2016 Fourth International Conference on 3D Vision

Robust Recovery of Heavily Degraded Depth Measurements

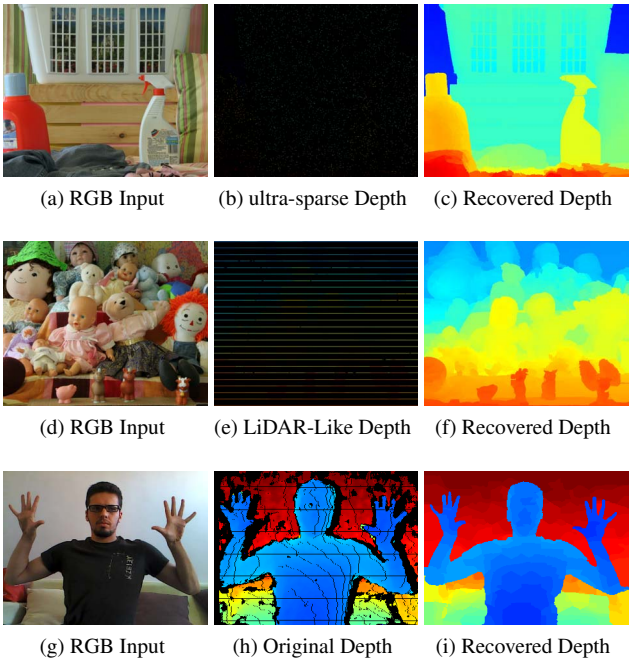
Gilad Drozdov, Yevgeny Shapiro and Guy Gilboa
Electrical Engineering Department, Technion - Israel Institute of Technology
Technion City, Haifa, Israel 32000
{sdgilad@tx, yevsh@tx, guy.gilboa@ee}.technion.ac.il

Abstract

The revolution of RGB-D sensors is advancing towards mobile platforms for robotics, autonomous vehicles and consumer hand-held devices. Strong pressures on power consumption and system price require new powerful algorithms that can robustly handle very low quality raw data. In this paper, we demonstrate the ability to reliably recover depth measurements from a variety of highly degraded depth modalities, coupled with standard RGB imagery. The method is based on a regularizer that fuses super-pixel information with the total-generalized-variation (TGV) functional. We examine our algorithm on several different degradations, including the new RealSense hand-held device by Intel, LiDAR-type data and ultra-sparse random sampling. In all modalities that are heavily degraded, our robust algorithm achieves superior performance over the state-of-the-art. Additionally, a robust error measure based on Tukeys biweight metric is suggested, which is better at ranking algorithm performance since it does not reward blurry non-physical depth results.

1. Introduction

Depth sensor technologies are rapidly improving, providing dense depth measurements at high frame rates, smaller sizes and lower costs. However, there are still significant limitations with respect to quality, accuracy, and resolution. These features vary with different technologies and modalities. In recent years, commercial companies were able to manufacture depth sensors in smaller form factors, allowing them to be integrated into hand-held devices. Most notable are prototype and research hand-held cameras, to be released soon to the market, by Intel (RealSense) and by Google (Project Tango). The requirement for extremely efficient power-consumption is a very heavy strain on hand-held devices. This inevitably leads to a severe compromise in raw depth quality and demands sophisticated and robust algorithms as compensation. A real-world example is shown in Fig. 1 (g)-(i).



(a) RGB Input (b) ultra-sparse Depth (c) Recovered Depth
(d) RGB Input (e) LiDAR-Like Depth (f) Recovered Depth
(g) RGB Input (h) Original Depth (i) Recovered Depth

Figure 1: An example of depth recovery using our algorithm, which can take as input highly degraded depth measurements and a standard registered RGB image: (a)-(c) are examples for depth sampled in an ultra-sparse manner (1% of pixels are kept); (d)-(f) show depth sampled in a LiDAR-like fashion (25-sample lines); (h)-(i) show real data captured by a mobile Intel-RealSense camera.

In recent years, technologies for autonomous vehicles are emerging as a very active field of research and development. Autonomous vehicles are equipped with several types of sensors for fast and accurate construction of the surrounding scene and avoiding obstacles and accidents. LiDAR is the depth sensor of choice for those applications. It is based on Time-of-Flight principles and uses a set of vertically aligned transmitter-receiver units that scan the environment continuously. Along the vertical axis, resolution is limited by the number of transmitters (commonly 32 or 64)

978-1-5090-5407-7/16 $31.00 © 2016 IEEE
DOI 10.1109/3DV.2016.15

and produces highly sparse data, illustrated in Fig. 1 (d)-(f).

This motivated us to propose a very general and robust algorithm for fusing color and depth. The goal is to maintain high quality performance at very extreme cases of heavily degraded data. We further explored new depth scenarios of randomly sampled measurements, which we refer to as *ultra-sparse*, in which only a small randomly spread proportion of the depth data was retained, as presented in Fig. 1 (a)-(c). We found these types of sampling as highly efficient, and advocate the development of depth camera technologies in this direction.

The contributions of this paper are as follows:

1. We propose a highly robust and flexible algorithm that can recover depth measurements well, given a registered RGB image.
   - The algorithm is based on high-order variational methods incorporating the color side information using superpixel (SP) processing.
   - Our method can handle large image regions without depth data, as well as depth outliers.
   - We demonstrate the superiority of our method over state-of-the-art, in terms of both mean-absolute-distance (MAD) and Tukey measures.
2. We discuss the problems of using MAD to account for depth errors, when evaluating depth inpainting algorithms. We suggest an alternative error measure based on the Tukey biweight metric. This measure does not reward blurred solutions, which are not physical in terms of world geometry.
3. We present new insights on preferred sampling patterns of LiDAR sensors and show the effectiveness of randomly sampled measurements.

