	33.31	39.05	39.24	36.15	39.00	36.36	37.19
Manifold	16×	32.79	40.18	42.80	37.07	40.74	37.04	38.44
Ours	16×	**33.45**	**41.01**	**43.36**	**38.11**	**41.21**	**37.50**	**39.11**
gain	16×	0.66	0.83	0.56	1.04	0.47	0.46	0.67

Note: the best performance is highlighted as a bold text.

space, the following equations are established:

$$\mathcal{M}_k = \{\hat{\boldsymbol{\rho}}_k(\mathbf{l}) \,|\mathbf{l} \in \mathbb{R}^k\}$$
$$(\hat{\boldsymbol{\rho}}_k, \hat{\boldsymbol{\varphi}}_k) = \underset{(\boldsymbol{\rho}_k, \boldsymbol{\varphi}_k)}{\arg\min} \sum_{p=1}^{P} \left\|\mathbf{f}_p - \boldsymbol{\rho}_k \boldsymbol{\varphi}_k(\mathbf{f}_p)\right\|_2^2 \quad (19)$$

where $\boldsymbol{\varphi}_k$: $\mathbb{R}^D \to \mathbb{R}^k$ is an encoder, $\boldsymbol{\rho}_k$: $\mathbb{R}^k \to \mathbb{R}^D$ is a decoder, whereas $\hat{\boldsymbol{\rho}}_k \hat{\boldsymbol{\varphi}}_k$: $\mathbb{R}^D \to \mathbb{R}^D$ is an auto-encoder (AE) constructed from $\{\boldsymbol{f}_p\}_{p=1}^P$. Typically, the AE approaches are widely used, since those are well-established methods for manifold learning [42]. The properties of the manifold $\mathcal{M}_k$

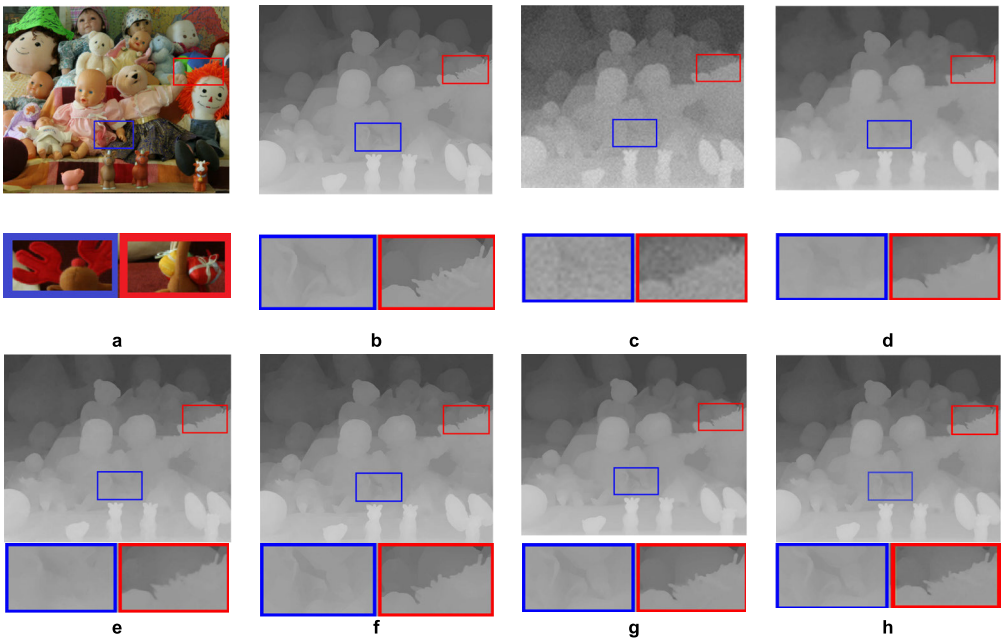


FIGURE 7. The comparison of ToF-like degradation on dataset dolls. (a) RGB, (b) the ground truth, (c) bicubic, (d) AR, (e) LN, (f) RCG, (g) manifold, and (h) the proposed method.

