Super-Pixel Sampler - a Data-driven Approach for Depth Sampling and Reconstruction

Adam Wolff, Shachar Praisler, Ilya Tcenov and Guy Gilboa, *Senior Member, IEEE*

***Abstract*— Depth acquisition, based on active illumination, is essential for autonomous and robotic navigation. LiDARs (Light Detection And Ranging) with mechanical, fixed, sampling templates are commonly used in today's autonomous vehicles. An emerging technology, based on solid-state depth sensors, with no mechanical parts, allows fast and adaptive scans.**

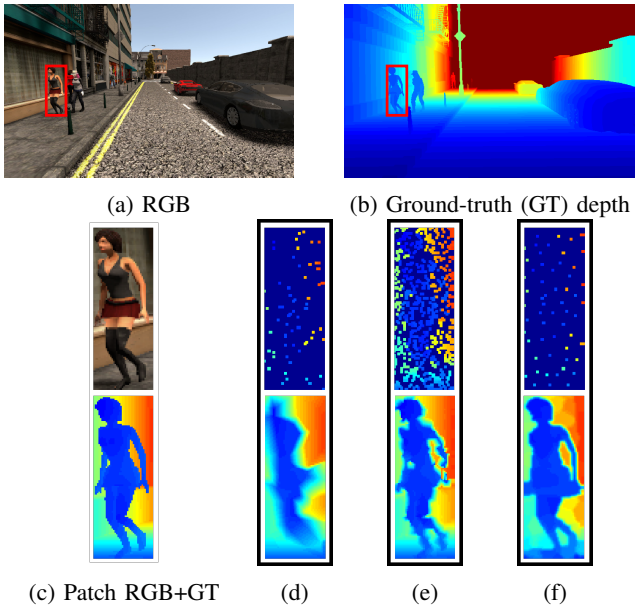
In this paper, we propose an adaptive, image-driven, fast, sampling and reconstruction strategy. First, we formulate a piece-wise planar depth model and estimate its validity for indoor and outdoor scenes. Our model and experiments predict that, in the optimal case, adaptive sampling strategies with about 20-60 piece-wise planar structures can approximate well a depth map. This translates to requiring a single depth sample for every 1200 RGB samples, providing strong motivation to investigate an adaptive framework. We propose a simple, generic, sampling and reconstruction algorithm, based on super-pixels. Our sampling improves grid and random sampling, consistently, for a wide variety of reconstruction methods. We propose an extremely simple and fast reconstruction for our sampler. It achieves state-of-the-art results, compared to complex image-guided depth completion algorithms, reducing the required sampling rate by a factor of 3-4. A single-pixel depth camera built in our lab illustrates the concept.

I. Introduction

In recent years, depth sensing has become essential for autonomous cars navigation and collision prevention [1]. The physical constraints on active depth sensing mobile devices, such as light detection and ranging (LiDAR), yield sparse depth measurements per scan. This results in a coarse point cloud and requires an additional estimation of missing data.

Traditional LiDARs have a restricted scanning mechanism. Those devices measure distance in specified angle intervals, using a fixed number of horizontal scan-lines (usually 16 to 64), depending on the number of transceivers. A new revolutionary technology is now emerging of solid-state depth sensors. They are based on optical phased-arrays with no mechanical parts, and can thus scan the scene fast in an adaptive manner (programmable scanning) [2], [3]. In addition, those innovative devices are much cheaper than those currently in use. This calls for the development of new, efficient, sampling strategies, which reduce the reconstruction error per sample. Since almost always autonomous platforms are equipped with RGB cameras, we investigate the possibility to improve the depth sampling process by taking the RGB information into account.

A. Wolff, S. Praisler, I. Tcenov and G. Gilboa are with the Department of Electrical Engineering, Technion - Israel Institue of Technology, Haifa, Israel, e-mail: {sadamwol,spraizler,ilya.tcenov}@campus.technion.ac.il, guy.gilboa@ee.technion.ac.il. We thank Yossi Bar-Erez for his work on the sampling device.



(a) RGB (b) Ground-truth (GT) depth (c) Patch RGB+GT (d) (e) (f)

Fig. 1: Adaptive (image-driven) vs. fixed (blind) sampling. It is shown how adaptive sampling and its respective reconstruction, developed by us, can dramatically improve fixed sampling. We can recover small and thin objects and reach higher resolution, with a given sampling budget. (d) Linear reconstruction based on 66 blind samples (random stencil), RMSE=0.52m. (f) Our method with 57 samples, RMSE=0.31m. (e) Fixed sampling, same RMSE as our method, requires 669 samples.