1.1. Previous work

In recent years high quality results in spatial enhancement of depth maps were achieved by performing fusion between registered depth and intensity images. In the pioneering work by Diebel and Thrun [9], an MRF model guided by texture derivatives was used for depth-maps denoising. A robust way of using second-order statistical models was suggested in [12].

A family of algorithms that employ data smoothness terms combined with additional regularization terms were successful in propagating sharp edges into depth missing regions. We briefly summarize the most recent methods which are today's state-of-the-art.

Herrera *et al.* [13] suggested an energy consisting of a data term and a second-order derivative smoothness term that promotes flat-planes. This is weighted by color gradients using a Euclidean distance in the color domain. Park *et al.* [24, 23] have incorporated a non-local structure term to retain fine local structures, using color segmentation and edge saliency through a edge weighting scheme.

Yang *et al.* [31, 32] proposed an auto-regressive (AR) color-guided model for generic map scenes. The algorithm minimizes the pixel-wise adaptive AR predictors for predefined degradation models. The minimization is with respect to a weighted L_2-norm constraining the similarity to the original depth image and the pre-estimation AR term.

Kwon *et al.* [17] present a data-driven approach that performs sparse-dictionary learning from a combined high-quality RGB-D training set. Their method is able to capture correlation between raw depth maps and RGB images. However, this approach lacks generality due to its dependence on a specific data set. A similar approach of joint sparsity models for depth and intensity appears in [15].

In [19], Lu *et al.* use the relationship between a graph-based color-image segmentation [10] and its registered depth-map for the task of depth super-resolution. Depth values are recovered per segment by means of spatial interpolation with radial basis functions, using depth maps reconstructed from other methods in case of ultra-sparse data and assume low noise in the original data together with other assumptions on depth-image structure.

Ferstl *et al.* [11] use the total generalized variation (TGV) minimization approach and color-depth alignment to create super-resolution depth-images from low-resolution depth maps by using the intensity level gradient. This yields good results and is most closely related to our approach. It is essentially based on the formulation of ITGV proposed by Ranftl *et al.* [26] modified to accommodate the problem of super-resolution. We found this method to be somewhat sensitive to intensity levels and ignorant of noisy regions, which may contain outliers during stereo acquisition. Practically, the transition between homogeneous image areas and an edge results in the assignment of intermediate values, which results in producing augmented structures.

Other methods [8, 20, 25, 18, 33] that we evaluated use various ways of weighted smoothing criteria, e.g., in the spirit of bilateral filtering. Their performance vary with applications and the ability to handle heavily degraded cases is less pronounced.

2. Preliminaries

2.1. TGV

Regularization techniques have been extensively used to solve many image processing and computer vision tasks such as denoising, deconvolution, segmentation, registration, or optical flow. In these cases, there is often insufficient data, or the data is too noisy to apply a direct solution. Imposing regularity on the solution, where the sense of regularity depends on the data and application, greatly

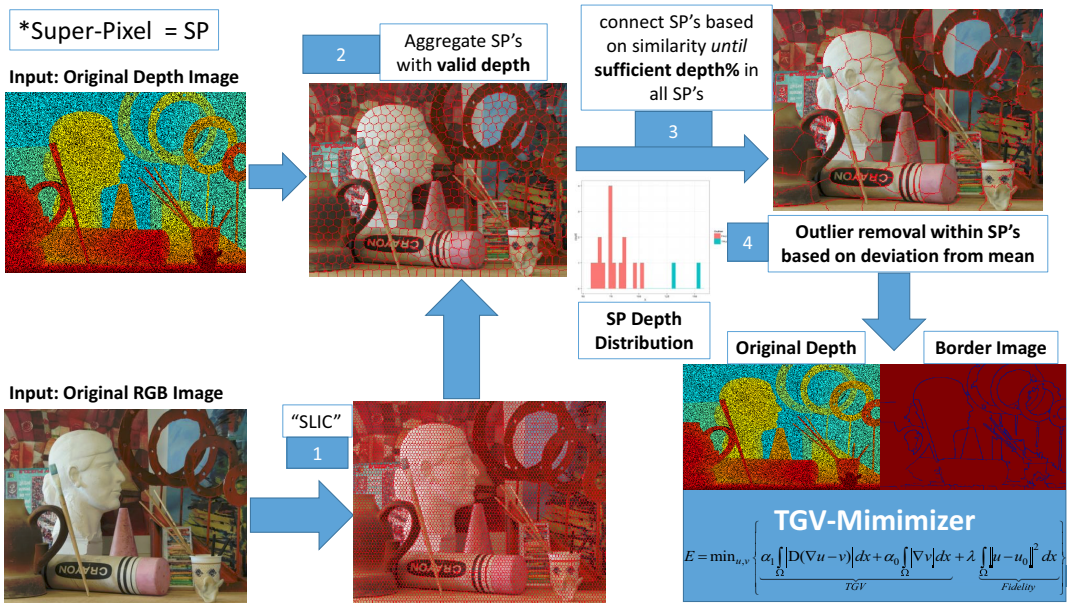


Figure 2: Algorithm concept diagram.

facilitates solving many inverse problems.

