Depth Image Reconstruction Using Low Rank and Total Variation Representations

Van Ha Tang[(✉)] and Mau Uyen Nguyen

Faculty of Information Technology, Le Quy Don Technical University,
Hanoi, Vietnam
hatv@lqdtu.edu.vn, nguyenitt2005@gmail.com

Abstract. Rapid advancement and active research in computer vision applications and 3D imaging have made a high demand for efficient depth image estimation techniques. The depth image acquisition, however, is typically challenged due to poor hardware performance and high computation cost. To tackle such limitations, this paper proposes an efficient approach for depth image reconstruction using low rank (LR) and total variation (TV) regularizations. The key idea is LR incorporates non-local depth information and TV takes into account the local spatial consistency. The proposed model reformulates the task of depth image estimate as a joint LR-TV regularized minimization problem, in which LR is used to approximate the low-dimensional structure of the depth image, and TV is employed to promote the depth sparsity in the gradient domain. Furthermore, this paper introduces an algorithm based on alternating direction method of multipliers (ADMM) for solving the minimization problem, whose solution provides an estimate of the depth map from incomplete pixels. Experimental results are conducted and the results show that the proposed approach is very effective at estimating high-quality depth images and is robust to different types of data missing models.

Keywords: Depth imaging · Depth image reconstruction · Low-rank matrix factorization · Sparse representation

1 Introduction

Recent development of numerous computer vision (CV) applications and 3D modeling have increased the research and development of depth sensing technologies. The capabilities of generating depth information together with the conventional 2D image of the desired scene, and thereby yielding the 3D model of the desired scene is very useful in a wide range in CV applications, from autonomous driving, robot navigation, to augmented reality and action recognition [6,7,18,19]. However, depth imaging acquisition faces with several technical difficulties, including poor hardware performance, prolonged data collection, and high computational cost, though much effort has been made to the development

of 3D cameras, such as Google Tango, Intel RealSense, and Microsoft Kinect [20,21]. Thus, new efficient techniques for fast data acquisition and efficient depth reconstruction are very vital for numerous depth technology applications.

Much interest in depth image estimation has been reported in several studies. Techniques exploiting the shapes of objects in the depth image were reported in [19] and [18]. In [19], the depth image quality has been improved by incorporating shading shapes. In [18] the defocusing shape of objects was considered to restore the missing values on the depth map. The other approaches that combine hand-tuned models and surface orientations were proposed in [14,17]. Image inpainting-based methods have also applied to depth estimate. In [5], missing depth values were restored by region growing and bilateral filtering. In [3], a Kalman filter was employed for enhancing smoothness of the depth image, whereas morphological operators were used in [8]. However, it is worth noting that the majority of such methods rely on color information of the scene. In other words, they require collecting the color image, which prolongs the time for data acquisition and makes a burden to data storage.

Recent approaches based on modern sensing and representations were proposed for depth estimate [11,12]. In [11], a sparse representation (SR) was used to estimate a depth map from the incomplete observed pixels. This approach relies on the theory of the powerful compressive sensing (CS) framework, which states that with high probability, a sparse signal/image can be reconstructed precisely from incomplete measurements/pixels provided that it is compressible or has a SR in proper bases [4,9]. In addition, CS enables the reconstruction and compression to be performed simultaneously, resulting in simple and cost-effective hardware sensing systems. Note that in the sparsity-based technique, the task of depth reconstruction is posed as a sparsity-regularized least squares (LS) minimization problem. A further extension of the sparsity-regularized model was presented in [12], where the color information was incorporated for enhancing the quality of image restoration.

Inspired by the SR-based models, this paper introduces a joint LR and TV regularizations for depth image estimation from incomplete observed pixels. Together with SR, the LR representation is incorporated into the imaging model to capture the low-dimensional structure of the depth images. Moreover, the TV regularizer is used to promote the local spatial consistency for the depth map. To this end, the task of depth image estimate is reformulated as a joint LR-TV regularized minimization problem, whose solution provides an estimate of the depth map. We present an algorithm based on ADMM technique for solving the minimization problem. The proposed model is validated by several experimental evaluations. Analysis and comparisons with other imaging models are also provided in this study.

The remainder of the paper is organized as follows. Section 2 introduces the depth image acquisition and the SR-based depth estimation. Section 3 describes the proposed LR-TV model for depth image reconstruction from incomplete measurements. Section 4 presents the experimental results. Section 5 gives concluding remarks.

2   Depth Sparse Imaging Model

This section first presents a brief mathematical model for depth image acquisition. It then describes a sparse representation approach for depth image reconstruction.

2.1   Depth Image Acquisition

Let $\mathbf{Z} \in \mathbb{R}^{h \times w}$ denote a depth map of a scene imaged by an active sensor. In the case of full sensing operations or ideal imaging, the sensor receives $m = h \times w$ depth pixels that represent the distances from the sensor to the surrounding obstacles. In fact, due to fast sensing or poor hardware performance, only n ($n \ll m$) measurements are acquired. Let $\mathbf{z} \in \mathbb{R}^m$ be a vector obtained by applying a vectorization operator to the unknown image $\mathbf{Z}$, $\mathbf{z} = \mathrm{vec}(\mathbf{Z})$, where $\mathrm{vec}(\cdot)$ is the vectorization operator forming a composite column vector by stacking the columns of a matrix in lexicographic order. Note also that the image $\mathbf{Z}$ can be obtained from its corresponding vector $\mathbf{z}$ through $\mathbf{Z} = \mathrm{mat}(\mathbf{z})$, where $\mathrm{mat}(\cdot)$ is the operator converting a column vector having $h \times w$ entries into a $h \times w$ matrix. Since the conversion between the image vector and matrix is simple, hereafter we use $\mathbf{z}$ to represent for both the vector and matrix depth image.