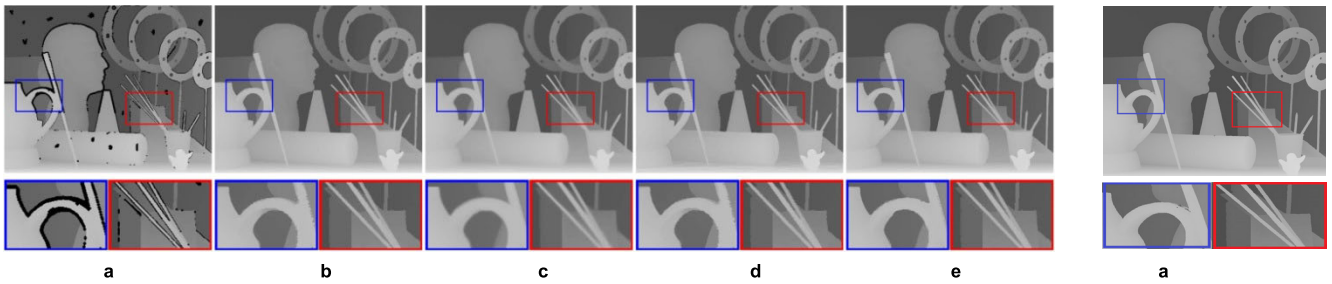


FIGURE 8. The comparison of kinect-like degradation on dataset art. (a) Degraded depth map, (b) AR, (c) LN, (d) RCG, (e) manifold, and (f) the proposed method.

are determined by using the properties of $\boldsymbol{\varphi}_k$ and $\boldsymbol{\rho}_k$. The encoder-decoder framework provided a smooth manifold.

IV. EXPERIMENTAL RESULTS

This section reports experimental results and accordingly makes a comparison between the proposed method and the other state-of-the-art ones in terms of the performance. The proposed method was analyzed in detail based on Middlebury Datasets [43] and NYU Depth Dataset [44], which are the most widely used datasets for depth recovery.

A. IMPLEMENTATION DETAILS

The proposed method was implemented in two phases. First, the DDPM method was executed on the PyTorch framework [37] with the CNN weights initialized to zero. The local manifold regularization parameters of the diffusion kernel were set as below:

$\sigma_p = 100, \sigma_x = 10, \sigma_r = 40$ when the rate is $2\times$.
$\sigma_p = 100, \sigma_x = 10, \sigma_r = 50$ when the rate is $4\times$.
$\sigma_p = 10, \sigma_x = 20, \sigma_r = 50$ when the rate is $8\times$.
$\sigma_p = 10, \sigma_x = 200, \sigma_r = 100$ when the rate is $16\times$.

The DIP network was used as a good feature extractor, and the DDPM was trained for 100 epochs by using the ADMM optimizer [39]. The execution time was 2 hours for each image per rate $\{2\times, 4\times, 8\times, 16\times\}$. The learning rate was firstly set to 10^{-4}. The second part (patches-manifold) of the proposed method was implemented on TensorFlow 2.2.0, and the patch size was set as $s = 5$. Furthermore, the low dimensionality of nonlocal manifold was set as $k = 6$, and the noise level was considered as $\sigma = 0.05$. The best number of iterations is 9000. It took 2.5 to 3 hours for each image per rate $\{2\times, 4\times, 8\times, 16\times\}$ on Google Colab Pro with NVIDIA® GeForceTesla P100-PCIE GPU and 16 GB of RAM. We used the source codes of DIP[1] framework which is modified to match to proposed idea. Since the proposed method uses single images, the depth estimation network is trained in a traditional way, with synthetic images and corresponding ground truth depth maps, and in the test phase, the trained network is applied directly to predict the depth maps. In the training phase, we have used many cases of input images

