A comparative review of plausible hole filling strategies in the context of scene depth image completion

Amir Atapour-Abarghouei[a,*], Toby P. Breckon[a]

[a]*Computer Science and Engineering, Durham University, Durham, UK*

ARTICLE INFO

Keywords: Depth Filling, Image Completion, RGB-D Completion, Surface Relief, Hole Filling, Surface Completion, 3D Texture, Depth Completion, Inpainting, Disparity Filling

ABSTRACT

Despite significant research focus on 3D scene capture systems, numerous unresolved challenges remain in relation to achieving full coverage scene depth estimation which is the key part of any modern 3D sensing system. This has created an area of research where the goal is to complete the missing 3D information post capture via a secondary depth filling process. In many downstream applications, an incomplete depth scene is of limited value, requiring many special cases for subsequent utilization, and thus techniques are required to "fill the holes" that exist in terms of both missing depth and color scene information. An analogous problem exists within the scope of scene filling post object removal in the same context. Although considerable research has resulted in notable progress in the synthetic expansion or reconstruction of missing color scene information in both statistical (texture synthesis) and structural (image completion) forms, work on the plausible completion of missing scene depth is contrastingly limited. This survey aims to provide a state of the art overview within this growing field of depth synthesis work whilst noting related solutions in the space of traditional texture synthesis and color image completion for hole filling. To these ends, we concentrate on the plausible completion of both underlying depth structure and relief texture to provide both greater understanding and future development in the area. Our analyses are in part supported by illustrative experimental examples of the comparative use of a subset of representative approaches over common depth completion examples.

1. Introduction

Three dimensional scene sensing is gaining an ever-increasing applicability and importance due to its wide-spread uses in real-world scenarios, including areas such as interactive entertainment, future vehicle autonomy, environment modeling, security surveillance, and future manufacturing in technologies. Despite extensive work on 3D sensing of late [1, 2, 3, 4, 5], a number of limitations pertaining to environmental conditions, inter-object occlusion, and sensor capabilities constrain fully-effective scene depth capture [1, 6]. As a result, significant research has been focused specifically on developing techniques to complete missing scene depth to increase the quality of the depth information for better applicability. Although many have attempted to use traditional texture synthesis and structural image completion techniques, (in whole or in part) to address the problem of scene depth completion, challenges remain in terms of efficiency, depth continuity, surface relief, and local feature preservation that have hindered flawless operation against high expectations of plausibility [7, 8, 9, 10, 11, 12]. This work aims to present a review of prior work in the domain focusing on current state of the art capabilities, shortcomings, and future

[*] Corresponding author.
e-mail: amir.atapour-abarghouei@durham.ac.uk (Amir Atapour-Abarghouei)

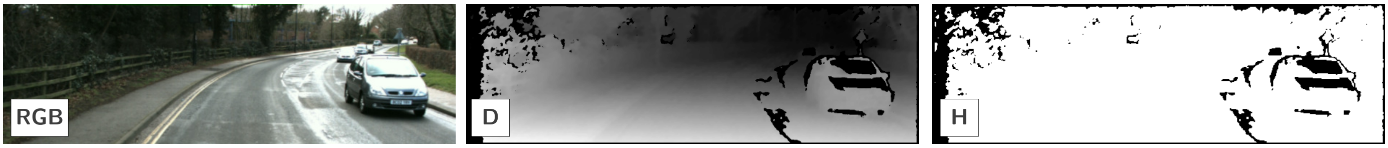

Fig. 1: Example of depth acquired via stereo correspondence in an urban driving scenario. Note the missing depth values despite accurate camera calibration.

challenges. Although the main focus of this study is on scene depth completion and hole filling, a summary of the most influential approaches within color image completion and synthesis are additionally presented to support this agenda.

In this vein, in Section 2, we present a short overview of the commonly-used approaches for capturing depth in an inexpensive widely accessible manner. Section 3 will provide a short description of a number of most relevant state-of-the-art color image completion techniques and Section 4 a taxonomy of recent advances in scene depth completion covering aspects of problem formulation, spatial consistency, temporal continuity and others. As appropriate, these sections are supported by comparative experimental results over common data examples (Figures 19 and 26; Table 2). Finally, we conclude with a summary of current themes, remaining limitations, and potential avenues for future investigation.