Let $\mathbf{y} \in \mathbb{R}^n$ be the vector containing the observed measurements. The incomplete image $\mathbf{y}$ can be related to the full unknown image $\mathbf{z}$ as

$$\mathbf{y} = \boldsymbol{\Phi} \, \mathbf{z}, \tag{1}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{n \times m}$ is the sampling matrix, mathematically representing the depth acquisition protocol. In this representation, $\boldsymbol{\Phi}$ is a diagonal binary matrix used to indicate which pixels are selected. Now, given the incomplete and thus corrupted image $\mathbf{y}$ and matrix $\boldsymbol{\Phi}$, our goal is to reconstruct a full and high-quality depth map $\mathbf{z}$.

2.2   Depth Estimation with Sparse Representation

The full depth image $\mathbf{z}$ can be reconstructed by exploring its structures. The most common structure that has been exploited is its sparsity in a transform domain. This is because the depth image $\mathbf{z}$ typically contains large homogenous objects with only a few discontinuity at their transitions. This characteristic leads to a SR for the depth image in proper domains, such as wavelets [12,13]. The sparsity and the wavelet representation can be justified in the aspect that large homogeneous regions can be compressively represented by only a small number of significant coefficients. Mathematically, the sparse representation of image $\mathbf{z}$ can be expressed as

$$\mathbf{z} = \boldsymbol{\Psi} \, \mathbf{x}, \tag{2}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{m \times q}$ is the dictionary matrix with its columns q ($q \geq m$) being the wavelet bases, and $\mathbf{x} \in \mathbb{R}^q$ is a vector of coefficients having only s nonzero

components, i.e., $s = \|\mathbf{x}\|_0$. Due to the sparseness of $\mathbf{x}$, s is much smaller than q ($s \ll q$). Here the ℓ_0-norm is used to measure the number of nonzero entries in vector $\mathbf{x}$. In practice, its counterpart, the ℓ_1-norm, is used to regularize the sparsity as it is a convex relaxation for the ℓ_0-norm [16].

It follows from (1) and (2) that $\mathbf{y} = \boldsymbol{\Phi}\,\boldsymbol{\Psi}\,\mathbf{x}$. Given the observation vector $\mathbf{y}$, an estimate of the coefficient vector $\mathbf{x}$ can be obtained by solving the following ℓ_1-regularized LS minimization problem:

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\,\boldsymbol{\Psi}\,\mathbf{x}\|_2^2 + \lambda\,\|\mathbf{x}\|_1, \tag{3}$$

where λ is a regularization parameter used to balance between the LS and the ℓ_1 penalty terms. Since the dictionary matrix $\boldsymbol{\Psi}$ is typically orthogonal, i.e., $\boldsymbol{\Psi}\,\boldsymbol{\Psi}^T = \mathbf{I}$. Thus, $\mathbf{z} = \boldsymbol{\Psi}\,\mathbf{x}$ and $\mathbf{x} = \boldsymbol{\Psi}^T\,\mathbf{z}$. This means that Problem (3) can be written equivalently as

$$\underset{\mathbf{z}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\,\mathbf{z}\|_2^2 + \lambda\,\|\boldsymbol{\Psi}^T\,\mathbf{z}\|_1. \tag{4}$$

The solution to Problem (4) gives an estimate for a full depth image. This sparse regularization approach, however, is effective only if the observed image $\mathbf{y}$ is corrupted with random missing values. In other words, the obtained estimate is good if the image does not contain large continuous missing regions. For example, if pixels are missing along entire rows or columns, then the sparse-regularized model cannot cope well with this issue. This limitation can be overcome by introducing more effective regularizations into the imaging model. In this study, we investigate two more regularizations of LR and TV. The new LR-TV model is described in the next section.

3   LR-TV Depth Reconstruction

This section first presents the proposed LR-TV model for depth estimate problem in Subsect. 3.1, followed by an algorithm based on ADMM to solve the LR-TV regularized minimization problem in Subsect. 3.2.

3.1   LR-TV Problem Formulation

LR and TV regularizers are introduced to further strengthen the depth imaging model. The motivation of using LR is due to the fact that natural images can be well approximated by their LR components. Furthermore, principal image structures typically reside in a low-dimensional subspace, thereby having a LR representation. Therefore, incorporating LR depth structure can improve the reconstruction quality. To this end, we extend the minimization problem (4) as

$$\underset{\mathbf{z}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\,\mathbf{z}\|_2^2 + \lambda\,\|\boldsymbol{\Psi}^T\,\mathbf{z}\|_1 + \beta\,\|\mathbf{z}\|_*. \tag{5}$$

Here, $\|\mathbf{z}\|_*$ is the nuclear-norm of the matrix $\mathbf{z}$, which the sum of its singular values, $\|\mathbf{z}\|_* = \sum_{i=1}^{p} \lambda_i(\mathbf{z})$ with $\lambda_i(\mathbf{z})$ being the ith largest singular value of matrix $\mathbf{z}$ of rank at most p. It is worth noting that the nuclear-norm is the convex relaxation for the LR constraint imposed on a matrix [2].