In this paper, we address the topic of image-guided depth sampling and reconstruction. First, we introduce the concept of adaptive depth sampling and develop an appropriate model of the data. Then, we introduce a fast and practical image-guided algorithm for depth sampling and reconstruction, based on super-pixels (SPs). A result example of our algorithm which was taken from the Synthia dataset [27], currently the only available dense RGB-D dataset for outdoors scenes, is shown in Fig. 1. We demonstrate in experiments that our framework outperforms state-of-the-art depth completion methods for both indoor and outdoor scenes. Finally, since current solid-state technology is not yet technically-open for reconfiguration of the sampling, we illustrate the concept in real life, by a single-pixel depth camera. The device consists of a 3D-printed motorized mount that was made in our lab (Fig. 13), a laser rangefinder, and a camera.

II. Related Work

Depth completion. The task of depth reconstruction from scattered sparse samples is being increasingly investigated. The main methods can be divided to those which require only the sparse depth input (unguided) and to those assisted by additional information, e.g. color image (guided).

Among the *unguided methods*, some use classical approach [4], [5]. For example, [6] exploited prior knowledge on indoor structures such as the second order depth derivative sparsity. Recent methods rely on deep neural nets [7], [8].

Guided methods exploit the connection between depth maps and their corresponding color image. Earlier methods used traditional image processing tools [9], [10]. Recently, several deep learning-based methods [11], [12], [13], [14], [15], [16], [17] achieved state-of-the-art results. [18] suggested a self-supervised training framework, relying on depth prediction from a monocular color image and model-based pose estimation.

Early guided depth sampling. Despite the intensive development in depth completion, the issue of adaptive sampling is yet little addressed. Only [19] and [20] have offered a non-trivial (i.e. uniformly random or grid) sampling pattern as a previous step to depth reconstruction. Both studies selected sampling at locations which are most probable to have strong depth gradient. Nonetheless, they failed dealing with very low ($< 5\%$ of GT pixels) sampling budget.

Nonuniform sampling. Over the years, the field of nonuniform sampling has been well established [21], [22], [23], [24]. However, these studies focus on the signal reconstruction for a given nonuniform sampling pattern and not on how to design data-driven patterns, based on side information.

III. The Space of Depth Images

As a motivation, we first examine the possible benefits of adaptive sampling. We would like to answer the following questions:

1) What is a reasonable functional space for depth images intended for navigation?
2) What is the dimensionality of this space? This induces a lower bound on the required number of samples.
3) How RGB side information is expected to improve the depth sampling?

Any sampling strategy is based on a model of the signal to be sampled. We would like to examine an appropriate model for depth scenes, as well as the relation to the RGB data of the same scene. We propose a simple depth model and try to validate it experimentally on benchmark data. We then relate it to RGB.

A. Piece-wise planar depth model

Our primary objective is to obtain depth information for autonomous navigation. Thus, an appropriate model should represent well the general geometrical setting (roads, walls, sidewalks) as well as the location of significant landmarks and obstacles (poles, signs, rocks and objects in a room).

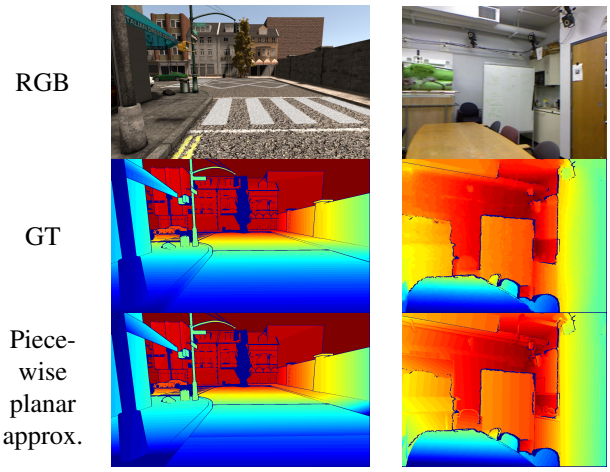


Fig. 2: Examples for piece-wise planar approximation from Synthia (left) and NYU-Depth-v2 (right) datasets.

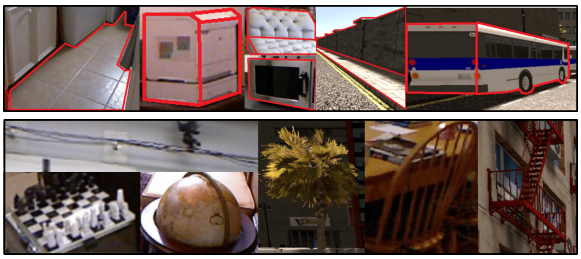


Fig. 3: Examples of approximately flat objects which fit (top) and non flat/convex objects which do not fit (bottom) the piece-wise planar depth model.

For objects, we would like to obtain their location but not necessarily their precise geometry.

This leads to a piece-wise planar model, which is a common assumption in depth estimation (see e.g. [25], [26]). However, we have not found in the literature a more precise formulation, along with an estimation of the parameters of that model (such as number of piece-wise regions and the expected error). We therefore briefly define and test our parametric model.

