

# Sparse Reconstruction of Depth Data: Representation, Algorithm, and Sampling

Lee-Kang Liu, *Student Member, IEEE*, Stanley H. Chan, *Member, IEEE* and Truong Q. Nguyen, *Fellow, IEEE*

**Abstract**—The rapid development of 3D technology and computer vision applications have motivated a thrust of methodologies for depth acquisition and estimation. However, most existing hardware and software methods have limited performance due to poor depth precision, low resolution and high computational cost. In this paper, we present a computationally efficient method to recover dense depth maps from sparse measurements. We make three contributions. First, we provide empirical evidence that depth maps can be encoded much more sparsely than natural images by using common dictionaries such as wavelets and contourlets. We also show that a combined wavelet-contourlet dictionary achieves better performance than using either dictionary alone. Second, we propose an alternating direction method of multipliers (ADMM) to achieve fast reconstruction. A multi-scale warm start procedure is proposed to speed up the convergence. Third, we propose a two-stage randomized sampling scheme to optimally choose the sampling locations, thus maximizing the reconstruction performance for any given sampling budget. Experimental results show that the proposed method produces high quality dense depth estimates, and is robust to noisy measurements. Applications to real data in stereo matching are demonstrated.

**Index Terms**—Sparse reconstruction, random sampling, wavelet, contourlet, disparity estimation, alternating direction method of multipliers, compressed sensing

## I. INTRODUCTION

The rapid development of 3D technology has created a new wave of visualization and sensing impacts to the digital signal processing community. From remote sensing [1] to preserving historical heritages [2], and from rescue [3] to 3D laparoscopic surgery [4], [5], the footprints of 3D have been influencing a broad aspect of the technological frontiers.

The successful development of 3D signal processing is fundamentally linked to a system's ability to acquire depth. To date, there are two major classes of depth acquisition techniques: hardware devices and computational procedures. Hardware devices are usually equipped with active sensors, for example, time-of-flight camera [6] and LiDAR [7]. While producing high quality depth maps, these hardware systems have high instrumentation cost. Moreover, the data acquisition time of the devices is long (10 fps as opposed to 60fps

on standard cameras [8]). Although speeding up is possible, spatial resolution has to be traded off in return.

An alternative solution to acquiring depth is to estimate the depth through a set of computational procedures. This class of computational methods, broadly referred to as disparity estimation algorithms [9–12], infer the depth by estimating the disparities between a pair of stereo images. However, most existing algorithms are based on calculating the spatial correspondences between the best matching features [13], [14]. Often times, the features could be sensitive to illumination, noise, stereo camera alignments, and other camera factors. Thus, the effective number of reliable feature matching points is much fewer than the dimension of the image [15], [16].

### A. Scope and Contributions

The objective of this paper is to present a sparse sampling and reconstruction framework to improve and speed up the depth acquisition process. Motivated by the literature of compressed sensing, we propose to sparsely sample a subset of depth values and reconstruct the dense depth map.

The merits of the new framework are two-fold. From the hardware perspective, the new method allows us to acquire fewer samples so that the acquisition time is reduced. From the computational perspective, the new method allows us to estimate a dense depth map with fewer feature points.

The three major contributions of this paper are as follows.

1) *Representation* (Section III). Unlike natural images which cannot be well approximated by a sparse set of dictionary atoms (e.g., wavelets), depth maps are significantly more appropriate to be encoded. We provide empirical evidence for this observation. Moreover, we show that a combined dictionary of wavelets and contourlets can further improve the reconstruction quality.

2) *Algorithm* (Section IV). We propose a fast numerical algorithm to reconstruct the depth map from a small set of samples. The reconstruction algorithm is based on the idea of alternating direction method of multipliers (ADMM). We derive novel splitting strategies that allow one to solve a sequence of parallelizable subproblems. We also present a multiscale implementation that utilizes the depth structures for efficient warm starts.

3) *Sampling* (Section V). We propose an efficient spatial sampling strategy that maximizes the reconstruction performance. In particular, we show that for a fixed sampling budget, a high quality sampling pattern can be obtained by allocating random samples with probabilities in proportional to the magnitudes of the depth gradients.

L. Liu and T. Nguyen are with Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093, USA. Emails: l7liu@ucsd.edu and nguyent@ece.ucsd.edu

S. Chan is with School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. Email: schan@seas.harvard.edu

This work was supported in part by a Croucher Foundation Post-doctoral Research Fellowship, and in part by the National Science Foundation under grant CCF-1065305. Preliminary material in this paper was presented at the 39th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, May 2014.

## B. Related Work

The focus of this paper lies in the intersection of two subjects: depth enhancement and compressed sensing. Prior works and the limitations are summarized as follows.

The goal of depth enhancement is to improve the resolution of a depth map. Some classical examples include Markov Random Field (MRF) [17], bilateral filter [18], and other approaches [19], [20]. One limitation of these methods is that the low-resolution depth maps are sampled uniformly. Also, the corresponding color images are usually assumed given. These are in contrast to the proposed method, which is applicable to any non-uniformly sampled low-resolution depth map and do not require additional color images. Thus, the new method allows for a greater flexibility for the enhancement.

Compressed sensing is a technique which recovers a signal from a small set of linear measurements [21]. In many cases, compressed sensing is applied to *natural images* – signals that are believed to exhibit sparse structures in wavelet domains. However, as will be discussed in Section III of this paper, natural images are indeed *not* sparse. If we compare natural images to depth maps, the latter would show a much sparser structure than the former. Furthermore, the theory of combined bases [22], [23] shows that a pair of incoherent bases are typically more effective for sparse reconstruction. Yet, the application of these theories to depth maps is not fully explored.

The most relevant work to this paper is a sparse reconstruction framework proposed in [24]. As a comparison, the method presented in this paper is more advanced in two aspects. First, instead of using a subgradient descent algorithm for the reconstruction, we propose a new ADMM algorithm for faster convergence (Section IV). Second, we present a sampling scheme to choose optimal sampling patterns that improve the depth reconstruction (Section V).

Finally, we note that there is a recent research interest in computational photography techniques to achieve fast depth acquisition, *e.g.*, [25], [26]. However, the problem settings of these work involve hardware designs and are thus different from this paper.

The rest of the paper is organized as follows. After elaborating the problem and clarifying notations in Section II, we discuss the representation of depth maps in Section III. A fast reconstruction algorithm is presented in Section IV. In Section V we discuss the design of optimal sampling patterns. Experimental results are shown in Section VI, and a concluding remark is given in Section VII.

## II. NOTATIONS AND PROBLEM FORMULATION

The purpose of this section is to introduce notations and elaborate the problem formulation.

### A. Depth and Disparity

The type of data that we are interested in studying is the depth map. Depth can be directly measured using active sensors, or *inferred* from the disparity of a pair of stereo images, because depth is inversely proportional to the disparities of a point in the stereo pair [27]. Since this correspondence is unique, in the rest of the paper we shall use depth and disparity interchangeably.

### B. Sparse Reconstruction Formulation

In this paper, we let  $\mathbf{x} \in \mathbb{R}^N$  be an  $N \times 1$  vector representing a disparity map. For simplicity we assume that  $\mathbf{x}$  is normalized so that  $0 \leq \mathbf{x} \leq 1$ .

To acquire a set of spatial samples of  $\mathbf{x}$ , we define a diagonal matrix  $\mathbf{S} \in \mathbb{R}^{N \times N}$  with the  $(i, i)$ th entry being

$$S_{ii} \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if the } i\text{th pixel is sampled,} \\ 0, & \text{if the } i\text{th pixel is not sampled.} \end{cases}$$

The sampled disparity map is then defined as

$$\mathbf{b} = \mathbf{S}\mathbf{x}. \quad (1)$$

The problem is to reconstruct  $\mathbf{x}$  from some measurement  $\mathbf{b}$ .

To properly formulate the reconstruction problem, we assume that the disparity map can be efficiently represented as a linear combination of basis vectors  $\{\varphi_j\}_{j=1}^M$ :

$$\mathbf{x} = \sum_{j=1}^M \langle \mathbf{x}, \varphi_j \rangle \varphi_j, \quad (2)$$

where  $\langle \cdot, \cdot \rangle$  denotes the standard inner product. Defining  $\alpha_j \stackrel{\text{def}}{=} \langle \mathbf{x}, \varphi_j \rangle$  as the  $j$ th basis coefficient,  $\boldsymbol{\alpha} \stackrel{\text{def}}{=} [\alpha_1, \dots, \alpha_M]^T$ , and  $\boldsymbol{\Phi} \stackrel{\text{def}}{=} [\varphi_1, \dots, \varphi_M]$ , the relationship in (2) can be equivalently written as  $\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\alpha}$ .

The reconstruction problem can be posed as an optimization problem in which the goal is to seek a sparse vector  $\boldsymbol{\alpha} \in \mathbb{R}^M$  such that the observed samples  $\mathbf{b}$  are best approximated. Mathematically, we consider the problem

$$\underset{\boldsymbol{\alpha}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{S}\boldsymbol{\Phi}\boldsymbol{\alpha} - \mathbf{b}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1, \quad (3)$$

where  $\lambda > 0$  is a regularization parameter, and  $\|\cdot\|_1$  is the  $\ell_1$ -norm of a vector.

In this paper, we are mainly interested in two types of  $\boldsymbol{\Phi}$  — the wavelet frame and the contourlet frame [28]. Frames are generalizations of the standard bases in which  $M$ , the number of bases, can be more than  $N$ , the dimension of  $\mathbf{x}$ . Moreover, for any frame  $\boldsymbol{\Phi}$ , it holds that  $\boldsymbol{\Phi}\boldsymbol{\Phi}^T = \mathbf{I}$ . Therefore,  $\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\alpha}$  if and only if  $\boldsymbol{\alpha} = \boldsymbol{\Phi}^T \mathbf{x}$ . Using this result, we can equivalently express (3) as

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{S}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\boldsymbol{\Phi}^T \mathbf{x}\|_1. \quad (4)$$

As discussed in [24], (4) is not an effective formulation because the  $\ell_1$  norm penalizes *both* the approximation (lowpass) and the detailed (highpass) coefficients. In reality, since disparity maps are mostly piecewise linear functions, the lowpass coefficients should be maintained whereas the highpass coefficients are desirable to be sparse. To this end, we introduce a binary diagonal matrix  $\mathbf{W} \in \mathbb{R}^{M \times M}$  where the  $(i, i)$ th entry is 0 if  $i$  is an index in the lowest passband, and is 1 otherwise. Consequently, we modify the optimization problem as

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{S}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{W}\boldsymbol{\Phi}^T \mathbf{x}\|_1. \quad (5)$$

Finally, it is desirable to further enforce smoothness of the reconstructed disparity map. Therefore, we introduce a total variation penalty so that the problem becomes

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{S}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{W}\Phi^T \mathbf{x}\|_1 + \beta \|\mathbf{x}\|_{TV}. \quad (6)$$

Here, the total variation norm is defined as

$$\|\mathbf{x}\|_{TV} \stackrel{\text{def}}{=} \sum_{i=1}^N \sqrt{[\mathbf{D}_x \mathbf{x}]_i^2 + [\mathbf{D}_y \mathbf{x}]_i^2}, \quad (7)$$

where  $\mathbf{D} = [\mathbf{D}_x; \mathbf{D}_y]$  is the first-order finite difference operator in the horizontal and vertical directions. The above definition of total variation is known as the isotropic total variation. The same formulation holds for anisotropic total variation, in which  $\|\mathbf{x}\|_{TV} = \|\mathbf{D}_x \mathbf{x}\|_1 + \|\mathbf{D}_y \mathbf{x}\|_1$ .