In [4], Bredies *et al.* proposed a very general regularizer, referred to as *total-generalized variation (TGV)*, that can have piecewise regularity of any derivative order. The TGV of order 2 balances in an optimal manner first- and second-order information of the image data

$$TGV_2(u) = \min_{v \in BD(\Omega)} \|\nabla u - v\|_{\mathcal{M}} + \alpha \|\mathcal{E} v\|_{\mathcal{M}}, \qquad (1)$$

where Ω is the image domain, ∇ is the distributional gradient, $\|\cdot\|_{\mathcal{M}}$ is the Radon norm [3], $BD(\Omega)$ is the space of functions of bounded deformation in Ω, $\mathcal{E}$ is a symmetric gradient acting on a vector field v, and $\alpha > 0$ is a weight, balancing the two energy terms. It can be shown that this functional is proper, convex, lower semi-continuous, rotationally-invariant, and edge preserving with no staircasing. Thus, it is adequate for image regularization and numerically powerful convex optimization techniques can be used as solvers. This type of regularization enables high quality denoising results and is used in various other image-processing algorithms like deconvolution or reconstruction of medical data [22, 5]. We present below a new algorithm based on a variant of the TGV energy for filling-in missing information in depth images.

3. Proposed algorithm

In this work we propose a super-pixel TGV algorithm that we refer to as *STGV*. We suggest a novel modification to the image-driven TGV regularization (ITGV) propsed by Ranftl *et al.* [26], where the Nagel-Enkelmann [21] operator was used to enforce a low degree of smoothness around edges. This is done by using the following anisotropic diffusion tensor,

$$D = \exp\left(-\beta \left|\nabla I_H\right|^{\gamma}\right) nn^T + n^{\perp} n^{\perp T}, \qquad (2)$$

where vector n is parallel to and $n^{\perp}$ is perpendicular to the normalized direction of the intensity-image gradient. We use $(\cdot)^T$ to denote transposition. The parameters γ and β are user-defined. The proposed method combines SP [1] and a variant of the TGV energy. In essence, an SP layer is composed from the input RGB image. An operator D is constructed such that it attains zero values at SP boundaries.

$$D = \Gamma nn^T + n^{\perp} n^{\perp T}, \qquad (3)$$

$$\Gamma = \begin{cases} 1 & \text{inside segment,} \\ 0 & \text{segment border,} \end{cases}$$

Based on this side information we obtain the following variant of the TGV energy:

$$TGV_2^D(u) = \min_{v \in BD(\Omega)} \|D(\nabla u - v)\|_{\mathcal{M}} + \alpha \|\mathcal{E} v\|_{\mathcal{M}}. \qquad (4)$$

A regularized solution u is found by minimizing the cost function

$$E(u) = TGV_2^D(u) + \lambda(x) \|u - u_0\|^2, \qquad (5)$$

where u_0 is the input depth-image and $\lambda(x) = 0$ in regions with unknown data (to be filled-in). As a numerical solver we use the Chambolle-Pock algorithm [6].

Our framework can operate in two separate regimes: either high confidence but sparse data (LiDAR-like) or dense

data containing outliers (often in stereo technologies). Regularizers, such as TGV, can also cope well with random noise. However, to simplify our setting, we disregard random depth variations in our degradation model.

Our algorithm generates a SP map that fuses iteratively RGB and depth information. The SP partition is initialized using SLIC [1] based on the RGB image only. SPs are then aggregated greedily based on combined color-depth similarity until all SPs have sufficient depth percentage γ_{DP} with respect to the initial degradation of the depth-map (e.g., 20% for RealSense completion, 10% LiDAR, and 1% for ultra-sparse). We define SP-similarity as the absolute distance between four channels: three color channels (RGB) and a fourth depth channel, i.e.,

$$\text{Sim}_{RGBD} = |SP_i^{RGBD} - SP_j^{RGBD}|. \tag{6}$$

The next step is outlier removal. We assume a Gaussian distribution model within a single SP, for which we calculate the standard deviation σ. Pixels that deviate by more than σ from the mean are deemed outliers and are replaced by the algorithm's estimate of the refined depth values.

We are now able to calculate D and perform the minimization of (5). Fig. 2 presents a high-level block-diagram of our framework. An outline of our STGV Algorithm 1 is described below.

Algorithm 1: STGV.

Initialize: with RGBD based SLIC [1];
repeat for each SP
  Compute depth-statistics ($SP_i \sim \mathcal{N}(m_i, \sigma_i^2)$)
  If (low-confidence depth)
    Invalidate pixels with depth deviation $> \sigma$
  Merge SP: use DBSCAN [30] by evaluating (6)
Until sufficient valid depth % per-SP is achieved
Generate border-image based on SP segmentation
Perform TGV minimization based on (5)

4. Robust error metric for depth inpainting

Despite the extensive work done in the depth-image enhancement field, in particular depth inpainting, reliable error measurements have drawn little attention. Evaluations tend to rely mostly on PSNR and MSE measures, and recently on the mean-absolute-difference (MAD). A common misconception is that MAD does not favor blurry results since it is based on the L^1 norm. In Fig. 4 we demonstrate that deviating from the ground truth in a blurry manner is rewarded by MAD. Thus, methods that recover depth by creating new objects with floating unattached point clouds of no physical meaning are preferred to algorithms that slightly mislocate the ending of the object or background,

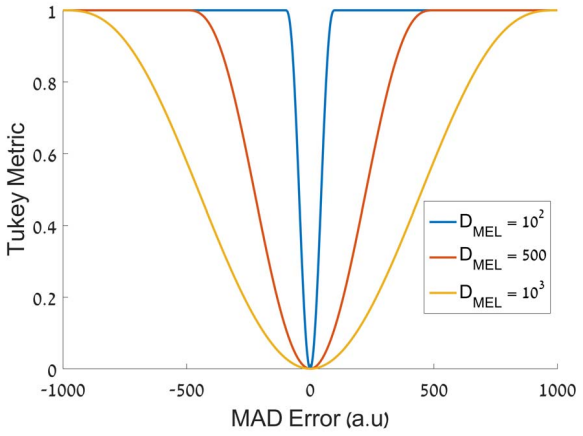


Figure 3: Tukey Metric function for 3 range scales. The dynamic error-range is controlled through the D_{MEL} parameter, which can be determined by camera noise-level.

but essentially retain the same *topology* of the scene. Therefore, we seek an error measure that does not reward algorithms that blur depth.