The LR is regarded as a non-local (global) regularizer since it considers the information throughout the image. As a result, LR may neglect the useful local information that often relates to the detail map. To fill this gap, a TV regularizer is added to guarantee the local spatial consistency. The proposed LR-TV model becomes

$$\underset{\mathbf{z}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\,\mathbf{z}\|_2^2 + \lambda\,\|\boldsymbol{\Psi}^T\,\mathbf{z}\|_1 + \beta\,\|\mathbf{z}\|_* + \gamma\,\|\mathbf{z}\|_{\text{TV}}. \tag{6}$$

Here, $\|\mathbf{z}\|_{\text{TV}}$ is the anisotropic TV defined as

$$\|\mathbf{z}\|_{\text{TV}} = \|\mathbf{D}\,\mathbf{z}\|_1 = \|\mathbf{D}_x\,\mathbf{z}\|_1 + \|\mathbf{D}_y\,\mathbf{z}\|_1, \tag{7}$$

where $\mathbf{D} = [\mathbf{D}_x; \mathbf{D}_y]$ is the first-order forward finite difference operator in the x-horizontal and y-vertical directions. In (6), the penalty parameters β and γ are used to control the importance of LR and TV terms, respectively. The remaining task is to solve Problem (6), which yields an estimate of the depth image.

3.2   ADMM-Based Algorithm

This subsection presents an algorithm to solve Problem (6) using ADMM technique [1,10]. The ADMM is a powerful framework for solving regularized optimization problems as it allows the entire problem to be decomposed into subproblems that can be handled more effectively. For efficiently handling the subproblems, it is vital to choose suitable auxiliary variables. Here, we introduce four auxiliary variables $\mathbf{s} = \mathbf{z}$, $\mathbf{r} = \boldsymbol{\Psi}^T\,\mathbf{z}$, $\mathbf{t} = \mathbf{z}$, and $\mathbf{u} = \mathbf{D}\,\mathbf{z}$. This way, the minimization problem in (6) is rewritten as

$$\begin{aligned}
&\underset{\mathbf{z},\mathbf{s},\mathbf{r},\mathbf{t},\mathbf{u}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\,\mathbf{s}\|_2^2 + \lambda\,\|\mathbf{r}\|_1 + \beta\,\|\mathbf{t}\|_* + \gamma\,\|\mathbf{u}\|_1, \\
&\text{subject to} \quad \mathbf{s} = \mathbf{z}, \quad \mathbf{r} = \boldsymbol{\Psi}^T\,\mathbf{z}, \quad \mathbf{t} = \mathbf{z}, \quad \mathbf{u} = \mathbf{D}\,\mathbf{z}.
\end{aligned} \tag{8}$$

Problem (8) has its augmented Lagrangian function given by

$$\begin{aligned}
&\mathcal{L}(\mathbf{z},\mathbf{s},\mathbf{r},\mathbf{t},\mathbf{u},\mathbf{w},\mathbf{b},\mathbf{h},\mathbf{v}) = \\
&\frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\,\mathbf{s}\|_2^2 + \lambda\,\|\mathbf{r}\|_1 + \beta\,\|\mathbf{t}\|_* + \gamma\,\|\mathbf{u}\|_1 \\
&- \mathbf{w}^T(\mathbf{s} - \mathbf{z}) - \mathbf{b}^T(\mathbf{r} - \boldsymbol{\Psi}^T\,\mathbf{z}) - \mathbf{h}^T(\mathbf{t} - \mathbf{z}) - \mathbf{v}^T(\mathbf{u} - \mathbf{D}\,\mathbf{z}) \\
&+ \frac{\mu}{2}\|\mathbf{s} - \mathbf{z}\|_2^2 + \frac{\rho}{2}\|\mathbf{r} - \boldsymbol{\Psi}^T\,\mathbf{z}\|_2^2 + \frac{\xi}{2}\|\mathbf{t} - \mathbf{z}\|_2^2 + \frac{\kappa}{2}\|\mathbf{u} - \mathbf{D}\,\mathbf{z}\|_2^2.
\end{aligned} \tag{9}$$

In (9), $\mathbf{w}$, $\mathbf{b}$, $\mathbf{h}$, and $\mathbf{v}$ are the Lagrange multipliers associated with their constraints, and μ, ρ, ξ, and κ are regularization parameters associated with the corresponding quadratic penalty terms.