The problem in (6) is generalizable to take into account of a combination of  $L$  dictionaries. In this case, one can consider a sum of  $L$  penalty terms as

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{S}\mathbf{x} - \mathbf{b}\|_2^2 + \sum_{\ell=1}^L \lambda_{\ell} \|\mathbf{W}_{\ell} \Phi_{\ell}^T \mathbf{x}\|_1 + \beta \|\mathbf{x}\|_{TV}. \quad (8)$$

For example, in the case of combined wavelet and contourlet dictionaries, we let  $L = 2$ .

### III. SPARSE REPRESENTATION OF DISPARITY MAP

The choice of the dictionary  $\Phi$  in (8) is an important factor for reconstruction performance. In this section we discuss the general representation problem of disparity maps. We show that disparity maps can be represented more sparsely than natural images. We also show that a combined wavelet-contourlet dictionary is more effective in representing disparity maps than using the wavelet dictionary alone.

#### A. Natural Images vs Depth Data

Seeking effective representations for *natural images* is a well-studied subject in image processing. In general, high quality representations can be obtained through dictionary learning, *e.g.*, K-SVD [29], [30]. However, for computational efficiency it is often more preferred to use pre-defined dictionaries, *e.g.*, wavelet [31], contourlet [28], [32], and many others [33–35]. In this paper we shall focus on the latter.

Representation for *disparity maps* is analogous to that of natural images. However, the effectiveness of these dictionaries for disparity maps is less understood. In particular, it is unclear how sparse can a predefined dictionary encode disparity maps as compared to natural images.

To address this question, we consider a  $128 \times 128$  cropped patch from a gray-scaled image and the corresponding patch in the disparity map. For each of the image and the disparity, we apply the wavelet transform with Daubechies 5/3 filter and 5 decomposition levels. Then, we truncate the wavelet coefficients to the leading 5% coefficients with the largest magnitudes. The reconstructed patches are compared and the results are shown in Figure 1.

The result indicates that for the same number of wavelet coefficients, the disparity map can be approximated with

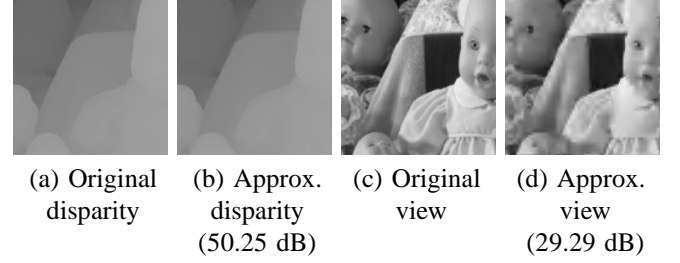


Fig. 1: PSNR values of approximating a disparity patch and a image patch using the leading 5% of the wavelet coefficients.

significantly lower approximation error than the image. While such result is not surprising, the big difference in the PSNRs provides evidence that sparse reconstruction of disparity maps should achieve better results than that of natural images.

#### B. Wavelet vs Contourlet

The above results indicate that wavelets are efficient representations for disparity maps. It is also important to ask if some of the dictionaries would do better than other dictionaries. In this section, we discuss how a combined wavelet-contourlet dictionary can improve the wavelet dictionary.

1) *Evaluation Metric*: To compare the performance of two dictionaries, it is necessary to first specify which metric we should use. For the purpose of this paper, we compare the mean squared error of the reconstructed disparity maps obtained by feeding different dictionaries into (8). For any fixed sampling pattern  $\mathbf{S}$ , we say that a dictionary  $\Phi_1$  is better than another dictionary  $\Phi_2$  if and only if the reconstruction result using  $\Phi_1$  has a lower mean squared error than using  $\Phi_2$ , for the best choice of parameters  $\lambda_1$ ,  $\lambda_2$  and  $\beta$ .

2) *Comparison Results*: We synthetically create a gray-scaled image consisting of a triangle overlapping with an ellipse to simulate a disparity map. We choose a uniformly random sampling pattern  $\mathbf{S}$  so that there is no bias caused by a particular sampling pattern. The  $(i, i)$ th entry of  $\mathbf{S}$  is defined to follow a Bernoulli process of fixed probability  $0 \leq \xi \leq 1$ , *i.e.*,  $S_{ii} \sim \text{Bernoulli}(\xi)$ . The probability  $\xi$  is referred to as the *sampling ratio*.

The goal of this comparison test is to evaluate the performance of a wavelet dictionary and a combined wavelet-contourlet dictionary. For the best performance of both dictionaries, we choose  $\lambda_1 = 4 \times 10^{-5}$ ,  $\lambda_2 = 0$  for the wavelet dictionary, and  $\lambda_1 = 4 \times 10^{-5}$ ,  $\lambda_2 = 2 \times 10^{-4}$  for the combined dictionary. For both cases, the total variation parameter is  $\beta = 2 \times 10^{-3}$ .

Using the proposed ADMM algorithm (See Section IV), we plot the performance of the reconstruction result as a function of the sampling ratio. For each point of the sampling ratio, we perform a Monte-Carlo simulation over 20 independent trials to reduce the fluctuation caused by the randomness in the sampling pattern. The result in Figure 2 indicates that the combined dictionary is consistently better than the wavelet dictionary alone. A snapshot of the result at  $\xi = 0.1$  is shown in (3). As observed, the reconstruction along the edges of the ellipse is better in the combined dictionary than the wavelet.

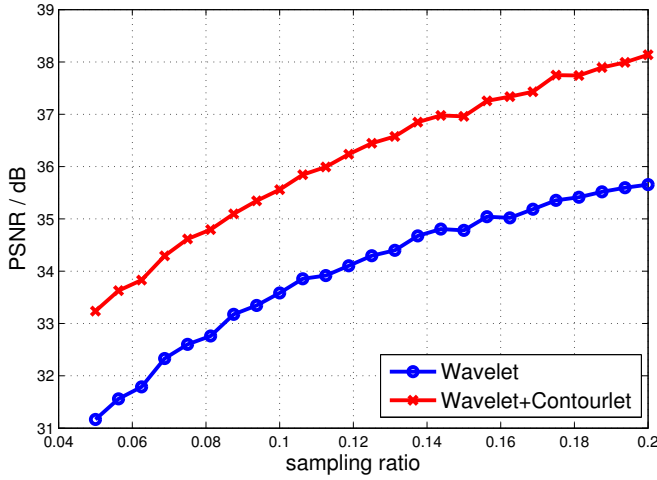


Fig. 2: ADMM reconstruction result as a function of sampling ratio  $\xi$ . Each point on the curves are averaged over 20 independent Monte-Carlo trials. The PSNR evaluates the performance of solving (8) using different combinations of dictionaries.



(a) Wavelet, 33.54 dB (b) Combined, 35.76 dB

Fig. 3: Snapshot of the comparison between wavelet dictionary and a combined wavelet-contourlet dictionary at  $\xi = 0.1$ .

#### IV. RECONSTRUCTION ALGORITHM

In this section we present an alternating direction method of multipliers (ADMM) algorithm to solve (8). For brevity we skip the introduction of ADMM algorithm because comprehensive tutorials are easily accessible [36], [37]. Instead, we highlight the unique contributions of this paper, which include a new operator splitting strategy that has not been discussed in the literature, and a multiscale implementation that utilizes image structures to facilitate warm starts.

For notational simplicity we consider a single dictionary so that  $L = 1$ . Generalization to  $L > 1$  is straight forward. Also, in our derivation we focus on the anisotropic total variation so that  $\|\mathbf{x}\|_{TV} = \|\mathbf{D}_x \mathbf{x}\|_1 + \|\mathbf{D}_y \mathbf{x}\|_1$ . Extension to isotropic total variation follows the same idea as presented in [5].

##### A. ADMM and Operator Splitting

A central question about ADMM algorithms is which of the variables should be splitted so that the subsequent subproblems can be efficiently solved. Inspecting (8), we observe that there are many possible choices. For example, we could split the quadratic term in (8) by defining an auxiliary variable  $\mathbf{u} = \mathbf{S}\mathbf{x}$ , or we could keep the quadratic term without a split. In what follows, we present an overview of our proposed splitting method and discuss the steps in subsequent subsections.

We begin the ADMM algorithm by introducing three auxiliary variables  $\mathbf{u}_1 = \Phi_\ell \mathbf{x}$ ,  $\mathbf{u}_2 = \mathbf{x}$ , and  $\mathbf{v} = \mathbf{D}\mathbf{x}$ . Consequently, we rewrite the optimization problem as

$$\begin{aligned} & \underset{\mathbf{x}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{v}}{\text{minimize}} && \frac{1}{2} \|\mathbf{b} - \mathbf{S}\mathbf{u}_2\|^2 + \lambda_\ell \|\mathbf{W}_\ell \mathbf{u}_1\|_1 + \beta \|\mathbf{v}\|_1 \\ & \text{subject to} && \mathbf{u}_1 = \Phi_\ell^T \mathbf{x}, \quad \mathbf{u}_2 = \mathbf{x}, \quad \mathbf{v} = \mathbf{D}\mathbf{x}. \end{aligned} \quad (9)$$

The ADMM algorithm is a computational procedure to find a stationary point of (9). The idea is to consider the augmented Lagrangian function defined as

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{v}, \mathbf{y}_1, \mathbf{y}_2, \mathbf{z}) &= \frac{1}{2} \|\mathbf{b} - \mathbf{S}\mathbf{u}_2\|^2 + \lambda_\ell \|\mathbf{W}_\ell \mathbf{u}_1\|_1 + \beta \|\mathbf{v}\|_1 \\ &\quad - \mathbf{y}_1^T (\mathbf{u}_1 - \Phi_\ell^T \mathbf{x}) - \mathbf{y}_2^T (\mathbf{u}_2 - \mathbf{x}) - \mathbf{z}^T (\mathbf{v} - \mathbf{D}\mathbf{x}) \\ &\quad + \frac{\rho_1}{2} \|\mathbf{u}_1 - \Phi_\ell^T \mathbf{x}\|^2 + \frac{\rho_2}{2} \|\mathbf{u}_2 - \mathbf{x}\|^2 + \frac{\gamma}{2} \|\mathbf{v} - \mathbf{D}\mathbf{x}\|^2. \end{aligned} \quad (10)$$

Here, the vectors  $\mathbf{y}_1$ ,  $\mathbf{y}_2$ ,  $\mathbf{z}$  are the Lagrangian multipliers, and  $\rho_1$ ,  $\rho_2$ ,  $\gamma$  are the penalty parameters. The stationary point of the augmented Lagrangian function can be determined by solving the following sequence of subproblems

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \underset{\mathbf{x}}{\text{argmin}} \mathcal{L}(\mathbf{x}, \mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)}, \mathbf{v}^{(k)}, \mathbf{y}_1^{(k)}, \mathbf{y}_2^{(k)}, \mathbf{z}^{(k)}), \\ \mathbf{u}_1^{(k+1)} &= \underset{\mathbf{u}_1}{\text{argmin}} \mathcal{L}(\mathbf{x}^{(k+1)}, \mathbf{u}_1, \mathbf{u}_2^{(k)}, \mathbf{v}^{(k)}, \mathbf{y}_1^{(k)}, \mathbf{y}_2^{(k)}, \mathbf{z}^{(k)}), \\ \mathbf{u}_2^{(k+1)} &= \underset{\mathbf{u}_2}{\text{argmin}} \mathcal{L}(\mathbf{x}^{(k+1)}, \mathbf{u}_1^{(k+1)}, \mathbf{u}_2, \mathbf{v}^{(k)}, \mathbf{y}_1^{(k)}, \mathbf{y}_2^{(k)}, \mathbf{z}^{(k)}), \\ \mathbf{v}^{(k+1)} &= \underset{\mathbf{v}}{\text{argmin}} \mathcal{L}(\mathbf{x}^{(k+1)}, \mathbf{u}_1^{(k+1)}, \mathbf{u}_2^{(k+1)}, \mathbf{v}, \mathbf{y}_1^{(k)}, \mathbf{y}_2^{(k)}, \mathbf{z}^{(k)}), \end{aligned}$$

and update the Lagrange multipliers as

$$\mathbf{y}_1^{(k+1)} = \mathbf{y}_1^{(k)} - \rho_1 (\mathbf{u}_1^{(k+1)} - \Phi_\ell^T \mathbf{x}^{(k+1)}), \quad (11a)$$

$$\mathbf{y}_2^{(k+1)} = \mathbf{y}_2^{(k)} - \rho_2 (\mathbf{u}_2^{(k+1)} - \mathbf{x}^{(k+1)}), \quad (11b)$$

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} - \gamma (\mathbf{v}^{(k+1)} - \mathbf{D}\mathbf{x}^{(k+1)}). \quad (11c)$$

In the following subsection we discuss how each subproblem is solved.

