REGION-BASED DEPTH RECOVERY FOR HIGHLY SPARSE DEPTH MAPS

Said Pertuz[⋆†]   *Joni Kamarainen*[⋆]

[⋆]Tampere University of Technology, Tampere, Finland.
[†]Universidad Industrial de Santander, Bucaramanga, Colombia.
{said.pertuz, joni.kamarainen}@tut.fi

ABSTRACT

The accurate recovery of missing values in depth maps is an important problem in computer vision and image processing. In depth maps with large, irregular missing regions (i.e., sparse depth maps) inaccuracies arise when depth values of known pixels are used to recover depth near object edges and depth discontinuities (leakage). In order to overcome this problem, we propose an iterative region-based depth recovery method. In the proposed approach, the depth recovery problem is solved iteratively for each region of the segmented image in order to reduce the effect of leakage. Quantitative and qualitative experiments conducted on real data sets show promising results when comparing the proposed approach with state-of-the-art methods.

Index Terms— Depth recovery, depth map, inpainting, RGBD, segmentation

1. INTRODUCTION

The accurate estimation of simultaneous color and depth information is important for many different applications, such as 3D scene reconstruction [1], object detection [2, 3], activity and scene recognition [4, 5] and segmentation [6, 7]. This problem is often solved by the simultaneous use of active depth estimation sensors and RGB cameras. For instance, the well known Microsoft's Kinect sensor combines a structured light sensor with an RGB camera. Recently, due to their high speed and compactness, time of flight (TOF) sensors are increasingly becoming a popular alternative for depth measurement [8].

Despite the promising results obtained by combining active depth sensors and RGB cameras, this approach has several limitations and challenges. Specifically, non-lambertian surfaces, noise, reflective surfaces, and occlusions often result in corrupted depth estimates and missing information. The problem of recovering missing information, namely *depth recovery*, is closely related to image inpainting [9, 10]. In simple terms, guided depth recovery methods try to infer missing depth values by leveraging texture or color priors found in RGB images. More specifically, most existing approaches exploit global optimization methods in order to propagate depth values into missing regions based on color and texture consistency. Utilized techniques include Laplace and mode filters [11, 12], Markov random fields (MRF) [13, 14, 15], multilateral filters [16, 17, 18] and non-linear diffusion and variational frameworks [19, 20], among others [21, 22, 23, 24, 25, 26].

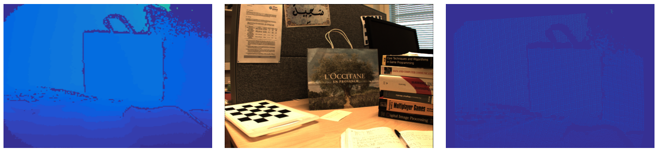


Fig. 1: Highly sparse depth maps. From left to right: original depth map, RGB image and warped sparse depth map.

The aforementioned methods yield compelling results for solving upsampling problems and filling in relatively small holes in the depth maps. However, as the unknown regions become large and irregular, these methods may not be able to produce accurate results and artifacts are introduced. In particular, in highly sparse depth maps, recovered depth becomes inaccurate as known depth estimates propagate through large unknown regions near object edges and depth discontinuities.

Highly sparse depth maps are typically obtained when there are large differences between the resolution of the depth sensor and the RGB camera, as well as due to large differences between the poses (location and orientation) of these devices. For illustration, Fig. 1 shows the result of warping a depth map captured by a TOF sensor into the frame of an RGB camera. Due to the difference in resolution and pose, the resulting depth map is highly sparse.

This paper proposes an iterative method for depth recovery in highly sparse depth maps. The proposed approach works by restricting the propagation of known depth values only into highly-consistent, independent image regions. For this purpose, the recovery problem is formulated based on the segmentation map obtained from the RGB image and is solved iteratively for each segmented region using an adjacency-based anisotropic diffusion framework. Quantitative and qualitative experiments on real datasets show that the proposed approach yields more accurate depth recovery than state-of-the-art methods for different levels of sparsity.