[1] https://github.com/DmitryUlyanov/deep-image-prior

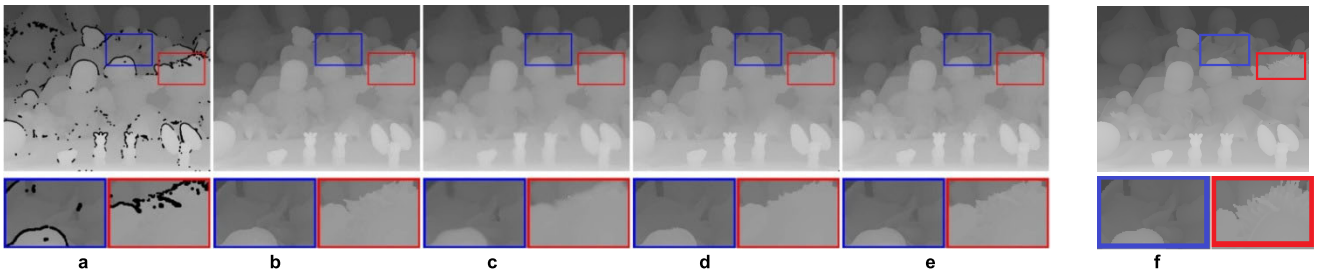


FIGURE 9. The comparison of kinect-like degradation on dataset dolls. (a) Degraded depth map, (b) AR, (c) LN, (d) RCG, (e) manifold, and (f) the proposed method.

TABLE 4. Quantitative evaluations on the NYU dataset.

` Method	δ <(thr=1.25)	δ < (thr=(1.25) ^{2})	δ < (thr=(1.25) ^{3})	Abs Rel	RMSE	RMSE log
	Higher value is better			Lower value is better		
Maker 3D [45]	0.601	0.820	0.926	0.280	8.734	0.361
Eigen et al. [46]	0.692	0.899	0.967	0.190	7.156	0.270
Liu et al. [47]	0.647	0.882	0.961	0.217	6.986	0.289
Godard et al. [48]	0.861	0.949	0.976	0.114	4.935	0.206
Kuznietsov et al. [49]	0.862	0.960	0.986	0.113	4.621	0.189
Gan et al. [50]	0.890	0.964	0.985	**0.098**	3.933	0.173
Song et al. [23]	0.893	0.976	0.992	0.108	2.454	0.160
Our	**0.909**	**0.989**	**0.995**	0.107	**1.916**	**0.030**

such as enlarged up to 100% and sometimes we cropped to the original size. In addition, input images are horizontally flipped, and the brightness of the image is randomly changed using a scale factor chosen from a range of [0.5, 2.0].

B. PERFORMANCE ANALYSIS

The performance analysis results indicated the efficiency and robustness of the proposed algorithm in comparison with the other five state-of-the-art depth recovery methods: 1) bicubic, 2) thresholding on manifolds [1], 3) color-guided through joint local and nonlocal structural low ranks (LN) [2], 4) robust color guided (RCG) [3], and 5) color-guided autoregressive (AR) [4]. The upsampling performance was tested in four different upsampling ratios: 2×, 4×, 8×, and 16×. The performance evaluation of PSNR was measured by considering error metrics to show differences between the proposed method and the other state-of-the-art techniques. The proposed method was also compared with eight repressive methods introduced by Saxena *et al.* [45], Eigen *et al.* [46], Liu *et al.* [47], Godard *et al.* [48], Kuznietsov *et al.* [49], Gan *et al.* [50], and Song and Kim [23]. According to Table 4, nearly the best result was obtained.

Comparisons were drawn in the standard four metrics defined as below:

- Average relative error (**Abs Rel**):

$$\frac{1}{y}\sum_{p=1}^{y}\frac{|x_p-\hat{x}_p|}{x_p}$$

- Root mean squared error (**RMSE**):

$$\sqrt{\frac{1}{y}\sum_{p=1}^{y}\left(x_p-\hat{x}_p\right)^2}$$

- Average (**RMSE log**) error:

$$\frac{1}{y}\sum_{p=1}^{y}|log_{10}(x_p)-log_{10}(\hat{x}_p)|$$

- Threshold accuracy (δ): Percentageof x_p s. t. $\max(\frac{x_p}{\hat{x}_p},\frac{\hat{x}_p}{x_p})=\delta<thr$ for $thr=1.25;\ 1.25^2;\ 1.25^3$

In all the above criteria x_p denotes the value of pixel p in depth image, $\hat{x}_p$ denotes the predicted value for pixel p in depth image, and y represents the total number of pixels.