##### B. Subproblems

1) *x-subproblem*: The  $\mathbf{x}$ -subproblem is obtained by dropping terms that do not involve  $\mathbf{x}$  in (10). This yields

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \underset{\mathbf{x}}{\text{argmin}} -\mathbf{y}_1^T (\mathbf{u}_1 - \Phi_\ell^T \mathbf{x}) - \mathbf{y}_2^T (\mathbf{u}_2 - \mathbf{x}) \\ &\quad - \mathbf{z}^T (\mathbf{v} - \mathbf{D}\mathbf{x}) + \frac{\rho_1}{2} \|\mathbf{u}_1 - \Phi_\ell^T \mathbf{x}\|^2 \\ &\quad + \frac{\rho_2}{2} \|\mathbf{u}_2 - \mathbf{x}\|^2 + \frac{\gamma}{2} \|\mathbf{v} - \mathbf{D}\mathbf{x}\|^2. \end{aligned} \quad (12)$$

Problem (12) can be solved by considering the first-order optimality condition, which yields a normal equation

$$\begin{aligned} & (\rho_1 \Phi_\ell \Phi_\ell^T + \rho_2 \mathbf{I} + \gamma \mathbf{D}^T \mathbf{D}) \mathbf{x}^{(k+1)} \\ &= \Phi_\ell (\rho_1 \mathbf{u}_1 - \mathbf{y}_1) + (\rho_2 \mathbf{u}_2 - \mathbf{y}_2) + \mathbf{D}^T (\gamma \mathbf{v} - \mathbf{z}). \end{aligned} \quad (13)$$

The matrix in (13) can be simplified as  $(\rho_1 + \rho_2) \mathbf{I} + \gamma \mathbf{D}^T \mathbf{D}$ , because for any frame  $\Phi_\ell$ , it holds that  $\Phi_\ell \Phi_\ell^T = \mathbf{I}$ . Now, since the matrix  $\mathbf{D}^T \mathbf{D}$  is a circulant matrix, the matrix  $(\rho_1 +$

$\rho_2)\mathbf{I} + \gamma\mathbf{D}^T\mathbf{D}$  is diagonalized by the Fourier transform. This leads to a closed form solution as

$$\mathbf{x}^{(k+1)} = \mathcal{F}^{-1} \left[ \frac{\mathcal{F}(\text{RHS})}{(\rho_1 + \rho_2)\mathbf{I} + \gamma|\mathcal{F}(\mathbf{D})|^2} \right], \quad (14)$$

where RHS denotes the right hand side of (13),  $\mathcal{F}(\cdot)$  denotes the 2-dimensional Fourier transform operation,  $\mathcal{F}^{-1}(\cdot)$  denotes the 2-dimensional inverse Fourier transform operation, and  $|\mathcal{F}(\mathbf{D})|^2$  denotes the magnitude square of the eigenvalues of the differential operator  $\mathbf{D}$ .

*Remark 1:* If we do not split the quadratic function  $\|\mathbf{b} - \mathbf{S}\mathbf{u}_2\|^2$  using  $\mathbf{u}_2 = \mathbf{x}$ , then the identity matrix  $\rho_2\mathbf{I}$  in (13) would become  $\rho_2\mathbf{S}^T\mathbf{S}$ . Since  $\mathbf{S}$  is a diagonal matrix containing 1's and 0's, the matrix  $\rho_1\Phi_\ell\Phi_\ell^T + \rho_2\mathbf{S}^T\mathbf{S} + \gamma\mathbf{D}^T\mathbf{D}$  is not diagonalizable using the Fourier transform.

2)  $\mathbf{u}_1$ -subproblem: The  $\mathbf{u}_1$ -subproblem is given by

$$\min_{\mathbf{u}_1} \lambda_\ell \|\mathbf{W}_\ell \mathbf{u}_1\|_1 - \mathbf{y}_1^T (\mathbf{u}_1 - \Phi_\ell^T \mathbf{x}) + \frac{\rho_1}{2} \|\mathbf{u}_1 - \Phi_\ell^T \mathbf{x}\|_2^2. \quad (15)$$

Since  $\mathbf{W}_\ell$  is a diagonal matrix, (15) is a separable optimization consisting of a sum of scalar problems. By using the standard shrinkage formula [5], it is easy to show that the closed-form solution of (15) exists and is given by

$$\mathbf{u}_1^{(k+1)} = \max \left( \left| \alpha_\ell + \frac{\mathbf{y}_1}{\rho_1} \right| - \frac{\lambda_\ell \mathbf{w}_\ell}{\rho_1}, 0 \right) \cdot \text{sign} \left( \alpha_\ell + \frac{\mathbf{y}_1}{\rho_1} \right). \quad (16)$$

where  $\alpha_\ell = \Phi_\ell^T \mathbf{x}$ , and  $\mathbf{w}_\ell = \text{diag}(\mathbf{W}_\ell)$  is a vector of the diagonal entries of  $\mathbf{W}_\ell$ .

*Remark 2:* If we do not split  $\mathbf{u}_1 = \Phi_\ell^T \mathbf{x}$ , then the  $\mathbf{u}_1$ -subproblem is not separable and hence the shrinkage formula cannot be applied. Moreover, if we split  $\mathbf{u}_1 = \mathbf{W}_\ell \Phi_\ell^T \mathbf{x}$ , i.e., include  $\mathbf{W}_\ell$ , then the  $\mathbf{x}$ -subproblem will contain  $\Phi_\ell \mathbf{W}_\ell \Phi_\ell^T$ , which is not diagonalizable using the Fourier transform.

3)  $\mathbf{u}_2$ -subproblem: The  $\mathbf{u}_2$ -subproblem is the standard quadratic minimization problem:

$$\min_{\mathbf{u}_2} \frac{1}{2} \|\mathbf{S}\mathbf{u}_2 - \mathbf{b}\|^2 - \mathbf{y}_2^T (\mathbf{u}_2 - \mathbf{x}) + \frac{\rho_2}{2} \|\mathbf{u}_2 - \mathbf{x}\|^2. \quad (17)$$

Taking the first-order optimality yields a normal equation

$$(\mathbf{S}^T \mathbf{S} + \rho_2 \mathbf{I}) \mathbf{u}_2 = (\mathbf{S}^T \mathbf{b} + \mathbf{y}_2 - \rho_2 \mathbf{x}). \quad (18)$$

Since  $\mathbf{S}$  is a diagonal binary matrix, (18) can be evaluated via an elementwise computation.

*Remark 3:* (18) shows that our splitting strategy of using  $\mathbf{u}_2 = \mathbf{x}$  is particularly efficient because  $\mathbf{S}$  is a diagonal matrix. If  $\mathbf{S}$  is a general matrix, e.g., i.i.d. Gaussian sampling matrix in the compressed sensing literature [38], then solving (18) will be less efficient.

4)  $\mathbf{v}$ -subproblem: The  $\mathbf{v}$ -subproblem is the standard total variation problem:

$$\min_{\mathbf{v}} \beta \|\mathbf{v}\|_1 - \mathbf{z}^T (\mathbf{v} - \mathbf{D}\mathbf{x}) + \frac{\gamma}{2} \|\mathbf{v} - \mathbf{D}\mathbf{x}\|^2. \quad (19)$$

The solution is given by

$$\mathbf{v}^{(k+1)} = \max \left( \left| \mathbf{D}\mathbf{x} + \frac{\mathbf{z}}{\gamma} \right| - \frac{\beta}{\gamma}, 0 \right) \cdot \text{sign} \left( \mathbf{D}\mathbf{x} + \frac{\mathbf{z}}{\gamma} \right). \quad (20)$$

The overall ADMM algorithm is shown in Algorithm 1.

---

#### Algorithm 1 ADMM Algorithm

---

**Require:**  $\mathbf{b}, \mathbf{S}$

- 1:  $\mathbf{x}^{(0)} = \mathbf{S}\mathbf{b}$ ,  $\mathbf{u}_1^{(0)} = \Phi_\ell^T \mathbf{x}^{(0)}$ ,  $\mathbf{u}_2^{(0)} = \mathbf{x}^{(0)}$ ,  $\mathbf{v}^{(0)} = \mathbf{D}\mathbf{x}^{(0)}$
  - 2: **while**  $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2 / \|\mathbf{x}^{(k)}\|_2 \geq \text{tol}$  **do**
  - 3:   Solve  $\mathbf{x}$ -subproblem by (14).
  - 4:   Solve  $\mathbf{u}_1$ ,  $\mathbf{u}_2$  and  $\mathbf{v}$  subproblems by (16), (18) and (20).
  - 5:   Update multipliers by (11a), (11b) and (11c).
  - 6: **end while**
  - 7: **return**  $\mathbf{x}^* \leftarrow \mathbf{x}^{(k+1)}$
- 

#### C. Parameters

The parameters  $\beta$  and  $\lambda_\ell$  are chosen empirically, and the optimal values are problem specific. Here, we provide the configurations for typical problem settings.

1) *Choosing  $\beta$ :* The parameter  $\beta$  determines the emphasis of the total variation penalty. Large  $\beta$  results in smooth depth maps. Our choice of  $\beta$  is based on empirical observations. In Figure 4, we show the MSE results by sweeping through  $\beta \in [10^{-5}, 10^{-1}]$ . As shown, the minimum of MSE lies in the range of  $\beta \in [10^{-3}, 10^{-2}]$ . We also find that the MSE is not sensitive to sampling ratio. Therefore, in the rest of the paper we choose  $\beta \in [10^{-3}, 10^{-2}]$ .

