

# Extended Patch Prioritization for Depth Filling within Constrained Exemplar-based RGB-D Image Completion

Amir Atapour-Abarghouei and Toby P. Breckon

Computer Science and Engineering, Durham University, UK

**Abstract.** We address the problem of hole filling in depth images, obtained from either active or stereo sensing, for the purposes of depth image completion in an exemplar-based framework. Most existing exemplar-based inpainting techniques, designed for color image completion, do not perform well on depth information with object boundaries obstructed or surrounded by missing regions. In the proposed method, using both color (RGB) and depth (D) information available from a common-place RGB-D image, we explicitly modify the patch prioritization term utilized for target patch ordering to facilitate improved propagation of complex texture and linear structures within depth completion. Furthermore, the query space in the source region is constrained to increase the efficiency of the approach compared to other exemplar-driven methods. Evaluations demonstrate the efficacy of the proposed method compared to other contemporary completion techniques.

**Keywords:** depth filling, image completion, exemplar-based inpainting, depth map, disparity hole filling.

## 1 Introduction

As three dimensional scene understanding based on scene depth is becoming ever more applicable, missing or invalid depth information has resulted in the need for special case facets of subsequent processing (e.g. semantic understanding, tracking, odometry and alike), and the prevalence of low cost, yet imperfect, depth sensing has seen depth completion emerge as an important research topic.

Despite significant prior work in color image completion [1–3, 6, 8, 11], depth filling is by contrast scanty present within the literature [4, 7, 9, 10, 13, 30] emerging as a relatively new research area posing significant challenges [12]. Although there have been many attempts to use structure-based or exemplar-based color image completion approaches for depth hole filling [1–3, 5], particular factors such as the absence of granular texture, clear object separation and the lack of in-scene transferability of varying depth sub-regions all create notable obstacles not present in the corresponding color completion case [29].

In this paper, we propose an improved exemplar-based inpainting approach [1] for depth completion (Fig. 1) that adds additional “*boundary*” and “*texture*” terms to aid in determining the priority of the sample patches used to propagate the structure and texture into the target region (Fig. 2). High computational demands, commonly associated with such approaches, are also reduced by dynamically constraining the query space

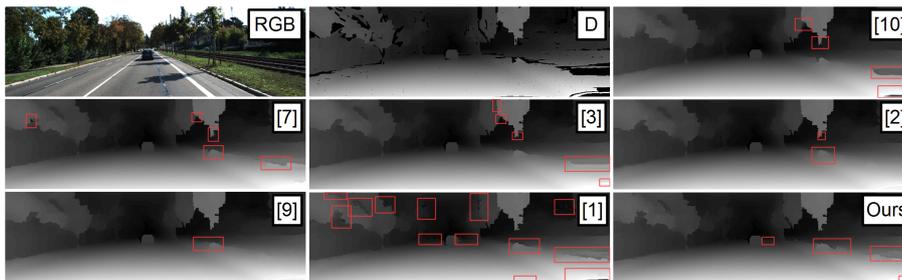
based on the location of spatially adjacent sample patch selections (Section 3). This is demonstrated by providing superior results within a traditional exemplar-based image completion paradigm against other leading contemporary approaches (Fig. 1).

## 2 Prior Work

Prior work in depth hole filling [7, 9, 12, 14, 23, 24] is not as comprehensive as color image completion. In the depth filling literature, there have been attempts to fill color and depth via depth-assisted texture synthesis in stereo images [15], a myriad of approaches utilizing filters [13, 16], temporal-based methods [17, 18], reconstruction-based methods [19, 20], and others [7, 9, 10, 21]. We focus on some of the most relevant [4, 7, 10, 21].

In a notable work, [7] improves upon the fast marching method-based inpainting proposed by [3] for depth filling. By assuming that the adjacent pixels with similar color values have a higher probability of having similar depth as well, they introduce an additional “*color*” term into the function to increase the contribution of the pixels with the same color.

By contrast, [21] uses a fusion-based method integrated with a non-local filtering strategy. Their framework follows [22], utilizing a scheme similar to non-local means to make accurate predictions for depth values based on image textures.



**Fig. 1.** Exemplar results on the KITTI dataset [25]. *RGB* denotes the color image, and *D* original (unfilled) disparity map. Results are compared with [1–3, 7, 9, 10]. Flaws are marked in red.