C. ANALYSIS OF ADVANTAGES AND DISADVANTAGES OF DIFFERENT MANIFOLD MODELING APPROACHES

This section discusses and analysis different parts and especially the role of manifold modeling in the proposed method. All of these remedies provide a final depth recovery. Several experiments are conducted to analyze the stability behavior systematically. When the local manifold per se was used, it yielded no strong results because it failed to refine the depth from noise definitively.

In contrast, the patch manifold approach achieved favorable denoising performance. However, it plays an effective role in restoration applications such as upsampling and reshaping. Figure 3 illustrates the performance of the proposed depth reconstruction model. We used six images, each image in four under-sampling rates ($2\times$, $4\times$, $8\times$, $16\times$), and the total test were 24 cases. If the number of iterations increases, the results might be more stable.

Table 1 shows the PNSR with and without the use of a patch manifold, which obviously shows the effectiveness of using patch manifold. Evidently, the proposed model not only managed to estimate the depth map but also used it for image super-resolution.

D. EXPERIMENTS ON SYNTHETIC DEGRADATIONS DATASETS

The test datasets included Art, Book, Dolls, Laundry, Moebius, and Reindeer selected from the Middlebury's benchmark [43]. The proposed model was simulated in three types of degradation, which are undersampling, ToF-like degradation, and Kinect-like degradation.

1) UNDERSAMPLING DEGRADATION

In this implementation, $\mathbf{H}$ is considered as a degradation matrix which is the blur operation on the ground truth through a Gaussian kernel after the downsampling process is performed in this case. The noise is assumed zero. The upsampling tests are four upsampling ratios including $2\times$, $4\times$, $8\times$, and $16\times$ as shown in Table 2, which explains and shows the comparison between different methods in PSNR values by four upsampling rates. The best result is boldfaced. The average of PSNR was obtained for the results of the proposed method reported as 0.53 dB ($2\times$), 2.10 dB ($4\times$), 1.59 dB ($8\times$), and 1.60dB ($16\times$).

Figure 4 and Figure 5 show the visual comparison between methods with an $8\times$ upsampling rate for Reindeer and Laundry images. The first result (b) in both Figures, which can be observed in the Bicubic, are very smooth. The results of the AR method (c) cannot be kept by the large scale edges. The LN result (d) is slightly jagged on the edges as shown by the blue box on the Laundry because of being mixed by two models using AR (d) and TV. The result of the RCG method was not satisfactory, since the manifold is still nearly over-smoothed.

FIGURE 10. The results of depth recovery on the dataset NYU. (a) Original color image, (b) ground truth, (c) GDN, and (d) proposed method.

2) TOF-LIKE DEGRADATION

In this implementation, $\mathbf{H}$ represents the degradation matrix similar to upsampling degradation discussed in Section 3. The noise $\mathbf{n}$ indicates the white Gaussian noise to simulate this degradation. The noise is first added with a variance of five to the original depth.

After that, it was downsampled with four rates on datasets.

Table 3 reports the quantitative performance of recovery to compare the proposed method with other models. Figure 6 and Figure 7 show the comparison results on Moebius and Dolls images under ToF-like degradation at an $8\times$ upsampling rate. The Bicubic method can be seen as noisy, and the result of the AR method is very blurring due to its coefficients being so sensitive to noise. The results of the LN model had no noise but led to over-smoothing with no sharpness. The RCG method produced noticeable distortions, especially in smooth regions. The manifold outperformed other models but was still over-smooth images. The proposed model obtained the best results even in the detailed regions and smoothness. The results of which were closer to the ground truth.

