

Depth Image Completion through Iterative Low-Pass Filtering

Tzu-Kai Wang¹, Yeh-Wei Yu¹, Tsung-Hsun Yang¹ , Pin-Duan Huang¹, Guan-Yu Zhu¹, Chi-Chung Lau²
and Ching-Cherng Sun^{1,*}

¹ Department of Optical and Photonics, National Central University, Chung-Li, Taoyuan City 320317, Taiwan

² Energy and Resource Laboratories, Industrial Technology Research Institute, Hsinchu 31041, Taiwan;
cclau999@gmail.com

* Correspondence: ccsun@dop.ncu.edu.tw

Abstract: This study introduces a spatial-modulated approach designed to recover missing data in in-depth images. Typically, commercial-grade RGB-D cameras utilize structured light or time-of-flight techniques for capturing scene depth. However, these conventional methods encounter difficulties in acquiring depth data from glossy, transparent, or low-reflective surfaces. Additionally, they are prone to interference from broad-spectrum light sources, resulting in defective areas in the captured data. The generation of dense data is further compromised by the influence of noise. In response to these challenges, we implemented an iterative low-pass filter in the frequency domain, effectively mitigating noise and restoring high-quality depth data across all surfaces. To assess the efficacy of our method, deliberate introduction of significant noise and induced defects in the generated depth images was performed. The experimental results unequivocally demonstrate the promising accuracy, precision, and noise-resilient capabilities of our approach. Our implementation is publicly available on the project's webpage.

Keywords: depth image; 3D sensing; iterative low-pass filter



Citation: Wang, T.-K.; Yu, Y.-W.; Yang, T.-H.; Huang, P.-D.; Zhu, G.-Y.; Lau, C.-C.; Sun, C.-C. Depth Image Completion through Iterative Low-Pass Filtering. *Appl. Sci.* **2024**, *14*, 696. <https://doi.org/10.3390/app14020696>

Academic Editor: Thomas Lindner

Received: 25 December 2023

Revised: 9 January 2024

Accepted: 9 January 2024

Published: 14 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Depth image completion has been a prominent focus in computer vision projects over the past few years, with significant advancements arising from both hardware improvements and developments in image processing. Hardware enhancements aim to achieve precise data detection, real-time data processing, and high filling rates for single frames. Consequently, technologies such as CMOS, compact laser diodes, projection units, and Diffractive Optical Elements (DOEs) have been developed and applied to depth sensors. Commercial-grade RGB-D cameras, including the Microsoft Kinect series, Intel RealSense, LUCID Helios, and Matterport 3D cameras, have found wide-ranging applications in scene reconstruction [1,2], robotic vision [3–5], and automotive systems [6,7].

These devices incorporate advanced optical technologies such as backside-illuminated CMOS [8] and current-assisted photonic demodulators [9], representing a significant leap in light sensitivity, gate-level signal processing, and overall performance. However, despite these technological strides, challenges persist, particularly in dealing with reflective targets, absorbing surfaces, and distant objects [10,11].

These issues stem from various factors affecting the light conditions, such as strong interference from infrared sources, reflective or transparent surfaces, and highly absorbing materials causing low contrast in infrared images or low Signal-to-noise ratio (SNR). Imaging geometry is an additional factor, causing shadows to be projected due to occlusion objects or the light source's position, resulting in blank areas or shadows in depth images. The pursuit of generating dense and accurate depth information faces many hurdles, with noise emerging as a formidable issue. Interference from broad-spectrum light sources introduces defects and noise, compromising the integrity of captured data. These disruptions can directly impede the structured light patterns and Time-of-Flight (ToF) signals to some

extent, resulting in inevitable defects in depth images and the generation of noisy scanning results [12].

This study aims to restore depth images derived from various 3D sensing techniques and accurately locate points in space, especially in areas affected by defects. Several methods have been explored for depth image completion, including extrapolation [13], LiDAR data inpainting with depth gradients [14,15], and surface reconstruction based on 3D structures [16–18]. In recent years, depth image completion using deep neural networks has received considerable attention, with representative research in monocular depth estimation [19–23], RGB-guided methods [24], learning-based surface normal-guided completion [25], and global and local depth prediction with RGB guidance [26]. While these neural network-based methods have achieved significant success in full-depth image processing with large dataset training, profound network structures, and fine-tuned parameters, they often suffer from the loss of fine details and malfunctions in the presence of noisy data [19,27].

In the realm of advanced depth completion techniques, RGB-guided methods such as Spatial Propagation Networks (SPN) have shown promise in capturing local affinities, resulting in improved depth completion scores. The Non-Local Spatial Propagation Network for Depth Completion (NLSPN) has emerged as a top performer in this regard [24,28–30]. Nevertheless, these approaches yield inaccurate results when confronted with glossy surfaces, absorbent surfaces, and noisy depth data. Depth errors occur when RGB data is incoherent with respect to depth data; for example, when the RGB reflects only part of the scene, overexposed areas, or dark scenes. Current advanced methods in depth estimation and completion achieve high scores primarily based on large dataset training, fine-tuning parameters, and the synthetic architecture of networks. However, dedicated-design networks still require further studies to tackle unknown scenes and complex distributed surfaces that do not exist in the dataset.

In light of these challenges, we introduce a novel depth completion methodology centered on frequency-domain manipulation. Spatial characteristics inherently encompass both the surface structure and noise, posing a formidable task in addressing both aspects effectively. To surmount this challenge, we have devised a method that robustly restores depth information within the frequency domain through an Iterative Low-Pass Filter (ILF). The ILF is designed to handle scenarios of single-shot chances and complex surfaces. We also provide an optimization characteristic of ILF using L1 norm analysis. As we process sparse and noisy depth data using our approach, noise is systematically eliminated, yielding a rich and accurate depth profile in each iteration. The resulting depth image encompasses the correct depth data, thereby resolving the longstanding issue of missing and noisy data.

