MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Motion-Adaptive Depth Superresolution

Kamilov, U.; Boufounos, P.T.

TR2017-040 January 2017

Abstract

Multi-modal sensing is increasingly becoming important in a number of applications, providing new capabilities and new processing challenges. In this paper we explore the benefit of combining of a low-resolution depth sensor with a high-resolution optical video sensor, in order to provide a highresolution depth map of the scene. We propose a new formulation that is able to incorporate temporal information and exploit the motion of objects in the video to significantly improve the results over existing methods. In particular, our approach exploits the space-time redundancy in the depth and intensity using motionadaptive low-rank regularization. We provide experiments to validate our approach and confirm that the quality of the estimated high-resolution depth is improved substantially. Our approach can be a first component in systems using vision techniques that rely on high resolution depth information.

IEEE Transactions on Image Processing

Motion-Adaptive Depth Superresolution

Ulugbek S. Kamilov, *Member, IEEE* and Petros T. Boufounos, *Senior Member, IEEE*

***Abstract*—Multi-modal sensing is increasingly becoming important in a number of applications, providing new capabilities and new processing challenges. In this paper we explore the benefit of combining of a low-resolution depth sensor with a high-resolution optical video sensor, in order to provide a high-resolution depth map of the scene. We propose a new formulation that is able to incorporate temporal information and exploit the motion of objects in the video to significantly improve the results over existing methods. In particular, our approach exploits the space-time redundancy in the depth and intensity using motion-adaptive low-rank regularization. We provide experiments to validate our approach and confirm that the quality of the estimated high-resolution depth is improved substantially. Our approach can be a first component in systems using vision techniques that rely on high resolution depth information.**

I. Introduction

One of the important challenges in computer vision applications is obtaining high resolution depth maps of observed scenes. A number of common tasks, such as object reconstruction, robotic navigation, and automotive driver assistance can be significantly improved by complementing intensity information from optical cameras with high resolution depth maps. However, with current sensor technology, direct acquisition of high-resolution depth maps is very expensive.

The cost and limited availability of such sensors imposes significant constraints on the capabilities of vision systems and has dampened the adoption of methods that rely on high-resolution depth maps. Thus, the literature has flourished with methods that provide numerical alternatives to boost the spatial resolution of the measured depth data.

One of the most popular and widely investigated class of techniques for improving the spatial resolution of depth is guided depth superresolution. These techniques jointly acquire the scene using a low-resolution depth sensor and a high-resolution optical camera. The information acquired from the camera is subsequently used to superresolve the low-resolution depth map. These techniques exploit the property that both modalities share common features, such as edges and joint texture changes. Thus, such features in the optical camera data provide information and guidance that significantly enhances the superresolved depth map.

To-date, most of these methods operate on a single snapshot of the optical image and the low-resolution depth map. However, most practical uses of such systems acquire a video from the optical camera and a sequence of snapshots of the depth map. The key insight in our paper is that information about one particular frame is replicated, in some form, in nearby frames. Thus, frames across time can be exploited to superresolve the depth map and significantly improve such methods. The challenge is finding this information in the presence of scene, camera, and object motion between frames. Figure 1 provides an example, illustrating the similarity of images and depth maps across frames.

U. S. Kamilov (e-mail: kamilov@merl.com) and P. T. Boufounos (e-mail: petrosb@merl.com) are with Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA 02139, USA

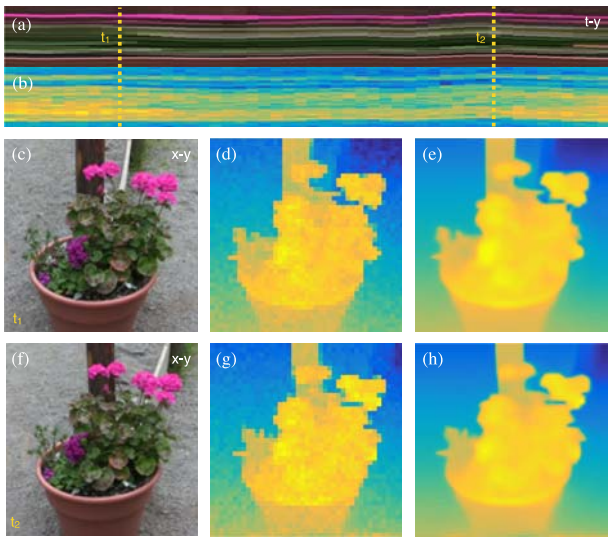

Fig. 1. Our motion adaptive method recovers a high-resolution depth sequence from high-resolution intensity and low-resolution depth sequences by imposing rank constraints on the depth patches: (a) and (b) t-y slices of the color and depth sequences, respectively, at a fixed x; (c)–(e) x-y slices at $t_1 = 10$; (f)–(h) x-y slices at $t_2 = 40$; (c) and (f) input color images; (d) and (g) input low-resolution and noisy depth images; (e) and (h) estimated depth images.

A key challenge in incorporating time into depth estimation is that depth images change significantly between frames. This results in abrupt variations in pixel values along the temporal dimension and may lead to significant degradation in the quality of the result. Thus, it is important to compensate for motion during estimation. To that end, the method we propose exploits space-time similarities in the data using motion adaptive regularization. Specifically, we identify and group similar depth patches, which we superresolve and regularize using a rank penalty.

A. Contributions

Our method builds upon prior work on patch-based methods and low-rank regularization, which were successfully applied to a variety of practical estimation problems. It further exploits the availability of optical images which provide a very robust guide to identify and group similar patches, even if the depth map has very low resolution. Thus, the output of our iterative algorithms is robust to operating conditions. Our key contributions are summarized as follows:

- We provide a new formulation for guided depth superresolution, incorporating temporal information. In this formulation, the high resolution depth is determined by solving an inverse problem that minimizes a cost. This cost includes a quadratic data-fidelity term, as well as a new motion adaptive regularizer based on a low-rank penalty on groups of similar patches.
- We develop two optimization strategies for solving our estimation problem. The first approach is based on exact optimization of the cost via alternating direction method of multipliers (ADMM). The second approach uses a simplified algorithm that alternates between enforcing data-consistency and low-rank penalty.
- We validate our approach experimentally and demonstrate it delivers substantial improvements. In particular, we compare several algorithmic solutions to the problem and demonstrate that: (a) availability of temporal information significantly improves the quality of estimated depth; (b) motion adaptive regularization is crucial for avoiding artifacts along temporal dimension; (c) using intensity during block matching is essential for optimal performance.

B. Outline

The next section provides a summary of related work in the literature. Section III describes the problem formulation and our regularization approach. Two different algorithmic strategies for computing the solution are described in Sec. IV. Extended experimental results validating our approach are presented in Sec. V. Finally, Sec. VI discusses our findings and concludes.

II. Related Work

In the last decade, guided depth superresolution has received significant attention. Early work by Diebel and Thrun [1] showed the potential of the approach by modeling the co-occurence of edges in depth and intensity with Markov Random Fields (MRF). Kopf *et al.* [2] and Yang *et al.* [3] have independently proposed an alternative approach based on joint bilaterial filtering, where intensity is used to set the weights of the filter. The bilaterial filtering approach was further refined by Chan *et al.* [4] who incorporated the local statistics of the depth and by Liu *et al.* [5] who used geodesic distances for determining the weights. Dolson *et al.* [6] extended the approach to dynamic sequences to compensate for different data rates in the depth and intensity sensors. He *et al.* [7], [8] proposed a guided image filtering approach, improving edge preservation. More recently, Lu and Forsyth [9] used sparse depth measurements to segment the intensity image, and then used a smoothing filter to reconstruct the depth within each segment. Their work also exploited samples from nearby temporal frames in videos.