3) KINECT-LIKE DEGRADATION

In this implementation, $\mathbf{H}$ represents the matrix of degradation like upsampling degradation without any noise to simulate Kinect-like degradation by missing structures with random missing on the original depth image. Figure 8 and Figure 9 demonstrate the results on two depth test images

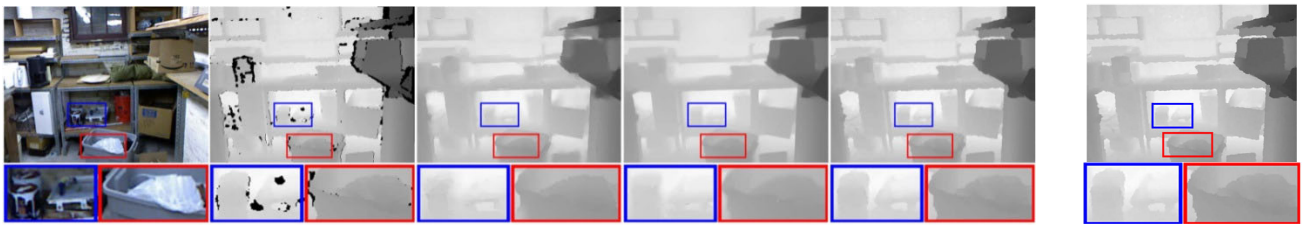


FIGURE 11. The comparison on real world depth on dataset from NYU RGB-D. (a) Color image, (b) degraded depth map, (c) AR, (d) LN, (e) manifold, and (f) the proposed method.

Dolls and Art. Nearly all methods yielded good recovery outputs by random missing in flat areas.

E. PERFORMANCE EVALUATION EXPERIMENTS ON OTHER DATASETS

In this implementation, the proposed method is tested on the NYU depth v2 dataset [44]. **H** represents the matrix of degradation like upsampling without any noise (**n** is zero). The method was applied on four test images and compared with the guided deep network (GDN), which was introduced by Song and Kim [23]. According to Figure 10, the results were better than those of GDN and closer to the ground truth. It is also possible to compare the proposed method with other models through the values of error metrics values. Table 4 shows that nearly the best result was obtained. The proposed model can also be implemented on the real world depth, which was taken by the Kinect camera as presented in Figure 11.

V. CONCLUSION

This study proposed a novel model for depth recovery from the low-quality/low-resolution depth images affected by various kinds of degradations. The proposed method developed the deep depth prior (DDP) framework by adding local and nonlocal (patch) manifold regularization frameworks. The hierarchy of convolution operations in DDP can be efficient in image recovery. First, better reconstruction outputs than those of several successful related methods were obtained by fully exploiting the input color and depth images through leveraging local manifolds used as regularizers. In the second step, the output of the first model called DDPM was used as the input in the patch-manifold stage. The optimization problem was finally solved in an ADMM framework. These methods were integrated to produce even better results and analyze the stability behavior systematically. Since the proposed approach is fast and highly efficient, it can be utilized in real-time applications such as self-driving vehicles (autonomous vehicles) and robotic navigation.

For future studies, there are still many more possible methods to implement encoder-decoder models, especially when the back propagation optimization algorithm for the depth recovery will yield high-quality outputs.

ACKNOWLEDGMENT

The authors would like to specially thank Dr. Tatsuya Yokota, the Esteemed Lecturer of Nagoya Institute of Technology, Nagoya, Japan, and a Senior Member of IEEE, for his undying support and advice. The author Abbas K. Ali would like to appreciate the help and consultation of Dr. Shaker K. Ali, who is the Iraqi Deputy Attaché in Romania.