Herrera et al. [10] propose an approach similarly guided by the color image based on the assumption that every surface is continuous and smooth within their energy function formulation. This “*smoothness*” term encourages flat depth planes in the completion process whilst ignoring the possibility of visible texture or relief in the filled depth region and hence limiting plausible (reasonable) completion characteristics. Zhang et al. [4] improve [1] by adding a “*level set distance*” term to the priority function. A joint trilateral filter performs smoothing post process.

Overall, although such exemplar-based methods have rarely been used in depth completion, they have the tendency to preserve texture. With increased granularity in modern depth sensing and increasing detail in depth scene rendering (e.g. illumination correction), the consideration of texture detail (relief) within any depth filling process is now paramount. As such, we propose an improved exemplar-based formulation capable of efficient and plausible depth texture completion.

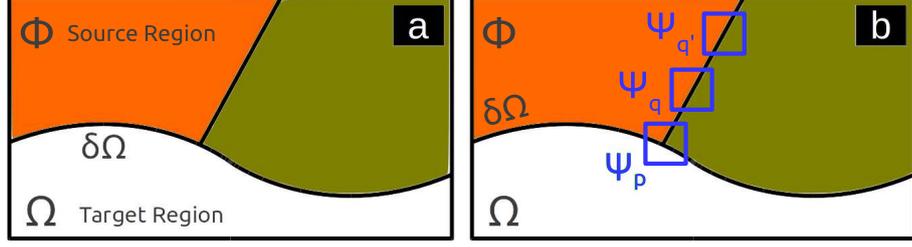


Fig. 2. Target region and target boundary (a), target/candidate patches (b).

### 3 Proposed Approach

In our approach, improvements are made to the framework of the exemplar-based inpainting [1] to create a more suitable and efficient depth filling approach. In the methodology of [1], the target region and its boundary are identified, a patch is selected to be inpainted and the source region is queried to find the best-matching patch via an appropriate error metric (e.g. sum of squared differences). After the candidate patch is found, all the information is updated and the process starts over. An extremely important factor in generating desirable results is the order in which these patches are selected for filling.

In [1], the priority of each patch is given by:

$$P(p) = C(p)D(p) \quad (1)$$

where  $C(p)$ , the “confidence” term, and  $D(p)$ , “data” term, are determined by:

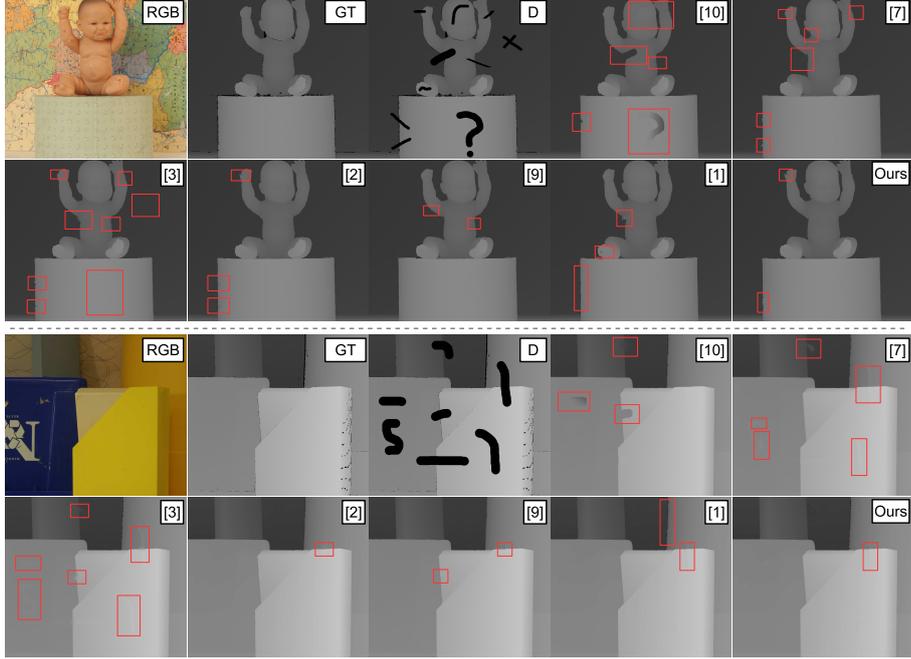
$$C(p) = \frac{\sum_{q \in \Psi_p \cap (\mathcal{I} - \Omega)} C(q)}{|\Psi_p|} \quad (2)$$

$$D(p) = \frac{|\nabla \mathcal{I}_p^\perp \cdot n_p|}{\alpha} \quad (3)$$

where  $|\Psi_p|$  is the area of the selected patch  $\Psi_p$ ,  $\mathcal{I}$  is the image,  $\Omega$  is the target region,  $\alpha$  is the normalization factor (255),  $n_p$  is a unit vector orthogonal to the target boundary, and  $\perp$  is the orthogonal operator (Fig. 2). Before the inpainting begins, the “confidence” term is initialized as:

$$C(p) = \begin{cases} 0, & \forall p \in \Omega \\ 1, & \forall p \in \Omega - \mathcal{I} \end{cases} \quad (4)$$

