DEPTH SUPERRESOLUTION USING MOTION ADAPTIVE REGULARIZATION

Ulugbek S. Kamilov and Petros T. Boufounos

Mitsubishi Electric Research Laboratories
201 Broadway
Cambridge, MA 02140, USA
{*kamilov,petrosb*}@*merl.com*

ABSTRACT

Spatial resolution of depth sensors is often significantly lower compared to that of conventional optical cameras. Recent work has explored the idea of improving the resolution of depth using higher resolution intensity as a side information. In this paper, we demonstrate that further incorporating temporal information in videos can significantly improve the results. In particular, we propose a novel approach that improves depth resolution, exploiting the space-time redundancy in the depth and intensity using motion-adaptive low-rank regularization. Experiments confirm that the proposed approach substantially improves the quality of the estimated high-resolution depth. Our approach can be a first component in systems using vision techniques that rely on high resolution depth information.

1. INTRODUCTION

A challenge 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, especially in outdoors environments.

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.

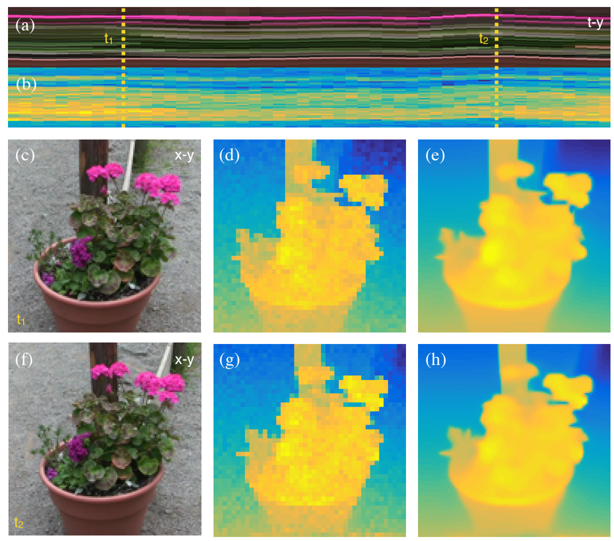

Fig. 1. Our motion adaptive method recovers a high-resolution depth sequence from high-resolution intensity and low-resolution depth sequences by imposing rank constraints on the depth patches: (a) and (b) t-y slices of the color and depth sequences, respectively, at a fixed x; (c)–(e) x-y slices at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

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.

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. 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.

Our method builds upon prior work on patch-based methods and low-rank regularization, successful in 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. This work provides three key contributions: (a) a new problem formulation incorporating temporal information, (b) two new optimization strategies to compute the resulting estimation problem, and (c) an experimental validation demonstrating that integrating temporal information is invaluable in the superresolution problem.

2. RELATED WORK

In the last decade, guided depth superresolution has received significant attention. Early work showed the potential of the approach by modeling the co-occurence of edges in depth and intensity with Markov Random Fields (MRF) [1]. An alternative approach based on joint bilateral filtering was proposed in [2, 3], where intensity is used to set the weights of the filter. This approach was further refined in [4], incorporating local depth statistics, and in [5], using geodesic distances to determine the weights. This approach has also been extended to dynamic sequences, compensating for different data rates in the depth and intensity sensors [6]. In addition, [7, 8] proposed a guided image filtering approach, improving edge preservation.

More recently, sparsity-promoting regularization—inspired by developments in compressive sensing [9, 10]—has provided more dramatic improvements in the quality of depth superresolution. For example [11] demonstrated improvements by combining dictionary learning and sparse coding algorithms. Instead, [12] relies on weighted total generalized variation (TGV) regularization for imposing a piecewise polynomial structure on depth. Sparsity regularization has also been combined with the conventional MRF approach in [13]. More recently, [14] uses the MRF model to jointly segment the objects and recover a higher quality depth. Similar tools are used in [15], performing 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) [16] and block matching and 3D filtering (BM3D) [17] take advantage of this redundancy by processing the image as a structured collection of patches. The original formulation of NLM was extended in [18] to more general inverse problems, introducing specific NLM regularizers. Similarly, [19] propsed a variational approach for general BM3D-based image reconstruction that inspired the current work. In the context of guided depth superresolution, NLM was used in [20, 21] to reduce noise in the estimated depth. In [22], block-matching is combined with low-rank constraints to enhance the resolution of a single depth image.