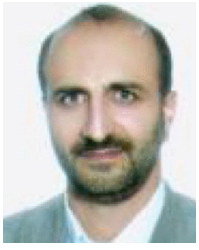
REFERENCES

[1] X. Liu, D. Zhai, R. Chen, X. Ji, D. Zhao, and W. Gao, "Depth restoration from RGB-D data via joint adaptive regularization and thresholding on manifolds," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1068–1079, Mar. 2019.
[2] W. Dong, G. Shi, X. Li, K. Peng, J. Wu, and Z. Guo, "Color-guided depth recovery via joint local structural and nonlocal low-rank regularization," *IEEE Trans. Multimedia*, vol. 19, no. 2, pp. 293–301, Feb. 2017.
[3] W. Liu, X. Chen, J. Yang, and Q. Wu, "Robust color guided depth map restoration," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 315–327, Jan. 2016.
[4] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3443–3458, Aug. 2014.
[5] X. Liu, D. Zhai, R. Chen, X. Ji, D. Zhao, and W. Gao, "Depth super-resolution via joint color-guided internal and external regularizations," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1636–1645, Apr. 2019.
[6] G. Peyré, "Image processing with nonlocal spectral bases," *Multiscale Model. Simul.*, vol. 7, no. 2, pp. 703–730, 2008.
[7] U. Hahne and M. Alexa, "Exposure fusion for time-of-flight imaging," *Comput. Graph. Forum*, vol. 30, no. 7, pp. 1887–1894, 2011.
[8] Y. Zhang, Y. Feng, X. Liu, D. Zhai, X. Ji, H. Wang, and Q. Dai, "Color-guided depth image recovery with adaptive data fidelity and transferred graph Laplacian regularization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 2, pp. 320–333, Feb. 2020.
[9] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, vol. 46. Hoboken, NJ, USA: Wiley, 2004.
[10] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*. Hoboken, NJ, USA: Wiley, 2009.
[11] J. Liu, Y. Sun, X. Xu, and U. S. Kamilov, "Image restoration using total variation regularized deep image prior," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, p. 7715.
[12] A. Kheradmand and P. Milanfar, "A general framework for regularized, similarity-based image restoration," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5136–5151, Dec. 2014.
[13] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[14] C. Li, B. Zhang, C. Chen, Q. Ye, J. Han, G. Guo, and R. Ji, "Deep manifold structure transfer for action recognition," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4646–4658, Sep. 2019.
[15] J. Tang, F.-P. Tian, W. Feng, J. Li, and P. Tan, "Learning guided convolutional network for depth completion," *IEEE Trans. Image Process.*, vol. 30, pp. 1116–1129, 2021.
[16] Y. Li, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Joint image filtering with deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1909–1923, Aug. 2019.
[17] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[18] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.