2. PROPOSED METHOD

The aim of the proposed approach is recovering missing depth values in a highly sparse depth map for which there is an RGB image available. From the RGB image, a segmentation map is obtained as a pre-processing step using either supervised or unsupervised segmentation methods [27, 28]. Subsequently, depth recovery is performed at two different levels: the region level (region-based inference) and the image level (iterative depth recovery). Each one of these stages are presented in the next sections. For the results presented in this paper, the GrabCut algorithm has been used in order to generate the segmented images [27, 29]. For the sake of brevity, it is assumed that the segmented image has been generated in a pre-processing step.

2.1. Iterative depth recovery

Let $\mathbf{S} = \{\mathbf{s}_i, i = 1, 2, \cdots, n\}$ bet the set of n regions in the segmented image (*segmented regions*), and $\mathbf{M} = \{\mathbf{m}_j, j = 1, 2, \cdots m\}$ be the set of m regions with missing values in the depth map (*missing regions*)[1]. Our hypothesis is that, for large regions with missing values, the accuracy of recovered pixels is reduced as known depth values leak into neighbor regions of the segmented images (*leakage*).

In order to reduce the effect of *leakage*, the depth recovery problem must be solved independently and iteratively for each segmented region. Specifically, we seek to generate a new set of p missing regions $\mathbf{P} = \{\mathbf{p}_k, k = 1, 2, \cdots, p\}$ that maximize the overlap between missing regions and segmented regions with $\cup_{k=1}^{p} \mathbf{p}_k = \mathbf{M}$. For this purpose, let us define the relative overlap $\mathcal{O} \in [0, 1]$ between two arbitrary regions, $\mathbf{a}$ and $\mathbf{b}$, as:

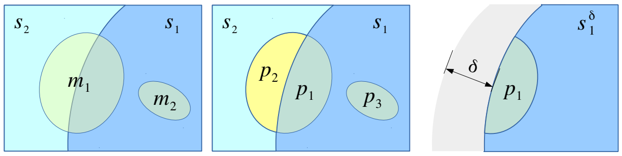
$$\mathcal{O}(\mathbf{a}, \mathbf{b}) = |\mathbf{a} \cap \mathbf{b}| / |\mathbf{a}| \qquad (1)$$

The new set of missing regions can be then generated using Algorithm 1. In this algorithm, the new set $\mathbf{P}$ is initialized as an empty set. There are basically two instances in which a new region is added to the new set: 1) when $\mathcal{O}(\mathbf{m}_j, \mathbf{s}_i) = 1$, which corresponds to missing regions fully contained in segmentation regions, and 2) when $0 < \mathcal{O}(\mathbf{m}_j, \mathbf{s}_i) < 1$, which corresponds to missing regions that overlap with more than one segmentation region. In this later case, the original missing region $\mathbf{m}_j$ is split into two or more new regions in $\mathbf{P}$ that are fully contained into one single region of the segmentation set.

For illustration, Fig. 2a shows a simple example with initial sets $\mathbf{S}$ and $\mathbf{M}$ with two regions each. In this example, it is clear that $\mathcal{O}(\mathbf{m}_2, \mathbf{s}_1) = 1$. Therefore, $\mathbf{m}_2$ is fully contained in $\mathbf{s}_1$ and gives origin to a unique new region in $\mathbf{P}$, namely $\mathbf{p}_3$ (Fig. 2b). In contrast, $0 < \mathcal{O}(\mathbf{m}_1, \mathbf{s}_1) < 1$. As a result, after applying 1 to Fig. 2a, $\mathbf{m}_1$ is split into $\mathbf{p}_1$ and $\mathbf{p}_2$.

Algorithm 1 Generation of new set of *missing regions*

```
$\mathbf{P} \leftarrow \{\emptyset\}$
for all $\mathbf{m}_j \in \mathbf{M}$ do
    for all $\mathbf{s}_i \in \mathbf{S}$ do
        if $0 < \mathcal{O}(\mathbf{m}_j, \mathbf{s}_i) \leq 1$ then
            $\mathbf{P} \leftarrow \mathbf{P} \cup (\mathbf{m}_j \cap \mathbf{s}_i)$
        end if
    end for
end for
return $\mathbf{P}$
```



(a) (b) (c)

Fig. 2: Iterative depth recovery. (a) Original sets with segmentation regions, $\mathbf{S}$, and missing regions, $\mathbf{M}$. (b) New region sets of missing regions $\mathbf{P}$ after the application of 1 to (a). (b) The expanded segmentation region $\mathbf{s}_i^\delta$.