Our paper extends prior work by introducing a new variational formulation that imposes low-rank constraints in the regularization. Furthermore, it incorporates motion adaptivity, with substantial improvement in the depth estimation quality.

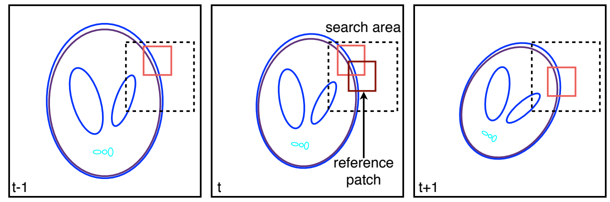

Fig. 2. Illustration of the block matching within a space-time search area. The area in the current frame t is centered at the reference patch. Search is also conducted in the same window position in multiple temporally adjacent frames. Similar patches are grouped together to construct a block $\boldsymbol{\beta}_p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

3. PROPOSED APPROACH

3.1. Problem 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 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. The regularization $\sum_{p=1}^{P} \mathcal{R}(\mathbf{B}_p\boldsymbol{\phi})$ imposes prior knowledge about the depth map.

The regularization term is applied on sets of patches from the image. Specifically, we define an operator $\mathbf{B}_p$, for each set of patches $p \in [1,\ldots,P]$, 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.

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.

3.2. Rank regularization

Each block, represented as a matrix, contains multiple similar patches, i.e., similar columns. Thus, the matrix should have low rank, making rank a natural regularizer for the problem. 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 resulting low-rank optimization problem non-convex and intractable.

The most popular approach convexification by replacing the rank with the nuclear norm [23]. Recent work has shown that nonconvex regularizers consistently outperform the nuclear norm by providing stronger denoising capability without losing important signal compoeenents [24–26]. In this paper, we use the nonconvex generalization in [24]

$$\mathcal{R}(\boldsymbol{\beta}) = \lambda \mathcal{G}_{\lambda,\nu}(\boldsymbol{\beta}) \triangleq \lambda \sum_{k=1}^{\min(B,L)} g_{\lambda,\nu}(\sigma_k(\boldsymbol{\beta})), \tag{3}$$

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

$$\min_{x\in\mathbb{R}} \left\{ \frac{1}{2}|x-y|^2 + \lambda g_{\lambda,\nu}(x) \right\} = h_{\lambda,\nu}(x), \tag{4}$$

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

$$h_{\lambda,\nu}(x) \triangleq \begin{cases} \frac{|x|^2}{2\lambda} & \text{if } |x| < \lambda^{1/(2-\nu)} \\ \frac{|x|^\nu}{\nu} - \delta & \text{if } |x| \geq \lambda^{1/(2-\nu)}, \end{cases} \tag{5}$$

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. Thus, the regularizer (3) 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. Note that due to nonconvexity of the regularizer, 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 4.

3.3. Iterative optimization

To solve the optimization problem (2) under the rank regularizer (3), we first simplify notation by defining an operator $\mathbf{B} \triangleq (\mathbf{B}_1, \dots, \mathbf{B}_P)$ and a vector $\boldsymbol{\beta} \triangleq \mathbf{B}\boldsymbol{\phi} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_P)$.

The minimization is performed using an augmented-Lagrangian (AL) method [27]. Specifically, we seek the critical points of the following AL

$$\begin{aligned} \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{s}) \triangleq\ & \frac{1}{2}\|\boldsymbol{\psi} - \mathbf{H}\boldsymbol{\phi}\|_{\ell_2}^2 + \sum_{p=1}^{P} \mathcal{R}(\boldsymbol{\beta}_p) \\ & + \frac{\rho}{2}\|\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\phi}\|_{\ell_2}^2 + \boldsymbol{s}^T(\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\phi}) \end{aligned} \tag{6a}$$