[19] P. Cascarano, A. Sebastiani, M. C. Comes, G. Franchini, and F. Porta, "Combining weighted total variation and deep image prior for natural and medical image restoration via ADMM," 2020, *arXiv:2009.11380*. [Online]. Available: http://arxiv.org/abs/2009.11380
[20] X. Tu, C. Xu, S. Liu, R. Li, G. Xie, J. Huang, and L. T. Yang, "Efficient monocular depth estimation for edge devices in Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2821–2832, Apr. 2021.
[21] S. Osher, Z. Shi, and W. Zhu, "Low dimensional manifold model for image processing," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1669–1690, Oct. 2017.
[22] T. Yokota, H. Hontani, Q. Zhao, and A. Cichocki, "Manifold modeling in embedded space: An interpretable alternative to deep image prior," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 4, 2021, doi: 10.1109/TNNLS.2020.3037923.
[23] M. Song and W. Kim, "Depth estimation from a single image using guided deep network," *IEEE Access*, vol. 7, pp. 142595–142606, 2019.
[24] I. Alhashim and P. Wonka, "High quality monocular depth estimation via transfer learning," 2018, *arXiv:1812.11941*. [Online]. Available: http://arxiv.org/abs/1812.11941
[25] Y. Zhong, Y. Pei, P. Li, Y. Guo, G. Ma, M. Liu, W. Bai, W. Wu, and H. Zha, "Depth-based 3D face reconstruction and pose estimation using shape-preserving domain adaptation," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 3, no. 1, pp. 6–15, Jan. 2021.
[26] W. Hu, G. Cheung, A. Ortega, and O. Au, "Multi-resolution graph Fourier transform for compression of piecewise smooth images," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 419–433, Jan. 2015.
[27] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.
[28] S. Gu, W. Zuo, S. Guo, Y. Chen, C. Chen, and L. Zhang, "Learning dynamic guidance for depth image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 712–721.
[29] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, p. 96, Jul. 2007.
[30] J. Xie, C.-C. Chou, R. Feris, and M.-T. Sun, "Single depth image super resolution and denoising via coupled dictionary learning with local constraints and shock filtering," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2014, pp. 1–6.
[31] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[32] Z. Jiang, Y. Hou, H. Yue, J. Yang, and C. Hou, "Depth super-resolution from RGB-D pairs with transform and spatial domain regularization," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2587–2602, May 2018.
[33] X. Liu, G. Cheung, X. Wu, and D. Zhao, "Random walk graph Laplacian-based smoothness prior for soft decoding of JPEG images," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 509–524, Feb. 2017.
[34] 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. (MMSP)*, Sep. 2013, pp. 1–6.
[35] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 341–349.
[36] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[37] D. Maclaurin, D. Duvenaud, and R. P. Adams, "Autograd: Effortless gradients in Numpy," in *Proc. ICML*, 2015, p. 5.
[38] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. San Diego, CA, USA, Jun. 2005, pp. 60–65.
[39] T. R. Shaham, T. Dekel, and T. Michaeli, "SinGAN: Learning a generative model from a single natural image," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4570–4580.
[40] S. Razavikia, A. Amini, and S. Daei, "Reconstruction of binary shapes from blurred images via Hankel-structured low-rank matrix recovery," *IEEE Trans. Image Process.*, vol. 29, pp. 2452–2462, 2020.
[41] M. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 2350, Copenhagen, Denmark, May 2002, pp. 447–460.
[42] T. Yokota, B. Erem, S. Guler, S. K. Warfield, and H. Hontani, "Missing slice recovery for tensors using a low-rank model in embedded space," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8251–8259.
[43] *Middlebury Datasets*. Accessed: Dec. 22, 2019. [Online]. Available: http://vision.middlebury.edu/stereo/data/
[44] *NYU Depth Dataset V2*. Accessed: Sep. 5, 2020. [Online]. Available: http://cs.nyu.deu/silberman/datasets
[45] A. Saxena, M. Sun, and A. Y. Ng, "Make3D: Learning 3D scene structure from a single still image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 824–840, May 2009.
[46] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2012, pp. 2366–2374.
[47] F. Liu, C. Shen, and G. Lin, "Deep convolutional neural fields for depth estimation from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5162–5170.
[48] C. Godard, O. M. Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6602–6611.
[49] Y. Kuznietsov, J. Stuckler, and B. Leibe, "Semi-supervised deep learning for monocular depth map prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2215–2223.
[50] Y. Gan, X. Xu, W. Sun, and L. Lin, "Monocular depth estimation with affinity, vertical pooling, and label enhancement," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 232–247.

ABBAS K. ALI received the B.S. degree from the Department of Computer Science, College of Science, University of Thi-Qar, Nasiriyah, Dhi Qar Governorate, Iraq, in 2007, and the M.S. degree from the Department of Information Science Engineering, Central South University, Changsha, Hunan, China, in 2011. Currently, he is pursuing the Ph.D. degree with the Artificial Intelligence Department, Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran. His current research interests include computer vision and augmented reality, depth estimation, and image reconstruction.

PEYMAN ADIBI received the B.S. degree in computer engineering from Isfahan University of Technology, in 1998, the M.S. degree in computer engineering from Amirkabir University of Technology, Tehran, Iran, in 2001, and the Ph.D. degree from the Faculty of Computer Engineering, Amirkabir University of Technology, in 2009. He is currently an Assistant Professor with the Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran, where he is the Head of the Artificial Intelligence Department. His current research interests include machine learning and pattern recognition, computer vision and image processing, computational intelligence and soft computing, and statistical signal processing.



MOHAMMAD-SAEED EHSANI received the B.Sc. and M.Sc. degrees in communication engineering from the School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran, the Ph.D. degree in computer science from the School of Engineering, Kennedy Western University, Cheyenne, WY, USA, and the D.L. degree (Hons.) from the University of Cambridge, Cambridge, U.K. He is currently a member of the Artificial Intelligence Department, Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran. His research interests include pattern recognition, neuro-fuzzy, rule-base control, next generation networks, big data, and cybernetics. He is a Fellow Member of ABI or IBC and a member of the IEEE's Technical Committees.