Given a depth image d our hypothesis is that most of the scene can be well represented by a piece-wise planar approximation. Formally, let $\Omega \subset \mathbb{R}^2$ be the image domain, where $|\Omega|$ is its area. Let $E = \{\Omega_i\}$, $i = 1,..,N$, be a set of N sub-domains which define a partition of Ω. Thus, $\Omega_i \subset \Omega$, $\Omega_i \cap \Omega_j = \emptyset$, $\forall i \neq j$, $\bigcup_{i=1..N} \Omega_i = \Omega$. Let $\hat{d}$ be a 2D piece-wise linear function, defined on Ω by

$$\hat{d}(x,y) = a_i x + b_i y + c_i, \quad (x,y) \in \Omega_i, \; i = 1..N, \qquad (1)$$

where a_i, b_i, c_i are some constants. Let v be a binary function in Ω which indicates validity of the model, where 1 indicates validity and 0 invalidity. We denote by V the set of valid points, $V = \{(x,y) \,|\, v(x,y) = 1\}$. We assume $|V| \geq (1-\delta)|\Omega|$, $0 \leq \delta \ll 1$. Our hypothesis is that given some small tolerance parameters δ, ϵ, a validity map v and a small number of regions N, for any depth map d there exists a

piece-wise planar approximation, defined by (1), such that

$$RMSE_v(d, \hat{d}) \equiv \sqrt{\frac{1}{|V|}\|v \cdot (d - \hat{d})\|_2^2} \leq \epsilon. \quad (2)$$

Thus, d can be well approximated by a 2D piece-wise linear function, almost everywhere, provided we know the partition set E and the plane parameters a_i, b_i, c_i for each Ω_i. Our aim is to approximate E from the RGB image. In order to recover $\hat{d}$ we need to sample each region Ω_i 3 times to estimate its coefficients (noiseless case). This gives us a lower bound on the number of samples required to obtain a depth image:

$$n_{min} \equiv Number - of - Samples \geq 3N. \quad (3)$$

We now turn to experimentally check this hypothesis and examine the values of δ, ϵ and N in indoor and outdoor scenes. To validate the proposed model, we made a piece-wise planar depth approximation for two datasets, Synthia [27] and NYU-Depth-v2 [28]. Altogether 1441 images were used. More details on the experiment can be found in [29].

Examples of piece-wise planar approximations are shown in Fig. 2 (bottom), compared to the ground truth (middle row). Dark-blue indicates regions not in the set V. It can be observed that the approximation is quite accurate. The average model parameters recovered for Synthia are $N = 66.6$, $\delta = 0.1$, $\epsilon = 1.35m$, and for NYU-Depth-v2 are $N = 18.5$, $\delta = 0.07$, $\epsilon = 0.18m$.

In these cases, according to Eq. (3), Synthia can be well approximated in an optimal scenario by an average of only 200 samples, whereas NYU-v2 by an average of 56 samples (in both cases, this translates to about 0.08% sampling ratio, compared to the ground-truth depth resolution). In Fig. 3 we show examples of objects which fit well a piece-wise planar approximation (top) and counter examples of highly non-convex structures or ones with high curvature.

B. Relation of RGB and depth

Next, we want to examine the possibility to estimate the partition set E from the RGB data. This is a very challenging task, which is an open problem at this point. We thus turn to a simpler problem of checking the relation between RGB edges and depth discontinuities. Given the set of RGB boundaries (edges) $B_{rgb} \subset \Omega$ and depth boundaries $B_d \subset \Omega$ we would like to calculate empirically, for each coordinate $\mathbf{x} = (x, y) \in \Omega$, the following conditional probabilities:

$$P_{rgb-d} := Prob(\mathbf{x} \in B_{rgb} \,|\, \mathbf{x} \in B_d), \quad (4)$$

$$P_{d-rgb} := Prob(\mathbf{x} \in B_d \,|\, \mathbf{x} \in B_{rgb}). \quad (5)$$

We estimated B_{rgb} by [30] and B_d by a normalized depth gradient threshold. More details can be found in [29].

For Synthia we obtain $P_{rgb-d} = 69.3\%$, $P_{d-rgb} = 33.7\%$, and for NYU-v2 $P_{rgb-d} = 80.8\%$, $P_{d-rgb} = 22.9\%$. The high values of P_{rgb-d} indicate the ability to predict well depth discontinuities, based on RGB edges. The relatively low values of P_{d-rgb} indicate that we should expect many false partitions (which appear only in the RGB, but not in the depth data). Thus we can expect to be able to approximate, to some extent, the partition set E based solely on the RGB image, by over-segmentation. As the partition is quite a rough approximation, additional samples are required, above the lower bound expressed in Eq. (3). In Fig. 4 we show examples of boundaries in the RGB and depth for both sets.

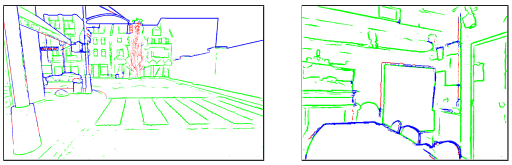


Fig. 4: Examples of depth discontinuities (red) and RGB edges (green) correlation. Blue pixels include both types.