More recently, sparsity-promoting regularization—an essential component of compressive sensing [10], [11]—has provided more dramatic improvements in the quality of depth superresolution. For example, Li *et al.* [12] demonstrated improvements by combining dictionary learning and sparse coding algorithms. Ferstl *et al.* [13] relied on weighted total generalized variation (TGV) regularization for imposing a piecewise polynomial structure on depth. Gong *et al.* [14] combined the conventional MRF approach with an additional term promoting transform domain sparsity of the depth in an analysis form. In their recent work, Huang *et al.* [15] use the MRF model to jointly segment the objects and recover a higher quality depth. Schuon *et al.* [16] performed depth superresolution by taking several snapshots of a static scene from slightly displaced viewpoints and merging the measurements using sparsity of the weighted gradient of the depth.

Many natural images contain repetitions of similar patterns and textures. Current state-the-art image denoising methods such as nonlocal means (NLM) [17] and block matching and 3D filtering (BM3D) [18] take advantage of this redundancy by processing the image as a structured collection of patches. The original formulation of NLM was extended by Yang and Jacob [19] to more general inverse problems via introduction of specific NLM regularizers. Similarly, Danielyan *et al.* [20] have proposed a variational approach for general BM3D-based image reconstruction that inspired the current work. In the context of guided depth superresolution, NLM was used by Huhle *et al.* [21] and Park *et al.* [22] for reducing the amount of noise in the estimated depth. Lu *et al.* [23] combined a block-matching procedure with low-rank constraints for enhancing the resolution of a single depth image.

Our paper extends prior work on depth supperresolution by introducing a new variational formulation that imposes low-rank constraints in the regularization. Furthermore, our formulation is motion-adaptive, resulting in substantial improvement of the quality of the estimated depth. This paper extends [24] to include more details on the formulation and algorithms, as well as more extensive experimental results.

III. Motion-adaptive Regularization

Our approach estimates the high-resolution depth map by minimizing a cost function that—as typical in such problems—combines a data-fidelity term and a regularizer. Specifically, we impose a quadratic data fidelity term that controls the error between the measured and estimated depth values. The regularizer groups similar depth patches from multiple frames and penalizes the rank of the resulting structure. Our method implicitly adapts to motion by using patches at different positions in multiple frames. Thus, by effectively combining multiple views of the scene, it yields improved depth estimates.

A. Formulation

The depth sensing system collects a set of measurements denoted $\{\boldsymbol{\psi}_t\}_{t\in[1\ldots T]}$. Each measurement is considered as a downsampled version of a higher resolution depth map $\boldsymbol{\phi}_t \in \mathbb{R}^N$ using a subsampling operator $\mathbf{H}_t$. Our end goal is to recover this high-resolution depth map $\boldsymbol{\phi}_t$ for all t.

In the remainder of this work, we use N to denote the number of pixels in each frame, T to denote the number of temporal frames, and M to denote the total number of depth measurements. Furthermore, $\boldsymbol{\psi} \in \mathbb{R}^M$ denotes the vector of all the measurements, $\boldsymbol{\phi} \in \mathbb{R}^{NT}$ the complete sequence of

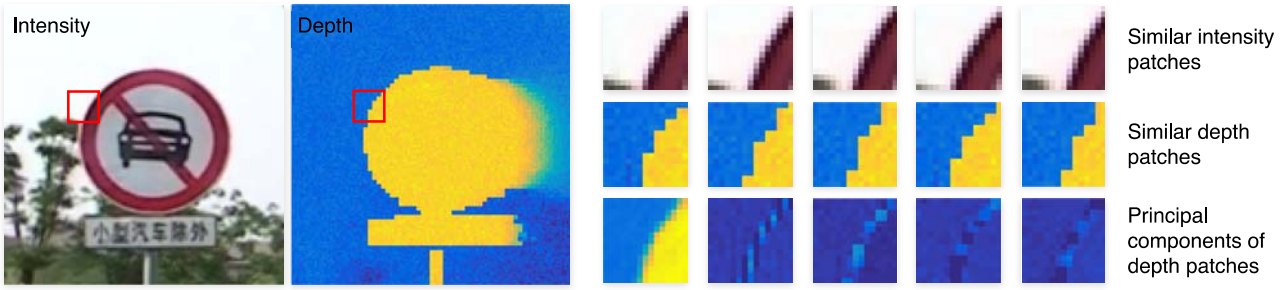

Fig. 3. Illustration of the way a rank penalty can improve the resolution and quality of the measured depth data. Left: A reference patch is selected from the region of interest of a dynamic sequence. Right: Collection of patches consisting of the reference patch as well as its similar patches in a 3D space-time neighborhood. The bottom row shows principal components extracted from the group of similar depth patches in the middle row. Note how most of the signal is in the first few components, which suggests that the rank penalty can effectively preserve important signal features while discarding most of the noise.

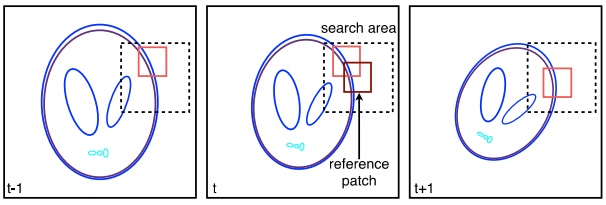

Fig. 2. Illustration of the block matching within a space-time search area. The area in the current frame t is centered at the reference patch. Search is also conducted in the same window position in multiple temporally adjacent frames. Similar patches are grouped together to construct a block $\boldsymbol{\beta}_p = \mathbf{B}_p\boldsymbol{\phi}$. Since patches from multiple frames are used, the final regularizer is implicitly motion adaptive.

high-resolution depth maps, and $\mathbf{H} \in \mathbb{R}^{M\times NT}$ the complete subsampling operator. We also have available the sequence of high-resolution intensity images from the optical camera, denoted $\mathbf{x} \in \mathbb{R}^{NT}$.

Using the above, a forward model for the depth recovery problem is given by

$$\boldsymbol{\psi} = \mathbf{H}\boldsymbol{\phi} + \mathbf{e}, \tag{1}$$

where $\mathbf{e} \in \mathbb{R}^M$ denotes the measurement noise. Thus, our objective becomes to recover high-resolution depth given the measured data $\boldsymbol{\psi}$ and $\mathbf{x}$, and the sampling operator $\mathbf{H}$.

As typical in such problems, we formulate the depth estimation task as an optimization problem

$$\widehat{\boldsymbol{\phi}} = \underset{\boldsymbol{\phi}\in\mathbb{R}^{NT}}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{\psi} - \mathbf{H}\boldsymbol{\phi}\|_{\ell_2}^2 + \sum_{p=1}^{P} \mathcal{R}(\mathbf{B}_p\boldsymbol{\phi}) \right\}, \tag{2}$$

where $\frac{1}{2}\|\boldsymbol{\psi}-\mathbf{H}\boldsymbol{\phi}\|_{\ell_2}^2$ enforces data fidelity and $\sum_{p=1}^{P}\mathcal{R}(\mathbf{B}_p\boldsymbol{\phi})$ is a regularization term that imposes prior knowledge about the depth map.

We form the regularization term by constructing sets of patches from the image. Specifically, we first define an operator $\mathbf{B}_p$, for each set of patches $p \in [1,\dots,P]$, where P is the number of such sets constructed. The operator extracts L patches of size B pixels from the depth image frames in $\boldsymbol{\phi}$. As illustrated in Fig. 2, each block $\boldsymbol{\beta}_p = \mathbf{B}_p\boldsymbol{\phi} \in \mathbb{R}^{B\times L}$ is obtained by first selecting a reference patch and then finding $L-1$ similar patches within the current frame as well as the adjacent temporal frames. Our method is motion adaptive since matching patches at different locations from multiple temporal frames are used. Thus our method automatically takes into account the motion of a patch, and its corresponding content, through the camera frames.