We propose a measure that does not penalize topology preserving algorithms, using Tukey's biweight estimation metric [2]. In Fig. 4 we illustrate the motivation behind the proposed metric:

$$\text{TukeyMetric} = \begin{cases} 1 & \text{MAD} > D_{\text{MEL}} \\ 1 - \left(1 - \left(\frac{\text{MAD}}{D_{\text{MEL}}}\right)^2\right)^3 & \text{otherwise,} \end{cases} \tag{7}$$

where D_{MEL} is a depth-range dependent parameter. As shown in Fig. 3, increasing values of D_{MEL} allow for a larger error-range, until a constant penalty is assigned.

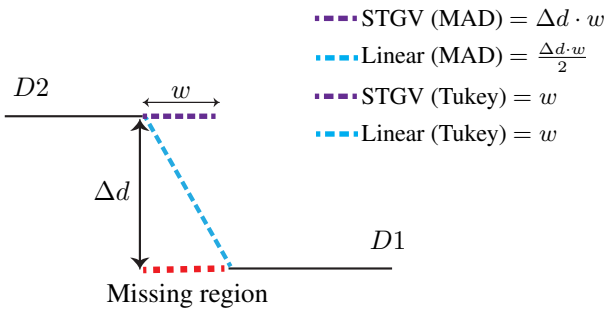


Figure 4: The scene contains depth values D_1, D_2 and a corrupted region (Red), originally corresponding to D_1. Comparing two algorithms for recreating the missing data: a bilinear interpolation (Blue) and a topology preserving method (Purple). MAD favors a linear (blurry) solution; under Tukey's metric identical performance is achieved.

5. Experimental Results

For comparison of our method, we use the widely accepted Middlebury-2005 dataset [14], comprised of 6 calibrated RGB-D images on which we perform quantitative evaluation and compare with state-of-the-art inpainting algorithms. we also perform qualitative evaluation over real-world data, captured by Intel's RealSense active-stereo camera-R200 [29]. We propose three degradation regimes,

1. LiDAR-like, where the depth-map is acquired by horizontal sampling bands, as LiDAR-RGB calibration is technically challenging.

2. Ultra-sparse acquisition, where depth-data is sampled in a random fashion.

3. 3D-imagery by Intel RealSense, which also contains falsely registered depth values – outliers.

We consider two quality criteria, the MAD error in cm and the Tukey error (7). The parameters for each evaluated method were selected to maximize performance in terms of MAD error. The focus of our work was not computational performance; we relied on non-optimized available code. For the numerical simulations we incorporated the basic SLIC [1] algorithm for color-image segmentation and the DBSCAN-clustering method [30]. For solving the proposed optimization problem (5) we employed the primal-dual energy minimization scheme, proposed in [7] and use the formulation and numerical implementation [11].

All experiments were carried out using a standard PC with Intel Core i7-4770 CPU with 3.4GHz and 16GB RAM.

Runtime varies between the degradation regimes (for 555x695 images): (a) non ultra-sparse: mean runtime 14 min. (b) ultra-sparse: mean runtime 20 min. Moreover, there exists a GPU-optimized implementation for SLIC and primal-dual energy minimization, which can dramatically reduce runtime to the order of seconds or less (e.g. gSLIC [28], Pock *et al.*, SSVM 2013 [27]).

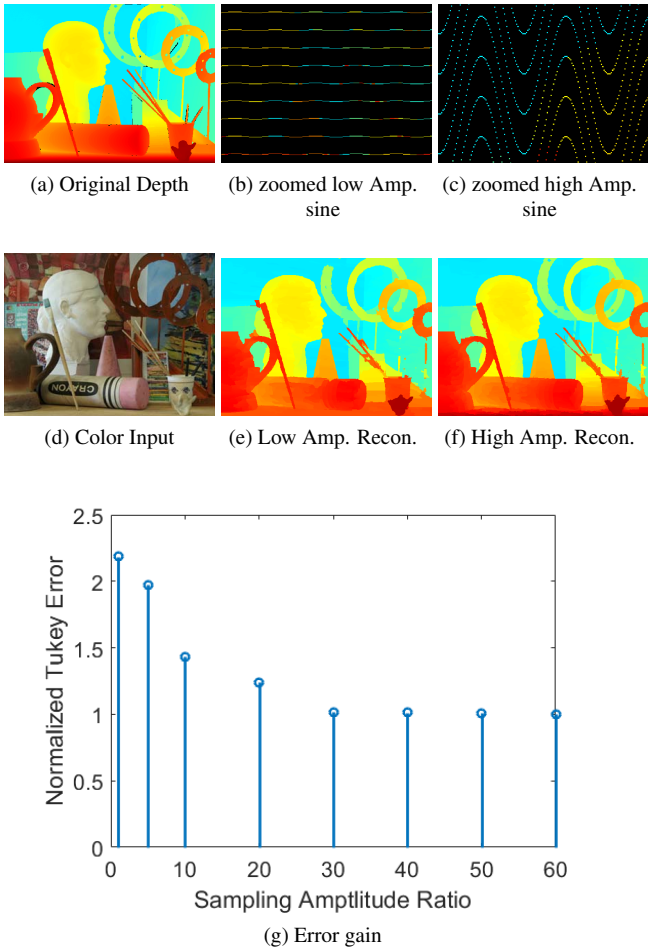
For the LiDAR-like scenario we examine 3 modes of operation using 16-25 sampling-bands; the inpainting results are in Table 1 and Fig. 6. The inpainting results for the ultra-sparse test case are presented in Table 2 and Fig. 7.