The idea of ADMM is to find a saddle point of the Lagrangian function $\mathcal{L}(\cdot)$, that is also the solution to Problem (6). The stationary point can be found by solving the following sequence of subproblems, in which the next estimate of each variable at the $(k+1)$th iteration is obtained by fixing the current estimates of the other counterparts,

$$
\begin{aligned}
\mathbf{z}_{k+1} &= \arg\min \mathcal{L}(\mathbf{z}, \mathbf{s}_k, \mathbf{r}_k, \mathbf{t}_k, \mathbf{u}_k, \mathbf{w}_k, \mathbf{b}_k, \mathbf{h}_k, \mathbf{v}_k), \\
\mathbf{s}_{k+1} &= \arg\min \mathcal{L}(\mathbf{z}_k, \mathbf{s}, \mathbf{r}_k, \mathbf{t}_k, \mathbf{u}_k, \mathbf{w}_k, \mathbf{b}_k, \mathbf{h}_k, \mathbf{v}_k), \\
\mathbf{r}_{k+1} &= \arg\min \mathcal{L}(\mathbf{z}_k, \mathbf{s}_k, \mathbf{r}, \mathbf{t}_k, \mathbf{u}_k, \mathbf{w}_k, \mathbf{b}_k, \mathbf{h}_k, \mathbf{v}_k), \\
\mathbf{t}_{k+1} &= \arg\min \mathcal{L}(\mathbf{z}_k, \mathbf{s}_k, \mathbf{r}_k, \mathbf{t}, \mathbf{u}_k, \mathbf{w}_k, \mathbf{b}_k, \mathbf{h}_k, \mathbf{v}_k), \\
\mathbf{u}_{k+1} &= \arg\min \mathcal{L}(\mathbf{z}_k, \mathbf{s}_k, \mathbf{r}_k, \mathbf{t}_k, \mathbf{u}, \mathbf{w}_k, \mathbf{b}_k, \mathbf{h}_k, \mathbf{v}_k).
\end{aligned}
\tag{10}
$$

The Lagrange multipliers are updated as

$$
\mathbf{w}_{k+1} = \mathbf{w}_k - \mu(\mathbf{s}_{k+1} - \mathbf{z}_{k+1}),
\tag{11}
$$

$$
\mathbf{b}_{k+1} = \mathbf{b}_{k+1} - \rho(\mathbf{r}_{k+1} - \boldsymbol{\Psi}^T \mathbf{z}_{k+1}),
\tag{12}
$$

$$
\mathbf{h}_{k+1} = \mathbf{h}_k - \xi(\mathbf{t}_{k+1} - \mathbf{z}_{k+1}),
\tag{13}
$$

$$
\mathbf{v}_{k+1} = \mathbf{v}_k - \kappa(\mathbf{u}_{k+1} - \mathbf{D}\,\mathbf{z}_{k+1}).
\tag{14}
$$

Now, our task is to solve Subproblems in (10). For notation simplicity, we drop the loop index k in the following description.

z-Subproblem: The **z**-subproblem is obtained by keeping only terms involving **z** in (9) yielding

$$
\begin{aligned}
\mathbf{z}_{k+1} = \arg\min\{&-\mathbf{w}^T(\mathbf{s}-\mathbf{z}) - \mathbf{b}^T(\mathbf{r}-\boldsymbol{\Psi}^T\mathbf{z}) - \mathbf{h}^T(\mathbf{t}-\mathbf{z}) - \mathbf{v}^T(\mathbf{u}-\mathbf{D}\,\mathbf{z}) \\
&+ \frac{\mu}{2}\|\mathbf{s}-\mathbf{z}\|_2^2 + \frac{\rho}{2}\|\mathbf{r}-\boldsymbol{\Psi}^T\mathbf{z}\|_2^2 + +\frac{\xi}{2}\|\mathbf{t}-\mathbf{z}\|_2^2 + \frac{\kappa}{2}\|\mathbf{u}-\mathbf{D}\,\mathbf{z}\|_2^2\}.
\end{aligned}
\tag{15}
$$

Applying the first-order optimal condition to Problem (15), we have

$$
\begin{aligned}
[(\mu+\rho+\xi)\mathbf{I} + \kappa\mathbf{D}^T\mathbf{D}]\mathbf{z}_{k+1} = \boldsymbol{\Psi}^T(\rho\mathbf{r} - \mathbf{b}) &+ (\mu\mathbf{s} - \mathbf{w}) \\
&+ (\xi\mathbf{t} - \mathbf{h}) + \mathbf{D}^T(\kappa\mathbf{u} - \mathbf{v}).
\end{aligned}
\tag{16}
$$

Note that $\mathbf{D}^T\mathbf{D}$ is a circulant matrix, and thus the matrix $(\mu+\rho+\xi)\mathbf{I}+\kappa\mathbf{D}^T\mathbf{D}$ is diagonalizable. Therefore, the solution to Problem (16) can be found efficiently using fast Fourier transforms. In particular, a closed form solution is obtained by

$$
\mathbf{z}_{k+1} = \mathcal{F}^{-1}\left[\frac{\mathcal{F}(\mathbf{g})}{(\mu+\rho+\xi)\mathbf{I} + \kappa|\mathcal{F}(\mathbf{D})|^2}\right].
\tag{17}
$$

In (17), **g** is the right hand side of (16), i.e., $\mathbf{g} = \boldsymbol{\Psi}^T(\rho\mathbf{r} - \mathbf{b}) + (\mu\mathbf{s} - \mathbf{w}) + (\xi\mathbf{t} - \mathbf{h}) + \mathbf{D}^T(\kappa\mathbf{u}-\mathbf{v})$. The notation $\mathcal{F}(\cdot)$ is the 2D Fourier transform, whereas $\mathcal{F}^{-1}(\cdot)$ is the 2D inverse Fourier transform. The term $|\mathcal{F}(\mathbf{D})|^2$ is the square magnitude of the eigenvalues of the differential operator matrix **D**.