2. Approach

This section provides insight into the functionality of the Iterative Low-Pass Filter (ILF) on a complement of depth images. The design concept aims to integrate features acquired from RGB-D scenes, mitigate noise interference, and generate accurate depth data. To achieve this, we employ an iteratively updated process combined with a dynamically adjustable low-pass filter. The RGB information collaboratively contributes to normalizing spatial frequencies, offering insights into the structure and 3D shape of the surface. Initially, we delineate the recovery surface and extract the spatial frequency of the surface structure from RGB images, utilizing watertight boundaries for precise definition. Following this, we extract the spatial features of the target surfaces from RGB images, expediting the overall process by highlighting surface trends. Figure 1 illustrates the Depth Image Completion Process (DICP), which comprises two integral components: (1) the Extraction of Boundaries and Surfaces (EBS) and (2) the ILF Process.

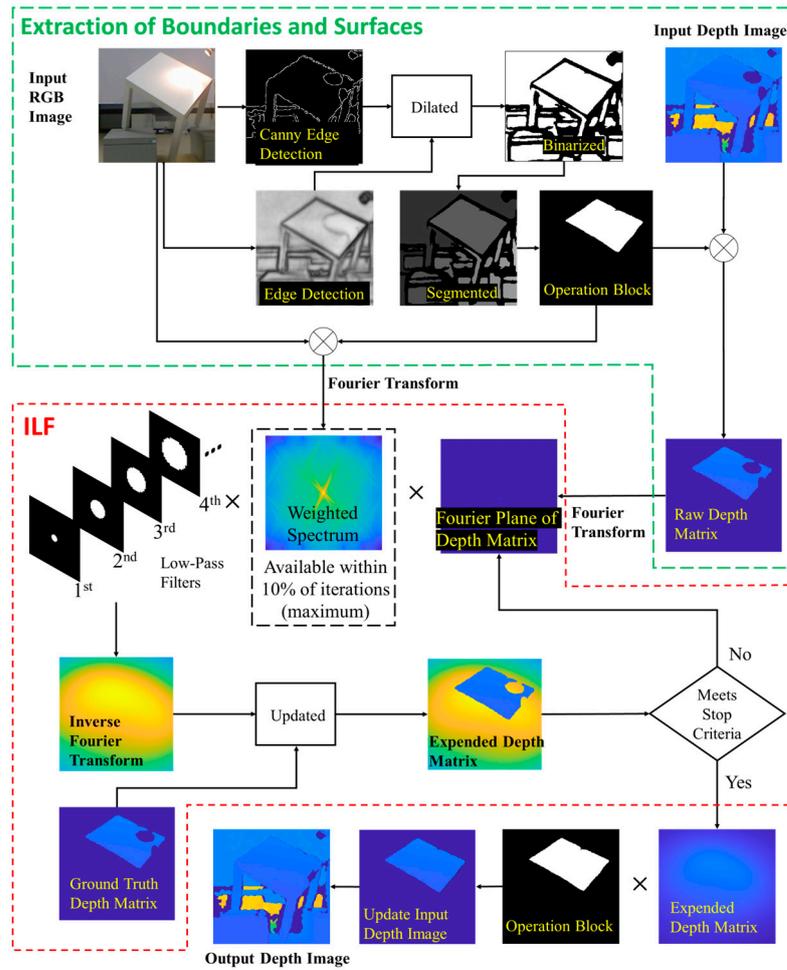


Figure 1. The flowchart of DICP.

2.1. Extraction of Boundaries and Surfaces

The primary objective of the Extraction of Boundaries and Surfaces (EBS) is to identify appropriate surfaces with watertight boundaries. The EBS process takes the RGB image and the depth image as inputs. The RGB image contributes edge information, which is processed through a combination of Canny edge detection [31] and other edge detection methods [32–35] to achieve a clear and watertight boundary. Among the edge detection methods, we opted for the deep network architecture introduced in holistic edge detection [32], known for its enhanced speed in edge detection and a reliable network structure. The network is a fully convolutional neural network (FCN) adopting VGGNet [36]. We utilized the BSD500 dataset [37] for training, as it demonstrated a high F-score and superior edge detection results [32] for our implementation.

To enhance surface extraction, we dilated the results of the edge detection with Canny Edge detection as a guide to contract the rough boundary, facilitating larger operation blocks. The contracted result represents the extracted surface, referred to as the operation block. The binarized shrunken results were then segmented to obtain operation blocks. Through this process, we obtained a reliable result for each surface. Finally, we multiplied the depth image by the operation block to generate the raw depth matrix. At this stage, the raw depth matrix included a defective area, and the subsequent step was to address and recover this defective area.

2.2. Iterative Low-Pass Filter (ILF)

After the extraction of boundaries and surfaces, the Iterative Low-Pass Filter (ILF) was applied to complete the depth matrices. In an ILF, depth matrices undergo the following process until the expanded depth matrix reaches the stopping criteria:

- (1) Multiplication with a low-pass filter in the Fourier domain, where the cut-off frequency changes from low to high in different iterations.
- (2) Updating of the expanded depth matrix with a raw depth matrix.

After several iterations, the low-frequency term reaches a steady state, and the incoming high frequency modifies the fine shape and detail of the surfaces. The expanded depth matrix is then cropped in the shape of the operation block, and the input depth image is updated in the same position. Consequently, the output depth image contains the updated depth matrix.

The initial spatial frequency of the low-pass filter starts from the baseband of the image, corresponding to the reciprocal of the most extended length of the target surface. Since the operation block can be seen as a 2D point distribution, we used Principal Component Analysis (PCA) to analyze the shortest length of the operation block. The shortest length was applied for the initial radius of the low-pass filter for the raw depth matrix. For instance, if the random shape of the operation block represents 10 pixels in the shortest length from the PCA, the starting spatial frequency will be 10 pixel units. Frequencies lower than the initial frequency contribute less to the shape of the surfaces. Thus, the information within these frequencies reaches a steady state in several epochs of the first low-pass filter. By applying the initial spatial frequency and dynamically adjusting the starting spatial frequency from zero to the initial frequency, we ensure a gradual restoration of information and accelerate the ILF process.

Following this, we iteratively expanded the low-pass filter's radius to obtain the surface's depth information in the shape of the operation block, moving from a low spatial frequency to a high spatial frequency. This step strengthens the included information from the previous iteration and minimizes interference from upcoming spatial frequencies. The depth information is updated iteratively to recover the defective area. As the depth information expands after each process with a low-pass filter, it becomes uniform due to the filter's smoothing effect. This uniformity makes the depth information inside the defective area similar to the depth information outside it. To address the compromise introduced by the low-pass filter, we reintegrated it with the raw depth matrix in each iteration.