To determine similarity and to group similar patches together we use the intensity image as a guide. To reduce the computational complexity of the search, we restrict it to a space-time window of fixed size around the reference patch. We perform the same block matching procedure for the whole space-time image by moving the reference patch and by considering overlapping patches in each frame. Thus, each pixel in the signal $\boldsymbol{\phi}$ may contribute to multiple blocks.

The adjoint $\mathbf{B}_p^T$ of $\mathbf{B}_p$ simply corresponds to placing the patches in the block back to their original locations in $\boldsymbol{\phi}$. It satisfies the following property

$$\sum_{p=1}^{P} \mathbf{B}_p^T\mathbf{B}_p = \mathbf{R}, \tag{3}$$

where $\mathbf{R} = \mathrm{diag}(r_1,\dots,r_N) \in \mathbb{R}^{NT\times NT}$ and r_n denotes the total number of references to the nth pixel by the matrices $\{\mathbf{B}_p\}_{p=1,\dots,P}$. Therefore, the depth image $\boldsymbol{\phi}$ can be expressed in terms of an overcomplete representation using the blocks

$$\boldsymbol{\phi} = \mathbf{R}^{-1}\sum_{p=1}^{P} \mathbf{B}_p^T\mathbf{B}_p\boldsymbol{\phi}. \tag{4}$$

B. Rank regularization

Dynamic image sequences consist of continuously varying objects with many similar features in adjacent temporal frames. This is illustrated in Fig. 3, where a reference patch is selected from the region of interest of a dynamic dataset containing both intensity and depth sequences. The top row shows a collection of patches within a 3D space-time neighborhood that are similar to the reference patch. As mentioned before, similar patches are found via block-matching, which is similar to block-matching algorithms used in motion estimation. The self-similarity within blocks can be effectively exploited to remove noise or subsampling artifacts in the depth data. For example, the bottom row in Fig. 3 shows principal components extracted from the group of similar depth patches. Note that most of the signal features are included in the first

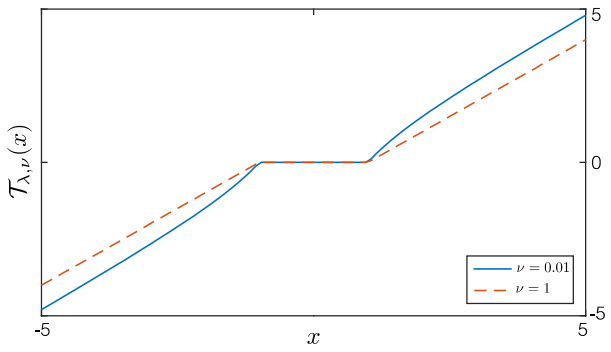

Fig. 4. Illustration of the ν-shrinkage operator $\mathcal{T}_{\nu,\lambda}$ for a fixed $\lambda = 1$ at two values of ν. For $\nu = 1$ the shrinkage is equivalent to soft-thresholding, while for $\nu \to 0$ it approaches hard-thresholding

few components, with other components containing noise. This suggests that a rank penalty applied on the block of similar patches can be used to capture the geometric information in the depth data.

Each block, represented as a matrix, contains multiple similar patches, *i.e.*, similar columns. Thus, we expect the matrix to have a low rank, making rank a natural regularizer for the problem

$$\mathcal{R}(\boldsymbol{\beta}) = \operatorname{rank}(\boldsymbol{\beta}). \quad (\boldsymbol{\beta} \in \mathbb{R}^{B\times L}) \tag{5}$$

By seeking a low-rank solution to (2), we exploit the similarity of blocks to guide superresolution while enforcing consistency with the observed data. However, the rank regularizer (5) is of little practical interest since its direct optimization is intractable. The most popular approach around this, first proposed by Fazel in [25], is to convexify the rank by replacing it with the nuclear norm:

$$\mathcal{R}(\boldsymbol{\beta}) = \lambda\|\boldsymbol{\beta}\|_* \triangleq \lambda \sum_{k=1}^{\min(B,L)} \sigma_k(\boldsymbol{\beta}), \tag{6}$$

where $\sigma_k(\boldsymbol{\beta})$ denotes the kth largest singular value of $\boldsymbol{\beta}$ and $\lambda > 0$ is a parameter controlling the amount of regularization.

In addition to its convexity, the nuclear norm is an appealing penalty to optimize because it also has a closed form proximal operator:

$$\begin{aligned}\operatorname{prox}_{\lambda\|\cdot\|_*}(\boldsymbol{\psi}) &\triangleq \operatorname*{arg\,min}_{\boldsymbol{\beta}\in\mathbb{R}^{B\times L}} \left\{ \frac{1}{2}\|\boldsymbol{\psi} - \boldsymbol{\beta}\|_F^2 + \lambda\|\boldsymbol{\beta}\|_* \right\} \\ &= \mathbf{u}\,\eta_\lambda(\boldsymbol{\sigma}(\boldsymbol{\psi}))\,\mathbf{v}^T, \end{aligned} \tag{7}$$

where $\boldsymbol{\psi} = \mathbf{u}\boldsymbol{\sigma}\mathbf{v}^T$ is the singular value decomposition (SVD) of $\boldsymbol{\psi}$ and η_λ is the soft-thresholding function applied to the diagonal matrix $\boldsymbol{\sigma}$.

Recent work has shown that nonconvex regularizers consistently outperform nuclear norm by providing stronger denoising capability without losing important signal compoenents [26]–[28]. In this paper, we use the nonconvex generalization to the nuclear norm proposed by Chartrand [26]

$$\mathcal{R}(\boldsymbol{\beta}) = \lambda\mathcal{G}_{\lambda,\nu}(\boldsymbol{\beta}) \triangleq \lambda \sum_{k=1}^{\min(B,L)} g_{\lambda,\nu}(\sigma_k(\boldsymbol{\beta})), \tag{8}$$

Here, the scalar function $g_{\lambda,\nu}$ is designed to satisfy

$$\min_{x\in\mathbb{R}} \left\{ \frac{1}{2}|x-y|^2 + \lambda g_{\lambda,\nu}(x) \right\} = h_{\lambda,\nu}(x), \tag{9}$$

where $h_{\lambda,\nu}$ is the ν-Huber function

$$h_{\lambda,\nu}(x) \triangleq \begin{cases} \frac{|x|^2}{2\lambda} & \text{if } |x| < \lambda^{1/(2-\nu)} \\ \frac{|x|^\nu}{\nu} - \delta & \text{if } |x| \geq \lambda^{1/(2-\nu)}, \end{cases} \tag{10}$$

with $\delta \triangleq (1/\nu - 1/2)\lambda^{\nu/(2-\nu)}$. Although $g_{\lambda,\nu}$ is nonconvex and has no closed form formula, its proximal operator does admit a closed form expression

$$\begin{aligned}\operatorname{prox}_{\lambda\mathcal{G}_{\lambda,\nu}}(\boldsymbol{\psi}) &\triangleq \operatorname*{arg\,min}_{\boldsymbol{\beta}\in\mathbb{R}^{B\times L}} \left\{ \frac{1}{2}\|\boldsymbol{\psi} - \boldsymbol{\beta}\|_F^2 + \lambda\mathcal{G}_{\lambda,\nu}(\boldsymbol{\beta}) \right\} \\ &= \mathbf{u}\,\mathcal{T}_{\lambda,\nu}(\boldsymbol{\sigma}(\boldsymbol{\psi}))\,\mathbf{v}^T, \end{aligned} \tag{11}$$

where $\mathcal{T}_{\lambda,\nu}$ is a poitwise ν-shrinkage operator defined as

$$\mathcal{T}_{\lambda,\nu}(x) \triangleq \max(0, |x| - \lambda|x|^{\nu-1})\frac{x}{|x|}. \tag{12}$$

For $\nu = 1$, ν-shrinkage (12) is equivalent to conventional soft thresholding (See illustration in Fig. 4). When $\nu \to 0$, it approaches hard thresholding, which is similar to principal component analysis (PCA) in the sense that it retains the significant few principal components.