IV. Method

We propose a generic and simple method for depth sparse sampling and reconstruction, based on the assumptions:

1) Measurements are of high quality, such that noise of the range measurement is negligible, compared to the global error of the dense reconstruction.
2) Sampling budget is limited to n samples.
3) An RGB image of the scene is available to guide the process. The reconstructed depth is registered to this image. Sensitive cameras may be used for night scenes.
4) Sampling is point-wise. The system can sample at any desired location of the RGB image. The sampling pattern can change for each image.

A. Algorithm design requirements

Several requirements are vital to the design of such an algorithm: It needs to capture well the shape and boundaries of objects, to be computationally fast and memory efficient, and to have control on the number of samples. Surprisingly, all these requirements coincide with those for super-pixels, e.g. Slic [31], an over-segmentation technique applied to RGB images. This led us to the following algorithm.

B. Proposed algorithm

The proposed algorithm is divided into two parts, sampling and reconstruction. It includes the following steps:

- **Sampling:**

S.1 A super-pixel map is generated from the RGB image using SLIC [31]. The desired number of SPs is set to n. The SPs compactness is adjusted to a high value to ensure regularly shaped SPs.

S.2 For each SP, the SP center of mass (CoM) is computed by calculating the mean of the (x, y) coordinates of all pixels in the SP. A depth sample is taken at the CoM location. If the CoM is located outside of the SP (for some non-convex SP), the depth sample is taken at the closest location to the CoM of the SP.

- **Reconstruction:** Our reconstruction is based on the samples and SPs of the sampling stage.

R.1 For each SP, a single depth measurement is available, thus a zero-order estimation is performed. That is, the

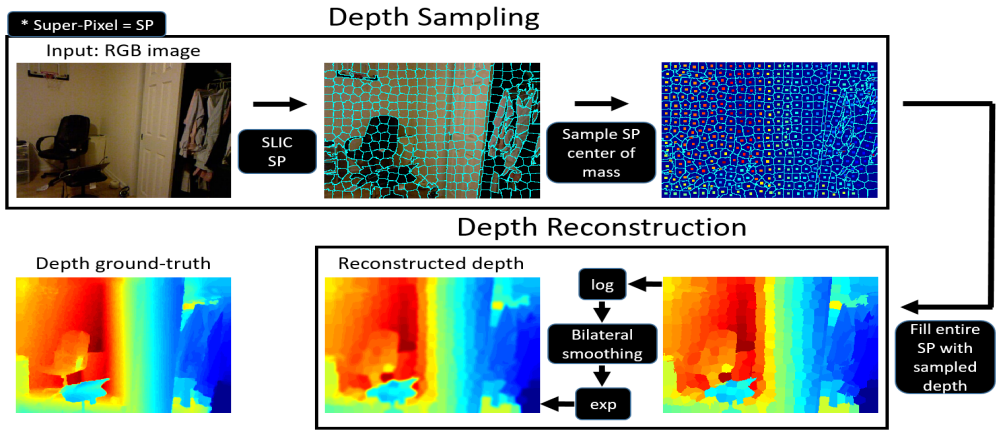


Fig. 5: Algorithm block diagram.

entire SP takes the depth value of the sample. Let d_0 be the resulting depth image.

R.2 d_{log} is calculated by $d_{log} = \log(d_0 + 1)$.

R.3 A bilateral filter [32] is applied over d_{log}. The filter's parameters are fixed for a given number of samples n and type of scene (road / room). Let d_{BF} be the bilateral filter result.

R.4 The final dense reconstructed depth image, d_r, is calculated by $d_r = \exp(d_{BF}) - 1$.

See Fig. 5 for a high-level diagram of our framework.

C. Principles of the algorithm

1) Sampling:

Sampling based on RGB segmentation. We attempt to recover the partition E through the RGB, following our model in the previous section.

Sampling at center of mass (CoM) of segments. Contrary to previous studies [19], [20], we aim to sample *as far as possible* from the RGB edge. There are two main reasons for this choice: First, for piece-wise planar depth regions, the CoM minimizes the RMSE. Secondly, it reduces inherent uncertainties near depth discontinuities, where measurements are inaccurate, in practice. Fig. 6 demonstrates that sampling at SP CoM leads to a more accurate depth reconstruction results than sampling a random pixel location inside the SP.

Why super-pixels. Sampling with super-pixels enables measuring small elements. It also limits reconstruction error since the size of the segment is limited. When the depth discontinuity is not well reflected in the RGB (we term it *camouflaged objects*) the sampling reduces to an approximate grid-sampling scheme, which provides a lower bound on the resolution. This is also illustrated in [29].