After several iterations, the depth information reaches a steady state, resulting in the recovery of depth information. For noise reduction, the ILF process starts from a low spatial frequency to a high spatial frequency, which contains most of the noise. The ILF effectively reduces the high spatial frequency component of noise while preserving most of the main signal, as the high-frequency noise does not contribute to the update of the initial expanded depth matrix.

Here, we established a stop criterion to enhance the optimized performance and quality of the recovered depths. As each epoch of the Iterative Low-Pass Filter (ILF) yields an expanded depth matrix, it becomes essential to evaluate the termination point of the ILF process. To determine the optimal iteration steps, we analyzed the alternation between standard deviations in each epoch and the number of iterations. The stop criterion was designed to satisfy two conditions: the deviation between the results of consecutive iterations was sufficiently small, and it reached the maximum iteration limit.

$$\sigma_i < S_{criteria} \parallel i < 1500, S_{criteria} = 0.005, \quad (1)$$

$$\sigma_i = \sqrt{\frac{\sum (P_i - P_{i-1})^2}{N}}, \quad (2)$$

where P_i is the expanded depth matrix after the i -th iteration, σ_i is the deviation between P_i and $P_{(i-1)}$, and $S_{criteria}$ is the stop criterion for the minimum deviation, set to 0.005 m. This deviation is aligned with the precision of the depth camera. The results are presented below.

Figure 2 illustrates the depth differences between each iteration, highlighting the nuanced evolution during the Iterative Low-Pass Filter (ILF) process. Notably, we employed a weighted spectrum derived from a single RGB image to stabilize the ILF process. The spatial frequency of the RGB image was extracted and utilized as a weight in the initial iterations. This strategic approach stems from the observation that, under certain illumination conditions, the spatial frequency of the depth image can be discerned in RGB images. The spatial frequency of the surface at low and mid frequencies serves as an insightful initial approximation.

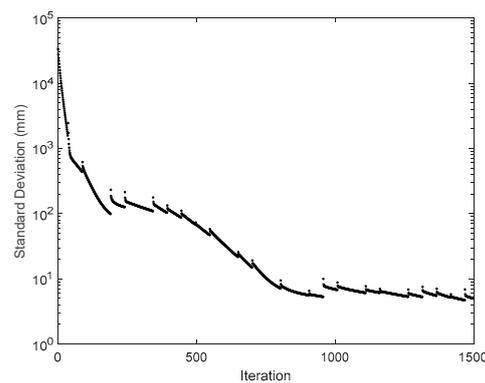


Figure 2. Standard deviation (log Y-axis)-iteration plot.

To account for noise in the high spatial frequencies present in RGB images, the weighted spectrum was incorporated for a maximum of 150 iterations in the initial stage of the ILF. This precaution was taken to prevent errors from emerging in the subsequent high spatial frequency information. During the initial process stage, the low-pass filter was multiplied by the Fourier spectrum of the depth matrix, steering the process towards generating depth data that aligned with the structural features of the RGB images. After a maximum of 150 iterations, the depth data accurately reflected the surface's trend. The weight of the RGB spatial frequency was then removed from the iterative process to mitigate potential side effects, such as overfitting and spatial noise from the RGB images.

We further delved into the convergence processes of the ILF by calculating the L1 norm value within operation blocks. To establish a reference to the ground truth data, RGB-D data extracted from the NYU Depth V2 dataset [38] was employed. In compressed sensing, the L1 norm is widely utilized for its capacity to encourage sparse representations of signals. The optimization objective is to ascertain a sparse representation by minimizing the L1 norm of the signal while adhering to the measurement constraints. The results demonstrated a decreasing trend in the L1 norm value throughout the iteration process, indicating the convergence of the expanded depth matrix during the ILF process. The L1 norm values of the depth matrices at each epoch are presented in the results.

Figure 3a illustrates a notable decrease in the L1 norm during the iteration steps, showcasing the convergence of depth matrices throughout the Iterative Low-Pass Filter (ILF) process. Meanwhile, Figure 3b presents the error between the depth matrix and the ground truth. The results indicate that, during the ILF process, depth matrices converge and are consistently updated with reference to residual data. It is worth noting the presence of several glitches in Figure 3a, which were attributed to the updated low-pass filters.

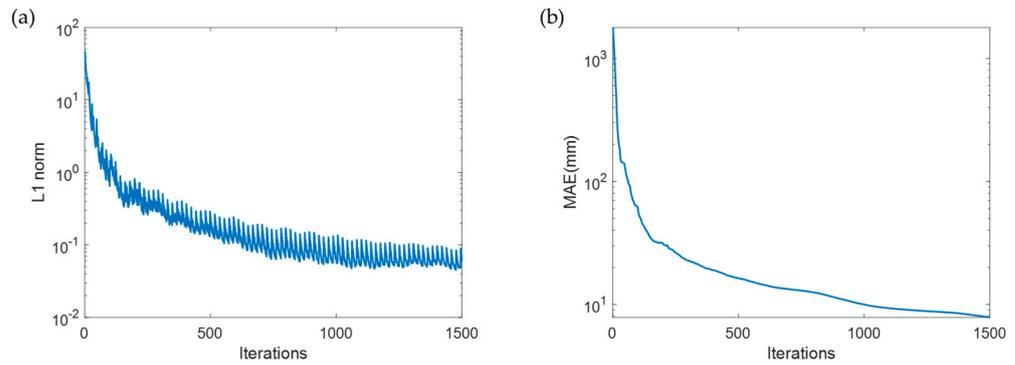


Figure 3. (a) L1 norm & Iterations analysis. (b) MAE & Iterations analysis.

When the ILF transitions to an updated low-pass filter, the L1 norm value experiences a temporary rise. This increase is a consequence of the updated low-pass filter providing additional frequency information due to its larger radius. As the new data, influenced by the updated low-pass filter, is incorporated, the L1 norm value initially rises and subsequently declines to a lower value in the subsequent epochs. This dynamic can be explained by the fact that the updated low-pass filter redirected the ILF towards the next optimization goal, thereby preventing local extrema that might lead to the ILF process being prematurely halted due to satisfaction of the stop criteria. Consequently, the ILF was adept at updating the depth matrix with appropriate values.