Thus, the regularizer (8) is a computationally tractable alternative to the rank penalty. While the regularizer is not convex, it can still be efficiently optimized due to closed form of its proximal operator (11). Note that due to nonconvexity of our regularizer for $\eta < 1$, it is difficult to theoretically guarantee global convergence. However, we have empirically observed that our algorithms converge reliably over a broad spectrum of examples presented in Section V.

IV. Algorithms

We consider two approaches to solve the problem and recover the signal. The first exactly optimizes (2) using the alternating direction method of multipliers (ADMM). The second approach uses a simplified algorithm that alternates between enforcing data-consistency and low-rank penalty. While an approximation of (2), it often performs better because it provides robustness to the choice of λ in (6). At the end of this section, we also provide a brief discussion on the complexity of the algorithms.

A. Iterative optimization

To solve the optimization problem (2) under the rank regularizer (8), we first simplify notation by defining an operator $\mathbf{B} \triangleq (\mathbf{B}_1, \ldots, \mathbf{B}_P)$ and a vector $\boldsymbol{\beta} \triangleq \mathbf{B}\boldsymbol{\phi} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_P)$.

The minimization is performed using an augmented-Lagrangian (AL) method [29]. Specifically, we seek the critical

points of the following AL

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{s}) \triangleq \frac{1}{2}\|\boldsymbol{\psi} - \mathbf{H}\boldsymbol{\phi}\|_{\ell_2}^2 + \sum_{p=1}^{P} \mathcal{R}(\boldsymbol{\beta}_p) + \frac{\rho}{2}\|\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\phi}\|_{\ell_2}^2 + \boldsymbol{s}^T(\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\phi}) \tag{13a}$$

$$= \frac{1}{2}\|\boldsymbol{\psi} - \mathbf{H}\boldsymbol{\phi}\|_{\ell_2}^2 + \sum_{p=1}^{P} \mathcal{R}(\boldsymbol{\beta}_p) + \frac{\rho}{2}\|\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\phi} + \boldsymbol{s}/\rho\|_{\ell_2}^2 - \frac{1}{2\rho}\|\boldsymbol{s}\|_{\ell_2}^2, \tag{13b}$$

where $\rho > 0$ is the quadratic parameter and $\boldsymbol{s}$ is the dual variable that imposes the constraint $\boldsymbol{\beta} = \mathbf{B}\boldsymbol{\phi}$. Traditionally, an AL scheme solves (2) by alternating between a joint minimization step and an update step as

$$(\boldsymbol{\phi}^k, \boldsymbol{\beta}^k) \leftarrow \underset{\boldsymbol{\phi}\in\mathbb{R}^{NT}, \boldsymbol{\beta}\in\mathbb{R}^{P\times B\times L}}{\arg\min} \left\{\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{s}^{k-1})\right\} \tag{14a}$$

$$\boldsymbol{s}^k \leftarrow \boldsymbol{s}^{k-1} + \rho(\boldsymbol{\beta}^k - \mathbf{B}\boldsymbol{\phi}^k). \tag{14b}$$

However, the joint minimization step (14a) is typically computationally intensive. To reduce complexity, we separate (14a) into a succession of simpler steps using the well-established by now alternating direction method of multipliers (ADMM) [30]. The steps are as follows

$$\boldsymbol{\phi}^k \leftarrow \underset{\boldsymbol{\phi}\in\mathbb{R}^{NT}}{\arg\min} \left\{\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\beta}^{k-1}, \boldsymbol{s}^{k-1})\right\} \tag{15a}$$

$$\boldsymbol{\beta}^k \leftarrow \underset{\boldsymbol{\beta}\in\mathbb{R}^{P\times B\times L}}{\arg\min} \left\{\mathcal{L}(\boldsymbol{\phi}^k, \boldsymbol{\beta}, \boldsymbol{s}^{k-1})\right\} \tag{15b}$$

$$\boldsymbol{s}^k \leftarrow \boldsymbol{s}^{k-1} + \rho(\boldsymbol{\beta}^k - \mathbf{B}\boldsymbol{\phi}^k). \tag{15c}$$

By ignoring the terms that do not depend on $\boldsymbol{\phi}$, (15a) amounts to solving a quadratic problem

$$\boldsymbol{\phi}^k \leftarrow \underset{\boldsymbol{\phi}\in\mathbb{R}^{NT}}{\arg\min} \left\{\frac{1}{2}\|\boldsymbol{\psi} - \mathbf{H}\boldsymbol{\phi}\|_{\ell_2}^2 + \frac{\rho}{2}\|\mathbf{B}\boldsymbol{\phi} - \mathbf{z}^{k-1}\|_{\ell_2}^2\right\} \leftarrow (\mathbf{H}^T\mathbf{H} + \rho\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{H}\boldsymbol{\psi} + \rho\mathbf{B}^T\mathbf{z}^{k-1}), \tag{16}$$

where $\mathbf{z}^{k-1} \triangleq \boldsymbol{\beta}^{k-1} + \boldsymbol{s}^{k-1}/\rho$. Solving this quadratic equation is efficient since the inversion is performed on a diagonal matrix. Similarly, (15b) is solved by

$$\boldsymbol{\beta}^k \leftarrow \underset{\boldsymbol{\beta}\in\mathbb{R}^{P\times B\times L}}{\arg\min} \left\{\frac{\rho}{2}\|\boldsymbol{\beta} - \mathbf{y}^k\|_{\ell_2}^2 + \sum_{p=1}^{P} \mathcal{R}(\boldsymbol{\beta}_p)\right\}, \tag{17}$$

with $\mathbf{y}^k \triangleq \mathbf{B}\boldsymbol{\phi}^k - \boldsymbol{s}^{k-1}/\rho$. This step can be solved via block-wise application of the proximal operator (11) as

$$\boldsymbol{\beta}_p^k \leftarrow \mathrm{prox}_{(\lambda/\rho)\mathcal{G}_{\lambda,\nu}}(\mathbf{B}_p\boldsymbol{\phi}^k - \boldsymbol{s}_p^{k-1}/\rho), \tag{18}$$

for all $p \in [1, \ldots, P]$.

B. Simplified algorithm

The algorithm in Section IV-A can be significantly simplified by decoupling the enforcement of the data-fidelity from the enforcement of the rank-based regularization. The simplified algorithm reduces computational complexity while making estimation more uniform across the whole space-time depth image.

In particular, Danielyan [31] has argued that, due to inhomogeneous distribution of pixel references generated by matching across the image, using a penalty with a single regularization parameter highly penalizes pixels with a large number of references. The resulting nonuniform regularization makes the algorithm potentially oversensitive to the choice of the parameter λ. Instead, we rely on the simplified algorithm

$$\boldsymbol{\beta}_p^k \leftarrow \underset{\boldsymbol{\beta}_p\in\mathbb{R}^{B\times L}}{\arg\min} \left\{\frac{1}{2}\|\boldsymbol{\beta}_p - \mathbf{B}_p\boldsymbol{\phi}^{k-1}\|_F^2 + \mathcal{R}(\boldsymbol{\beta}_p)\right\} \tag{19a}$$

$$\boldsymbol{\phi}^k \leftarrow \underset{\boldsymbol{\phi}\in\mathbb{R}^{NT}}{\arg\min} \left\{\frac{1}{2}\|\boldsymbol{\psi} - \mathbf{H}\boldsymbol{\phi}\|_{\ell_2}^2 + \frac{\rho}{2}\|\boldsymbol{\phi} - \tilde{\boldsymbol{\phi}}^k\|_{\ell_2}^2\right\} \tag{19b}$$

where $\tilde{\boldsymbol{\phi}}^k \triangleq \mathbf{R}^{-1}\mathbf{B}^T\boldsymbol{\beta}^k$, and $\lambda > 0$ is the regularization and $\rho > 0$ is the quadratic parameters.