2) Reconstruction:

0-order vs. 2D linear reconstruction in each segment. At a first glance, it seems natural to estimate by SPs the subdomains Ω_i of the model, which require 3 samples to obtain a plane approximation in the region of the SP. However, we found out that this is not an optimal strategy. A better approach is to increase the number of segments by a factor of 3 and to sample once each segment. This allows to increase the overall resolution (or *smallest object size*) of the system while still being able to recover reasonably well large planar segments, with a proper nonlinear filtering operation (see below). The depth of the smallest objects that the system can measure are estimated by a constant value. This facilitates the detection of poles, signs and small obstacles at a low sampling cost. Fig. 6 demonstrates that 0-order reconstruction is much more accurate than 2D linear reconstruction in terms of RMSE, for a given sampling budget n. In Fig. 7 the ability to detect well small objects by 0-order estimation is illustrated.

Bilateral filtering. Having more SPs with zero-order estimation allows to sample well small objects. However, now large flat regions are heavily degraded by staircasing artifacts. Our proposed solution is to apply a fast, nonlinear, edge preserving filter [32]. It is designed such that actual depth discontinuities are preserved, whereas false edges, which stem from the 0-order estimation, are smoothed out. Due to the log function, smoothing is relative to the depth. This approximates well the piece-wise planar model for large regions, such as walls and roads, yielding also lower RMSE, as seen in Fig. 6. As can be seen in Fig. 7, the artifacts in the reconstruction of the large planar background region are quite minimal.

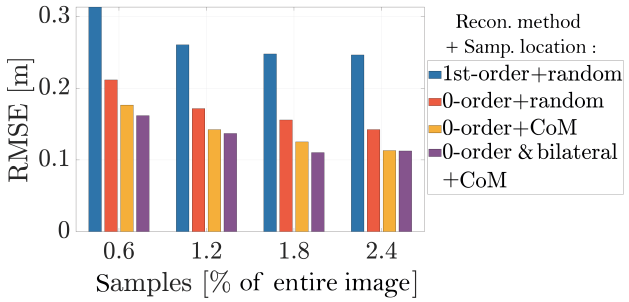
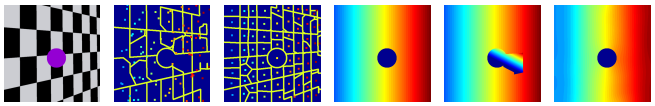


Fig. 6: Reconstruction variants on NYU-Depth-v2 pre-processed test set. We compare 1st-order reconstruction with 3 random depth samples inside each SP, to three variants of 0-order reconstruction with 1 sample inside each SP (in a random location or in its center of mass).



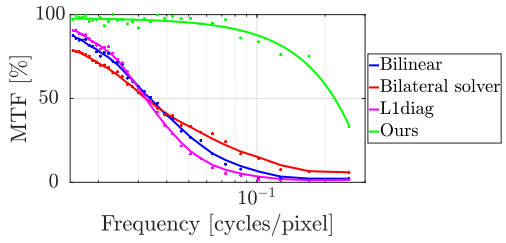
(a) RGB input (b) Large seg. samp. (c) Small seg. samp. (d) Depth GT (e) 1st order recon. (f) Our recon.

Fig. 7: Toy example 1: Comparing 1st-order estimation of larger segments (middle row, RMSE=55.0) to zero-order estimation for smaller segments, following nonlinear smoothing (bottom, RMSE=7.7). In both cases 75 samples were used.

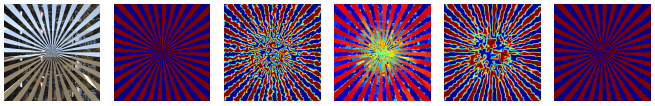
3) MTF analysis:

We aim to measure the spatial resolution of our sampling and reconstruction strategy. We use modulation transfer function (MTF) - a standard tool for characterizing the resolution of imaging systems [33]. Our chart is based on the Siemens Star testchart modified for RGB-guided depth sampling. We modified the original testchart by a road scene, replacing the white regions with typical background content and the black regions with typical foreground content (See Fig. 8b). Depth is assumed to take binary values, as in the original chart (Fig. 8c). The MTF calculation is explained in detail in [34].

Fig. 8 presents the results (a) computed from the reconstructed images (d-g). One can observe the clear increase of resolution of the proposed method, compared to RGB-guided and non-guided depth completion approaches.



(a) MTF plot



(b) RGB (c) GT (d) Bilin. (e) [9] (f) [6] (g) Ours

Fig. 8: MTF comparison between bilinear[1], bilateral solver [9], L1diag [6] and ours. $n = 5000$ (1% ratio).