The “confidence” term prioritizes patches constrained by more valid depth values (fewer missing neighbors) and the “data” term encourages the filling of patches into which isophotes (lines of equal intensity) flow. This framework creates a balance between these two terms for a more plausible inpainting [1]. However, when completing real-world depth images with large holes covering entire objects, boundaries, and isophotes, the information in the accompanying color image (within RGB-D) can be used to create a suitable depth filling approach.



**Fig. 3.** Exemplar results on the Middlebury dataset [28]. *RGB* denotes the original color images, *GT* the ground truth depth, and *D* the original (unfilled) depth maps. Results are compared with [1–3, 7, 9, 10]. All flaws are marked in red.

In our approach, the “*confidence*” term is initialized and updated based on the depth image while the “*data*” term is calculated over the corresponding color image region (from RGB-D). To ensure a better flow of dominant linear structures into the target region, a “*boundary*” term is added based on the color image:

$$B(p) = \frac{\sum_{q \in \Psi_p \cap (\mathcal{I} - \Omega)} (|G_{x \geq \tau}(q)| + |G_{y \geq \tau}(q)|)}{|\Psi_p|} \quad (5)$$

where  $G_{x \geq \tau}$  and  $G_{y \geq \tau}$  are strong intensity gradients in the color image in the  $x$  and  $y$  directions respectively, with  $\tau$  being the gradient threshold (e.g.  $\tau = 0.7$ ). This term essentially prioritizes patches that contain a larger number of pixels that are part of a significant edge or gradient structure in the color image. This ensures a better propagation of object boundaries into the target region. As seen in Fig. 4, the original exemplar-based approach [1] gives equal priority to points A, B, and C (Fig. 4, result of [1]) while the proposed method prioritizes points B and C because of the “*boundary*” term (Fig. 4, proposed approach), which greatly effects the quality of the results.

Additionally, a “*texture*” term is introduced to guarantee a better propagation of texture into the target region. Since the color and depth gradients in certain parts of an image do not always match due to factors such as lighting and perspective, color information is not always a great indicator of texture. However, soft depth gradients always point to texture and relief, even though a depth image might appear smooth to the human eye. The “*texture*” term, which is applied to the depth image, determines which

Method	Plastic (1270 × 1110) [28]			Baby (1240 × 1110) [28]			Bowling (1252 × 1110) [28]		
	RMSE	PBMP	Run-time (s)	RMSE	PBMP	Run-time (s)	RMSE	PBMP	Run-time (s)
GIF [7]	0.7947	0.0331	3.10800	0.6008	0.0095	2.58000	0.9436	0.0412	4.87500
SSI [10]	1.7573	0.0102	42.3600	2.9638	0.0180	41.2000	6.4936	0.0455	71.1200
FMM [3]	0.9580	0.0435	0.9390	0.83490	0.0120	0.79400	1.2422	0.054	1.1190
FEI [2]	<b>0.6952</b>	<b>0.0032</b>	1641.33	<b>0.6755</b>	<b>0.0024</b>	995.250	<b>0.4857</b>	<b>0.0035</b>	1937.47
FBF [9]	0.8643	<b>0.0023</b>	>20000	<b>0.6238</b>	0.0081	>20000	<b>0.5918</b>	0.0072	>20000
EBI [1]	<b>0.4081</b>	0.0066	2145.71	0.9053	<b>0.0025</b>	1196.49	0.8733	<b>0.0045</b>	2921.15
Ours	<b>0.3843</b>	<b>0.0051</b>	1538.16	<b>0.6688</b>	<b>0.0021</b>	879.730	<b>0.7021</b>	<b>0.0037</b>	1606.75

**Table 1.** Comparing the RMSE (root-mean-square error), PBMP (percentage of bad matching pixels), and mean run-time of the methods over the Middlebury dataset [28].

parts of the image surrounding the target boundary contain texture and encourages the process to fill them earlier to propagate texture in the target region:

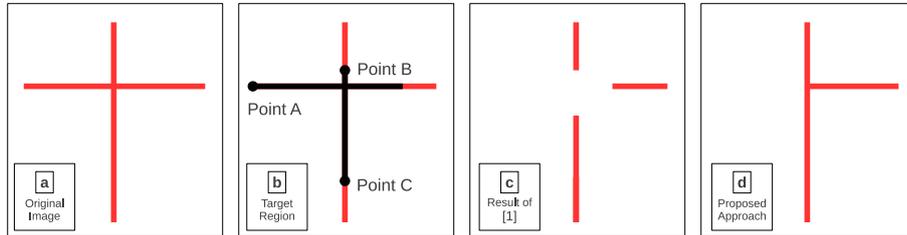
$$T(p) = \frac{\sum_{q \in \Psi_p \cap (\mathcal{I} - \Omega)} |G_{x < \tau}(q)| + |G_{y < \tau}(q)|}{|\Psi_p|} \quad (6)$$

where  $G_{x < \tau}$  and  $G_{y < \tau}$  are slight intensity gradients in the depth image in the  $x$  and  $y$  directions respectively, with  $\tau$  being the gradient threshold (e.g.  $\tau = 0.3$ ). Smallest changes in the depth image are identified and taken into account for a better relief texture propagation. As seen in Fig. 5, in which significant edges and linear structures are hard to find, the proposed method correctly prioritizes patches with slight depth changes and functions better than the original approach [1]. After adding the two aforementioned terms, the priority evaluation function is transformed to:

$$P(p) = C(p)D(p)B(p)T(P) \quad (7)$$

where  $C(p)$ ,  $D(p)$ ,  $B(p)$ ,  $T(P)$  are the “confidence” term (based on the depth image), “data” term (based on the color image), “boundary” (based on the color image), and the “texture” term (based on the depth image) respectively.

Finally, in most exemplar-based methods [1, 4, 6, 11], the entire source region is queried for candidate patches. However, our analysis shows that most suitable candidates for any patch are located close to where the best-matching candidates were found for adjacent patches in previous patch filling iterations. As a result, a dynamic search perimeter is created when sampling candidates for a patch with previously filled neighbors (Fig. 6). The maximum and minimum of  $x$  and  $y$  indices of the selected candidates



**Fig. 4.** A demonstration of the effect of the “boundary” term.

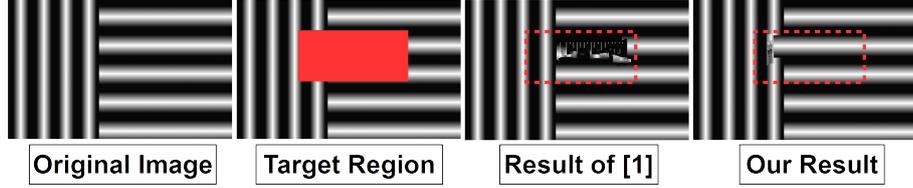


Fig. 5. A demonstration of the effect of the “texture” term.

for the previously-filled adjacent patches are used to determine the perimeter. Tests run over 20 different color and depth image pairs indicate that in 91.2% of queries, the best matching patch was found inside perimeter. Although this can negatively effect the quality of the results for the remaining 8.8% of patches, the efficiency is improved by an average of 31% (with negligible standard deviation), which is significant.

## 4 Experimental Results

Hole filling is fraught with constant compromises between efficiency and accuracy. The proposed approach is an example of this, as it outperforms many of its predecessors qualitatively and quantitatively [1, 3, 7, 10] while being faster than others [1, 2, 9]. Results were evaluated using a number of images, but in the interest of space, only a few are presented here. We utilize the Middlebury dataset [28] to provide qualitative and quantitative evaluation. Fig. 3 demonstrates that the proposed method generates plausible results without significant invalid outliers, blurring, jaggging or other artefacts compared to other approaches [1–3, 7, 9, 10]. All flaws and artefacts are marked in Fig. 3. Table 1 provides quantitative evaluation of the proposed approach against the same comparator set (GIF is the guided inpainting and filtering [7], SSI the second-order smoothness inpainting [10], FMM the fast marching method [3], FEI the framework for exemplar based inpainting [2], FBF the Fourier basis for filling [9], and EBI the exemplar-based inpainting [1]). As shown in Table 1, the method is in balance between efficiency and accuracy. While it is more efficient than other exemplar-based methods [1, 2], it has a smaller root-mean-square error and fewer bad pixels (based on the evaluation methodology of [27]) than faster comparators [3, 7]. Experiments were performed on a 2.30GHz CPU (Table 1).