s-Subproblem: the **s**-subproblem is obtained by holding only **s**-related terms in (9):

$$\frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\,\mathbf{s}\|_2^2 - \mathbf{w}^T(\mathbf{s} - \mathbf{z}) + \frac{\mu}{2}\|\mathbf{s} - \mathbf{z}\|_2^2. \tag{18}$$

Performing the first-order optimal condition produces

$$(\boldsymbol{\Phi}^T\,\boldsymbol{\Phi} + \mu\mathbf{I})\,\mathbf{s} = \boldsymbol{\Phi}^T\,\mathbf{y} + \mathbf{w} + \mu\mathbf{z}. \tag{19}$$

As $\boldsymbol{\Phi}$ is a diagonal binary matrix, the solution for $\mathbf{s}_{k+1}$ is computed efficiently through an element-wise manner.

r-Subproblem: The **r**-subproblem is obtained by keeping only **r**-related terms in (9) yielding

$$\lambda\,\|\mathbf{r}\|_1 - \mathbf{b}^T(\mathbf{r} - \boldsymbol{\Psi}^T\,\mathbf{z}) + \frac{\rho}{2}\|\mathbf{r} - \boldsymbol{\Psi}^T\,\mathbf{z}\|_2^2. \tag{20}$$

Define an element-wise soft-thresholding operator,

$$\mathcal{T}(x,\tau) = \text{sign}(x)\max(|x| - \tau, 0) = \frac{x}{|x|}\,\max(|x| - \tau, 0). \tag{21}$$

This way, Problem (20) has a closed-form solution,

$$\mathbf{r}_{k+1} = \mathcal{T}(\boldsymbol{\Psi}^T\,\mathbf{z} + \frac{\mathbf{b}}{\rho}, \frac{\lambda}{\rho}) = \text{sign}(\boldsymbol{\Psi}^T\,\mathbf{z} + \frac{\mathbf{b}}{\rho})\,\max(\left|\boldsymbol{\Psi}^T\,\mathbf{z} + \frac{\mathbf{b}}{\rho}\right| - \frac{\lambda}{\rho}, 0). \tag{22}$$

Hereafter, the soft-thresholding operator $\mathcal{T}(\cdot)$ is used in element-wise manner.

t-Subproblem: likewise, the **t**-subproblem is obtained by holding only **t**-terms in (9) having

$$\beta\,\|\mathbf{t}\|_* - \mathbf{h}^T(\mathbf{t} - \mathbf{z}) + \frac{\xi}{2}\|\mathbf{t} - \mathbf{z}\|_2^2. \tag{23}$$

This problem can be solved efficiently through a singular value soft-thresholding (SVT) operator $\mathcal{S}(\cdot)$,

$$\mathbf{t}_{k+1} = \mathcal{S}(\mathbf{z} + \frac{\mathbf{h}}{\xi}, \frac{\beta}{\xi}). \tag{24}$$

Here, the SVT operator in the form $\mathcal{S}(\mathbf{x}, \tau)$ is computed by applying the soft-thresholding operator with τ to the singular values of the input matrix $\mathbf{x}$,

$$\mathcal{S}(\mathbf{x}, \tau) = \mathbf{u}\,\mathcal{T}(\boldsymbol{\lambda}, \tau)\,\mathbf{v}^T, \tag{25}$$

where $\mathbf{x} = \mathbf{u}\,\boldsymbol{\lambda}\,\mathbf{v}^T$ is the singular value decomposition of the input matrix $\mathbf{x}$.

u-Subproblem: the **u**-subproblem is obtained by keeping only **u**-related terms in (9) yielding

$$\gamma\,\|\mathbf{u}\|_1 - \mathbf{v}^T(\mathbf{u} - \mathbf{D}\,\mathbf{z}) + \frac{\kappa}{2}\|\mathbf{u} - \mathbf{D}\,\mathbf{z}\|_2^2. \tag{26}$$

The solution to Problem (26) is obtained by,

$$\mathbf{u}_{k+1} = \mathcal{T}(\mathbf{D}\,\mathbf{z} + \frac{\mathbf{v}}{\kappa}, \frac{\gamma}{\kappa}) = \text{sign}(\mathbf{D}\,\mathbf{z} + \frac{\mathbf{v}}{\kappa})\,\max(\left|\mathbf{D}\,\mathbf{z} + \frac{\mathbf{v}}{\kappa}\right| - \frac{\gamma}{\kappa}, 0). \tag{27}$$

In summary, the ADMM-based algorithm is presented in Algorithm 1. This algorithm starts by the initialization of the variables and iteratively updates them until convergence. Here, the stopping condition is satisfied if the relative change of the image $\mathbf{z}$ is negligible, i.e., $\|\mathbf{z}_{k+1} - \mathbf{z}_k\|_2 / \|\mathbf{z}_k\|_2 < \text{tol}$. The algorithm is computational efficient because the variable updates have closed-form solutions with element-wise computation.

Algorithm 1. LR-TV regularized depth estimation.