2) *Choosing  $\lambda_\ell$ :* Similar to  $\beta$ , the parameters  $\lambda_\ell$  can be chosen according to Figure 4. More specifically, we observe that an acceptable range of  $\lambda_1$  is  $[3.1623 \times 10^{-5}, 10^{-3}]$ . For  $\lambda_2$ , we keep  $\beta = 2 \times 10^{-3}$ ,  $\lambda_1 = 4 \times 10^{-5}$ , and search for the range that produces the minimum MSE. The optimal value of  $\beta$  that works robustly is  $\lambda_2 = 2 \times 10^{-4}$ .

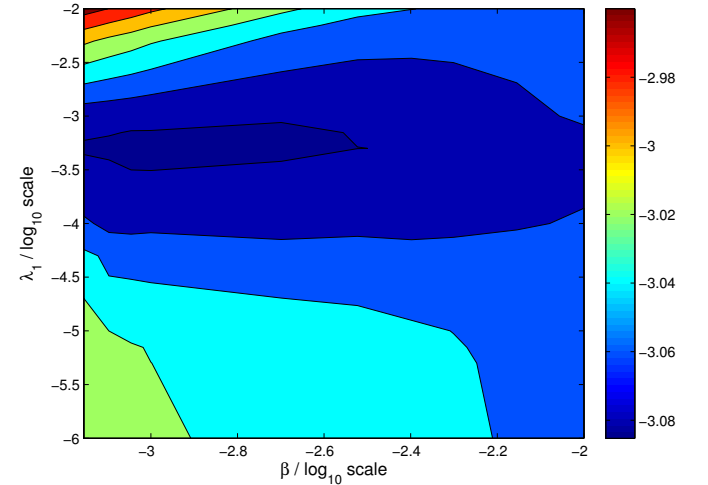


Fig. 4: Log scale MSE contours with sweeping parameters  $\lambda_1$  and  $\beta$  using fixed 10% uniformly random samples. The acceptable range for  $\beta$  is  $[10^{-3}, 10^{-2}]$ , and the range for  $\lambda_1$  is  $[10^{-4.5}, 10^{-3}] = [3.1623 \times 10^{-5}, 10^{-3}]$ .

#### D. Convergence Comparison

The general convergence of the ADMM algorithm has been discussed in [37], which is equally applicable to our problem

formulation. Thus, instead of reviewing the convergence theory, we compare our proposed algorithm with a subgradient algorithm proposed by Hawe et al. [24], which is an alternative algorithm that solves a simplified version of (8).

To set up the experiment, we consider a uniformly random sampling pattern  $\mathcal{S}$  with sampling ratios of 10%, 15% and 20%. For our algorithm and the subgradient algorithm proposed in [24], we consider a single wavelet dictionary using Daubechies wavelet “db2” with 2 decomposition levels. Other choices of wavelets are possible, but we observe that the difference is not significant.

Figure 5 shows the convergence results of our proposed algorithm and the subgradient algorithm. It is evident from the figure that the ADMM algorithm converges at a significantly higher rate than the subgradient algorithm. In particular, we see that the ADMM algorithm reaches a steady state at 200 iterations, whereas the subgradient algorithm does not converge in 400 iterations.

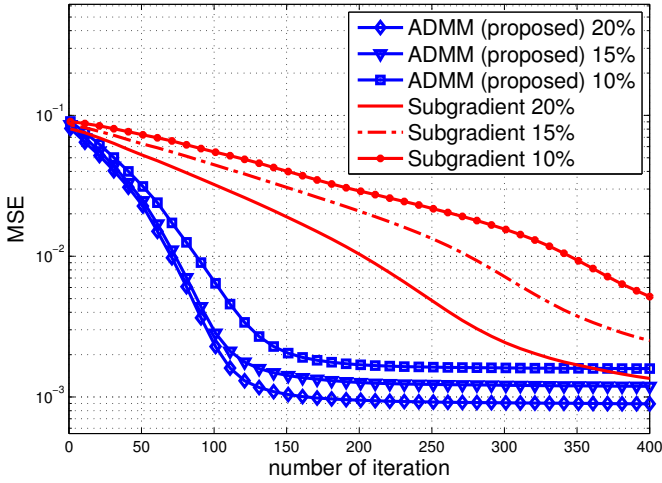


Fig. 5: Comparison of the rate of convergence between ADMM (proposed) and subgradient [24] using the disparity map “Aloe”. The ADMM algorithm requires approximately 200 iterations to reach steady state. The subgradient algorithm requires  $2 \times$  number of iterations than the ADMM algorithm to reach steady state.

### E. Multiscale ADMM

The ADMM algorithm shown in Algorithm 1 can be modified to incorporate a multiscale warm start. The idea works as follows.

First, given the observed data  $\mathbf{b}$ , we construct a multiscale pyramid  $\{\mathbf{b}_q \mid q = 0, \dots, Q-1\}$  of  $Q$  levels, with a scale factor of 2 across adjacent levels. Mathematically, by assuming without loss of generality that  $N$  is a power of 2, we define a downsampling matrix  $\mathbf{A}_q$  at the  $q$ th level as

$$\mathbf{A}_q = [\mathbf{e}_1, \mathbf{0}, \mathbf{e}_2, \mathbf{0}, \dots, \mathbf{0}, \mathbf{e}_{N/2^q}],$$

where  $\mathbf{e}_k$  is the  $k$ th standard basis. Then, we define  $\mathbf{b}_q$  as

$$\mathbf{b}_q = \mathbf{A}_q \mathbf{b}_{q-1}, \quad (21)$$

for  $q = 1, \dots, Q-1$ , and  $\mathbf{b}_0 = \mathbf{b}$ . Correspondingly, we define a pyramid of sampling matrices  $\{\mathcal{S}_q \mid q = 0, \dots, Q-1\}$ , where

$$\mathcal{S}_q = \mathbf{A}_q \mathcal{S}_{q-1}, \quad (22)$$

with the initial sampling matrix  $\mathcal{S}_0 = \mathcal{S}$ .

The above downsampling operation allows us to solve (8) at different resolution levels. That is, for each  $q = 0, \dots, Q-1$ , we solve the problem

$$\mathbf{x}_q = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathcal{S}_q \mathbf{x} - \mathbf{b}_q\|_2^2 + \sum_{\ell=1}^L \lambda_\ell \|\mathbf{W}_\ell \Phi_\ell^T \mathbf{x}\|_1 + \beta \|\mathbf{x}\|_{TV}, \quad (23)$$

where  $\Phi_\ell$  and  $\mathbf{W}_\ell$  are understood to have appropriate dimensions.

Once  $\mathbf{x}_q$  is computed, we feed an upsampled version of  $\mathbf{x}_q$  as the initial point to the  $(q-1)$ th level’s optimization. More specifically, we define an upsampling and averaging operation:

$$\mathbf{B}_q = [\mathbf{e}_1^T; \mathbf{e}_1^T; \mathbf{e}_2^T; \mathbf{e}_2^T; \dots; \mathbf{e}_{N/2^q}^T; \mathbf{e}_{N/2^q}^T], \quad (24)$$

and we feed  $\mathbf{x}_q$ , the solution at the  $q$ th level, as the initial guess to the problem at the  $(q-1)$ th level:

$$\mathbf{x}_{q-1}^{(0)} = \mathbf{x}_q. \quad (25)$$

A pictorial illustration of the operations of  $\mathbf{A}_q$  and  $\mathbf{B}_q$  is shown in Figure 6. The algorithm is shown in Algorithm 2.

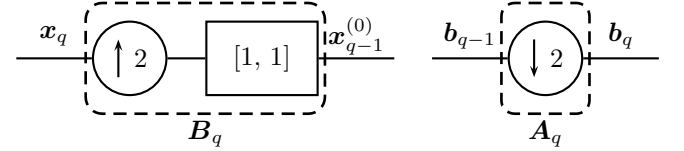


Fig. 6: Schematic diagram showing the operations of  $\mathbf{A}_q$  and  $\mathbf{B}_q$ :  $\mathbf{A}_q$  downsamples the observed data  $\mathbf{b}_q$  by a factor of 2;  $\mathbf{B}_q$  upsamples the solution  $\mathbf{x}_q$  by a factor of 2, followed by a two-tap filter of impulse response  $[1, 1]$ .

---

### Algorithm 2 Multiscale ADMM Algorithm

---

**Require:**  $\mathcal{S}_0, \dots, \mathcal{S}_{Q-1}$  and  $\mathbf{b}_0, \dots, \mathbf{b}_{Q-1}$   
1: **for**  $q = Q-1$  **to** 0 **do**  
2:    $\mathbf{x}_q = \text{ADMM}(\mathbf{b}_q, \mathcal{S}_q)$  with initial guess  $\mathbf{x}_q^{(0)}$   
3:   Let  $\mathbf{x}_{q-1}^{(0)} = \mathbf{B}_q \mathbf{x}_q$ , if  $q \geq 1$ .  
4: **end for**  
5: **Output**  $\mathbf{x} = \mathbf{x}_0$ .

---

*Remark 4:* When propagating the  $q$ th solution,  $\mathbf{x}_q$ , to the  $(q-1)$ th level, we should also propagate the corresponding auxiliary variables  $\mathbf{u}_1$ ,  $\mathbf{u}_2$ ,  $\mathbf{v}$  and the Lagrange multipliers  $\mathbf{y}_1$ ,  $\mathbf{y}_2$  and  $\mathbf{z}$ . The auxiliary variables can be updated according to  $\mathbf{x}_{q-1}^{(0)}$  as  $\mathbf{u}_{1,q-1}^{(0)} = \Phi_1 \mathbf{x}_{q-1}^{(0)}$ ,  $\mathbf{u}_{2,q-1}^{(0)} = \mathbf{x}_{q-1}^{(0)}$ , and  $\mathbf{v}_{q-1}^{(0)} = \mathbf{D} \mathbf{x}_{q-1}^{(0)}$ . For the Lagrange multipliers, we let  $\mathbf{y}_{1,q-1}^{(0)} = \mathbf{B}_q \mathbf{y}_{1,q}^{(0)}$ ,  $\mathbf{y}_{2,q-1}^{(0)} = \mathbf{B}_q \mathbf{y}_{2,q}^{(0)}$ , and  $\mathbf{z}_{q-1}^{(0)} = \mathbf{B}_q \mathbf{z}_q^{(0)}$ .

*Remark 5:* The choice of the up/down sampling factor is not important. In our experiment, we choose a factor of 2 for simplicity in implementation. Other sampling factors such as  $\sqrt{2}$  are equally applicable. Furthermore, the two-tap average



filter  $[1, 1]$  in Figure 6 can be replaced by any valid averaging filter. However, experimentally we find that other choices of filters do not make a significant difference comparing to  $[1, 1]$ .

To validate the effectiveness of the proposed multiscale warm start, we compare the convergence rate against the original ADMM algorithm. In Figure 7, we observe that the multiscale ADMM converges at a significantly faster rate than the original ADMM algorithm. For example, using a sampling ratio of 20%, the multiscale ADMM algorithm converges in 40 seconds whereas the original ADMM algorithm converges in 120 seconds which corresponds to a factor of 3 in runtime reduction. Here, we plot the mean squared error as a function of runtime instead of the number of iterations, because the number iterations of the multiscale ADMM varies at different scale levels. For fairness, both algorithms are tested under the same platform of MATLAB 2012b / 64-bit Windows 7 / Intel Core i7 / CPU 3.2GHz (single thread) / 12 GB RAM.