In Figure 3b, the error consistently decreased with each iteration. Consequently, when the input raw depth matrix underwent processing with the ILF, the depth matrix converged, and was updated with data that closely aligned with the ground truth.

3. Experimental Results

In this section, we present a comprehensive overview of the experiments conducted to validate the proposed depth-completion methodology. Employing Kinect cameras, specifically Kinect for Xbox 360 and Kinect for Xbox One, RGB and depth images were captured to assess the approach's performance under diverse conditions. Notably, Kinect for Xbox 360 utilizes the structured light technique, while Kinect for Xbox One utilizes the time-of-flight technique.

This study investigated the maximum defective area rate, estimated the continuous defective area, and evaluated the denoising capability of the Depth Image Completion Process (DICP) for a real-scene depth camera. Through a series of tests with a real-scene captured using Kinect for Xbox One, the achieved precision is discussed concerning the extent of the defective area. Additionally, we ascertain the maximum denoising ability and evaluate the overall performance of our method. The denoising ability test simulates common noise encountered in RGB-D cameras, creating rigorous noise conditions by manually removing continuous depth data and introducing normally distributed noise.

For a comprehensive assessment of the results, this study calculated the accuracy and precision of each depth matrix. Accuracy error (ACC) and precision error (PRS) were defined and employed for quantitative evaluation, with the definitions as follows [39]:

$$ACC = \frac{1}{N} \sum P_{fit_i} - \frac{1}{N} P_{raw_i}, \quad (3)$$

$$PRS = \sqrt{\frac{1}{N} \left[\sum (P_{fit_i} - P_{raw_i}) - \left(\frac{1}{N} \sum P_{fit_i} - \frac{1}{N} P_{raw_i} \right) \right]^2}, \quad (4)$$

where P_{fit_i} is the depth matrix generated from the DICP; P_{raw_i} is the raw depth matrix; and N is the total number of pixels inside the operation block.

We also present benchmark results, comparing our method with advanced approaches [24,28], utilizing the NYU Depth V2 dataset [39]. The evaluation metrics included the Root Mean

Square Error (RMSE), Mean Absolute Error (MAE), and Absolute Relative Error (Rel), following established methodologies [28,39,40]. The definitions for these evaluation metrics are provided below:

$$RMSE = \sqrt{\frac{1}{N} \sum (P_{fit_i} - P_{raw_i})^2}, \tag{5}$$

$$MAE = \frac{1}{N} \sum |P_{fit_i} - P_{raw_i}|, \tag{6}$$

$$Rel = \frac{1}{N} \sum \left| \frac{P_{fit_i} - P_{raw_i}}{P_{raw_i}} \right| \tag{7}$$

3.1. Rate of the Maximum Defected Area

In this section, we assess the maximum continuous defected area and denoising capability of the Depth Image Completion Process (DICP) for a real-scene depth camera. Figure 4 illustrates a real-scene cylinder captured as a raw depth matrix using the Kinect for Xbox One at an observation distance of 1.5 m. The detected depth data inherently carry noise, distributed normally with a standard deviation of 1.5 mm. Subsequently, we systematically removed depth data and applied the DICP. By progressively increasing the continuous defected area along the vertical direction, we created different scenarios. Figure 4c presents the accuracy error and precision error corresponding to various defected areas. Notably, when the defected area constituted 80.22% of the cylinder’s detection area, the precision was 10.28 mm.

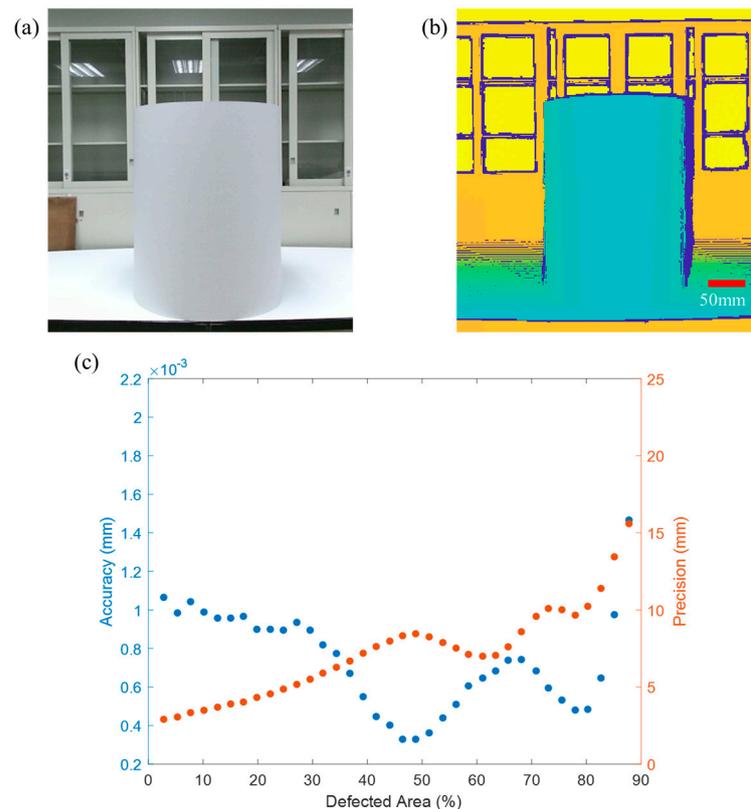


Figure 4. (a) Cylinder RGB Image. (b) Cylinder depth Image. (c) Defect area accuracy and precision analysis.

3.2. Denoising Ability

Figure 5c depicts the creation of a spherical depth image designed to assess the denoising capability of the Depth Image Completion Process (DICP). The dimensions of the depth image were 2.6×1.2 m at a distance of 2 m, with a matrix size of 540×960 pixels. Deliber-

ately introducing high noise, distributed within the range of a 65 mm standard deviation, we systematically increased the continuous defected area along the x -axis. Subsequently, the DICP was executed to evaluate its performance.