For a qualitative evaluation, we compare our method to the best performing techniques considered in this paper, presented in Figs. 10-11. Clearly, our method outperforms with respect to original shape-preservation and sharpness.

Finally, the evaluation for Intel-RealSense is shown in Figs. 8-9, which demonstrates the superiority of our method, in both edge preservation and outliers correction.

5.1. Optimal Sampling Lattice

Following the above experiments we advocate for sensors that can obtain random sparse depth-data acquisition.



(a) Original Depth (b) zoomed low Amp. sine (c) zoomed high Amp. sine

(d) Color Input (e) Low Amp. Recon. (f) High Amp. Recon.

(g) Error gain

Figure 5: Visual comparison between two depth-sampling modes are compatible with the error analysis, achieving roughly 2.5-fold decrease in error between the two sampling regimes with the same percentage of missing data.

To enhance LiDAR technology using standard mechanisms, one can periodically move the LiDAR acquisition system vertically, creating sine-wave like acquisition patterns. We simulated these cases for various sine amplitudes to see the effect on reconstruction performance. This test case shows that the distribution of depth-measurements coupled with the intensity information can significantly improve results under an oscillating sampling regime. By using this capturing procedure, one can obtain an improved depth image quality, while maintining the same number of LiDAR-beams, cost, and power-consumption; all that is needed is the addition of a minor motor. Fig. 5 shows the performance and reduction in error as the amplitude grows. This procedure reduces the sparsity ratio between the x- and y-axes, and yields more homogeneous sampling of the space.

6. Conclusion

In this paper we address the problem of recovering heavily degraded depth measurements for RGBD-type sensors. This problem is becoming extremely relevant as mobile sensors for hand-held devices and autonomous vehicles are developed. In these cases there is strong pressure to reduce size and power, at the expense of quality. Our algorithm is based on using color for guiding TGV, a convex second-order regularizing functional. This is by using SP information that gives reliable sharp segmentation queues between objects and avoids blurring of the depth results when filling-in missing data. We show that the algorithm is highly robust, flexible, and performs well in challanging scenarios.

In addition, we show that in order to evaluate the algorithms, the use of MAD is suboptimal and may lead to erroneous performance assumptions, as introduction of mid-range values is not penalized and might be rewarded. We suggest an alternative error measure, based on the Tukey biweight metric. This measure does not reward blurred solutions, which are not physical in terms of world geometry.

Finally, new insights on preferred sampling patterns of LiDAR sensors are examined. for a fixed percentage of the depth pixels measured, it is shown that random samples are more efficient than straight-line scan patterns, as in LiDAR technology. We show through simulations that scans with a sine pattern can reduce the error by a factor of two. This motivates research for depth-sensor technology that can acquire pseudo-random sampled measurements, which follows a reconstruction algorithm.

ACKNOWLEDGEMENT

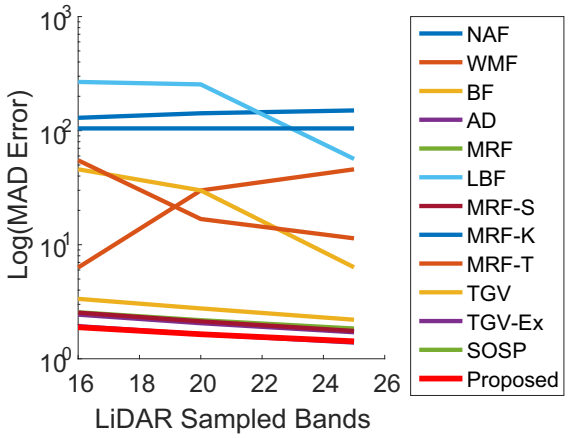
We acknowledge support of the Magnet program of the OCS, Israel Ministry of Economy, in the framework of Omek Consortium.

Method \ Image	20		16	
	MAD	Tukey	MAD	Tukey
NAF[8]	150.345	0.843	129.565	0.737
WMF[20]	45.811	0.526	6.307	0.216
BF[25]	6.338	0.220	45.827	0.528
AD[18]	1.708	0.272	2.436	0.393
MRF[9]	1.841	0.309	2.557	0.424
LBF[33]	56.537	1.000	267.328	0.999
MRF-S[12]	1.752	0.280	2.536	0.403
MRF-K[12]	104.547	0.995	104.604	0.995
MRF-T[12]	11.370	0.274	54.953	0.548
SOSP[13]	-	-	-	-
TGV[4]	2.199	0.282	3.352	0.421
TGV-Ex[11]	**1.387**	0.195	1.931	0.291
Proposed	1.422	**0.027**	**1.640**	**0.048**