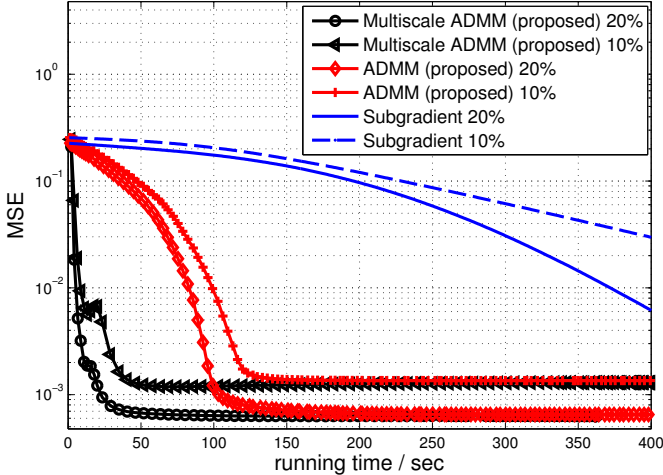


Fig. 7: Runtime comparison of original ADMM algorithm, multiscale ADMM algorithm and subgradient algorithm. All algorithms use a combined wavelet-contourlet dictionary. The disparity map “Art” is tested, with 10% and 20% uniformly random samples, and  $Q = 3$  multiscale levels.

## V. SAMPLING SCHEME

In the above sections, we assume that the sampling matrix  $S$  is given and is fixed. However, in practice, there is usually freedom in choosing an optimal  $S$ . The purpose of this section is to present an efficient design procedure based on a random sampling concept.

### A. Motivating Example

To motivate our discussion on how to find an optimal sampling pattern, it is useful to first consider the following example. In Figure 8 we consider a simple disparity map consisting of an ellipse of constant intensity and a plain background. We consider three sampling patterns of approximately equal sampling ratios  $\xi$ : (a) a sampling pattern defined according to the magnitude of the disparity gradient; (b) a

uniform grid with specified sampling ratio  $\sqrt{\xi}$  along both directions; (c) a random sampling pattern drawn from a uniform distribution with probability  $\xi$ . The three sampling patterns correspondingly generate three sampled disparity maps. For each sampled disparity map, we run the proposed ADMM algorithm and record the reconstructed disparity map. In all experiments, we use a wavelet dictionary for demonstration.

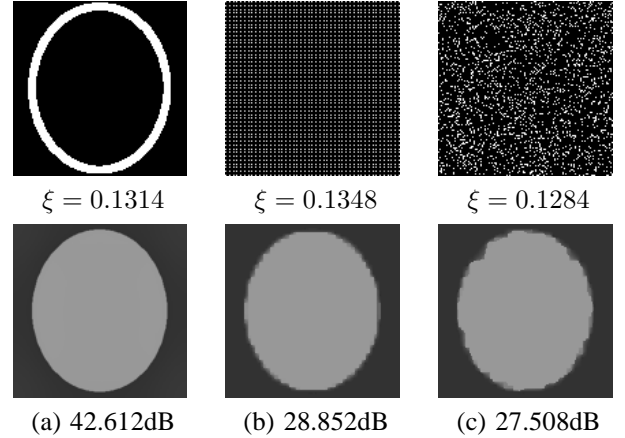


Fig. 8: Three sampling patterns and the corresponding reconstruction results using the proposed ADMM algorithm. Here,  $\xi$  denotes the sampling ratio. (a) Sampling along the gradient; (b) Sampling from a grid; (c) Sampling from a uniformly random pattern.

The results shown in Figure 8 illustrate that for a fixed sampling budget  $\xi$ , it is more efficient to allocate the samples along the strong gradient locations. The theoretical support of this observation is the theory of finite rate of innovation (FRI) [39] (and its multi-dimensional extensions [40], [41]). FRI suggests that optimal samples should be picked at places where there are sufficient innovations (*e.g.*, gradient if we use wavelet basis).

Unfortunately, allocating samples along the gradient is not feasible for two reasons. First, the gradient of the disparity map is not available prior to reconstructing the disparity. Therefore, all gradient information can only be inferred from the view image. Second, the gradients of a view image could be very different from the gradients of the corresponding disparity map. Thus, inferring the disparity gradient from the view image gradient is a challenging task. In the followings, we present a randomized sampling scheme to address these two issues.

### B. Oracle Random Sampling Scheme

Our proposed random sampling scheme is based on two simple hypotheses: (1) it is desired to allocate samples along the disparity gradient (if we knew it); (2) it is desired to randomly allocate a small portion of the samples throughout the entire disparity map besides the strong gradient locations, because there is always uncertainty about the exact locations of the strong gradients. To illustrate the effectiveness of this hypothesis, we first consider an oracle scenario where the ground truth disparity map is available.

Let  $\mathbf{a} = [a_1, \dots, a_N]^T$  be a vector denoting the magnitude of the ground truth disparity map's gradient. Ideally, for a fixed sampling budget  $0 \leq \xi \leq 1$ , one would choose the  $\xi N$  largest magnitude samples from  $\{a_j\}_{j=1}^N$ .

However, this greedy sampling process is not desirable because the disparity boundaries between two objects having similar depth are likely to be missed if there is another object with sharp depth transition. To overcome this issue, we propose to consider a soft decision process where we allow a pixel to be sampled with probability defined according to some functions of  $\{a_j\}_{j=1}^N$ . More formally, we define a sequence of independent Bernoulli random variables  $\{I_j\}_{j=1}^N$ , and the probability of having  $I_j \in \{0, 1\}$  is given by

$$\Pr[I_j = 1] = p_j, \quad \text{and} \quad \Pr[I_j = 0] = 1 - p_j, \quad (26)$$

where  $0 \leq p_j \leq 1$  is the *sampling probability*. Our goal is to determine  $\{p_j\}_{j=1}^N$ .

Our randomized scheme exploits the fact that while it is not desirable to greedily pick samples according to  $\{a_j\}_{j=1}^N$ , it is possible to pose a soft criteria that the *average* gradient of *all*  $N$  samples should be similar to the average gradient of a *subset* of  $\xi N$  samples. To illustrate this more precisely, we denote the average of  $\{a_j\}_{j=1}^N$  as

$$\mu \stackrel{\text{def}}{=} \frac{1}{N} \sum_{j=1}^N a_j. \quad (27)$$

Now, we also consider the average of a randomly selected subset as

$$Y \stackrel{\text{def}}{=} \frac{1}{N} \sum_{j=1}^N \frac{a_j}{p_j} I_j, \quad (28)$$

where the division by  $p_j$  ensures that  $Y$  is unbiased, *i.e.*,  $\mathbb{E}[Y] = \mu$ . Therefore, if we want to minimize the discrepancy between  $\mu$  and  $Y$ , then it is equivalent to minimize the variance  $\mathbb{E}[(Y - \mu)^2]$ . Observing that

$$\mathbb{E}[(Y - \mu)^2] = \frac{1}{N} \sum_{j=1}^N \frac{a_j^2}{p_j^2} \text{Var}[I_j] = \frac{1}{N} \sum_{j=1}^N a_j^2 \left( \frac{1 - p_j}{p_j} \right),$$

where the last equality holds because  $\text{Var}[I_j] = p_j(1 - p_j)$ , we obtain the following optimization problem

$$(P): \quad \begin{aligned} & \underset{\mathbf{p}}{\text{minimize}} && \frac{1}{N} \sum_{j=1}^N \frac{a_j^2}{p_j} \\ & \text{subject to} && \sum_{j=1}^N p_j = \xi N, \text{ and } 0 \leq p_j \leq 1. \end{aligned}$$

Here, the equality constraint forces the average number of samples  $\mathbb{E}[\sum_{j=1}^N I_j]$  to be equal to the sampling budget  $\xi N$ . (P) has a closed-form solution given by [42, Lemma 2]

$$p_j = \min(\tau a_j, 1), \quad (29)$$

where  $\tau$  is the root of the equation

$$g(\tau) \stackrel{\text{def}}{=} \sum_{j=1}^N \min(\tau a_j, 1) - \xi N. \quad (30)$$



Fig. 9: Comparison between a deterministic sampling pattern by selecting samples greedily according to the magnitude of  $\{a_j\}$ , and a randomized sampling pattern using the proposed scheme.

It is interesting to compare this new random sampling scheme versus the aforementioned greedy sampling scheme. Figure 9 shows a comparison. For the greedy sampling scheme, we first compute the gradient of the disparity map  $\nabla \mathbf{x} \stackrel{\text{def}}{=} \sqrt{(\mathbf{D}_x \mathbf{x})^2 + (\mathbf{D}_y \mathbf{x})^2}$  and threshold it to obtain a set of samples  $\Omega \stackrel{\text{def}}{=} \{j \mid [\nabla \mathbf{x}]_j > \alpha \|\nabla \mathbf{x}\|_\infty\}$ , where  $\alpha = 0.1$  is the threshold. The actual sampling ratio is then  $\xi = |\Omega|/N$ . For the randomized scheme, we let  $\mathbf{a} = \nabla \mathbf{x}$  and we compute  $p_j$  according to (29). In this particular example, we observe that the randomized sampling scheme achieves a PSNR improvement of more than 3 dB.

### C. Practical Random Sampling Scheme

The sampling scheme we studied thus far is oracle because we assume that the ground truth disparity is given. Therefore, it is important to find a practically implementable alternative. In particular, we need to overcome the fact that the gradient information of the disparity is not available. To solve this problem, we propose a two-stage sampling process.

Our proposed sampling scheme consists of two stages - a pilot stage to obtain a rough estimate of the disparity, and a refinement stage to improve the disparity. In the first step pilot stage, we pick  $\xi N/2$  samples (*i.e.*, half of the desired number of samples) using a uniformly random sampling pattern. This gives a sampling pattern  $\{I_j^{(1)}\}_{j=1}^N$ , where the superscript denotes the first stage. Correspondingly, we have a sampling matrix  $\mathbf{S}^{(1)}$  and the sampled data  $\mathbf{b}^{(1)}$ . Given  $\mathbf{S}^{(1)}$  and  $\mathbf{b}^{(1)}$ , we apply the ADMM algorithm to obtain a pilot estimate  $\mathbf{x}^{(1)}$ .

In the second stage, we use the pilot estimate  $\mathbf{x}^{(1)}$  as a guide to compute the gradient  $\nabla \mathbf{x}^{(1)}$ . By (29), this suggests that the optimal sampling probability is  $p_j = \min(\tau [\nabla \mathbf{x}^{(1)}]_j, 1)$ . However, in order to ensure that the  $\xi N/2$  samples picked at the second stage *do not overlap* with those picked in the first stage, instead of letting  $p_j = \min(\tau [\nabla \mathbf{x}^{(1)}]_j, 1)$ , we let  $p_j = \min(\tau a_j, 1)$ , where

$$a_j = \begin{cases} [\nabla \mathbf{x}^{(1)}]_j, & I_j^{(1)} = 0, \\ 0, & I_j^{(1)} = 1. \end{cases} \quad (31)$$

In words,  $a_j$  defined by (31) forces  $p_j = 0$  when the  $j$ th pixel is picked in the first step. Thus, the non-zero entries of  $\{I_j^{(1)}\}$  and  $\{I_j^{(2)}\}$  are mutually exclusive, and hence we can now apply the ADMM algorithm to recover  $\mathbf{x}^{(2)}$  from



$S_1 + S_2$  and  $b_1 + b_2$ . The overall method is summarized in Algorithm 3.