The next step, after the generation of the new set $\mathbf{P}$, is the computation of missing depth values for each $\mathbf{p}_k \in \mathbf{P}$. The proposed method generates the recovered depth map $\mathbf{D}$ by solving the depth recovery problem individually for each region $\mathbf{s}_i \in \mathbf{S}$ by means of the algorithm 2.

As illustrated in Fig. 2c, the function *expand*$(\mathbf{s}_i, \delta)$ in algorithm 2 takes a segmentation region $\mathbf{s}_i$ and expands it by δ pixels beyond its boundaries using nearest neighbor interpolation. The aim of this operation is reducing the effect of leakage in boundary regions due to the initialization of the depth values in missing regions. In turn, the function *solve*$(\mathbf{p}_k, \mathbf{s}_i^\delta)$ finds the depth values in missing regions by means of an anisotropic diffusion framework as described in next section.

Algorithm 2 Iterative depth recovery

```
$\mathbf{D} \leftarrow \mathbf{S}$
for all $\mathbf{s}_i \in \mathbf{S}$ do
    $\mathbf{s}_i^\delta \leftarrow$ expand $(\mathbf{s}_i, \delta)$
    for all $\mathbf{p}_k \in \mathbf{P}$ do
        if $\mathcal{O}(\mathbf{p}_k, \mathbf{s}_i^\delta) \neq 0$ then
            $\hat{\mathbf{p}}_k \leftarrow$ solve $(\mathbf{p}_k, \mathbf{s}_i)$
            $\mathbf{D} \leftarrow \mathbf{D} \cup \hat{\mathbf{p}}_k$
        end if
    end for
end for
return $\mathbf{D}$
```

[1] In the sequel, it is assumed that the depth map is warped into the RGB camera frame

2.2. Region-based estimation

Given a missing region $\mathbf{p}_k$ corresponding to an expanded segmentation region $\mathbf{s}_i^\delta$, the recovered depth values $\hat{\mathbf{p}}_k$ can be estimated by finding the steady state solution to the diffusion equation:

$$\mathrm{div}(\Omega(\mathbf{x})\nabla D(\mathbf{x})) = 0 \qquad \text{s.t. } D(\mathbf{x}) = D_0(\mathbf{x}) \forall \mathbf{x} \notin \mathbf{p}_k \quad (2)$$

where $\mathbf{x} = [x, y]^T$ are pixel coordinates, $\Omega(\mathbf{x})$ is a conductance term, $D(\mathbf{x})$ is the sought solution and $D_0(\mathbf{x})$ the initial depth map with missing values.

Following [19], we solve (2) by reformulating the diffusion problem as a linear system of equations $\mathbf{AD} = \mathbf{D_0}$, where $\mathbf{D}_0$ is the vectorized initial depth map generated by setting the missing values to 0, and $\mathbf{A} \in \mathbb{R}^{N \times N}$ is a sparse positive-definite matrix whose elements are given by:

$$A_{i,j} = \begin{cases} I_{i,j} & \text{if } D_0(i) \neq 0 \\ I_{i,j} - \gamma W_{i,j} & \text{otherwize} \end{cases} \qquad (3)$$

with γ being a constant conductivity term, $I_{i,j}$ are the elements of the identity matrix, N is the number of pixels in the depth map, and $W_{i,j}$ is an adjacency term that equals 1 when the i-th and j-th pixels are 4-connected neighbors.