To solve (19a) we apply the proximal operator

$$\boldsymbol{\beta}_p^k \leftarrow \mathrm{prox}_{\lambda\mathcal{G}_{\lambda,\nu}}(\mathbf{B}_p\boldsymbol{\phi}^{k-1}), \tag{20}$$

for all $p \in [1, \ldots, P]$. Next, (19b) reduces to a linear step

$$\boldsymbol{\phi}^k \leftarrow (\mathbf{H}^T\mathbf{H} + \rho\mathbf{I})^{-1}(\mathbf{H}^T\boldsymbol{\psi} + \rho\tilde{\boldsymbol{\phi}}^k). \tag{21}$$

There are substantial similarities between algorithms (15) and (19). The main differences are that we eliminated the dual variable $\boldsymbol{s}$ and simplified the quadratic subproblem (16).

C. Computational Complexity

To conclude this section, we described two methods for improving the resolution of depth data in a motion adaptive fashion. While our approach has nontrivial complexity, it is easily manageable.

Specifically, the main sub-units of our algorithm are block-matching and iterative optimization. Among the two, block-matching is the costlier, involving pairwise comparisons of patch differences within a search window. It is performed only once per frame and requires $O(PWB)$ floating point operations, where P is the total number of patches in the frame, W is the number of patches in the search window, and B is the size of the patch. The cost is similar to stereo matching, with the caveat that some algorithms might use heuristics to limit the search scope, *i.e.*, W.

The main cost of each iteration is an SVD computed on each block generated for each patch, with cost $O(B^2L + L^3)$, where L is the number of patches included in each block. The total iteration cost is P times that, and the method needs approximately 10-20 iterations to converge.

Furthermore, both block-matching and iterative optimization operate independently on each block. They can thus be easily parallelized, which is something our implementation does not exploit. Of course, GPUs can further speed up computation and exploit parallelism, for example, by performing real-time block-matching [32]. Also, note that there is no need to store matrices $\mathbf{H}$ and $\mathbf{B}$ explicitly in the memory, since they are implemented as a subsampling and an extraction operators, respectively.

	Flower				Lawn				Road			
	2×	3×	4×	5×	2×	3×	4×	5×	2×	3×	4×	5×
Linear	25.62	22.81	21.07	20.15	28.32	25.89	24.32	23.05	25.44	22.78	21.16	20.18
Guided filter	23.72	22.57	21.32	20.52	27.15	25.81	24.58	23.45	24.19	22.76	21.43	20.62
TV-2D	26.52	23.17	21.30	20.32	29.97	26.56	24.70	23.31	26.30	23.21	21.44	20.38
WTGV-2D	26.73	23.54	21.68	20.66	30.16	26.87	24.94	23.66	26.44	23.20	21.38	20.57
WTV-3D	26.84	23.56	21.69	20.72	30.45	27.00	25.09	23.68	26.54	23.49	21.69	20.73
GDS-2D	27.76	23.91	21.78	20.58	31.27	27.58	25.36	23.88	27.39	23.89	21.87	20.70
DS-3D	28.00	23.82	21.79	20.64	31.37	27.34	25.23	23.69	27.30	23.92	21.75	20.56
ADMM-3D	29.76	25.07	22.58	21.26	33.06	**28.62**	**26.07**	**24.39**	28.58	25.18	22.74	21.39
GDS-3D	**30.04**	**25.34**	**22.79**	**21.42**	**32.54**	28.51	26.02	24.36	**29.10**	**25.52**	**22.96**	**21.66**

TABLE I

QUANTITATIVE COMPARISON ON THREE VIDEO SEQUENCES WITH ADDED NOISE OF 30 DB. THE QUALITY OF FINAL DEPTH IS EVALUATED IN TERMS OF SNR FOR FOUR DIFFERENT DOWNSIZING FACTORS OF 2, 3, 4, AND 5. THE BEST RESULT FOR EACH SCENARIO IS HIGHLIGHTED.

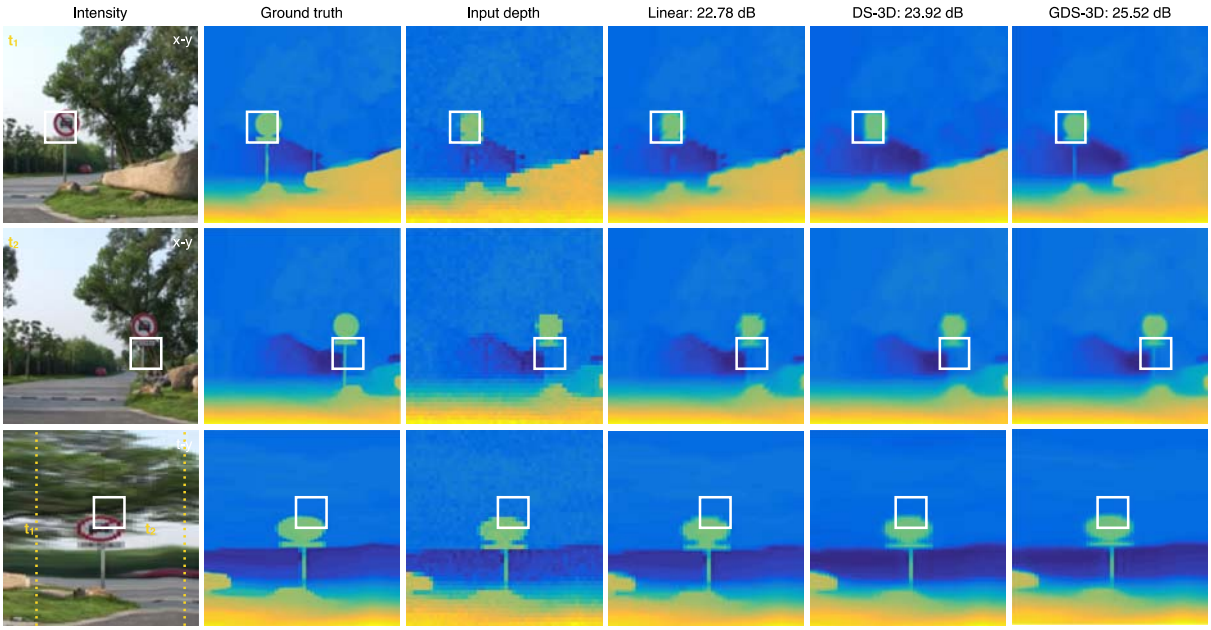

Fig. 7. Visual evaluation on *Road* video sequence. Estimation of depth from its 3× downsized version at 30 dB input SNR. Row 1 shows the data at time instance $t = 9$. Row 2 shows the data at the time instance $t = 47$. Row 3 shows the t-y profile of the data at $x = 64$. Highlights indicate some of the areas where depth estimated by GDS-3D recovers details missing in the depth estimate of DS-3D that does not use intensity information.

V. EXPERIMENS

To verify our development, we report results on extensive simulations using our guided depth superresolution algorithms. In particular, we compare results of both the ADMM approach (denoted *ADMM-3D*) and its simplified variant (denoted *GDS-3D*) against six alternative methods.