V. Experiments

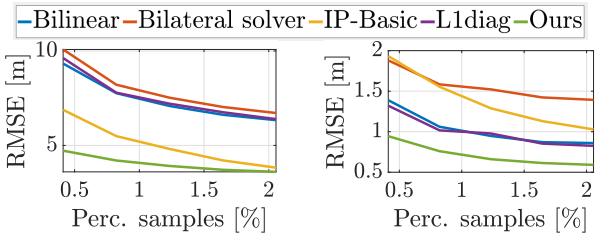
We evaluate the performance of our depth sampling and reconstruction method and compare it to other approaches with non-adaptive sampling (uniform random). We define pixel density as the ratio between the number of sampled pixels to the total number of pixels in the image. To demonstrate the generalization of our algorithm, we use two distinct datasets for evaluation - Synthia [27] for outdoor scenarios and NYU-Depth-v2 [28] for indoor scenarios. For the outdoor dataset we also evaluate over a subset of image areas focusing on small obstacles in the image.

[1]We use Delaunay triangulation [35] to perform a bivariate linear interp.

A. Outdoor data (Synthia)

The Synthia dataset [27] provides synthetic RGB, depth and semantic images for urban driving scenarios. More technical details are given in [29]. We use synthetic data as for now there is no large non-synthetic dataset that provides a dense and accurate depth map. We need a full dense map data to be able to sample freely in any desired coordinate. We make the evaluation on 0-100m depth range, which is similar to the range of a typical vehicle-mounted LiDAR.

Full scene experiment. Quantitative results are presented in Fig. 9a. We achieve a 30% lower RMSE than the second best at 0.45% density, and keep having the best result through all densities. Qualitative comparison is shown in Fig. 10, exhibiting precise and sharp reconstruction, especially of small objects.



(a) Synthia (b) Synthia obstacles

Fig. 9: Quantitative comparison on Synthia (left) and obstacles (right) datasets between bilinear interpolation, bilateral solver [9], IP-Basic [4], L1diag [6] and ours.

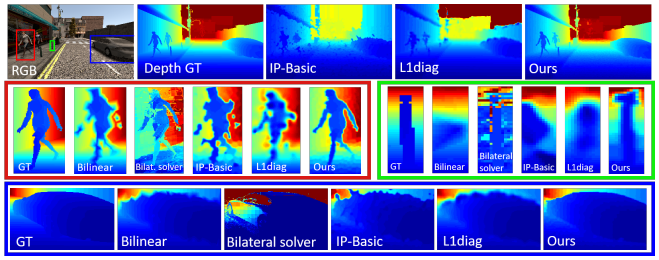


Fig. 10: Qualitative results of depth completion on Synthia and our obstacles dataset at 2% density (5000 samples).

Obstacles set. To enable evaluation over important objects in the image and to reduce the impact of far background on the performance measurement, we derived from Synthia a set of 100 obstacles, which we refer to as *the obstacles dataset*. We applied sampling and reconstruction over the entire image, but now evaluate over the obstacle mask. Quantitative results are presented in Fig. 9b, and qualitative comparison is shown in Fig. 10. Table I compares the number of samples n required to achieve certain levels of accuracy. *We require 3-4 times less samples for a given RMSE.*

Sampling only. We claim that using only our proposed sampling pattern, any completion method can achieve better results than using other existing sampling patterns, especially for small objects in the scene. Fig. 11 proves this qualitatively for 3 distinct reconstruction methods.

RMSE [m]	0.6	0.75	0.95
Bilinear interp.	7.87%	2.90%	1.23%
Bilateral solver [9]	$> 8\%$	$> 8\%$	$> 8\%$
IP-Basic [4]	$> 8\%$	4.78%	2.42%
L1diag [6]	7.09%	3.08%	1.33%
Ours	**1.89%**	**0.86%**	**0.40%**

TABLE I: Quantitative comparison for depth completion of required sampling percentage per RMSE on our Synthia obstacles dataset. Our method is 3-4 times more economic than second best.

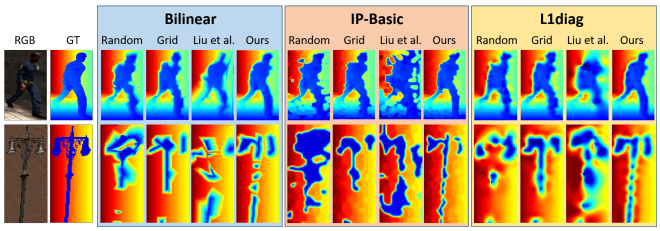


Fig. 11: **Sampling methods comparison:** 4 sampling patterns are compared for different reconstruction method: random, grid, Liu *et al* [20] and ours (sampling only). Recon. methods are bilinear interp., IP-Basic [4] and L1diag [6]

B. Indoor data (NYU-Depth-v2)

The NYU-Depth-v2 dataset [28] includes labeled pairs of aligned RGB and dense depth images collected from different indoor scenes. More technical details are given in [29]. Quantitative results are listed in Table II. Our method outperforms all other methods. Note that 4 out of the other 6 methods are based on deep learning, while ours is not. A qualitative comparison is shown in Fig. 12. Although suffering from slight staircase artifacts, our result preserves edges better and stays precise even in small items.