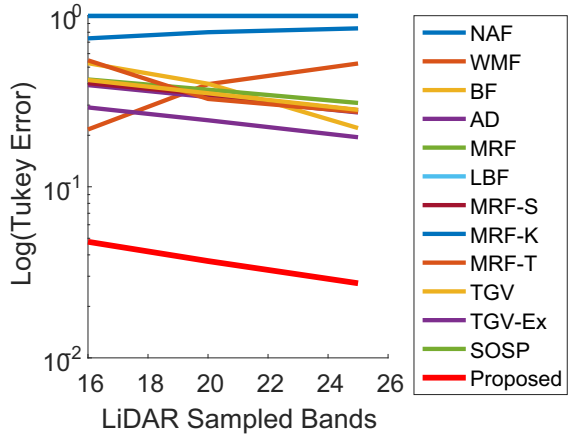
Table 1: Quantitative Inpainting Results (in MAD (cm) and Tukey) for LiDAR sampling. with 20 and 16 horizontal sampled bands. The best results are highlighted in bold.

Method \ Image	99%		99.5%	
	MAD	Tukey	MAD	Tukey
JBF[16]	113.524	0.663	143.517	0.812
NAF[8]	113.524	0.663	143.517	0.812
WMF[20]	3.479	0.182	14.532	0.306
BF[25]	3.480	0.180	14.529	0.304
AD[18]	1.784	0.308	2.296	0.410
MRF[9]	2.260	0.421	2.872	0.526
LBF[33]	63.031	1.000	128.687	1.000
MRF-S[12]	1.804	0.309	2.334	0.414
MRF-K[12]	104.566	0.995	104.567	0.995
MRF-T[12]	51.597	0.492	34.650	0.428
SOSP[13]	-	-	-	-
TGV[4]	3.630	0.488	5.785	0.673
TGV-Ex[11]	1.623	0.265	2.423	0.393
Proposed	**1.380**	**0.026**	**2.001**	**0.051**

Table 2: Quantitative Inpainting Results (in MAD (cm) and Tukey) for Ultra-Sparse sampling, with 1% and 0.5% of valid data. The best results are highlighted in bold.

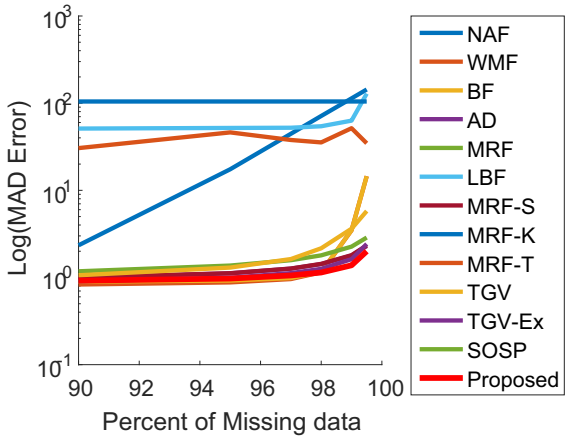


(a) Average MAD LiDAR Error (cm)

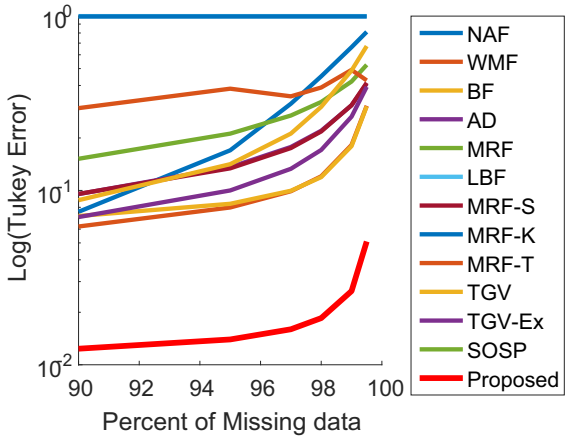


(b) Average Tukey LiDAR Error

Figure 6: (a) LiDAR Average MAD, (b) Tukey errors for a range of sampling bands.

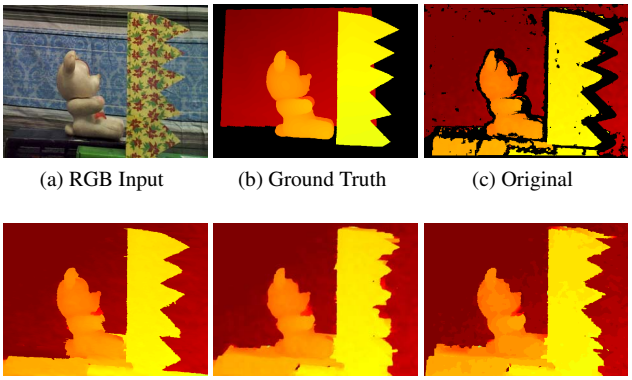


(a) Average MAD Sparse Error (cm)



(b) Average Tukey Sparse Error

Figure 7: (a) Average MAD (cm),(b) Tukey error for sparse sampling.



(a) RGB Input (b) Ground Truth (c) Original

(d) Proposed (e) TGV-Ex (f) SOSP

Figure 8: Qualitative results for 3 state-of-the-art methods evaluating the quality of reconstruction for depth-measurements captured by Intel-RealSense camera, with complementary ground-truth data from IR-scanner.