As the first and the simplest reference method, we consider standard linear interpolation (*Linear*). Another baseline method is the guided image filtering approach (*Guided filter*) by He *et al.* [8]. Additionally, we consider methods relying in some form of total variation (TV) regularization, one of the most widely used regularizers in the context of image reconstruction due to its ability to reduce noise while preserving image edges [33]. Specifically, we consider depth interpolation using TV-regularized least squares on a frame-by-frame basis (*TV-2D*) [34]. We also consider the weighted-TV formulation proposed by Ferstl *et al.* [13] (*WTGV-2D*), also operating on a frame-by-frame basis. This formulation uses a weighted anisotropic total generalized variation, where weighting is computed using the guiding intensity image, thus promoting edge co-occurrence in both modalities.

Finally, we consider a weighted-TV formulation which includes time, *i.e.*, multiple frames (*WTV-3D*), with weights computed using the guiding intensity image, as before. Specifically, the depth image is computed by solving an optimization problem

$$\widehat{\boldsymbol{\phi}} = \underset{\boldsymbol{\phi}\in\mathbb{R}^{NT}}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{\psi} - \mathbf{H}\boldsymbol{\phi}\|_{\ell_2}^2 + \mathcal{R}_{\text{TV}}(\boldsymbol{\phi}) \right\}, \tag{22}$$

where

$$\mathcal{R}_{\text{TV}}(\boldsymbol{\phi}) \triangleq \sum_{i=1}^{NT} \left(w_{xi}|[\mathbf{D}_x\boldsymbol{\phi}]_i + w_{yi}|[\mathbf{D}_y\boldsymbol{\phi}]_i| + w_{ti}|[\mathbf{D}_t\boldsymbol{\phi}]_i| \right). \tag{23}$$

Here, the linear operators $\mathbf{D}_x, \mathbf{D}_y$, and $\mathbf{D}_t$ are the forward finite-difference operators along the horizontal, vertical, and temporal directions, respectively. The weights $\mathbf{w}_x, \mathbf{w}_y$, and $\mathbf{w}_t$ are specified using the camera image, $w_{di} = \lambda \exp(-\gamma|[\mathbf{D}_d\mathbf{x}]_i|)$ with $d \in \{x, y, t\}$, thus promoting

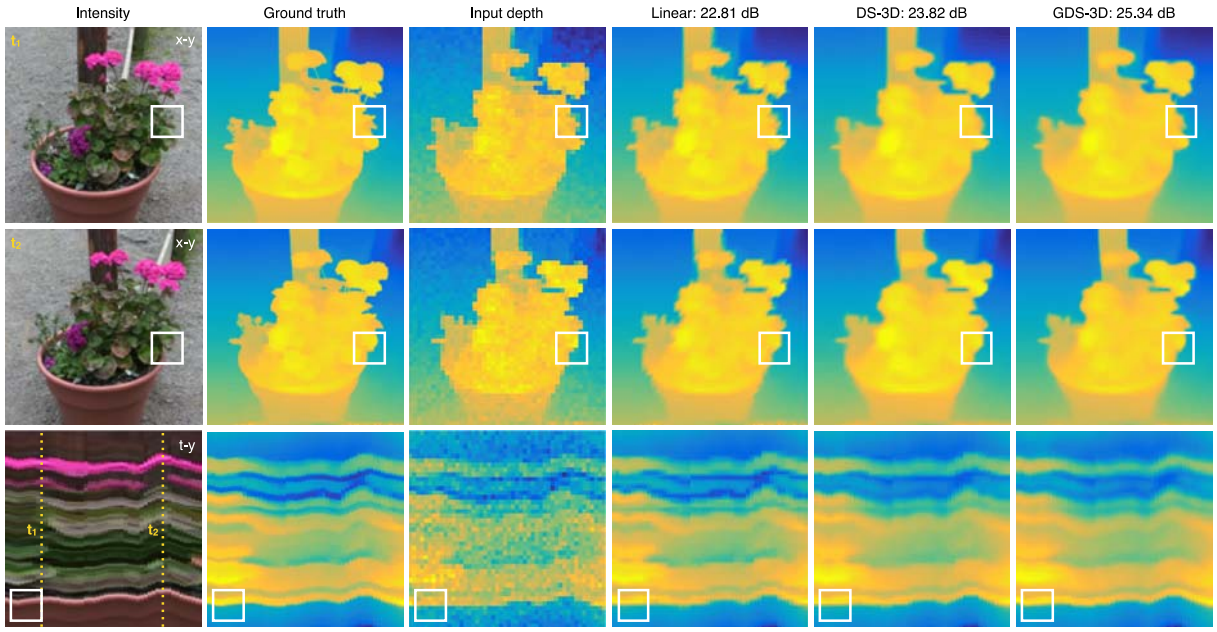

Fig. 8. Visual evaluation on *Flower* video sequence. Estimation of depth from its $3\times$ downsized version at 30 dB input SNR. Row 1 shows the data at time instance $t = 10$. Row 2 shows the data at the time instance $t = 40$. Row 3 shows the t-y profile of the data at $x = 64$. Highlights indicate some of the areas where depth estimated by GDS-3D recovers details missing in the depth estimate of DS-3D that does not use intensity information.

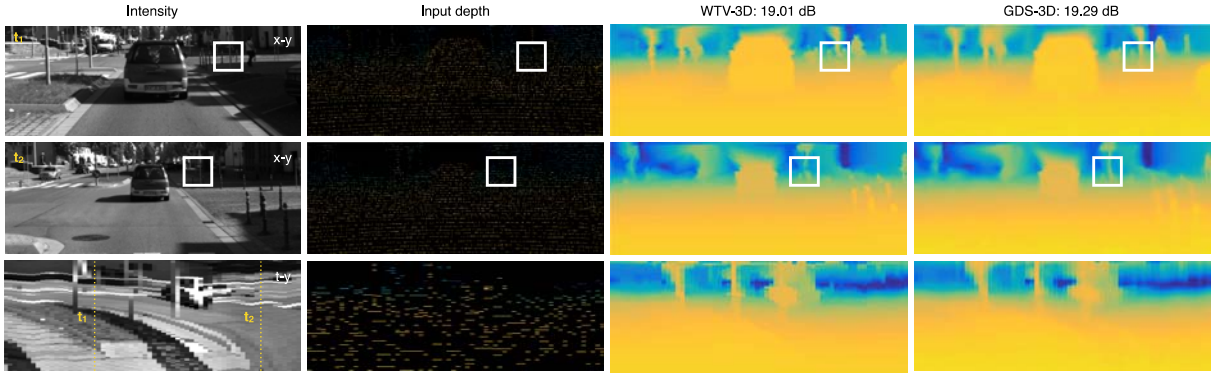

Fig. 9. Visual evaluation on *KITTI* dataset. Estimation of depth images of size 192×512 with 64 time frames from 247794 lidar measurements, which corresponds to a measurement rate of just 3.94%. Row 1 shows the data at time instance $t = 20$. Row 2 shows the data at the time instance $t = 53$. Row 3 shows the t-y profile of the data at $x = 96$. Highlights indicate some of the areas where depth estimated by GDS-3D recovers details missing in the WTV-3D estimate.

edge alignment between the camera image and the depth-map, as controlled by the regularization parameters $\lambda, \gamma > 0$. The history of space-time application of TV goes all the way to the seminal works for optical-flow estimation in video [35], [36], with more recent applications in video deconvolution, denoising, and disparity refinement [37], [38]. The minimization (22) is solved with the fast iterative shrinkage/thresholding algorithm (FISTA), which is particularly well-suited for such large-scale optimization problems [39], [40].