---

**Algorithm 3** Two-Stage Algorithm
 

---

- 1: Input:  $N, \xi, \mathbf{b}$
  - 2: Output:  $\mathbf{x}^{(2)}$
  - 3: **Stage 1:**
  - 4: Let  $I_j^{(1)} = 1$  with probability  $\xi/2$ , for  $j = 1, \dots, N$ .
  - 5: Define  $\mathbf{S}^{(1)}$  and  $\mathbf{b}^{(1)}$  according to  $\{I_j^{(1)}\}$ .
  - 6: Compute  $\mathbf{x}^{(1)} = \text{ADMM}(\mathbf{S}^{(1)}, \mathbf{b}^{(1)})$ .
  - 7: **Stage 2:**
  - 8: Compute  $\nabla \mathbf{x}^{(1)}$ .
  - 9: For  $j = 1, \dots, N$ , define  $a_j = \begin{cases} [\nabla \mathbf{x}^{(1)}]_j, & I_j^{(1)} = 0, \\ 0, & I_j^{(1)} = 1. \end{cases}$
  - 10: Compute  $\tau$  such that  $\sum_{j=1}^N \min\{\tau a_j, 1\} = N\xi/2$ .
  - 11: Let  $p_j = \min\{\tau a_j, 1\}$ , for  $j = 1, \dots, N$ .
  - 12: Let  $I_j^{(2)} = 1$  with probability  $p_j$ , for  $j = 1, \dots, N$ .
  - 13: Define  $\mathbf{S}^{(2)}$  and  $\mathbf{b}^{(2)}$  according to  $\{I_j^{(2)}\}$ .
  - 14: Compute  $\mathbf{x}^{(2)} = \text{ADMM}(\mathbf{S}^{(1)} + \mathbf{S}^{(2)}, \mathbf{b}^{(1)} + \mathbf{b}^{(2)})$ .
- 

#### D. Further Improvement by PCA

The two-stage sampling procedure can be further improved by utilizing the prior information of the view image. The intuition is that since both view image and disparity map are captured for the same scene, strong gradients in the disparity map should align with those in the view image. However, since a view image typically contains complex gradients which are irrelevant to the disparity reconstruction, it is important to filter out these unwanted gradients while preserving the important ones. To this end, we consider the following patch-based principal component analysis.

Given a view image  $\mathbf{y} \in \mathbb{R}^N$ , we partition the image into  $N$  overlapping patches of dimension  $d$ , i.e.,  $\{\mathbf{y}_j\}_{j=1}^N \in \mathbb{R}^d$ . We then vectorize the patches and form a data matrix  $\mathbf{Y} \stackrel{\text{def}}{=} [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$ . This data matrix leads to a principal component decomposition as

$$\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T = \mathbf{Y}\mathbf{Y}^T, \quad (32)$$

where  $\mathbf{U}$  is the eigenvector matrix, and  $\mathbf{\Lambda}$  is the eigenvalue matrix. Geometrically, the projection of any patch  $\mathbf{y}_j$  onto the subspace spanned by any eigenvector  $\mathbf{u}_i$  is equivalent to applying a finite impulse response filter to the patch, i.e.,  $\mathbf{u}_i^T \mathbf{y}_j$ . In many cases, except for the first eigenvector  $\mathbf{u}_1$ , all remaining eigenvectors  $\mathbf{u}_2, \dots, \mathbf{u}_d$  are in the form of differential operators (of different orders and orientations, see examples in Figure 10). More interestingly, these filters are typically *bandpass* filters, which suggests that both low frequency components (e.g., smooth regions) and high frequency components (e.g., complex textures) of the view image can be filtered by applying the projections. Consequently, we consider the following filtered signal

$$a_j = \sum_{i=1}^{d'} |\langle \mathbf{u}_i, \mathbf{y}_j \rangle|, \quad j = 1, \dots, N, \quad (33)$$

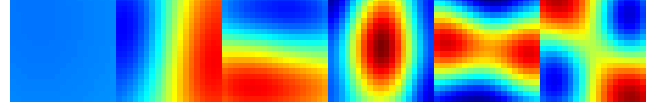


Fig. 10: The first 6 eigenvectors of the data matrix  $\mathbf{Y}\mathbf{Y}^T$ , where  $\mathbf{Y}$  is obtained from the view image corresponding to Figure 9. In this example we set the patch size as  $19 \times 19$  so that  $d = 361$ . The range of the color index of this figure is  $[-0.1, 0.1]$ .

where  $d' < d$  is a tunable parameter (which was set to  $d' = 16$  for  $d = 49$  in this paper). Here, the absolute value in (33) is used to get the magnitude of  $\langle \mathbf{u}_i, \mathbf{y}_j \rangle$ , as  $a_j$  must be a non-negative number.

To see how this PCA concept can be incorporated into our two-stage sampling scheme, we make the following observations. First, the uniform sampling in Stage-1 can well be replaced by the PCA approach. In particular, instead of setting  $I_j^{(1)} = 1$  with probability  $\xi/2$ , we can define  $a_j$  according to (33), and let  $p_j = \min(\tau a_j, 1)$  for  $\tau$  being the root of (30). Consequently, we let  $I_j^{(1)} = 1$  with probability  $p_j$ .

In Stage-2, since we have already had a pilot estimate of the disparity map, it is now possible to replace  $\mathbf{Y}$  in (32) by  $\mathbf{X}$ , where  $\mathbf{X} = [\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_N^{(1)}]$  is a data matrix formed by stacking the  $N$  non-overlapping patches of the pilot disparity map  $\mathbf{x}^{(1)}$ . Thus, instead of setting  $a_j = [\nabla \mathbf{x}^{(1)}]_j$  in (31), we can set  $a_j = \sum_{i=1}^{d'} |\langle \mathbf{u}_i, \mathbf{x}_j^{(1)} \rangle|$  using (33). The advantage of this new  $a_j$  is that it softens the sampling probability at the object boundaries to a neighborhood surrounding the boundary. This reduces the risk of selecting irrelevant samples because of a bad pilot estimate.

#### E. Comparisons

As a comparison between various sampling patterns, we consider a disparity map shown in Figure 11. Setting a target sampling ratio of  $\xi = 0.1$  (i.e., 10%), we study four sampling patterns including two versions of our proposed two-stage method. The results shown in Figure 11(c) are generated using the original two-stage sampling scheme without PCA improvement, whereas the results shown in Figure 11(d) are generated using an improved two-stage sampling scheme where the first stage is uniform and the second stage is PCA. These results indicate that for the same sampling ratio  $\xi$ , the choice of the sampling pattern has some strong influence to the reconstruction quality. For example, the original two-stage sampling can improve both uniform random sampling and grid sampling about 1.60 dB, and can be further improved by almost 3.44 dB using the PCA idea.

## VI. EXPERIMENTAL RESULTS

In this section we present additional results to illustrate the performance of the proposed method.

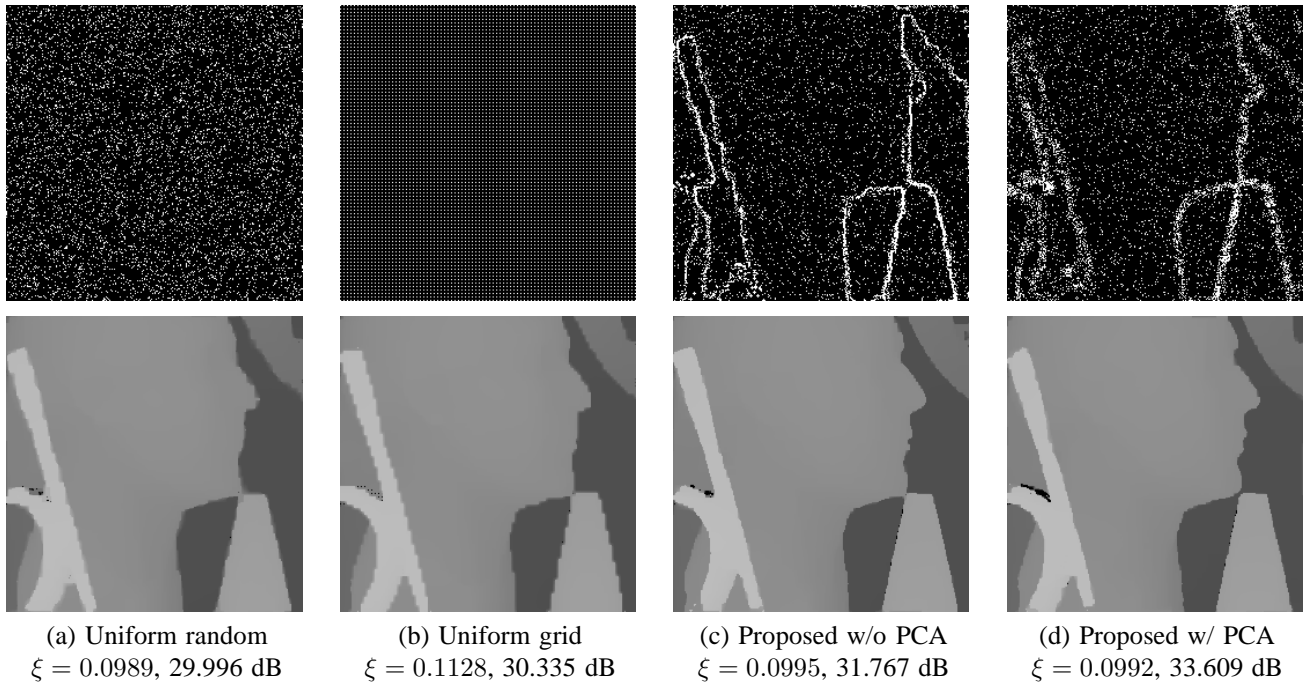


Fig. 11: Comparison between four sampling patterns. (a) Uniformly random sampling pattern; (b) Uniform grid; (c) Proposed two-stage sampling without PCA improvement; (d) Proposed two-stage sampling with PCA improvement. For the two-stage sampling in (c)-(d), we pick  $\xi N/2$  uniformly random samples in stage 1, and pick the remaining  $\xi N/2$  samples according to the pilot estimate from Stage 1.

#### A. Synthetic Data

We first compare the proposed algorithm with existing methods on the Middlebury dataset<sup>1</sup> where ground truth disparities are available. We consider two versions of the proposed algorithm: “Proposed WT+CT Grid” and “Proposed WT+CT 2-Stage”. “Proposed WT+CT Grid” is the ADMM algorithm presented in Section IV using both wavelet and contourlet basis. Here, “Grid” refers to using a deterministic uniform grid sampling pattern and “2-stage” refers to using the 2-stage randomized sampling scheme presented in Section V. In both cases, the parameters of the ADMM algorithm are:  $\lambda_1 = 4 \times 10^{-5}$ ,  $\lambda_2 = 2 \times 10^{-4}$ ,  $\beta = 2 \times 10^{-3}$  and  $Q = 3$ . We use Daubchies wavelet “db2” with 2 decomposition levels for wavelet dictionary, and we set “bior9-7” wavelet function with [5 6] directional decompositions for contourlet dictionary.

As a comparison, we also consider the method proposed in [24], which has three differences with our algorithm: (1) [24] uses a subgradient descent algorithm whereas we use an ADMM algorithm; (2) [24] considers only a wavelet basis whereas we consider a combined wavelet-contourlet basis; (3) [24] uses a combination of canny edges and uniformly random samples whereas we use a principled design process to determine samples.