1: Input: $\mathbf{y}$, $\boldsymbol{\Phi}$
2: Initialize $\mathbf{z}_0 = \boldsymbol{\Phi}^T \mathbf{y}$, $\mathbf{s}_0 = \mathbf{z}_0$, $\mathbf{r}_0 = \boldsymbol{\Psi}^T \mathbf{z}_0$, $\mathbf{t}_0 = \mathbf{z}_0$, $\mathbf{u}_0 = \mathbf{D}\, \mathbf{z}_0$.
3: **repeat**
4:    Update $\mathbf{z}$ using (17).
5:    Update $\mathbf{s}$ by solving (19).
6:    Update $\mathbf{r}$, $\mathbf{t}$, and $\mathbf{u}$ using (22), (24), and (27), respectively.
7:    Update the multipliers $\mathbf{w}$, $\mathbf{b}$, $\mathbf{h}$, and $\mathbf{v}$ using (11)–(14), respectively.
8: **until** $\|\mathbf{z}_{k+1} - \mathbf{z}_k\|_2 / \|\mathbf{z}_k\|_2 < \text{tol}$.

---

4   Experimental Results

This section gives experimental evaluations on depth image benchmark datasets. Subsection 4.1 presents the setup for experiments, followed by imaging results, performance analysis and comparisons in Subsects. 4.2 and 4.3.

4.1   Experimental Setup

The proposed approach is evaluated using the Middlebury Stereo Dataset[1] [15], where the ground truth depth maps are available. To measure the performance accuracy, the peak signal-to-noise ratio (PSNR) is used (in dB):

$$\text{PSNR} = 10 \log_{10} \left(\frac{I_{\text{peak}}^2}{\text{MSE}} \right), \tag{28}$$

where I_{peak} is the peak intensity of reconstructed image I, and MSE is the mean-square-error between the reconstructed image I and the ground-truth image I_g defined as

$$\text{MSE} = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} |I(i,j) - I_g(i,j)|^2 . \tag{29}$$

The algorithm requires a set of parameters that need to be set appropriately. The basis matrix $\boldsymbol{\Psi}$ is constructed from wavelets with Daubechies 2 and 3 decomposition levels. Although the algorithm gives satisfactory results for a range of parameters, the typical settings for the regularization parameters are given in Table 1. The algorithm converges if the relative change of the reconstructed image is smaller than $\text{tol} = 10^{-4}$ (see Step 8 in Algorithm 1).
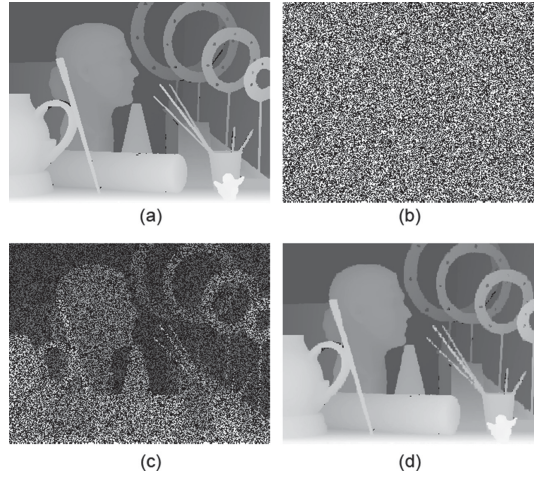
---

[1] http://vision.middlebury.edu/stereo/data/.

Fig. 1. Depth image reconstructed by the proposed LR-TV algorithm: (a) ground-truth depth image of art, (b) the random missing mask keeping only 50% pixels, (c) corrupted depth image with only 50% data measurements (PSNR = 7.08 dB), and (d) reconstructed image by the proposed LR-TV model (PSNR = 31.28 dB).
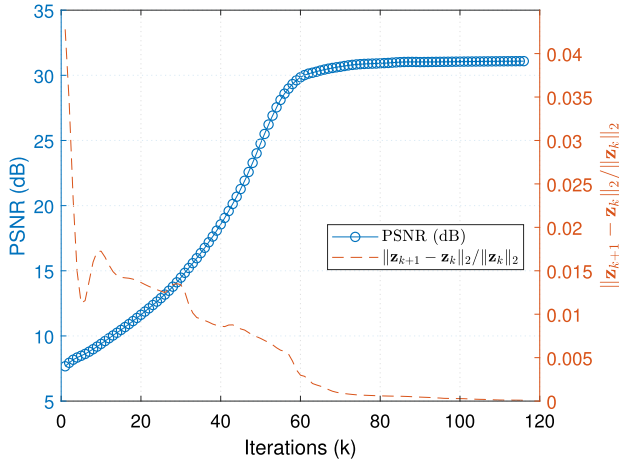

Fig. 2. PSNRs of the depth image reconstructed by the proposed LR-TV algorithm (solid line) and the relative change of the image $\mathbf{z}$, $\|\mathbf{z}_{k+1} - \mathbf{z}_k\|_2 / \|\mathbf{z}_k\|_2$ (dashed line) recorded during the minimization using 50% of total measurements.

Table 1. Parameter settings.