Fig. 1 demonstrates the results of the proposed method in comparison with [1–3, 7, 9, 10] when applied to examples from the KITTI dataset [25] (resolution,  $1242 \times 375$ ).

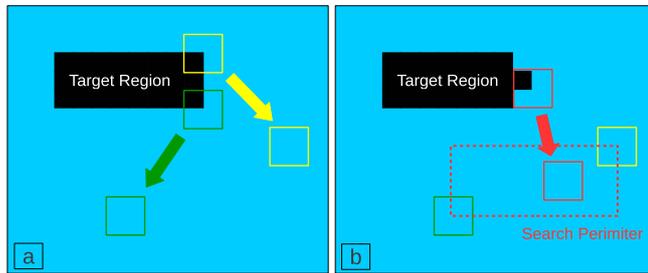


Fig. 6. Constraining the query space to improve efficiency.

Depth is calculated using [26] with significant disparity speckles filtered out. The proposed method results in sharp images with fewer additional artefacts (Fig. 1). The closest performing approach, the variational framework for exemplar-based approach of [2], shows comparable quantitative performance (RMSE/PBMP, Table 1) in some aspects but our approach offers a mean computational saving of 15.2% over [2]. The faster approaches [3, 7, 10] have significantly worse completion performance (Table 1, Fig. 1 & 3) than our approach. We have created a video displaying the results of the work. We invite you to view the video, which can be found here: <https://vimeo.com/251792601>.

## 5 Conclusions

In this paper, the problem of depth completion is addressed in an exemplar-based framework with a focus on a balance between efficiency and attention to surface (relief) detail accuracy. While exemplar-based methods, are mostly used for color images, their ability to preserve texture in the target region makes them suitable for depth filling when texture is of importance. Here, the priority term that determines the order of patch sampling has been modified to allow for a better propagation of strong linear structures and texture into the target region. Moreover, by constraining the query space, the method performs more efficiently than other exemplar-based approaches. Our evaluation demonstrates that while the efficiency of the proposed method is better than other exemplar-based frameworks, the plausibility and statistical relevance of the depth filled results compete against the accuracy of contemporary filling approaches in the field.

## References

1. Criminisi, A. and Pérez, P. and Toyama, K.: Region Filling and Object Removal by Exemplar-based Image Inpainting. In: *IEEE Trans. Image Processing*. 13(9), 1200–1212 (2004)
2. Arias, P. and Facciolo, G. and Caselles, V. and Sapiro, G.: A Variational Framework for Exemplar-based Image Inpainting. In: *Int. J. Computer Vision*. 93(3), 319–347. Springer (2011)
3. Telea, A.: An Image Inpainting Technique Based on the Fast Marching Method. In: *Graphics Tools*. 9(1), 23–24 (2004)
4. Zhang, L. and Shen, P. and Zhang, S. and Song, J. and Zhu, G.: Depth Enhancement with Improved Exemplar-based Inpainting and Joint Trilateral Guided Filtering. In: *Int. Conf. Image Processing*, pp. 4102–4106 (2016)
5. Hervieu, A. and Papadakis, N. and Bugeau, A. and Gargallo, P. and Caselles, V.: Stereoscopic Image Inpainting: Distinct Depth Maps and Images Inpainting. In: *Int. Conf. Pattern Recognition*, pp. 4101–4104 (2010)
6. Cheng, W. and Hsieh, C. and Lin, S. and Wang, C. and Wu, J.: Robust Algorithm for Exemplar-based Image Inpainting. In: *Int. Conf. Computer Graphics, Imaging and Visualization*, pp. 64–69 (2005)
7. Liu, J. and Gong, X. and Liu, J.: Guided Inpainting and Filtering for Kinect Depth Maps. In: *Int. Conf. Pattern Recognition*, pp. 2055–2058 (2012)
8. Goyal, P. and Diwakar, S. and others: Fast and Enhanced Algorithm for Exemplar-based Image Inpainting. In: *Symp. Image and Video Technology*, pp. 325–330 (2010)
9. Atapour-Abarghouei, A. and Payen de La Garanderie, G. and Breckon, T.P.: Back to Butterworth - a Fourier Basis for 3D Surface Relief Hole Filling within RGB-D Imagery. In: *Int. Conf. Pattern Recognition*, pp. 2813–2818 (2016)