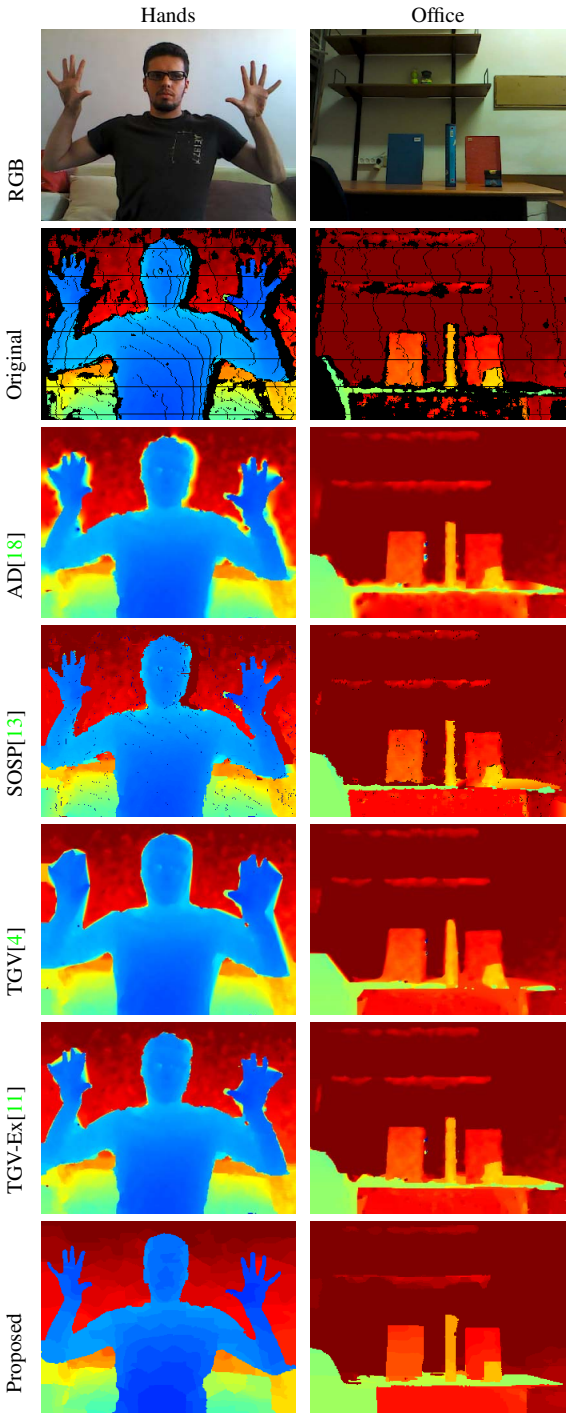


Figure 9: Qualitative results evaluating the quality of reconstruction for depth-measurements captured by Intel-RealSense camera.

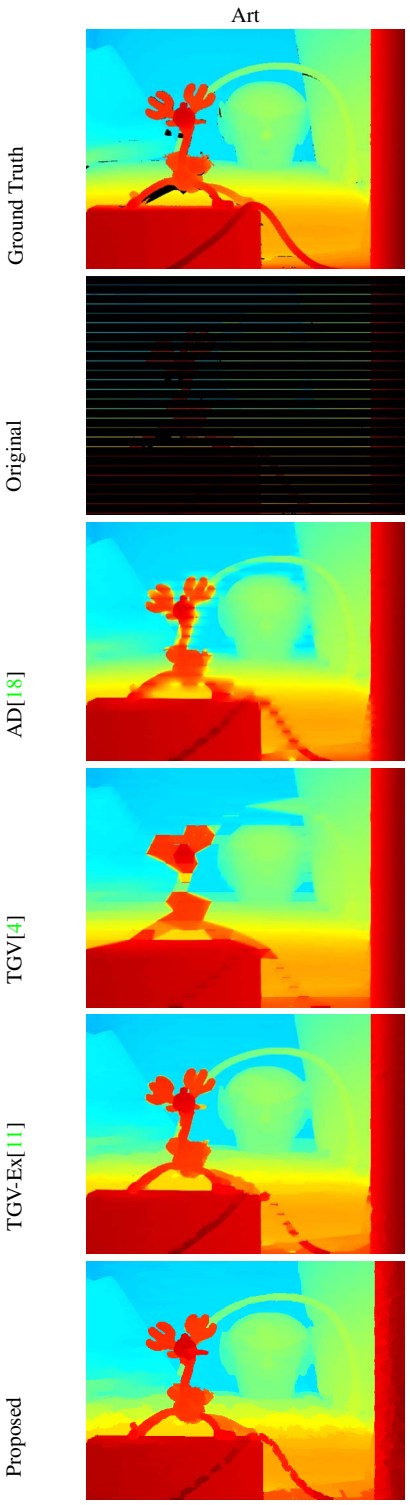


Figure 10: Qualitative results for *Reindeer* for LIDAR-like sampling with 25 sample-lines.