We also compare these methods to two variations of our algorithm. To illustrate the potential gains of our method due to temporal information, we run our simplified algorithm on a frame-by-frame basis (*GDS-2D*), *i.e.*, using no temporal information. Similarly, to illustrate the gains due to intensity information, we also run the simplified algorithm by performing block matching on the low-resolution depth as opposed to intensity (*DS-3D*). This last approach conceptually corresponds to denoising the initial depth using our motion adaptive low-rank prior.

To improve convergence speed, we initialized all the iterative algorithms with the solution of linear interpolation. The parameters of the Guided filter, TV-2D, WTGV-2D, and WTV-3D were optimized for the best signal-to-noise ratio (SNR) performance. Similarly, the regularization parameters λ for GDS-2D, DS-3D, ADMM-3D, and GDS-3D were optimized for the best SNR performance. However, in order to reduce the computational time, the selection was done from a restricted set of three predetermined parameters. The methods TV-2D, WTV-3D, GDS-2D, DS-3D, ADMM-3D, and GDS-3D were all run for a maximum of $t_{\max} = 100$ iterations with an

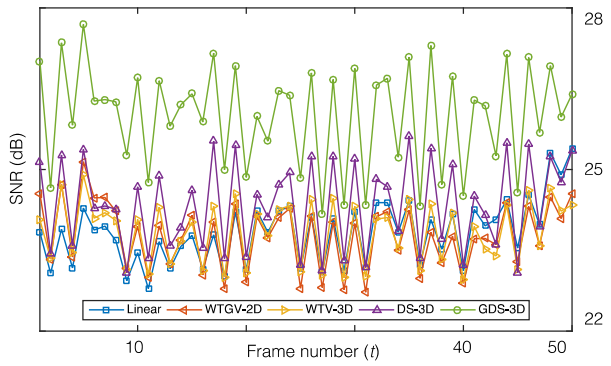

Fig. 5. Quantitative evaluation on *Road* video sequence. Estimation of depth from its 3× downsized version at 30 dB input SNR. We plot the reconstruction SNR against the video frame number. The plot illustrates potential gains that can be obtained by fusing intensity and depth information in a motion-adaptive way.

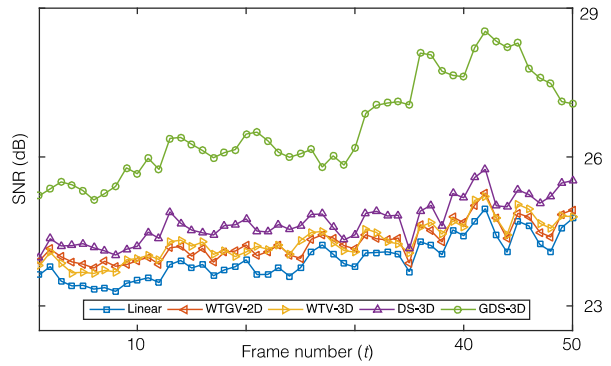

Fig. 6. Quantitative evaluation on *Flower* video sequence. Estimation of depth from its 3× downsized version at 30 dB input SNR. We plot the reconstruction SNR against the video frame number. The plot illustrates potential gains that can be obtained by fusing intensity and depth information in a motion-adaptive way.

additional stopping criterion based on measuring the relative change of the solution in two successive iterations

$$\frac{\|\boldsymbol{\phi}^k - \boldsymbol{\phi}^{k-1}\|_{\ell_2}}{\|\boldsymbol{\phi}^{k-1}\|_{\ell_2}} \leq \delta, \tag{24}$$

where $\delta = 10^{-4}$ is the desired tolerance level. We selected the neighborhood size parameter of Guided filter to 5×5, which is the default value provided in the implementation by MATLAB. Similarly, other parameter of WTGV-2D were selected as suggested in the code provided by the authors; in particular, we run the algorithm for a maximum of 1000 iterations with the stopping tolerance of 0.1. In all experiments, the patch size was set to 5×5 pixels, the space-time window size to $11 \times 11 \times 3$ pixels, and the number of similar patches was fixed to 10. Parameters ν and ρ were hand selected to 0.02 and 1, respectively.

For quantitative comparison of the algorithms, we rely on the data-set by Zhang *et al.* [41], which consists of three video sequences *Flower*, *Lawn*, and *Road*, containing both intensity and depth information on the scenes. We considered images of size 128×128 with 50 time frames. The ground truth depth was downsized by factors of 2, 3, 4, and 5, and was corrupted by additive Gaussian noise corresponding to SNR of 30 dB. Table I reports the SNR of superresolved depth for all the algorithms and downsizing factors. Figures 5 and 6 illustrate the evolution of SNR against the frame number for *Road* and *Flower*, respectively, at downsizing factor of 3. The effectiveness of the proposed approach can also be appreciated visually in Figs. 7 and 8. For these experiments, our unoptimized MATLAB code required about 8 seconds per frame for block matching, and about 5 seconds per frame per iteration for reconstruction. As mentioned in Section IV-C these times can be substantially reduced by code optimization and by exploiting the inherent parallelism of the algorithm.

For additional validation, we test the algorithm on the KITTI dataset [42]. The dataset contains intensity images from PointGray Flea2 cameras and 3D point clouds from a Velodyne HDL-64E, which have been calibrated a priori using specific known targets. We consider a region of interest of 192×512 pixels with 64 time frames. The 495588 total lidar measurements are randomly and uniformly split into a reconstruction and a validation sets. Note that this implies a depth measurement rate of just 3.94%. We then estimate depth from the reconstruction set using WTV-3D, as well as GDS-3D, and use the validation set to evaluate the quality of the results; these are illustrated in Fig. 9. Here, our MATLAB code required about 50 seconds per frame for block matching, and about 28 seconds per frame per iteration for reconstruction.

The examples and simulations in this section, validate our claim: the quality of estimated depth can be considerably boosted by properly incorporating temporal information into the reconstruction procedure. Comparison of GDS-2D against GDS-3D highlights the importance of additional temporal information. The approach we propose is implicitly motion adaptive and can thus preserve temporal edges substantially better than alternative approaches such as WTV-3D. Moreover, comparison between DS-3D and GDS-3D highlights that the usage of intensity significantly improves the performance of the algorithm. Note also the slight improvement in SNR of GDS-3D over ADMM-3D. This is consistent with the arguments in [31] that suggest to decouple data-fidelity from the enforcement of prior constraints when using block-matching-based methods.

VI. Conclusion

We presented a novel motion-adaptive approach for guided superresolution of depth maps. Our method identifies and groups similar patches from several frames, which are then supperresolved using a rank regularizer. Using this approach, we can produce high-resolution depth sequences from significantly down-sized low-resolution ones. Compared to the standard techniques, the proposed method preserves temporal edges in the solution and effectively mitigates noise in practical configurations.

While our formulation has higher computational complexity than standard approaches that process each frame individually, it allows us to incorporate a very effective regularization for stabilizing the inverse problem associated with

superresolution. The algorithms we describe enable efficient computation and straightforward implementation by reducing the problem to a succession of straightforward operations. Our experimental results demonstrate the considerable benefits of incorporating time and motion adaptivity into inverse-problems for depth estimation.

REFERENCES

[1] J. Diebel and S. Thrun, "An application of markov random ffield to range sensing," in *Proc. Advances in Neural Information Processing Systems 18*, vol. 18, Vancouver, BC, Canada, December 5-8, 2005, pp. 291–298.

[2] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," in *ACM Trans. Graph. (Proc. SIGGRAPH 2007)*, vol. 26, San Diego, CA, USA, August 5-9, 2007, p. 96.

[3] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, MN, USA, June 17-22, 2007, pp. 1–8.