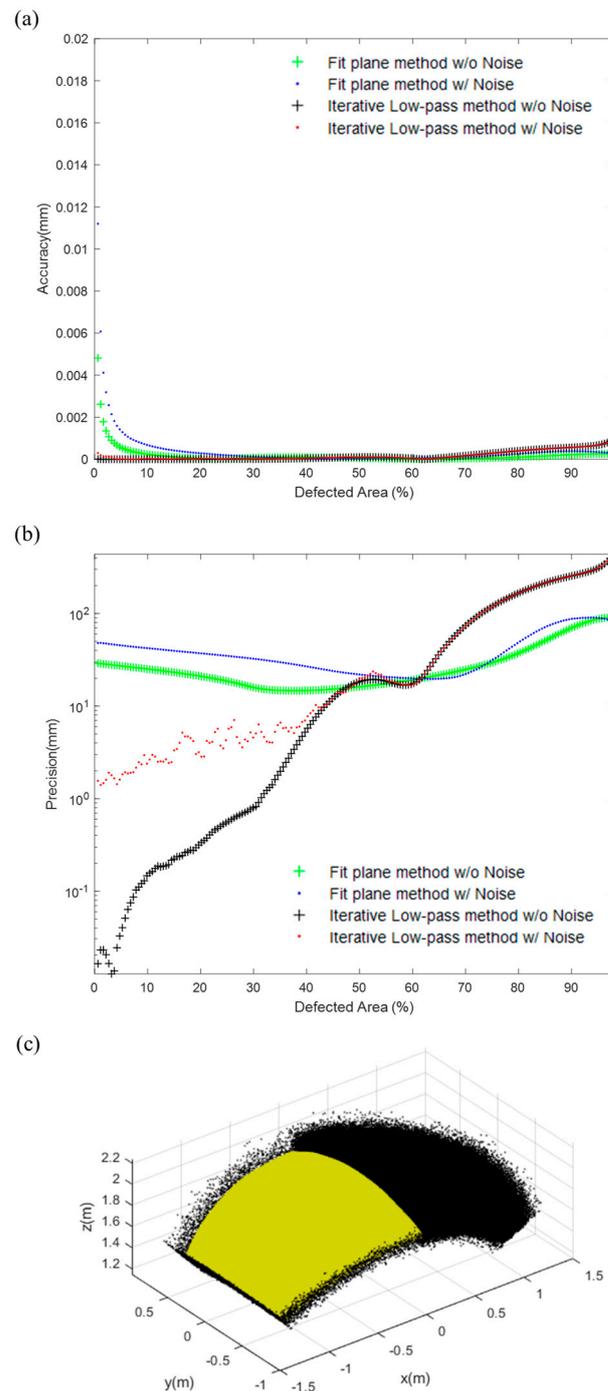


Figure 5. Analysis of a spherical depth image with a 65 mm standard deviation of noise. (a) Accuracy error for the sphere depth image with and without noise. (b) Precision for the sphere depth image with and without noise. (c) Point cloud of a spherical depth image with a 51% defected area and a 65 mm standard deviation of noise.

We also conducted a comparative analysis between the Depth Image Completion Process (DICP) and polynomial surface fitting methods [41,42]. The polynomial surface fitting method considers available inlier data, iteratively seeking the best-fitting surface

to fill the defected area. Figure 5a,b presents the accuracy error and precision error of the DICP in the presence of added high-interruption noise.

The results indicate that the introduced noise did not significantly impact the completion result, as evidenced by the calculated accuracy error. Furthermore, the precision error remained relatively stable even when the defected area exceeded 30%. Notably, the DICP outperformed polynomial surface fitting in the 0–42.77% defected area range, suggesting that noise had a more pronounced effect on the polynomial surface fitting recovery process.

Figure 6 illustrates the accuracy error and precision error resulting from noise simulation. The added noise was sourced from a Kinect for Xbox One (Kinect 2.0), measured with a standard deviation of 6.5 mm at a distance of 5 m. The precision results demonstrate that our method surpassed polynomial surface fitting in the 0–46.88% defected area range. These simulations affirm that the Iterative Low-Pass Filter (ILF) can effectively recover spherical surfaces within a standard deviation of 6.5 mm and a 65 mm range of noise. This capability positions ILF as a versatile solution for addressing noisy depth images captured by commercial-grade RGB-D cameras.

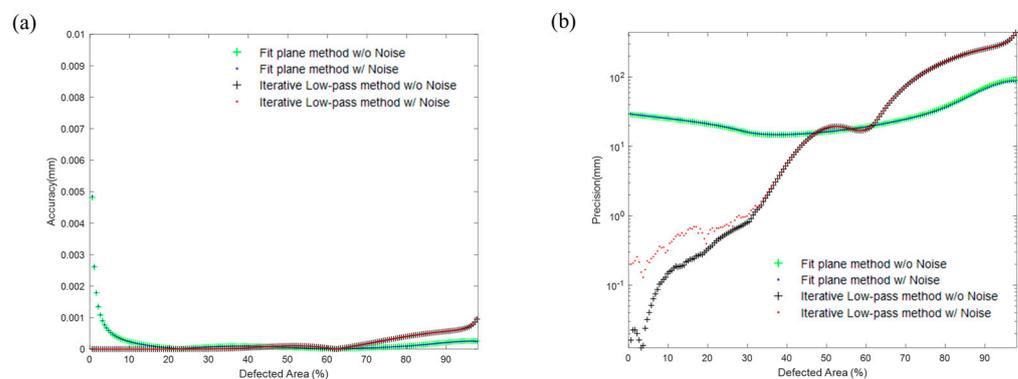


Figure 6. Analysis of a cylinder with a 6.5 mm standard deviation of noise. (a) Accuracy for the cylinder depth image with and without noise. (b) Precision for the cylinder depth image with and without noise.

3.3. Real-Scene Results

In this section, we showcase completion depth images obtained from a Kinect for Xbox 360 (Kinect 1.0) and a Kinect 2.0. Kinect 1.0 employs IR structured light to acquire depth images [11].

Figure 7a–c showcases the completion results obtained from the Kinect 1.0. In Figure 7a, captured in an indoor setting, we deliberately introduced disruption by placing a halogen lamp to interfere with the infrared structured light. The halogen lamp, visible as overexposed in the infrared image, caused signal loss, creating a hole with a diameter of 15 cm in the depth image. This resulted in a 9.46% loss in the operation block. Figure 7b presents the completion result, and Figure 7c displays the point cloud of the completion result, with yellow points representing accurate completion alignment.