2. Depth Acquisition

While high-end depth sensing technologies, including light field cameras and LIDAR, exist that are capable of capturing accurate scene depth with relatively fewer anomalies (missing or invalid depth, undesirable artefacts, and depth inhomogeneity) compared to consumer devices, they remain expensive and difficult to operate in terms of size, weight, and power. As a result, both industry and academia have gravitated toward more easily accessible technologies such as stereo correspondence [13, 14, 15], structured light [16, 17, 18, 2], and time-of-flight cameras [19, 20, 21].

Stereo imaging as a passive scene acquisition method has long been used as a reliable source of depth sensing, but not without certain issues. Although stereo correspondence is better equipped than structured light and time-of-flight cameras to estimate depth where highly granular texture is present compared, smoothing still occurs. Additionally slightest miscalibration or issues in the setup and synchronization can lead to invalid or missing depth information. Moreover, missing depth (holes) are often observed in the scene depth where absence of camera overlap, featureless surfaces, sparse information for a scene object such as shrubbery, unclear object boundaries, very distant objects, and alike are present. Such issues can be seen in Figure 1, where "*RGB*" denotes the left color image, "*D*" the estimated depth, and "*H*" is a binary mask marking where depth holes are (in black).

Structured light devices and time-of-flight cameras are active range sensors, and while they can suffer from mis-calibration issues, they are more-widely utilized for a variety of purposes due to their low-cost availability in the commercial market with factory calibration settings [22, 23, 24, 25].

Despite this, structured light sensors are subject to a wide range of issues including but not limited to over-saturation due to ambient light [22], external active illumination source interference [23, 24], active light path error caused by reflective surfaces, occlusion, fronto-parallel angle of the object to the sensor [25, 26], erroneous light pattern detection in dynamic scenes [25], and others.

Similarly, time-of-flight cameras have their own flaws that lead to invalid or missing depth, noise, and other additional artefacts, such as depth error caused by light scattering or semi-transparent surfaces [27, 28], external illumination interference [29], depth offset for non-reflective objects [30], and alike [25].

It must be noted that not all of such issues will lead to missing depth information (holes), but invalid depth and noise are essentially detriments in practice and are best handled through removal and subsequent filling. Figure 2 depicts examples of depth images obtained using a structured light camera (left; "RGB_1" denotes the color image, "D_1" the depth, and "H_1" a binary mask depicting the location of depth holes in black) and a time-of-flight camera (right; "RGB_2" denotes the color image,

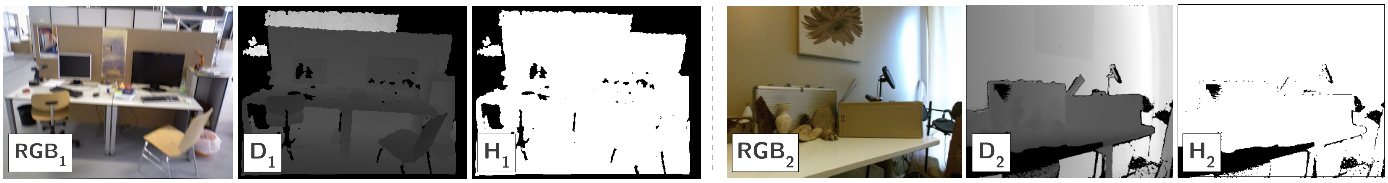

Fig. 2: Examples of depth acquired via a structured light device (left) and a time-of-flight camera (right). Depth is missing in images captured using both devices.

"D_2" the depth, and "H_2" a binary mask depicting the location of depth holes in black).