2.3. Parameter selection

The proposed algorithm depends on two parameters, namely the region expansion δ (measured in pixels) and the conductivity term γ. The region expansion is aimed at reducing the effect of leakage thus preserving edges in the recovered depth map. The selection of this parameter is a trade-off between accuracy and computational cost. For the results presented in this paper, a value of $\delta = 25$ has been used. As for the conductivity term, this parameter defines the strength of the influence of pixels with known depth values in the estimation of depth for missing pixels. In the limit, when $\gamma \rightarrow 0$, the solution of the diffusion problem in (2) converges to $D_0(\mathbf{x})$. A value of $\gamma = 0.01$ has been selected in this work.

3. EXPERIMENTS AND RESULTS

In order to validate the proposed approach, both quantitative and qualitative experiments are performed. For this purpose, a set of depth maps are acquired using a *Camboard Picoflex* TOF sensor for 15 different scenes with distance ranges between 0.15 and 4 meters. Simultaneously, RGB images of each scene are captured using a *Lytro Illum* camera. The intrinsic and extrinsic parameters of the TOF sensor and the RGB camera are pre-calibrated using standard camera calibration.

For quantitative experiments, the captured depth maps are deteriorated at different sparsity levels by randomly removing k% of pixels. Depth recovery methods are then used in order to obtain depth values of missing pixels and the root-mean-squared-error (RMSE) is computed using the orignal depth map as ground truth. The proposed method is compared with state-of-the-art approaches based on different principles, such as the original anisotropic diffusion approach (AD) [19], bilateral filtering [16], kernel-based MRF [21], MRF [13], and second order smoothing (SOS) [22]. For all the experiments and methods, the segmented image is used as guidance for depth recovery. As shown in Table. 1, the proposed method yields better performance in the test set for sparsity levels between 20% and 60%.

Table 1: Average RMSE (mm) of recovered depth maps at different levels of sparsity

	Depth map sparsity *k* (%)				
Method	20	30	40	50	60
Proposed	**1.38**	**2.08**	**2.93**	**4.03**	**5.24**
AD [19]	1.62	2.49	3.56	4.95	6.69
BF [16]	18.3	18.5	18.5	18.6	18.7
k-MRF [21]	6.03	6.94	8.02	9.49	11.39
MRF [13]	11.0	12.1	13.4	15.1	17.3
SOS [22]	3.39	5.89	10.4	14.4	14.8

For qualitative evaluation, the original depth maps are warped into the RGB camera frame[2]. Due to differences in pose and resolution (RGB images are 434×625 and TOF images are 171×224 pixels), the resulting depth maps are highly sparse (see second row of Fig. 3). The studied depth recovery algorithms are then used to estimate full depth maps in camera frame. As illustrated in Fig. 3, the proposed algorithm yields recovered depth maps with sharper edges. These results support the claim that the proposed iterative depth recovery method reduces the effect of leakage in large missing regions close to depth discontinuities.

4. CONCLUSIONS

This paper presents a method for depth recovery in highly sparse depth maps. The proposed method solves the depth recovery problem iteratively by applying an anisotropic diffusion framework to each region of the segmented image. Qualitative experiments show that the proposed approach outperform existing alternatives in depth maps with $20 - 60\%$ of missing data. In contrast to previous existing methods, the proposed algorithm better takes advantage of the regions found in the segmentation image in order to prevent oversmoothing effects near image boundaries and depth discontinuities. These results are experimentally validated using real data for the registration of depth maps captured with TOF sensor with color images captured using an RGB camera.

[2] Notice that, in this case, the ground truth is not available in the RGB camera frame and, therefore, quantitative evaluation is not possible.