C. Prototype: Single-pixel mechanical sampler

Finally, to demonstrate the applicability of the concept, we performed an experiment over real scene designed in our laboratory. To enable controllable sampling, we built a sampling device (Fig. 13) composed of a laser rangefinder, a camera, motors and printed parts. A video demonstrating the device is provided. We generated ground-truth images for comparison with a Kinect 2 sensor. Note that ground-truth is in real-world coordinate, while our system measures range values (we did not perform a post-correction, thus, for instnce, the wall is not flat).

We created two scenes and sampled them with 3 different patterns to demonstrate the superiority of our method. Results are presented in Fig. 14. While the first scene (top) is a toy example for testing the sampling resolution, the second scene (bottom) is more realistic. As opposed to other patterns, our method is able to sample all object (even the thinner ones) in the scene and to reconstruct them well.

Samples	Method	RMSE [m]	REL
200	Bilinear interp.	0.257	0.047
200	L1diag [6]	0.236	0.044
225	Liao *et al* [16]	0.442	0.104
200	Ma *et al* [17]	0.230	0.044
200	HMS-Net [13]	0.233	0.044
200	Li *et al* [15]	0.256	0.046
200	Ours	**0.211**	**0.035**

TABLE II: Quantitative comparison with state-of-the-art on the NYU-Depth-v2 dataset.

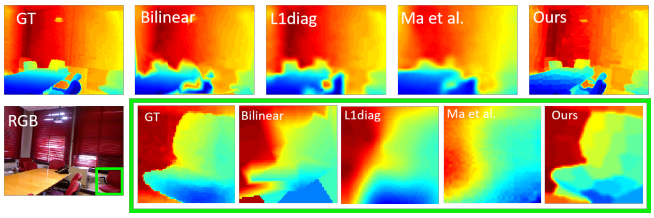


Fig. 12: Qualitative results of depth completion on NYU-Depth-v2 dataset at 0.7% density (500 samples).

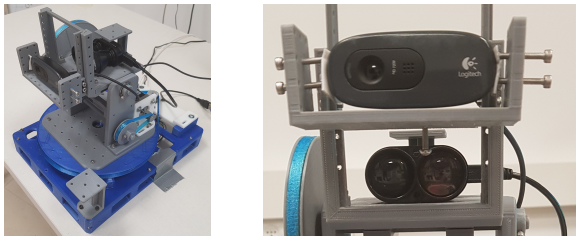
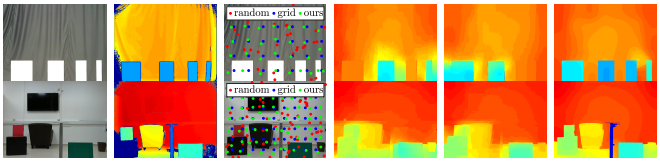


Fig. 13: Our mechanical sampler.



(a) RGB (b) GT (c) Samp. (d) Rand (e) Grid (f) Ours

Fig. 14: Experimental results on simple (top) and challenging (bottom) scenes taken in our lab. Bilateral solver [9] was applied over the random and grid samples.

VI. Conclusion

In this paper, we introduced a novel approach for image-based sparse depth sampling and dense reconstruction. We suggested a parametric piece-wise planar model and have shown its validity for indoor and outdoor datasets. We demonstrated that the correlation between depth and color domains allows to approximate well depth scenes using only an RGB image and a low number of carefully chosen depth samples. A single-pixel depth sampler was constructed as a proof-of-concept, verifying our predictions. We believe that this new direction calls for additional extensive research, in order to develop advanced, cheap and accurate depth sensing systems. In future work, we plan to combine classical and modern learning methods to further improve the performance and accuracy.

REFERENCES

[1] B. Schwarz, "Lidar: Mapping the world in 3d," *Nature Photonics*, vol. 4, no. 7, p. 429, 2010.

[2] P. Cheben, R. Halir, J. H. Schmid, H. A. Atwater, and D. R. Smith, "Subwavelength integrated photonics," *Nature*, vol. 560, no. 7720, p. 565, 2018.

[3] C. V. Poulton, A. Yaacobi, D. B. Cole, M. J. Byrd, M. Raval, D. Vermeulen, and M. R. Watts, "Coherent solid-state lidar with silicon photonic optical phased arrays," *Optics letters*, vol. 42, no. 20, pp. 4091–4094, 2017.

[4] J. Ku, A. Harakeh, and S. L. Waslander, "In defense of classical image processing: Fast depth completion on the cpu," *arXiv preprint arXiv:1802.00036*, 2018.

[5] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 349–356.

[6] F. Ma, L. Carlone, U. Ayaz, and S. Karaman, "Sparse depth sensing for resource-constrained robots," *arXiv preprint arXiv:1703.01398*, 2017.