While many 3D computer vision applications continue to move forward as they cope with the issues caused by depth holes, performance can be improved in many respects if accurate hole-free depth information is readily available for processing, hence the creation of the entire literature on depth completion. In this study, we aim to encapsulate the essence of the research conducted on this subject matter to provide a better understanding of the approaches so future researchers may benefit by choosing the right technique for their purposes. Furthermore, we aim to bring together a sample of illustrative experimental results to both support current performance trends and identified future research directions.

Since depth information is represented and processed in the form of images, and many researchers still apply more classical color image completion methods to depth maps, a brief overview of image completion within the context of scene color maps (RGB) can be beneficial for a better understanding of the many-facet subject of depth filling.

3. Image Completion

A long-standing and analogous challenge to depth filling problem has been to complete a color image after a selected object or region is removed or alternatively to create a plausible synthesis of the image over a larger spatial area. As there already exists an extensive literature on the subject, in this section, we only focus on methods that have been or can potentially be used for depth filling.

Early color image inpainting techniques (focusing on the geometry of the shapes), attempted to smoothly propagate the isophotes (lines where the intensity value is the same) into the target area that is to be inpainted. However, most structure-based inpainting approaches overlook one of the most important image components which plays a significant role in what the observer senses as reality: high fidelity (spatial frequency) texture. As a result, many subsequent inpainting techniques began to incorporate ideas from the field of texture synthesis (in which the objective is to generate a large texture region given a smaller sample of texture without visible artefacts of repetition within the larger region [31, 32, 33, 34, 35]) into their inpainting techniques [36, 37, 38], which resulted in more plausible better quality outputs (exemplar-based image completion).

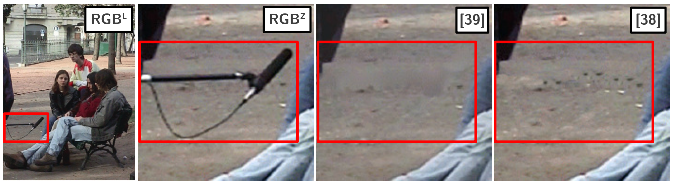

Fig. 3: Results of [38] compared to [39]. The foreground microphone has been removed and inpainted, but compared to [38], the texture in the result of [39] is not accurate, leading to a perception of blurring (reproduced from [38]).

With their focus on structure rather than texture, Bertalmio et al. [39] attempt to solve the problem of inpainting in a pioneering work using higher order partial differential equations and anisotropic diffusion to propagate pixel values along isophote directions (Figure 3). After consulting various experts on scene composition in the artistic sense, they created a general set of inpainting principles that have henceforth become widely-used standard guidelines for how inpainting algorithms should function, which remain highly relevant even in depth completion cases:

- **Rule 1:** after the inpainting process is completed, the inpainted target region must be consistent with the known region of the image to preserve global continuity.
- **Rule 2:** the structures present within the known region must be propagated and linked into the target regions.
- **Rule 3:** the structures formed within the target region must be filled with color consistent with the known regions.

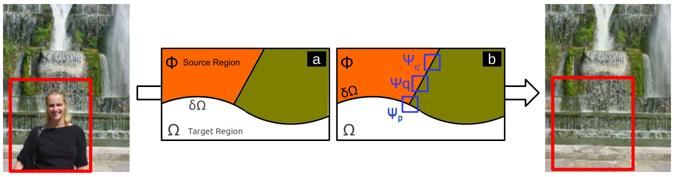

Fig. 4: The process within the framework of [36] (reproduced from [36]).

- **Rule 4:** texture must be added into the target region after or while the structures are filled.

In [39], these rules are used in an iterative approach to fill the target region and mimic the principles of generalized 3D object completion as identified in [40] with reference to the psychological literature on human visual perception. To achieve visually convincing and plausible results, these rules should also be followed within the context of depth completion, except for *rule 3*, since no color information is contained within a depth image.

While [39] works well for small areas or smooth and untextured background regions, inpainting was by no means a solved problem, and in the presence of fine texture, the approach fails to generate satisfactory results as it mainly focuses on structure, failing to preserve texture (Figure 3).