Parameters	Related terms	Values
λ	Sparsity penalty	4×10^{-5}
β	LR penalty	1×10^{-2}
γ	TV penalty	1×10^{-3}
μ	**s**-quadratic penalty	1×10^{-2}
ρ	**r**-quadratic penalty	1×10^{-3}
ξ	**t**-quadratic penalty	1×10^{-3}
κ	**u**-quadratic penalty	1×10^{-1}

4.2 Evaluation and Analysis of LR-TV Model

In the first experiment, we aim to evaluate the performance of the LR-TV model for depth image reconstruction. In doing so, Fig. 1(a) shows the ground-truth depth image of art with a size of $h \times w = 277 \times 347$ used to evaluate the performance. Now, only 50% measurements of the total pixels are randomly selected using the mask shown in Fig. 1(b). Because only 50% measurements are kept, the depth image is corrupted as shown in Fig. 1(c) with a PSNR = 7.08 dB. Using the 50% data measurements, the depth image reconstructed by the LR-TV model is presented in Fig. 1(d). It can be observed that the LR-TV model estimates the depth image well and yields a high quality image with a PSNR = 31.28 dB.

For further insights into the proposed algorithm, we report here the PSNR values of the estimated image and the relative change of the depth image, $\|\mathbf{z}_{k+1} - \mathbf{z}_k\|_2 / \|\mathbf{z}_k\|_2$, during the minimization. Figure 2 shows the PSNRs of the images estimated during minimization as a function of iterations. It can be observed from the figure that the image quality is enhanced during the update, and the quality in terms of PSNR is not changed much after 60 iterations. Furthermore, it can be seen that the algorithm is deemed to converge after 116 iterations when the relative change reaches 9.5×10^{-5}.

4.3 Comparison with Other Imaging Approaches

In the second experiment, we aim to compare the performance of the proposed LR-TV model with those by other imaging models. Two other imaging models are considered here. The first approach is the sparsity-based method for depth estimate that exploits only the sparseness of depth representation [11]. The second method considers both the sparsity and TV regularization for depth reconstruction [12]. We evaluate these models using two missing scenarios: random missing and entire (large) column missing. The random missing is generally related to compressive sensing data acquisition, whereas the large column missing is typically due to poor hardware performance. The ground-truth depth image and the corrupted images due to the masks for the two missing cases are shown
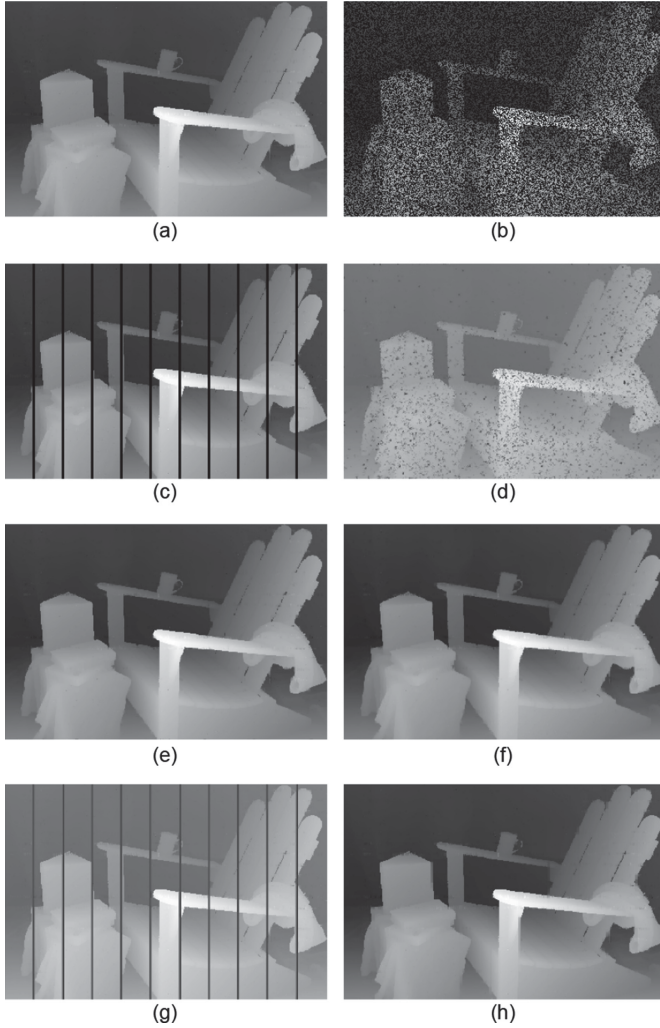
Fig. 3. Depth image reconstructed by the different imaging models with two missing data patterns: (a) the ground-truth depth image, (b) the corrupted image by the missing mask representing missing values (black pixels), (c) the corrupted image by the columns-missing mask representing missing values (black pixels); for the random missing case, the depth images reconstructed by (d) the sparsity-based model, (e) the sparsity-TV model, and (f) the LR-TV model; for the large column missing case, the depth images reconstructed by (g) the sparsity-based model, and (g) the proposed LR-TV model.

in Figs. 3(a)–(c), respectively. Here, the random missing mask is generated by keeping 50% pixels, and the missing column width is 5 pixels. The same input corrupted data and the same input parameters are used for all the evaluated models.