[7] N. Chodosh, C. Wang, and S. Lucey, "Deep convolutional compressed sensing for lidar depth completion," *arXiv preprint arXiv:1803.08949*, 2018.

[8] J. Uhrig, N. Schneider, L. Schneider, U. Franke, T. Brox, and A. Geiger, "Sparsity invariant cnns," *arXiv preprint arXiv:1708.06500*, 2017.

[9] J. T. Barron and B. Poole, "The fast bilateral solver," in *European Conference on Computer Vision*. Springer, 2016, pp. 617–632.

[10] G. Drozdov, Y. Shapiro, and G. Gilboa, "Robust recovery of heavily degraded depth measurements," in *3D Vision (3DV), 2016 Fourth International Conference on*. IEEE, 2016, pp. 56–65.

[11] Z. Chen, V. Badrinarayanan, G. Drozdov, and A. Rabinovich, "Estimating depth from rgb and sparse sensing," *arXiv preprint arXiv:1804.02771*, 2018.

[12] A. Eldesokey, M. Felsberg, and F. S. Khan, "Propagating confidences through cnns for sparse data regression," *arXiv preprint arXiv:1805.11913*, 2018.

[13] Z. Huang, J. Fan, S. Yi, X. Wang, and H. Li, "Hms-net: Hierarchical multi-scale sparsity-invariant network for sparse depth completion," *arXiv preprint arXiv:1808.08685*, 2018.

[14] M. Jaritz, R. De Charette, E. Wirbel, X. Perrotton, and F. Nashashibi, "Sparse and dense data with cnns: Depth completion and semantic segmentation," in *2018 International Conference on 3D Vision (3DV)*. IEEE, 2018, pp. 52–60.

[15] Y. Li, K. Qian, T. Huang, and J. Zhou, "Depth estimation from monocular image and coarse depth points based on conditional gan," in *MATEC Web of Conferences*, vol. 175. EDP Sciences, 2018, p. 03055.

[16] Y. Liao, L. Huang, Y. Wang, S. Kodagoda, Y. Yu, and Y. Liu, "Parse geometry from a line: Monocular depth estimation with partial laser observation," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 5059–5066.

[17] F. Ma and S. Karaman, "Sparse-to-dense: Depth prediction from sparse depth samples and a single image," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–8.

[18] F. Ma, G. V. Cavalheiro, and S. Karaman, "Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera," *arXiv preprint arXiv:1807.00275*, 2018.

[19] S. Hawe, M. Kleinsteuber, and K. Diepold, "Dense disparity maps from sparse disparity measurements," in *13th International Conference on Computer Vision*, 2011.

[20] L.-K. Liu, S. H. Chan, and T. Q. Nguyen, "Depth reconstruction from sparse samples: Representation, algorithm, and sampling," *IEEE Transactions on Image Processing*, vol. 24, no. 6, pp. 1983–1996, 2015.

[21] A. Aldroubi and K. Gröchenig, "Nonuniform sampling and reconstruction in shift-invariant spaces," *SIAM review*, vol. 43, no. 4, pp. 585–620, 2001.

[22] P. Babu and P. Stoica, "Spectral analysis of nonuniformly sampled data–a review," *Digital Signal Processing*, vol. 20, no. 2, pp. 359–378, 2010.

[23] F. Marvasti, *Nonuniform sampling: theory and practice*. Springer Science & Business Media, 2012.

[24] J. Yen, "On nonuniform sampling of bandwidth-limited signals," *IRE Transactions on circuit theory*, vol. 3, no. 4, pp. 251–257, 1956.

[25] S. Baker, R. Szeliski, and P. Anandan, "A layered approach to stereo reconstruction," in *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231)*. IEEE, 1998, pp. 434–441.

[26] H. Tao, H. S. Sawhney, and R. Kumar, "A global matching framework for stereo computation," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1. IEEE, 2001, pp. 532–539.

[27] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez, "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3234–3243.

[28] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *European Conference on Computer Vision*. Springer, 2012, pp. 746–760.

[29] A. Wolff, S. Praisler, I. Tcenov, and G. Gilboa, "Image-guided depth sampling and reconstruction," *arXiv preprint arXiv:1908.01379*, 2019.

[30] P. Dollár and C. L. Zitnick, "Structured forests for fast edge detection," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1841–1848.

[31] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk *et al.*, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[32] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on*. IEEE, 1998, pp. 839–846.

[33] G. D. Boreman, *Modulation transfer function in optical and electro-optical systems*. SPIE press Bellingham, WA, 2001, vol. 21.

[34] C. Loebich, D. Wueller, B. Klingen, and A. Jaeger, "Digital camera resolution measurement using sinusoidal siemens stars," in *Digital Photography III*, vol. 6502. International Society for Optics and Photonics, 2007, p. 65020N.

[35] I. Amidror, "Scattered data interpolation methods for electronic imaging systems: a survey," *Journal of electronic imaging*, vol. 11, no. 2, pp. 157–177, 2002.