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—typically computationally intensive—joint minimization step and an update step. To reduce complexity, we separate this step into a succession of simpler steps using the well-established by now alternating direction method of multipliers (ADMM) [28]. The steps are as follows

$$\boldsymbol{\phi}^k \leftarrow \underset{\boldsymbol{\phi}\in\mathbb{R}^{NT}}{\arg\min} \left\{ \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\beta}^{k-1}, \boldsymbol{s}^{k-1}) \right\} \tag{7a}$$

$$\boldsymbol{\beta}^k \leftarrow \underset{\boldsymbol{\beta}\in\mathbb{R}^{P\times B\times L}}{\arg\min} \left\{ \mathcal{L}(\boldsymbol{\phi}^k, \boldsymbol{\beta}, \boldsymbol{s}^{k-1}) \right\} \tag{7b}$$

$$\boldsymbol{s}^k \leftarrow \boldsymbol{s}^{k-1} + \rho(\boldsymbol{\beta}^k - \mathbf{B}\boldsymbol{\phi}^k). \tag{7c}$$

Step (7a) is a simple quadratic problem, while (7b) can be computed using the proximal operator in [24].

3.4. Simplified algorithm

The algorithm above 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, 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 [29] . The resulting nonuniform regularization makes the algorithm potentially oversensitive to the choice of the parameter λ.

Instead, we rely on the simplified algorithm

$$\boldsymbol{\beta}_p^k \leftarrow \underset{\boldsymbol{\beta}_p\in\mathbb{R}^{B\times L}}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{\beta}_p - \mathbf{B}_p\boldsymbol{\phi}^{k-1}\|_F^2 + \mathcal{R}(\boldsymbol{\beta}_p) \right\} \tag{8a}$$

$$\boldsymbol{\phi}^k \leftarrow \underset{\boldsymbol{\phi}\in\mathbb{R}^{NT}}{\arg\min} \left\{ \frac{1}{2}\|\boldsymbol{\psi} - \mathbf{H}\boldsymbol{\phi}\|_{\ell_2}^2 + \frac{\rho}{2}\|\boldsymbol{\phi} - \tilde{\boldsymbol{\phi}}^k\|_{\ell_2}^2 \right\} \tag{8b}$$

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. Again, (8a) is computed using the proximal operator, and (8b) reduces to a linear step.

There are substantial similarities between algorithms (7) and (8). The main differences are that we eliminated the dual variable $\boldsymbol{s}$ and simplified the quadratic subproblem.

4. 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

	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
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 1. 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.