For the random missing case, Figs. 3(d)–(f) show the images recovered by the sparsity-based, the sparsity-TV, and the proposed LR-TV method, respectively. It can be observed that all the three models can reconstruct the depth images well, but the sparsity-TV and the proposed LR-TV produce better reconstructions compared with the sparsity method. For the large column missing case, the imaging result of the sparsity-based method is very poor, as demonstrated in Fig. 3(g). The sparsity-based model is unable to recover the missing column pixels. The proposed LR-TV, on the other hand, reconstructs the image well, as shown in Fig. 3(h).

To quantify the performances of the different imaging approaches, the PSNRs of the reconstructed images are computed and listed in Table 2. The most noticeable feature from the table is the considerable improvement of the sparsity + TV and the LR+TV over the sparsity model, especially for the large missing data case. Furthermore, for both missing data patterns, the proposed LR+TV model has the highest PSNR values among the tested methods; it archives 38.22 dB for the random missing data case, and 40.39 dB for large missing data case.

Table 2. PSNRs in dB of the depth images reconstructed by the different imaging models for two missing data patterns.

Imaging methods	Random missing pattern	Large missing pattern
Sparsity	22.07	17.65
Sparsity + Total variation	37.81	39.01
Low rank + Total variation	**38.22**	**40.39**

5   Conclusion

This paper presented a new LR-TV imaging model for depth image reconstruction from incomplete data measurements. The proposed approach formulates the task of depth image estimate as a joint LR and TV regularized minimization problem and proposes an algorithm based on ADMM to solve it, yielding a reconstructed depth image. By incorporating the LR and TV regularizers, the proposed model yields high quality image reconstruction for different missing data cases. Experimental evaluations are provided and the results show the proposed model is very promising that it enhances the quality of image estimation and outperforms the other evaluated state-of-the-art imaging models in terms of PSNR metric.

References

1. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
2. Cai, J.F., Candes, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. **20**(4), 1956–1982 (2010)
3. Camplani, M., Salgado, L.: Adaptive spatio-temporal filter for low-cost camera depth maps. In: IEEE International Conference on Emerging Signal Processing Applications, pp. 33–36 (2012)
4. Candes, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Signal Process. Mag. **25**(2), 21–30 (2008)
5. Chen, L., Lin, H., Li, S.: Depth image enhancement for Kinect using region growing and bilateral filter. In: IEEE International Conference on Pattern Recognition, pp. 3070–3073 (2012)
6. Chen, W., Fu, Z., Yang, D., Deng, J.: Single-image depth perception in the wild. In: Advances in Neural Information Processing Systems, pp. 730–738 (2016)
7. Chodosh, N., Wang, C., Lucey, S.: Deep convolutional compressed sensing for lidar depth completion. In: Jawahar, C.V., Li, H., Mori, G., Schindler, K. (eds.) ACCV 2018. LNCS, vol. 11361, pp. 499–513. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20887-5_31
8. Dimitrievski, M., Veelaert, P., Philips, W.: Learning morphological operators for depth completion. In: Blanc-Talon, J., Helbert, D., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2018. LNCS, vol. 11182, pp. 450–461. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01449-0_38
9. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
10. Han, D., Yuan, X.: A note on the alternating direction method of multipliers. J. Optim. Theory Appl. **155**(1), 227–238 (2012)
11. Hawe, S., Kleinsteuber, M., Diepold, K.: Dense disparity maps from sparse disparity measurements. In: IEEE International Conference on Computer Vision, pp. 2126–2133 (2011)
12. Liu, L.K., Chan, S.H., Nguyen, T.Q.: Depth reconstruction from sparse samples: representation, algorithm, and sampling. IEEE Trans. Image Process. **24**(6), 1983–1996 (2015)
13. Mallat, S.: A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way. Academic Press, Inc., Orlando (2008)
14. Saxena, A., Chung, S.H., Ng, A.Y.: Learning depth from single monocular images. In: Advances in Neural Information Processing Systems, pp. 1161–1168 (2006)
15. Scharstein, D., et al.: High-resolution stereo datasets with subpixel-accurate ground truth. In: Jiang, X., Hornegger, J., Koch, R. (eds.) GCPR 2014. LNCS, vol. 8753, pp. 31–42. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11752-2_3
16. Scherzer, O.: Handbook of Mathematical Methods in Imaging. Springer, New York (2011). https://doi.org/10.1007/978-0-387-92920-0
17. Sun, M., Ng, A.Y., Saxena, A.: Make3D: learning 3D scene structure from a single still image. IEEE Trans. Pattern Anal. Mach. Intell. **31**(5), 824–840 (2009)
18. Suwajanakorn, S., Hernandez, C., Seitz, S.M.: Depth from focus with your mobile phone. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3497–3506 (2015)

19. Zhang, R., Tsai, P.S., Cryer, J.E., Shah, M.: Shape from shading: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **21**(8), 690–706 (1999)
20. Zhu, Z.Y., Chan, S., Shum, H.Y.: Real-time depth image acquisition and restoration for image based rendering and processing systems. J. Sig. Process. Syst. **79**, 1–18 (2013)
21. Zou, N., Xiang, Z., Chen, Y.: RSDCN: a road semantic guided sparse depth completion network. Neural Process. Lett. **51**(3), 2737–2749 (2020). https://doi.org/10.1007/s11063-020-10226-7