Moving on to Figure 7d–f, we present the completion results from the Kinect 2.0, which utilizes the Infrared Time-of-Flight (IR ToF) technique for depth image acquisition. Capturing both RGB and depth images under sunlight illumination and on a glossy granite floor tile, which disrupts the Kinect 2.0 IR signal, presents challenges. However, with proper segmentation of the operation blocks, our method proved effective in recovering highly sparse data. It is important to note that the experiments considered a defect rate below 42.77%. The point cloud results affirm the correct positioning of the recovered point cloud, demonstrating our method's capability to recover depth information accurately from both shiny surfaces and non-reflective materials in Figure 7d.

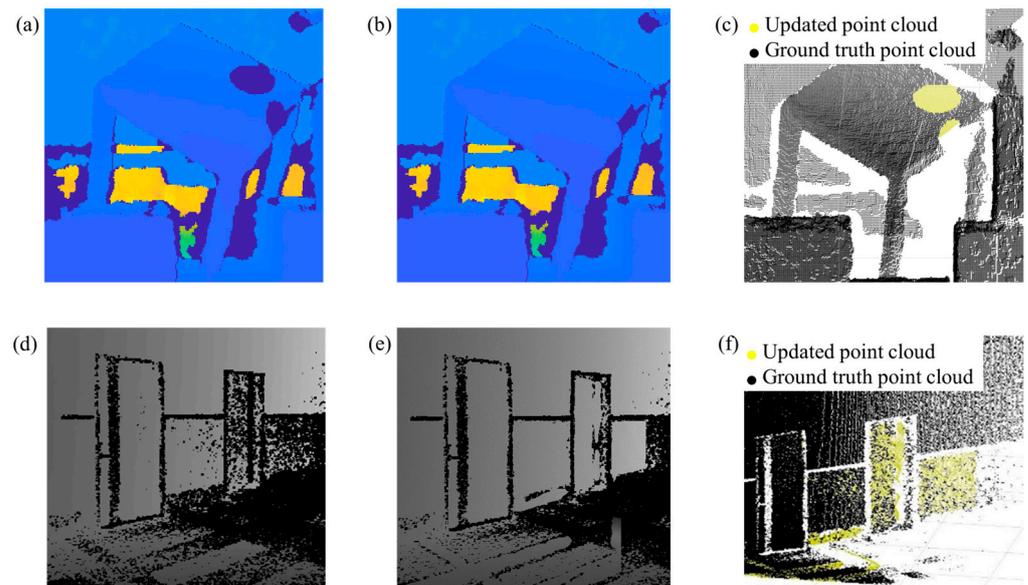


Figure 7. Real Scene completion results. (a) Input depth image for Kinect 1.0. (b) Completion depth image for Kinect 1.0. (c) Point cloud of the completion depth image for Kinect 1.0. (d) Input depth image for Kinect 2.0. (e) Completion depth image for Kinect 2.0. (f) Point cloud of the completion depth image for Kinect 2.0.

Figure 8, captured by the Kinect 1.0, illustrates the depth image completion of a random surface converted into a point cloud. The random surface, obtained from curved wallpaper illuminated by a halogen lamp, showcases the correct alignment of the completion result on the surface.

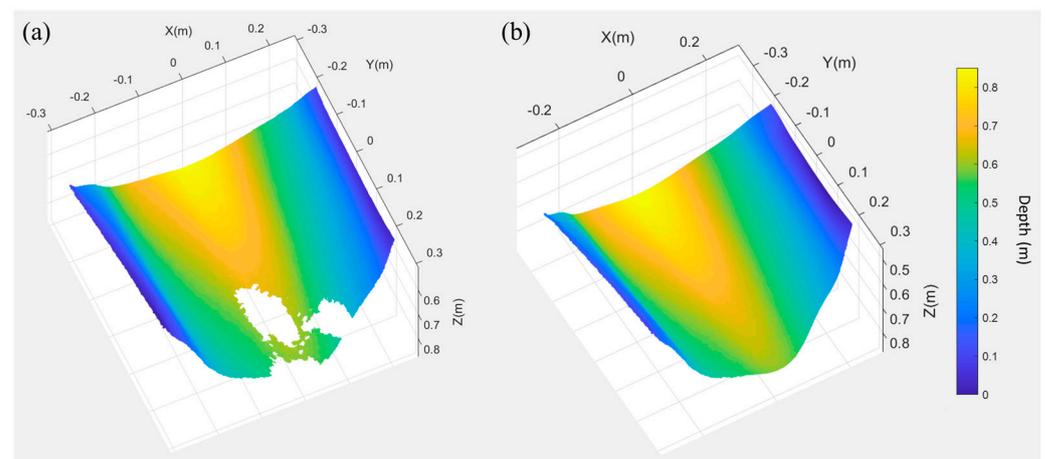


Figure 8. Curved wallpaper completion results. (a) Rendered point cloud of raw depth matrix. (b) Rendered point cloud of completion result.

3.4. Benchmark Results

In this section, we present a benchmark analysis comparing an Iterative Low-Pass Filter (ILF) with a Non-Local Spatial Propagation Network (NLSPN) [24], utilizing the NYU Depth V2 dataset [39]. The dataset comprises video sequences capturing various indoor scenes recorded by both the RGB and Depth cameras from the Microsoft Kinect. It includes labeled pairs of aligned RGB and depth images, along with ground-truth depth images. The input depth images for the benchmark were sampled at a 42.77% filling rate.

For our analysis, we selected 9409 raw depth matrices with clearly defined boundaries using the Extraction of Boundaries and Surfaces (EBS) method. These matrices include defect areas affected by infrared sources, reflecting surfaces, absorbing materials, overex-

posed scenes, and dusky scenes. To ensure robustness, each depth matrix was confined by the operation block, and the boundaries of operation blocks were carefully adjusted to avoid segmentation errors. In this comparison, ILF emerges as superior in recovering data characterized by these challenging conditions.

Figure 9 showcases three examples: a plush floor, a glossy floor, and a black reflective table. The Iterative Low-Pass Filter (ILF) excels in performance, offering superior details in these challenging areas. The highlighted scenes include reflective surfaces, absorbent furnishings, dim illumination, and overexposed RGB signals. Notably, the Non-Local Spatial Propagation Network (NLSPN) is susceptible to lighting conditions and surface properties, as it extracts features from both RGB and depth signals.