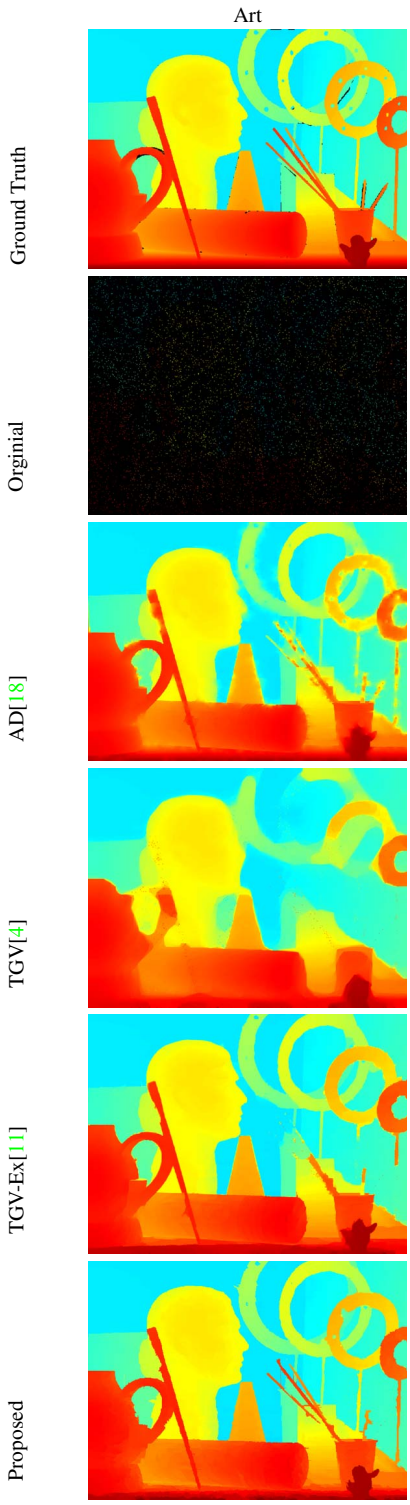


Figure 11: Qualitative results for *Art* ultra-sparse sampling with 2% of valid data.

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.

[2] A. E. Beaton and J. W. Tukey. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16(2):147–185, 1974.

[3] K. Bredies and M. Holler. Regularization of linear inverse problems with total generalized variation. *Journal of Inverse and Ill-posed Problems*, 22(6):871–913, 2014.

[4] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM J. Imaging Sciences*, 3(3):492–526, 2010.

[5] K. Bredies and T. Valkonen. Inverse problems with second-order total generalized variation constraints. *Proc. SampTA*, pages 1–4, 2011.

[6] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[7] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[8] D. Chan, H. Buisman, C. Theobalt, and S. Thrun. A noise-aware filter for real-time depth upsampling. In *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications-M2SFA2 2008*, 2008.

[9] J. Diebel and S. Thrun. An application of markov random fields to range sensing. In *NIPS*, volume 5, pages 291–298, 2005.

[10] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2):167–181, Sept. 2004.

[11] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof. Image guided depth upsampling using anisotropic total generalized variation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 993–1000, 2013.

[12] A. Harrison and P. Newman. Image and sparse laser fusion for dense scene reconstruction. In *Field and Service Robotics*, pages 219–228. Springer, 2010.

[13] D. Herrera C., J. Kannala, L. Ladicky, and J. Heikkilä. Depth map inpainting under a second-order smoothness prior. In *Image Analysis, 18th Scandinavian Conference, SCIA 2013, Espoo, Finland, June 17-20, 2013. Proceedings*, pages 555–566, 2013.

[14] H. Hirschmüller and D. Scharstein. Evaluation of cost functions for stereo matching. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[15] M. Kiechle, S. Hawe, and M. Kleinsteuber. A joint intensity and depth co-sparse analysis model for depth map super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1545–1552, 2013.

[16] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. *ACM Transactions on Graphics (TOG)*, 26(3):96, 2007.

[17] H. Kwon, Y.-W. Tai, and S. Lin. Data-driven depth map refinement via multi-scale sparse representation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 159–167, June 2015.

[18] J. Liu and X. Gong. Guided depth enhancement via anisotropic diffusion. In *Advances in Multimedia Information Processing–PCM 2013*, pages 408–417. Springer, 2013.

[19] J. Lu and D. Forsyth. Sparse depth super resolution. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2245–2253, June 2015.

[20] D. Min, J. Lu, and M. N. Do. Depth video enhancement based on weighted mode filtering. *Image Processing, IEEE Transactions on*, 21(3):1176–1190, 2012.

[21] H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(5):565–593, Sept. 1986.

[22] K. Papafitsoros and C.-B. Schönlieb. A combined first and second order variational approach for image reconstruction. *Journal of mathematical imaging and vision*, 48(2):308–338, 2014.

[23] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon. High quality depth map upsampling for 3d-tof cameras. In *2011 International Conference on Computer Vision*, pages 1623–1630. IEEE, 2011.

[24] J. Park, H. Kim, Y. W. Tai, M. S. Brown, and I. S. Kweon. High-quality depth map upsampling and completion for rgb-d cameras. *IEEE Transactions on Image Processing*, 23(12):5559–5572, Dec 2014.

[25] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, pages 664–672, New York, NY, USA, 2004. ACM.

[26] R. Ranftl, S. Gehrig, T. Pock, and H. Bischof. Pushing the limits of stereo using variational stereo estimation. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 401–407. IEEE, 2012.

[27] R. Ranftl, T. Pock, and H. Bischof. Minimizing tgv-based variational models with non-convex data terms. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 282–293. Springer, 2013.

[28] C. Y. Ren and I. Reid. gslic: a real-time implementation of slic superpixel segmentation. *University of Oxford, Department of Engineering, Technical Report*, 2011.

[29] D. Rotman and G. Gilboa. A depth restoration occlusionless temporal dataset. In *International Conference on 3D Vision (3DV)*. IEEE, 2016 in press.

[30] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data Min. Knowl. Discov.*, 2(2):169–194, June 1998.

[31] J. Yang, X. Ye, K. Li, and C. Hou. Depth recovery using an adaptive color-guided auto-regressive model. In *European Conference on Computer Vision*, pages 158–171. Springer, 2012.

[32] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang. Color-guided depth recovery from rgb-d data using an adaptive autoregressive model. *Image Processing, IEEE Transactions on*, 23(8):3443–3458, 2014.

[33] Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-depth super resolution for range images. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.