We note that there are other depth super resolution algorithms in the literature, *e.g.*, [18], [43], [44]. However, these methods require a view image to guide the resolution enhancement process. This setting is different from what is presented in this paper, in the sense that we use the view

image to guide the design of the sampling pattern whereas the other methods use the view image for the reconstruction. From this perspective, we do not include the comparison with these methods. However, as a benchmark we show results of a bicubic interpolation using uniform grid sampling pattern.

Table I shows the performance of various methods at different sampling ratios and sampling methods. Referring to PSNR, the results show that “Proposed WT+CT 2-Stage” outperforms the other methods because 2-Stage algorithm efficiently determines the sampling points while the sampling budget is given. Furthermore, the results show that 2-Stage algorithm has prominent improvement as sampling rate increases. Yet, “Bicubic Grid” has limited improvement as sampling rate increases. Considering the case of Baby disparity, “Proposed WT+CT 2-Stage” has approximately 14.3 dB improvement from 5% to 25% samples, and “Proposed WT+CT Grid” has around 5.7 dB improvement. However, “Bicubic Grid” only has 3.5 dB improvement.

It is also instructive to compare the percentage of bad pixels (% Bad Pixel), which is a popular metric to measure the quality of disparity estimates. Given a threshold  $\tau > 0$ , the percentage of bad pixel is defined as

$$\% \text{ Bad Pixel} \stackrel{\text{def}}{=} |\{j : |\hat{x}_j - x_j^*| > \tau\}|, \quad (34)$$

where  $\hat{x}$  is the reconstructed disparity and  $x^*$  is the ground truth disparity.

In Table I, we show the % Bad Pixel with  $\tau = 2$ . The results show that “Proposed WT+CT 2-Stage” suppresses errors. We denote that comparing to other methods, the “Proposed WT+CT 2-Stage” has prominent error reduction while  $\tau$  varies

<sup>1</sup><http://vision.middlebury.edu/stereo/data/>

TABLE I: Comparisons of reconstruction algorithms in terms of PSNR and % Bad Pixels. We denote N/A when the algorithm does not converge in 1000 iterations.

Disparity Name	Method Algorithm / Sampling Strategy	Percentage of Samples {PSNR (dB) / % Bad Pixels (%) [ $\tau = 2$ ]}				
		5%	10%	15%	20%	25%
Aloe	Proposed WT+CT 2-Stage	27.2493 / <b>19.84</b>	<b>30.9526</b> / <b>7.30</b>	<b>32.5546</b> / <b>3.54</b>	<b>36.7395</b> / <b>2.15</b>	<b>38.2949</b> / <b>1.09</b>
	Proposed WT+CT Grid	25.7967 / 21.84	29.1086 / 10.32	30.3861 / 7.17	31.4282 / 5.44	32.4174 / 4.19
	[24] Grid	25.2561 / 20.19	27.7682 / 8.86	28.8843 / 6.10	30.2323 / 4.73	31.6519 / 3.56
	Bicubic Grid	<b>27.8899</b> / 23.51	29.3532 / 17.49	30.1019 / 13.78	31.0031 / 11.96	31.8908 / 10.04
Art	Proposed WT+CT 2-Stage	<b>30.4187</b> / 16.84	<b>34.9997</b> / <b>5.97</b>	<b>37.1935</b> / <b>2.50</b>	<b>42.3322</b> / <b>1.52</b>	<b>42.1885</b> / <b>0.42</b>
	Proposed WT+CT Grid	27.5524 / <b>12.55</b>	29.3706 / 7.19	31.2610 / 5.31	32.6743 / 4.05	33.7395 / 3.02
	[24] Grid	27.0025 / 14.03	N/A / N/A	N/A / N/A	N/A / N/A	N/A / N/A
	Bicubic Grid	29.1550 / 19.49	30.3536 / 14.28	31.1098 / 11.33	31.9473 / 9.56	32.8366 / 7.97
Baby	Proposed WT+CT 2-Stage	<b>39.0051</b> / 8.19	<b>44.5248</b> / <b>2.74</b>	<b>48.3065</b> / <b>0.96</b>	<b>51.3144</b> / <b>0.12</b>	<b>53.3443</b> / <b>0.06</b>
	Proposed WT+CT Grid	33.8826 / 7.02	36.6908 / 3.08	37.6589 / 2.25	38.6715 / 1.71	39.5831 / 1.27
	[24] Grid	33.5585 / <b>6.45</b>	34.8044 / 2.75	36.0139 / 1.94	37.4481 / 1.47	38.7448 / 1.10
	Bicubic Grid	34.8368 / 7.89	36.2385 / 5.63	37.1749 / 4.34	37.5973 / 3.73	38.3961 / 3.10
Dolls	Proposed WT+CT 2-Stage	<b>29.6914</b> / 21.21	<b>32.6440</b> / 9.73	<b>34.3968</b> / <b>4.70</b>	<b>37.1573</b> / <b>2.51</b>	<b>37.1239</b> / <b>1.34</b>
	Proposed WT+CT Grid	28.5086 / 14.88	29.3275 / <b>7.61</b>	30.3519 / 5.22	31.2155 / 3.92	32.1821 / 2.88
	[24] Grid	28.4952 / <b>12.27</b>	N/A / N/A	N/A / N/A	N/A / N/A	32.0500 / 2.44
	Bicubic Grid	29.0612 / 12.64	30.0475 / 9.44	30.4374 / 7.76	31.0053 / 6.78	31.8800 / 5.48
Moebius	Proposed WT+CT 2-Stage	<b>31.8184</b> / 13.34	<b>34.7836</b> / <b>3.87</b>	<b>38.4404</b> / <b>1.63</b>	<b>40.5738</b> / <b>0.84</b>	<b>43.1185</b> / <b>0.46</b>
	Proposed WT+CT Grid	27.7468 / 12.13	29.1431 / 6.72	30.4755 / 4.95	31.4821 / 3.74	32.6680 / 2.83
	[24] Grid	27.7278 / 12.05	21.5485 / N/A	N/A / N/A	N/A / N/A	32.0975 / 2.31
	Bicubic Grid	28.3987 / <b>11.70</b>	29.9338 / 8.39	30.6607 / 6.87	30.9427 / 5.92	32.0143 / 4.92
Rocks	Proposed WT+CT 2-Stage	<b>31.5251</b> / <b>7.61</b>	<b>35.2405</b> / <b>2.50</b>	<b>37.9699</b> / <b>0.96</b>	<b>40.2570</b> / <b>0.40</b>	<b>43.3501</b> / <b>0.20</b>
	Proposed WT+CT Grid	25.9819 / 7.88	29.3224 / 4.35	30.8119 / 3.14	31.7405 / 2.52	33.2075 / 1.88
	[24] Grid	25.7976 / 8.40	28.6000 / 4.10	29.4776 / 2.89	30.1535 / 2.24	32.1658 / 1.62
	Bicubic Grid	28.7241 / 10.20	30.4212 / 7.95	30.7552 / 6.46	31.6722 / 5.26	32.6706 / 4.61

from 1 to 2. The results further infer that “Proposed WT+CT 2-Stage” has fewer outliers; therefore, it has higher PSNR and less % Bad Pixel ( $\tau = 2$ ). Similarly, comparing “Proposed WT+CT Grid” to “[24] Grid”, the higher PSNR infers that “Proposed WT+CT Grid” has error suppression.

Finally, we show the performance of the proposed algorithm towards additive i.i.d. Gaussian noise. The purpose of this experiment is to demonstrate the sensitivity and robustness of the algorithm in the presence of noise. While in reality the noise in disparity estimates is not i.i.d. Gaussian, the result presented here serves as a reference for the algorithm’s performance. A more realistic experiment on real data will be illustrated in the next subsection.

The results are shown in Figure 12. Using “Bicubic Grid” as the baseline, we observe that “Proposed WT+CT 2-Stage” on average has 4.8 dB improvement, “Proposed WT+CT Grid” has 3.6 dB improvement, whereas “[24] Grid” only has 2.08 dB improvement. This provides a good indicator of the robustness of the proposed methods.

### B. Real Data

In this experiment we study the performance of the proposed algorithm for real data. The top left part of Figure 13 shows a snapshot of a stereo video (with resolution  $320 \times 240$ , 30 fps). For this video sequence, we apply the block matching algorithm by Lee et al. [45] to obtain the initial disparity estimates. However, instead of running the algorithm for all pixels in the disparity map, we only estimate 10% of the total pixels. The choice of these 10% samples are decided by the “Proposed WT+CT 2-Stage” algorithm. That is, we draw 5%

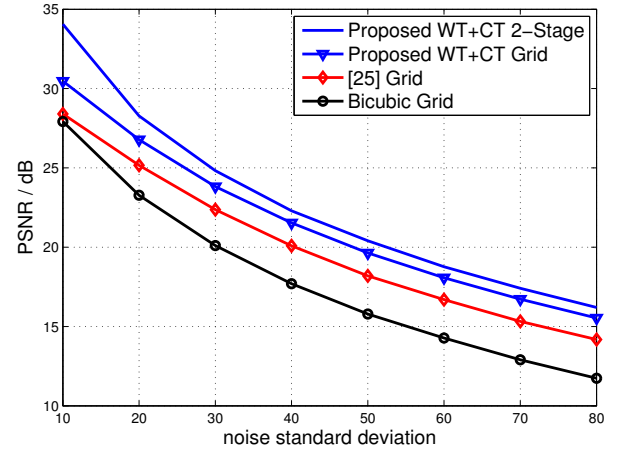


Fig. 12: Comparison of reconstruction performance with noisy samples. Sampling ratio is 20%.

samples uniformly at random in the first stage to make a pilot estimate of the disparity map, and draw another 5% of the samples according to the pilot estimate in the second stage. The results shown in the middle row of Figure 13 illustrate that the “Proposed WT+CT 2-Stage” generates the closest results compared to an ideal dense estimate.