[4] D. Chan, H. Buisman, C. Theobalt, and S. Thrun, "A noise-aware filter for real-time depth upsampling," in *ECCV Workshop on Multicamera and Multimodal Sensor Fusion Algorithms and Applications*, 2008.

[5] M.-Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Portland, OR, USA, June 23-28, 2013, pp. 169–176.

[6] J. Dolson, J. Baek, C. Plagemann, and S. Thrun, "Upsampling range data in dynamic environments," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, vol. 23, San Francisco, CA, USA, June 13-18, 2010, pp. 1141–1148.

[7] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. ECCV*, Hersonissos, Greece, September 5-11, 2010, pp. 1–14.

[8] ——, "Guided image filtering," *IEEE Trans. Patt. Anal. and Machine Intell.*, vol. 35, no. 6, pp. 1397–1409, June 2013.

[9] J. Lu and D. Forsyth, "Sparse depth super resolution," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, June 7-12, 2015, pp. 2245–2253.

[10] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, February 2006.

[11] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[12] Y. Li, T. Xue, L. Sun, and J. Liu, "Joint example-based depth map super-resolution," in *Proc. IEEE Int. Con. Multi.*, Melbourne, VIC, Australia, July 9-13, 2012, pp. 152–157.

[13] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Sydney, NSW, Australia, December 1-8, 2013, pp. 993–1000.

[14] X. Gong, J. Ren, B. Lai, C. Yan, and H. Qian, "Guided depth upsampling via a cosparse analysis model," in *Proc. IEEE CVPR WKSHP*, Columbus, OH, USA, June 23-28, 2014, pp. 738–745.

[15] W. Huang, X. Gong, and M. Y. Yang, "Joint object segmentation and depth upsampling," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 192–196, February 2015.

[16] S. Schouon, C. Theobalt, J. Davis, and S. Thrun, "LidarBoost: Depth superresolution for ToF 3D shape scanning," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, June 20-25, 2009, pp. 343–350.

[17] A. Buades, B. Coll, and J. M. Morel, "Image denoising methods. A new nonlocal principle." *SIAM Rev*, vol. 52, no. 1, pp. 113–147, 2010.

[18] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 16, pp. 2080–2095, August 2007.

[19] Z. Yang and M. Jacob, "Nonlocal regularization of inverse problems: A unified variational framework," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3192–3203, August 2013.

[20] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1728, April 2012.

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

[22] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3D-TOF cameras," in *Proc. IEEE Int. Conf. Comp. Vis.*, Barcelona, Spain, November 6-13, 2011, pp. 1623–1630.

[23] S. Lu, X. Ren, and F. Liu, "Depth enhancement via low-rank matrix completion," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, June 23-28, 2014, pp. 3390–3397.

[24] U. S. Kamilov and P. T. Boufounos, "Depth superresolution using motion adaptive regularization," in *Proc. 2015 IEEE Int. Conf. Multimedia & Expo Workshops (ICMEW 2016)*, Seattle, WA, USA, July 11-15, 2016.

[25] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Stanford University, 2002.

[26] R. Chartrand, "Nonconvex splitting for regularized Low-Rank + Sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5810–5819, November 2012.

[27] Y. Hu, S. G. Lingala, and M. Jacob, "A fast majorize-minimize algorithm for the recovery of sparse and low-rank matrices," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 742–753, February 2012.

[28] H. Yoon, K. S. Kim, D. Kim, Y. Bresler, and J. C. Ye, "Motion adaptive patch-based low-rank approach for compressed sensing cardiac cine MRI," *IEEE Trans. Med. Imag.*, vol. 33, no. 11, pp. 2069–2085, November 2014.

[29] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. Springer, 2006.

[30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[31] A. Danielyan, "Block-based collaborative 3-D transform domain modeing in inverse imaging," Publication 1145, Tampere University of Technology, May 2013.

[32] E. Monteiro, M. Maule, F. Sampaio, C. Diniz, B. Zatt, and S. Bampi, "Real-time block matching motion estimation onto GPGPU," in *Proc. IEEE Int. Conf. Image Proc. (ICIP'12)*, Orlando, FL, USA, September 30-October 3, 2012, pp. 1693–1696.

[33] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, November 1992.

[34] J. Castorena, U. S. Kamilov, and P. T. Boufounos, "Autocalibration of LIDAR and optical cameras via edge alignment," in *IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP 2016)*, Shanghai, China, March 20–25 2016, pp. 2862–2866.

[35] J. Weickert and C. Schnörr, "Variational optic flow computation with a spatio-temporal smoothness constraint," *J. Math. Imaging Vision*, vol. 14, no. 3, pp. 245–255, May 2001.

[36] C. Zach, T. Pock, and H. Bischof, *A duality based approach for realime TV-ℓ_1 optical flow*, ser. Lecture Notes in Computer Science. Springer, 2007, vol. 4713, ch. Pattern Recognition, pp. 214–223.

[37] 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, November 2011.

[38] M. S. Hosseini and K. N. Plataniotis, "High-accuracy total variation with application to compressed video sensing," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3869–3884, September 2014.

[39] A. Beck and M. Teboulle, "Fast gradient-based algorithm for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, November 2009.

[40] U. S. Kamilov, "A parallel proximal algorithm for anisotropic total variation minimization," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 539–548, February 2017.

[41] G. Zhang, J. Jia, T. Wong, and H. Bao, "Consistent depth maps recovery from a video sequence," *IEEE Trans. Patt. Anal. and Machine Intell.*, vol. 31, no. 6, pp. 974–988, June 2009.

[42] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, November 2013.

Ulugbek S. Kamilov (S'11–M'15) is a Research Scientist in the Computational Sensing team at Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA. Dr. Kamilov obtained his B.Sc. and M.Sc. in Communication Systems, and Ph.D. in Electrical Engineering from the École polytechnique fédérale de Lausanne (EPFL), Switzerland, in 2008, 2011, and 2015, respectively. In 2007, he was an Exchange Student at Carnegie Mellon University (CMU), Pittsburgh, PA, USA, in 2010, a Visiting Student at Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, and in 2013, a Visiting Student Researcher at Stanford University, Stanford, CA, USA.

Dr. Kamilov's research focus is computational imaging with an emphasis on the development of large-scale computational techniques for biomedical and industrial applications. His research interests cover large-scale optimization, multimodal imaging, machine learning, through-the-wall imaging, and distributed radar sensing. He has co-authored 17 journal and 30 conference publications in these areas. His Ph.D. thesis work on Learning Tomography (LT) was selected as a finalist for EPFL Doctorate Awards 2016 and was featured in the "News and Views" of Nature. Since 2016, Dr. Kamilov is a member IEEE Special Interest Group on Computational Imaging.

Petros T. Boufounos (S'02-M'06-SM'13) is a Senior Principal Research Scientist and the Computational Sensing Team Leader at Mitsubishi Electric Research Laboratories (MERL), and a visiting scholar at the Rice University Electrical and Computer Engineering department. Dr. Boufounos completed his undergraduate and graduate studies at MIT. He received the S.B. degree in Economics in 2000, the S.B. and M.Eng. degrees in Electrical Engineering and Computer Science (EECS) in 2002, and the Sc.D. degree in EECS in 2006. Between September 2006 and December 2008, he was a postdoctoral associate with the Digital Signal Processing Group at Rice University. Dr. Boufounos joined MERL in January 2009. Dr. Boufounos' immediate research focus includes signal acquisition and processing, frame theory, quantization and data representations. He is also interested into how signal acquisition interacts with other fields that use sensing extensively, such as machine learning, robotics and mechatronics. Dr. Boufounos is a Senior Area Editor at IEEE Signal Processing Letters. He is also a senior member of the IEEE and a member of Sigma Xi, Eta Kappa Nu, and Phi Beta Kappa.