10. Herrera, D. and Kannala, J. and Heikkilä, J.: Depth Map Inpainting under a Second-Order Smoothness Prior. In: Scandinavian Conf. Image Analysis, pp. 555–566 (2013)
11. Kumar, V. and Mukhopadhyay, J. and Mandal, S.K.D.: Modified Exemplar-Based Image Inpainting via Primal-Dual Optimization. In: Int. Conf. Pattern Recognition and Machine Intelligence, pp. 116–125. Springer (2015)
12. Breckon, T.P. and Fisher, R.B.: Amodal Volume Completion: 3D Visual Completion. In: Computer Vision and Image Understanding, 99(3), 499–526 (2005)
13. Camplani, M. and Salgado, L.: Efficient Spatio-Temporal Hole Filling Strategy for Kinect Depth Maps. In: IS&T/SPIE Electronic Imaging, pp. 82900E–82900E (2012)
14. Breckon, T.P. and Fisher, R.B.: A Hierarchical Extension to 3D Non-parametric Surface Relief Completion. In: Pattern Recognition, 45, 172–185 (2012)
15. Wang, L. and Jin, H. and Yang, R. and Gong, M.: Stereoscopic Inpainting: Joint Color and Depth Completion from Stereo Images. In: Int. Conf. Computer Vision and Pattern Recognition, pp. 1–8 (2008)
16. Camplani, M. and Salgado, L.: Adaptive Spatio-Temporal Filter for Low-Cost Camera Depth Maps. In: Int. Conf. Emerging Signal Processing Applications, pp. 33–36 (2012)
17. Matyunin, S. and Vatolin, D. and Berdnikov, Y. and Smirnov, M.: Temporal Filtering for Depth Maps Generated by Kinect Depth Camera. In: 3DTV Conference, pp. 1–4 (2011)
18. Berdnikov, Y. and Vatolin, D.: Real-Time Depth Map Occlusion Filling and Scene Background Restoration for Projected-Pattern Based Depth Cameras. In: Graphic Conf. IETP (2011)
19. Chen, C. and Cai, J. and Zheng, J. and Cham, T. and Shi, G.: Kinect Depth Recovery Using a Color-Guided, Region-Adaptive, and Depth-Selective Framework. In: ACM Trans. Intelligent Systems and Technology, 6(2), 12 (2015)
20. Wang, Z. and Hu, J. and Wang, S. and Lu, T.: Trilateral Constrained Sparse Representation for Kinect Depth Hole Filling. In: Pattern Recognition Letters, 65, 95–102 (2015)
21. Qi, F. and Han, J. and Wang, P. and Shi, G. and Li, F.: Structure Guided Fusion for Depth Map Inpainting. In: Pattern Recognition Letters, 34(1), 70–76 (2013)
22. Bugeau, A. and Bertalmio, M. and Caselles, V. and Sapiro, G.: A Comprehensive Framework for Image Inpainting. In: IEEE Trans. Image Processing, 19(10), 2634–2645 (2010)
23. Bevilacqua, M. and Aujol, J. and Brédif, M. and Bugeau, A.: Visibility Estimation and Joint Inpainting of Lidar Depth Maps. In: Int. Conf. Image Processing, pp. 3503–3507 (2016)
24. Zuo, Y. and Wu, Q. and An, P. and Zhang, J.: Explicit Measurement on Depth-Color Inconsistency for Depth Completion. In: Int. Conf. Image Processing, pp. 4037–4041 (2016)
25. Geiger A. and Lenz P. and Stiller C. and Urtasun R.: Vision meets Robotics: The KITTI Dataset. In: Robotics Research (2013)
26. Yamaguchi, K. and McAllester, D. and Urtasun, R.: Efficient Joint Segmentation, Occlusion Labeling, Stereo and Flow Estimation. In: European Conf. Computer Vision, pp. 756–771. Springer (2014)
27. Scharstein, D. and Szeliski, R.: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. In: Int. J. Computer Vision, 47, 7–42 (2002)
28. Hirschmuller, H. and Scharstein, D.: Evaluation of Cost Functions for Stereo Matching. In: Conf. Computer Vision and Pattern Recognition, pp. 1–8 (2007)
29. Atapour-Abarghouei, A. and Breckon, T.P.: A Comparative Review of Plausible Hole Filling Strategies in the Context of Scene Depth Image Completion. In: J. Computers and Graphics, 72, 39–58 (2018)
30. Atapour-Abarghouei, A. and Breckon, T.P.: DepthComp: Real-time Depth Image Completion Based on Prior Semantic Scene Segmentation. In: British Machine Vision Conference, pp. 208.1–208.13 (2017)