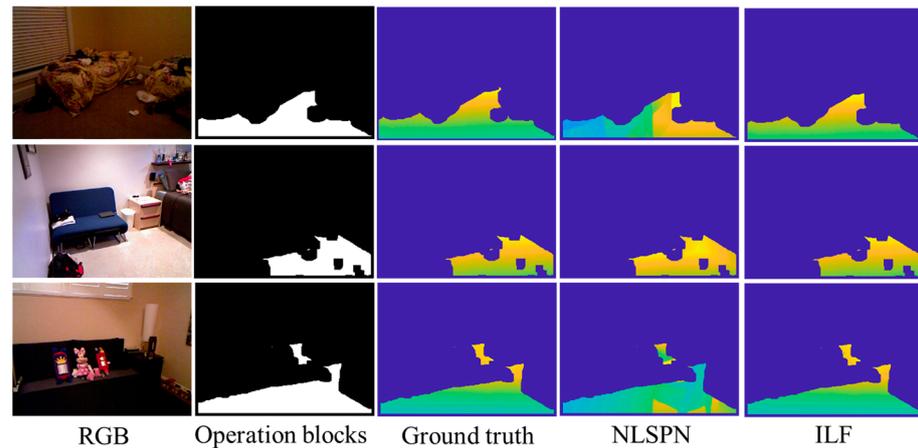


Figure 9. NYU Depth V2 dataset [39] DICI results.

In dim lighting conditions, NLSPN yields inaccurate results compared to the ground truth, as evident in the color maps. The glossy floor, reflecting the scene's RGB signal, leads to the mirroring or creation of the scene's depth in the reflective surface, resulting in higher values in quantitative evaluation. In the case of the absorbent black carpet, the system erroneously derived depth from the coffee table. In contrast, ILF demonstrated minimal effects from these specialized surfaces, as evidenced by the color maps, where ILF's results closely aligned with the ground truth.

The depth-image completion results of the 9409 raw depth matrices were evaluated using the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Absolute Relative Error (Rel), presented in Table 1. In direct comparisons, ILF consistently achieved smaller values in RMSE, MAE, and Rel, underscoring its superior performance.

Table 1. Quantitative evaluation on the NYU Depth V2 dataset [39].

Methods	RMSE (mm)	MAE (mm)	Rel
NLSPN [24]	1962.779	1935.296	0.701
ILF	300.224	173.014	0.008

4. Conclusions

This study introduced a Depth Image Completion Process method employing an Iterative Low-Pass Filter (ILF). Grounded in spatial frequency modulation and guided by RGB information, our approach leverages RGB images to predict depth data. RGB images play a crucial role in providing an initial estimate of the target surface structure. These initial estimations, coupled with frequency domain weights, significantly expedite the iterative process. The depth images underwent multiple low-pass filter iterations, gradually shifting the cut-off frequency from low to high. Filtered depth images outside the defected area were updated with a raw depth matrix in each iteration. The low-pass filter imparted a smoothing effect on the boundaries of the defected area. After a series of

iterations, the frequency domain variation converged, resulting in the recovery of depth data within the defected area. L1 norm analysis indicated the optimization characteristics of the ILF, making it applicable in depth completion.

Additionally, a benchmarking analysis comparing the ILF method with the Non-Local Spatial Propagation Network (NLSPN) using the NYU Depth V2 dataset was conducted. Depth errors arise when inconsistency occurs between RGB data and depth information, often stemming from issues such as reflections in the RGB representation, overexposed regions, or scenes with low illumination. Our evaluation consistently revealed that the ILF method outperforms the NLSPN in key metrics such as the Root Mean Square Error, Mean Absolute Error, and Relative Error. This underscores the superior performance of the ILF approach in-depth image completion.

Author Contributions: T.-K.W. wrote the paper and mainly developed the techniques used, including the completion algorithm, system setup, and experiments. Y.-W.Y. and T.-H.Y. participated in the discussion of the study and contributed to the principle of the iterative low-pass filter and the development of the completion algorithm. P.-D.H. contributed to the acquisition of depth images. G.-Y.Z. experimentally demonstrated the benchmarks of the advanced completion methods. C.-C.L. participated in the development of the completion algorithm. C.-C.S. served as the team leader and submitted the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the National Science and Technology Council of Taiwan, grant numbers 111-2218-E-008-004-MBK, 108-2221-E-008-097-MY3, 108-2221-E-008-084-MY3, and 111-2221-E-008-100.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The benchmark datasets we used are available from the NYU Depth dataset V2 website: https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html. The NLSPN code is available at: https://github.com/zzangjinsun/NLSPN_ECCV20. The ILF code and the benchmark result are available at https://github.com/Whachudoing/Depth-image-completion-using-iterative-low-pass-filter_Benchmark.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Geiger, A.; Ziegler, J.; Stiller, C. StereoScan: Dense 3d reconstruction in real-time. In Proceedings of the 2011 IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, Germany, 5–9 June 2011.
- Rusu, R.B.; Marton, Z.C.; Blodow, N.; Dolha, M.; Beetz, M. Towards 3D point cloud-based object maps for household environments. *Robot. Auton. Syst.* **2008**, *56*, 927–941. [\[CrossRef\]](#)
- Marapane, S.B.; Trivedi, M.M. Region-based stereo analysis for robotic applications. *IEEE Trans. Syst. Man Cybern.* **1989**, *19*, 1447–1464. [\[CrossRef\]](#)
- Nalpantidis, L.; Gasteratos, A. Stereo vision for robotic applications in the presence of non-ideal lighting conditions. *Image Vis. Comput* **2010**, *28*, 940–951. [\[CrossRef\]](#)
- Murray, D.; Little, J.J. Using real-time stereo vision for mobile robot navigation. *Auton. Robots* **2000**, *8*, 161–171. [\[CrossRef\]](#)
- Huang, Y.K.; Liu, Y.C.; Wu, T.H.; Su, H.T.; Hsu, W.H. Expanding Sparse Guidance for Stereo Matching. *arXiv* **2020**, arXiv:2005.02123.
- Shaked, A.; Wolf, L. Improved stereo matching with constant highway networks and reflective confidence learning. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
- Saito, Y.; Yamamoto, Y.; Kan, T.; Tsukagoshi, T.; Noda, K.; Shimoyama, I. Electrical detection SPR sensor with grating coupled backside illumination. *Opt. Express* **2019**, *27*, 17763–17770. [\[CrossRef\]](#) [\[PubMed\]](#)
- Dalla Betta, G.F.; Donati, S.; Hossain, Q.D.; Martini, G.; Pancheri, L.; Saguatti, D. Design and Characterization of Current-Assisted Photonic demodulators. *IEEE Trans. Electron Devices* **2011**, *58*, 1702–1709. [\[CrossRef\]](#)
- Khoshelham, K.; Elberink, S.O. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors* **2012**, *12*, 1437–1454. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mineo, C.; Cerniglia, D.; Ricotta, V.; Reitingner, B. Autonomous 3D geometry reconstruction through robot-manipulated optical sensors. *Int. J. Adv. Manuf. Technol.* **2021**, *116*, 1895–1911. [\[CrossRef\]](#)
- Mallick, T.; Das, P.P.; Majumdar, A.K. Characterizations of noise in Kinect depth images. *IEEE Sens. J.* **2014**, *6*, 1731–1740. [\[CrossRef\]](#)