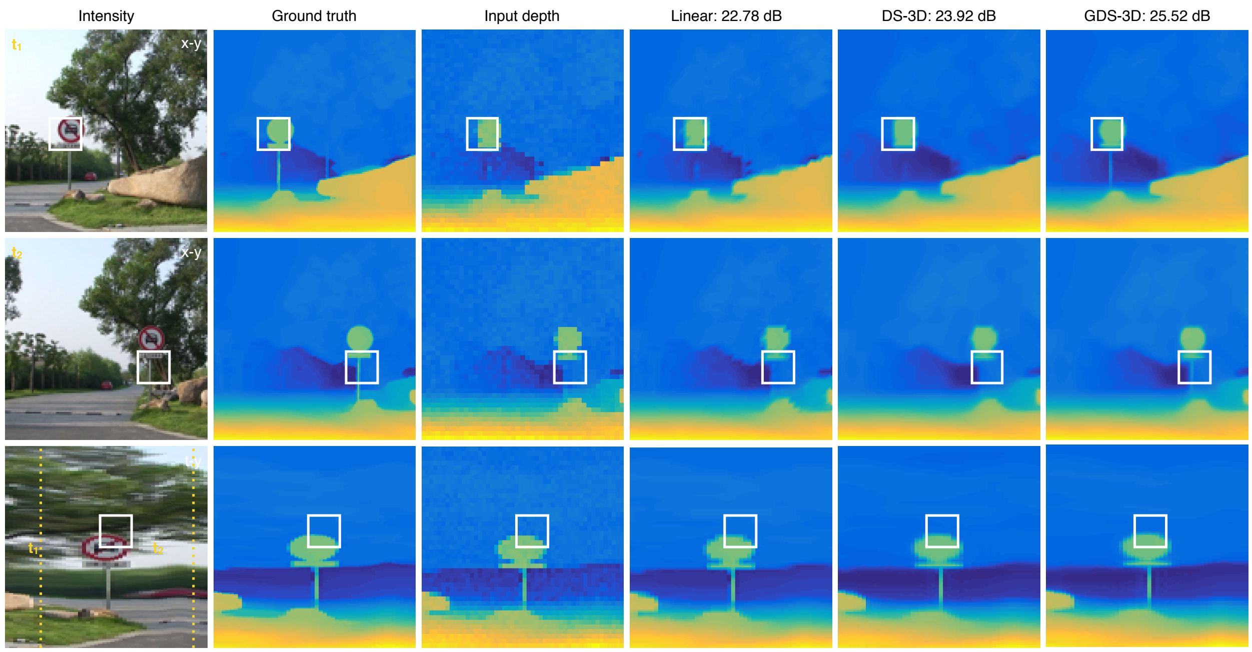

Fig. 4. Visual evaluation on *Road* video sequence. Estimation of depth from its 3× downsized version at 30 dB input SNR. Row 1 shows the data at time instance $t = 9$. Row 2 shows the data at the time instance $t = 47$. Row 3 shows the t-y profile of the data at $x = 64$. Highlights indicate some of the areas where depth estimated by GDS-3D recovers details missing in the depth estimate of DS-3D that does not use intensity information.

(denoted *ADMM-3D*) and its simplified variant (denoted *GDS-3D*) against six alternative methods.

As a reference method, we consider standard linear interpolation (*Linear*). In addition, we consider methods relying in some form of total variation (TV) regularization, one of the most widely used regularizers for depth [30]. Specifically, we consider depth interpolation using TV-regularized least squares on a frame-by-frame basis (*TV-2D*). We also consider the weighted-TV formulation proposed in [12] (*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.

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 conceptually corresponds to denoising the initial depth using our motion adaptive low-rank prior.

For faster convergence, we initialized all iterative algorithms with the solution of linear interpolation. Regularization parameters were optimized by selecting for the best signal-to-noise ratio (SNR) performance from a restricted set. Methods TV-2D, WTV-3D, GDS-2D, DS-3D, ADMM-3D, and GDS-3D were allowed $t_{\text{max}} = 100$ iterations, stopping early if the relative change of the solution in two successive iterations was lower than 10^{-4}. WTGV-2D was tuned as suggested in the code provided by the authors; in particular, we run the

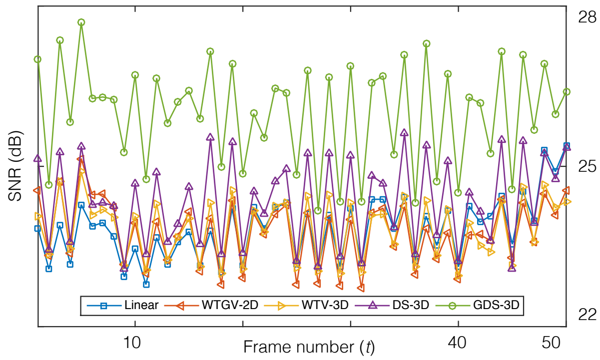

Fig. 3. Quantitative evaluation on *Road* video sequence. Estimation of depth from its $3\times$ downsized version at 30 dB input SNR. We plot the reconstruction SNR against the video frame number. The plot illustrates potential gains that can be obtained by fusing intensity and depth information in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

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.

We performed quantitative comparison using the data-set in [31], consisting 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 1 reports the SNR of superresolved depth for all the algorithms and downsizing factors. Figure 3 illustrate the evolution of SNR against the frame number for *Road*, at downsizing factor of 3. The effectiveness of our approach can also be appreciated visually in Fig. 4.

The examples and simulations results, 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 [29] that suggest to decouple data-fidelity from the enforcement of prior constraints when using block-matching-based methods.

5. 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.

6. 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] 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.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

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

[25] 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.

[26] 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.

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

[28] 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.

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

[30] 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.

[31] L. Zhang, S. Vaddadi, H. Jin, and S. K. Nayar, "Multiple view image denoising," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, June 20-25, 2009, pp. 1542–1549.