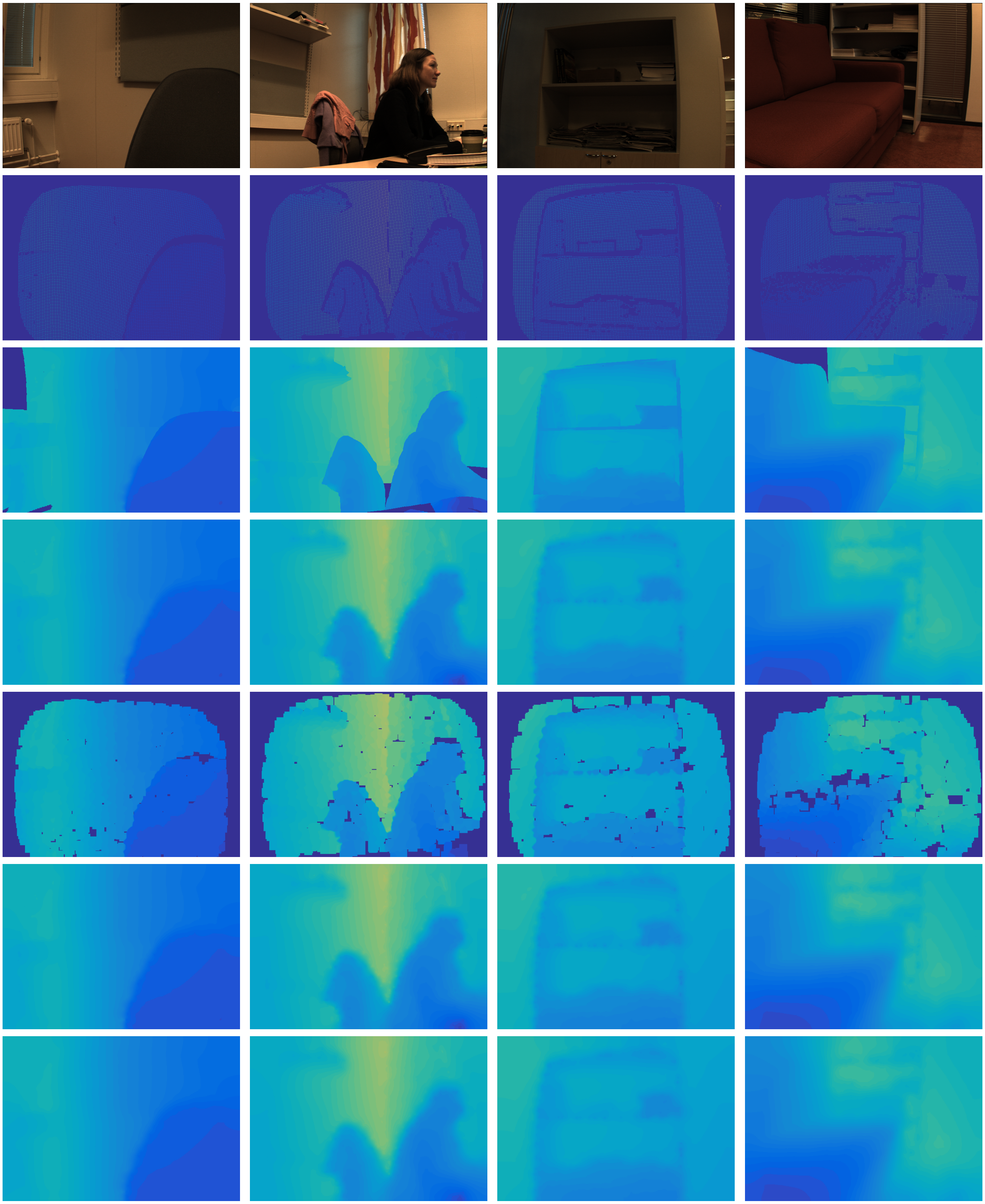


Fig. 3: Qualitative comparison. From top to bottom: RGB image, original depth map (after warping), recovered depth map using proposed method, AD [19], Bilateral filter [16], k-MRF [21] and MRF [13]. The SOS method [22] was not included since it failed to converge due to the high sparsity of the depth maps.

5. REFERENCES

[1] G. Hu, S. Huang, L. Zhao, A. Alempijevic, and G. Dissanayake, "A robust RGB-D SLAM algorithm," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 1714–1719.

[2] S. Gupta, P. Arbelez, and J. Malik, "Perceptual organization and recognition of indoor scenes from RGB-D images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 564–571.

[3] D. Lin, S. Fidler, and R. Urtasun, "Holistic scene understanding for 3D object detection with RGBD cameras," in *IEEE International Conference on Computer Vision*, 2013, pp. 1417–1424.