13. Matsuo, K.; Aoki, Y. Depth image enhancement using local tangent plane approximations. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
14. Doria, D.; Radke, R.J. Filling large holes in lidar data by inpainting depth gradients. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops), Providence, RI, USA, 16–21 June 2012.
15. Zhang, F.; Prisacariu, V.; Yang, R.; Torr, P.H. Ga-net: Guided aggregation net for end-to-end stereo matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.
16. Park, S.; Guo, X.; Shin, H.; Qin, H. Shape and appearance repair for incomplete point surfaces. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 17–21 October 2005.
17. Kochanowski, M.; Jenke, P.; Straßer, W. Analysis of texture synthesis algorithms with respect to usage for Hole-Filling in 3D geometry. In Proceedings of the ITCS, Beijing, China, 12–17 October 2008.
18. Hanocka, R.; Metzger, G.; Giryas, R.; Cohen-Or, D. Point2Mesh: A Self-Prior for Deformable Meshes. *arXiv* **2020**, arXiv:2005.11084. [[CrossRef](#)]
19. Watson, J.; Mac Aodha, O.; Prisacariu, V.; Brostow, G.; Firman, M. The temporal opportunist: Self-supervised multi-frame monocular depth. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021.
20. Eigen, D.; Puhersch, C.; Fergus, R. Depth map prediction from a single image using a multi-scale deep network. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2366–2374.
21. Godard, C.; Mac Aodha, O.; Brostow, G.J. Unsupervised monocular depth estimation with left-right consistency. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
22. Hu, M.; Wang, S.; Li, B.; Ning, S.; Fan, L.; Gong, X. Towards precise and efficient image guided depth completion. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021.
23. Casser, V.; Pirk, S.; Mahjourian, R.; Angelova, A. Unsupervised monocular depth and ego-motion learning with structure and semantics. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 15–20 June 2019.
24. Park, J.; Joo, K.; Hu, Z.; Liu, C.K.; So Kweon, I. Non-local spatial propagation network for depth completion. In Proceedings of the European Computer Vision–ECCV 2020: 16th European Conference, Part XIII 16, Glasgow, UK, 23–28 August 2020.
25. Zhang, Y.; Funkhouser, T. Deep depth completion of a single rgb-d image. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
26. Van Gansbeke, W.; Neven, D.; De Brabandere, B.; Van Gool, L. Sparse and noisy lidar completion with rgb guidance and uncertainty. In Proceedings of the 2019 16th International Conference on Machine Vision Applications (MVA), Tokyo, Japan, 27–31 May 2019.
27. Izadi, S.; Kim, D.; Hilliges, O.; Molyneaux, D.; Newcombe, R.; Kohli, P. KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera. In Proceedings of the UIST 2011—24th Annual ACM Symposium on User Interface Software and Technology, Santa Barbara, CA, USA, 16–19 October 2019.
28. Cheng, X.; Wang, P.; Yang, R. Depth estimation via affinity learned with convolutional spatial propagation network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
29. Liu, S.; De Mello, S.; Gu, J.; Zhong, G.; Yang, M.H.; Kautz, J. Learning affinity via spatial propagation networks. *Adv. Neural Inf. Process Syst.* **2017**, *30*.
30. Cheng, X.; Wang, P.; Guan, C.; Yang, R. Cspn++: Learning context and resource aware convolutional spatial propagation networks for depth completion. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, New York, USA, 7–12 February 2020.
31. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *6*, 679–698. [[CrossRef](#)]
32. Xie, S.; Tu, Z. Holistically-nested edge detection. In Proceedings of the IEEE International Conference on Computer vision, Santiago, Chile, 7–13 December 2015.
33. Ganin, Y.; Lempitsky, V. -Fields: Neural network nearest neighbor fields for image transforms. In *Asian Conference on Computer Vision*; Springer International Publishing: Cham, Switzerland, 2014.
34. Bertasius, G.; Shi, J.; Torresani, L. Deepedge: A multi-scale bifurcated deep network for top-down contour detection. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
35. Hwang, J.J.; Liu, T.L. Pixel-wise deep learning for contour detection. *arXiv* **2015**, arXiv:1504.01989.
36. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
37. Martin, D.R.; Fowlkes, C.C.; Malik, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 530–540. [[CrossRef](#)] [[PubMed](#)]
38. Smith, S.W. *The Scientist and Engineer's Guide to Digital Signal Processing*; California Technical Pub.: San Diego, CA, USA, 1997; pp. 32–34.
39. Silberman, N.; Hoiem, D.; Kohli, P.; Fergus, R. Indoor segmentation and support inference from rgb-d images. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012.
40. Ma, F.; Karaman, S. Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018.

41. Franke, R. Scattered data interpolation: Tests of some methods. *Math. Comput.* **1982**, *38*, 181–200.
42. Franke, R.; Nielson, G. Smooth interpolation of large sets of scattered data. *Int. J. Numer. Methods Eng.* **1980**, *15*, 1691–1704. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.