In addition to real video sequences, we also test the proposed algorithm on a stereo system we developed. The system consists of a low cost stereo camera with customized block matching algorithms. The bottom row of Figure 13 shows the results of the reconstructed disparity maps. Looking at the results with “[24] Grid” and “Bicubic Grid”, we observe stair-like artifacts at object boundaries. In contrast, the two

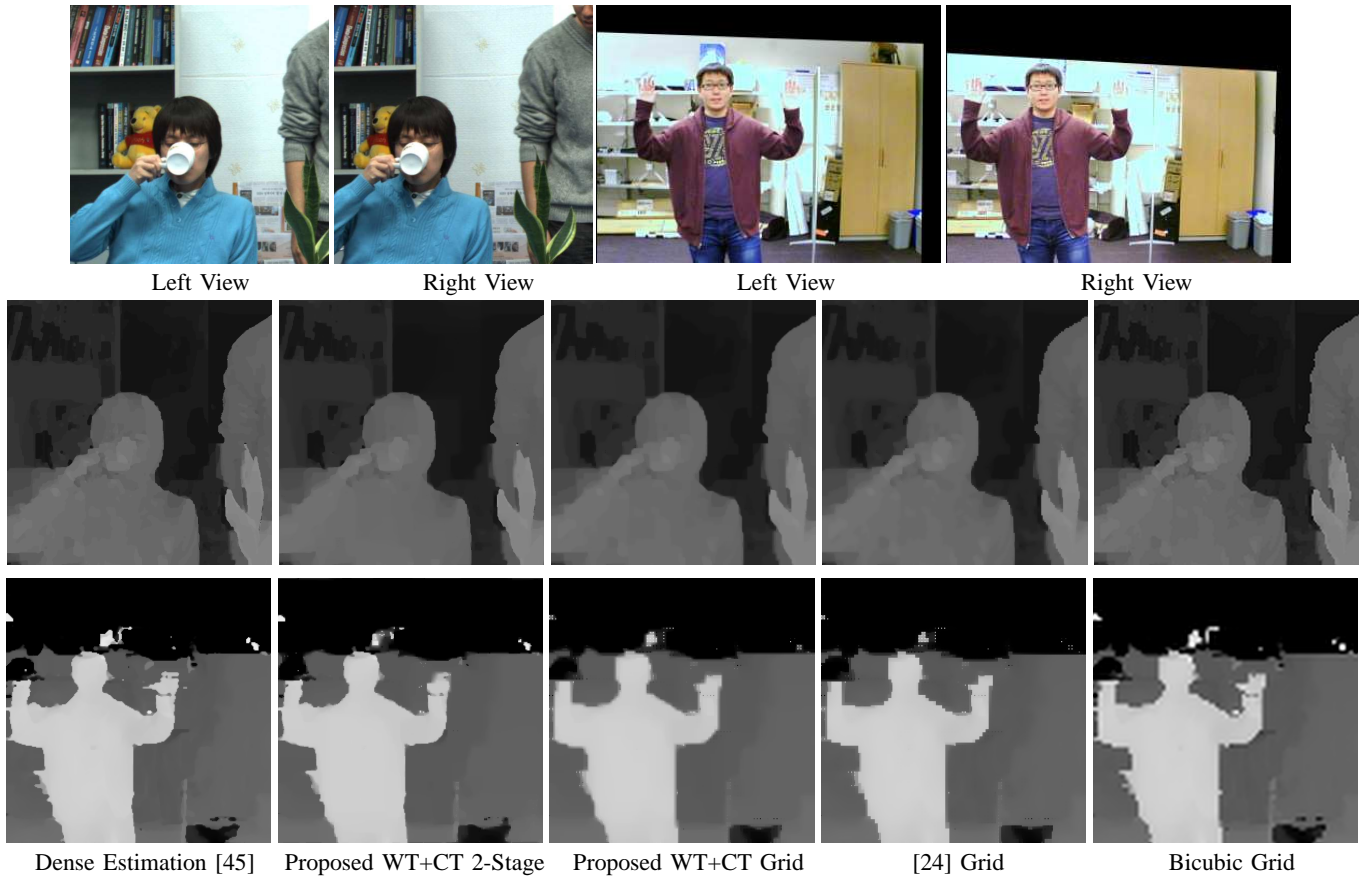


Fig. 13: Examples of disparity map reconstruction from 10% measured samples using real data. [Top] Left and right view images of the “Newspaper” dataset, and a sequence captured by a stereo system we developed. [Middle] The reconstructed disparity maps of “Newspaper”. [Bottom] The reconstructed disparity maps of our sequence. For the reconstructed disparity maps, we show the zoom-in results of size  $256 \times 256$  for better visualization. Methods under comparisons include: a dense disparity estimation [45] to acquire initial estimate; “Proposed WT+CT 2-Stage” which applies the 2-Stage randomized scheme to determine sampling locations; “Proposed WT+CT Grid” which picks samples from a uniform grid; “[24] Grid” which applies a subgradient algorithm to samples picked from a uniform grid; “Bicubic Grid” which applies bicubic interpolation to samples picked from a uniform grid.

proposed methods in general produce much smoother object boundaries, thanks to the superior modeling and the optimized sampling scheme. More interestingly, we observe that “Proposed WT+CT 2-Stage” indeed removes some undesirable noisy estimates in the recovered disparity map. This shows that the proposed method could potentially further developed as a depth enhancement method.

## VII. CONCLUSION

A framework for dense depth reconstruction from sparse samples was presented in this paper. Three contributions are made. First, we provided empirical evidence that depth data can be more sparsely encoded by a combination of wavelet and contourlet dictionaries. This provides a better understanding of the structures of depth data. Second, we proposed a general optimization formulation. An alternating direction method of multipliers (ADMM) with a multi-scale warm start was proposed to achieve fast reconstruction. The ADMM algorithm demonstrated faster rate of convergence than the existing subgradient descent algorithms. Third, we proposed

an efficient method to select samples by a randomized sampling scheme. The proposed sampling scheme achieves high quality reconstruction results at a given sampling budget. The new tools developed in this paper are applicable to many depth data processing tasks, with applications in acquisition, compression, and enhancement. Future work shall be focused on extending the methods to space-time data volume to further improve consistency of the estimates.

## REFERENCES

- [1] M.A. Lefsky, W.B. Cohen, G.G. Parker, and D.J. Harding, “LIDAR remote sensing of above-ground biomass in three biomes,” *Global Ecology and Biogeography*, vol. 11, pp. 393–399, Oct. 2002.
- [2] S. Agarwal, N. Snavely, I. Simon, S.M. Seitz, and R. Szeliski, “Building Rome in a day,” in *Proc. IEEE Int. Conf. Computer Vision (ICCV’09)*, Sep. 2009, pp. 72–79.
- [3] S. Burion, “Human detection for robotic urban search and rescue,” M.S. thesis, Carnegie Mellon Univ., 2004, available at <http://www.cs.cmu.edu/afs/cs/project/retsina-31/www/Report/Final%20Report.pdf>.
- [4] R. Khoshabeh, J. Juang, M.A. Talamini, and T.Q. Nguyen, “Multiview glasses-free 3-D laparoscopy,” *IEEE Trans. Bio. Eng.*, vol. 59, no. 10, pp. 2859–2865, Oct. 2012.

- [5] S.H. Chan, R. Khoshabeh, K.B. Gibson, P.E. Gill, and T.Q. Nguyen, "An augmented Lagrangian method for total variation video restoration," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3097–3111, Nov. 2011.
- [6] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (ToF) cameras: A survey," *IEEE Sensors Journal*, vol. 11, no. 9, pp. 1917–1926, Sep. 2011.
- [7] B. Schwarz, "LIDAR: Mapping the world in 3D," *Nature Photonics*, vol. 4, pp. 429–430, Jul. 2010.
- [8] C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami, "A 100-m range 10-frame/s  $340 \times 96$ -pixel time-of-flight depth sensor in 0.18- $\mu\text{m}$  cmos," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 2, pp. 559–572, Feb. 2013.
- [9] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," in *Proc. IEEE Int. Conf. Computer Vision (ICCV'11)*, Nov. 2011, pp. 467–474.
- [10] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *Proc. IEEE Int. Conf. on Pattern Recognition (ICPR'06)*, Aug. 2006, vol. 3, pp. 15–18.
- [11] Z. Wang and Z. Zheng, "A region based stereo matching algorithm using cooperative optimization," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR'08)*, Jun. 2008, pp. 1–8.
- [12] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 3, pp. 492–504, Mar. 2009.
- [13] J. Heikkilä and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR'97)*, Jun. 1997, pp. 1106–1112.
- [14] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proc. IEEE Int. Conf. Computer Vision (ICCV'99)*, Sep. 1999, vol. 1, pp. 666–673.
- [15] R.S. Feris, J. Gemmell, K. Toyama, and V. Kruger, "Hierarchical wavelet networks for facial feature localization," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition (FG'02)*, May 2002, pp. 118–123.
- [16] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR'04)*, Jun. 2004, vol. 2, pp. 506–513.
- [17] J. Diebel and S. Thrun, "An application of Markov random field to range sensing," in *Advances in Neural Info. Process. System (NIPS'05)*, Dec. 2005, pp. 291–298.
- [18] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR'07)*, Jun. 2007, pp. 1–8.
- [19] J. Li, T. Xue, L. Sun, and J. Liu, "Joint example-based depth map super-resolution," in *IEEE Int. Conf. Multimedia and Expo (ICME'12)*, Jul. 2012, pp. 152–157.
- [20] O. M. Aodha, N. D. F. Campbell, A. Nair, and G. J. Brostow, "Patch based synthesis for single depth image super-resolution," in *Proc. European Conf. Computer Vision (ECCV'12)*, Oct. 2012, pp. 71–84.
- [21] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [22] D.L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Info. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.
- [23] M. Elad and A.M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Info. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.
- [24] S. Hawe, M. Kleinstueber, and K. Diepold, "Dense disparity maps from sparse disparity measurements," in *Proc. IEEE Int. Conf. Computer Vision (ICCV'11)*, Nov. 2011, pp. 2126–2133.
- [25] A. Kirmani, A. Colaco, F.N.C. Wong, and V.K. Goyal, "CODAC: A compressive depth acquisition camera framework," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP'12)*, Mar. 2012, pp. 5425–5428.
- [26] A. Kirmani, D. Venkatraman, D. Shin, A. Colaco, F.N. C. Wong, J.H. Shapiro, and V.K. Goyal, "First photon imaging," *Science Magazine*, vol. 343, no. 6166, pp. 58–61, Nov. 2013.
- [27] R.I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Mar. 2004.
- [28] M.N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
- [29] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [30] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53–69, Jan. 2008.
- [31] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, Dec. 2008.
- [32] D.D.-Y. Po and M.N. Do, "Directional multiscale modeling of images using the contourlet transform," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1610–1620, Jun. 2006.
- [33] E.J. Candes and D.L. Donoho, "Recovering edges in ill-posed inverse problems: Optimality of curvelet frames," *Annals of Statistics*, , no. 3, pp. 784–842, Aug. 2002.
- [34] E.J. Candes and D.L. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise  $C^2$  singularities," *Communications on Pure and Applied Mathematics*, vol. 57, no. 2, pp. 219–266, Feb. 2004.
- [35] E. Le Pennec and S. Mallat, "Bandelet image approximation and compression," *Multiscale Model. Simul.*, vol. 4, no. 3, pp. 992–1039, 2005.
- [36] D. Han and X. Yuan, "A note on the alternating direction method of multipliers," *Journal of Optim. Theory and Applications*, vol. 155, no. 1, pp. 227–238, Oct. 2012.
- [37] 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.
- [38] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Info. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
- [39] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.
- [40] P. Shukla and P.L. Dragotti, "Sampling schemes for multidimensional signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3670–3686, Jul. 2007.
- [41] C. Chen, P. Marziliano, and A.C. Kot, "2D finite rate of innovation reconstruction method for step edge and polygon signals in the presence of noise," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2851–2859, Jun. 2012.
- [42] S.H. Chan, T. Zickler, and Y.M. Lu, "Monte Carlo non-local means: Random sampling for large-scale image filtering," *IEEE Trans. Image Process.*, 2014, in press. [Online] arXiv:1312.7366.
- [43] F. Li, J. Yu, and J. Chai, "A hybrid camera for motion deblurring and depth map super-resolution," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR'08)*, Jun. 2008, pp. 1–8.
- [44] J. Park, H. Kim, Y. Tai, M.S. Brown, and I. Kweon, "High quality depth map upsampling for 3D-TOF cameras," in *Proc. IEEE Int. Conf. Computer Vision (ICCV'11)*, Nov. 2011, pp. 1623–1630.
- [45] Z. Lee, J. Juang, and T.Q. Nguyen, "Local disparity estimation with three-moded cross census and advanced support weight," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1855–1864, Dec. 2013.