[4] H. Zhu, J. B. Weibel, and S. Lu, "Discriminative multimodal feature fusion for RGBD indoor scene recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2969–2976.

[5] Jaeyong Sung, C. Ponce, B. Selman, and A. Saxena, "Unstructured human activity detection from RGBD images," in *IEEE International Conference on Robotics and Automation*, 2012, pp. 842–849.

[6] H. Fu, D. Xu, and S. Lin, "Object-based multiple foreground segmentation in RGBD video," *IEEE Transactions on Image Processing*, vol. PP, no. 99, pp. 1–1, 2017.

[7] J. Feng, B. Price, S. Cohen, and S. F. Chang, "Interactive segmentation on RGBD images via cue selection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 156–164.

[8] A. Kolb, E. Barth, R. Koch, and R. Larsen, "Time-of-flight cameras in computer graphics," *Computer Graphics Forum*, vol. 29, no. 1, pp. 141–159, 2010.

[9] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[10] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 882–889, 2003.

[11] D. Min, J. Lu, and M. N. Do, "Depth video enhancement based on weighted mode filtering," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1176–1190, 2012.

[12] S. Liu, Y. Wang, H. Wang, and C. Pan, "Kinect depth inpainting via graph laplacian with TV21 regularization," in *IAPR Asian Conference on Pattern Recognition*, 2013, pp. 251–255.

[13] James Diebel and Sebastian Thrun, "An application of Markov Random Fields to range sensing," in *International Conference on Neural Information Processing Systems*, 2005, pp. 291–298.

[14] J. Zhu, L. Wang, J. Gao, and R. Yang, "Spatial-temporal fusion for high accuracy depth maps using dynamic MRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 899–909, 2010.

[15] J. Park, H. Kim, Yu-Wing Tai, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3d-TOF cameras," in *International Conference on Computer Vision*, 2011, pp. 1623–1630.

[16] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama, "Digital photography with flash and no-flash image pairs," in *ACM SIGGRAPH 2004 Papers*. 2004, pp. 664–672, ACM.

[17] Johannes Kopf, Michael F. Cohen, Dani Lischinski, and Matt Uyttendaele, "Joint bilateral upsampling," in *ACM SIGGRAPH 2007 Papers*. 2007, ACM.

[18] S. W. Jung, "Enhancement of image and depth map using adaptive joint trilateral filter," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 2, pp. 258–269, 2013.

[19] Junyi Liu and Xiaojin Gong, "Guided depth enhancement via anisotropic diffusion," in *Advances in Multimedia Information Processing*, 2013, pp. 408–417.

[20] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *IEEE International Conference on Computer Vision*, Dec 2013, pp. 993–1000.

[21] Ruigang Yang, Qingxiong Yang, James Davis, and David Nister, "Spatial-depth super resolution for range images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[22] Carlos Herrera, Juho Kannala, Ladicky Lubor, and Janne Heikkila, "Depth map inpainting under a second–order smoothness prior," in *Scandinavian Conference on Image Analysis*, 2013, pp. 555–566.

[23] Xiaojin Gong, Junyi Liu, Wenhui Zhou, and Jilin Liu, "Guided depth enhancement via a fast marching method," *Image and Vision Computing*, vol. 31, no. 10, pp. 695–703, 2013.

[24] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3443–3458, 2014.

[25] M. Stommel, M. Beetz, and W. Xu, "Inpainting of missing values in the Kinect sensor's depth maps based on background estimates," *IEEE Sensors Journal*, vol. 14, no. 4, pp. 1107–1116, 2014.

[26] M. Bevilacqua, J. F. Aujol, M. Brdif, and A. Bugeau, "Visibility estimation and joint inpainting of lidar depth maps," in *IEEE International Conference on Image Processing*, Sept 2016, pp. 3503–3507.

[27] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, ""GrabCut": Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.

[28] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "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.

[29] OpenCV, "Interactive foreground extraction using GrabCut algorithm," online, Dec 2015, visited on 23/01/2017.