Later on, improved inpainting approaches began to emerge based on a range of techniques including fast marching method [41], Total Variational (TV) models [42, 43, 44], and exemplar-based methods that focus on "synthesizing" fine texture in the target region along with propagating structure [38, 36, 45, 46].

The notable work in [36], which is regularly used within color and depth filling ([47, 48]) followed traditional exemplar-based texture synthesis methods [32] by prioritizing the order of filling based on the strength of the gradient along the target region boundary. Although there have previously been attempts to complete images via exemplar-based synthesis [45, 46], they are all lacking in either structure propagation or defining an explicit filling order that could prevent the introduction of blurring or distortion in shapes and structures (see Figure 4). The method in [36] demonstrates that exemplar-based methods are not only well-suited for two-dimensional texture but also capable of propagating isophotes and linear structures. An example of the results of this method is seen in Figure 4, where we see that plausible water texture has been synthesized in the target region after the person is removed from the original image. However, even though the algorithm is able to deal with texture and linear structure, it cannot handle curved structures and is highly dependent on the existence of similar pixel neighborhoods in the sample region for plausible completion. Additionally, the approach highly relies on the existence of fine reflectance texture to prioritize patches and can fail when dealing with large objects in more smooth depth images (Figure 5). These issues will be discussed further is Section 4.

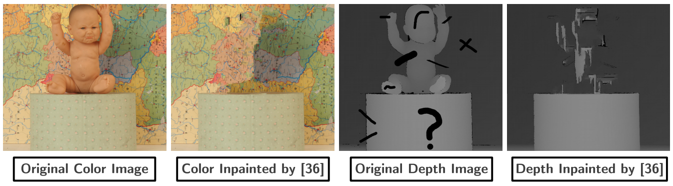

Fig. 5: An example of [36] applied to depth and color images. The goal is to remove an object (*the baby*) from both color and depth and fill the extisting holes in the depth image (represented by the black marks) at the same time.

While exemplar-based inpainting (sampling and copying patches from the known regions of the image) have been proven successful in many respects, there are limitations regarding the amount of samples available, but more importantly, performance is significantly degraded when dealing with scenes that are not of a fronto-parallel view, which can create issues such as perspective handling within the completion process. Some methods have been proposed to combat this issue [49, 50] by including the transformed version of patches in the sample search space. The transformation can include rotation, scale, gain, and bias color adjustments. Although this can solve the problem of perspective and view angle and improve the performance of the completion process, it exponentially increases the size of the search space from 2 degrees of freedom per output pixel or patch to 8 (i.e. equivalent to a homography transformation) or more (photometric variations, e.g. bias/gain of intensity channels). Not only is the efficiency and speed of the process thus affected, but also an elevated probability of taking a *local optimum* as the result can be expected.

To combat these limitations, many other image completion techniques [51, 52, 53, 54, 55, 56, 57, 37, 58, 59] have been proposed that are capable of filling large portions of an image successfully. For instance, certain methods use schemes such as

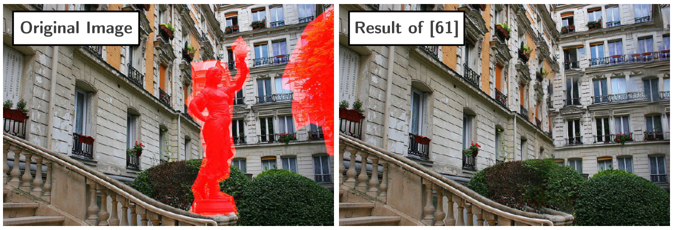

Fig. 6: Results of [61], achieved by adding the mid-level scene understanding constraints of translational regularity (via scale-invariant feature (SIFT) matching [72]) and planar perspective to guide the process (reproduced from [61]).

the reformulation of the exemplar-based inpainting problem as a metric labeling problem [51] subsequently solved using simulated annealing, energy minimization methods [52, 53, 54], Markov Random Field models with labels assigned to patches [55] using belief propagation via priority scheduling, models represented as an optimal graph labeling problem, where the shift-map (the relative shift of every pixel in the output from its source in the input) represents the selected label and solved by graph cuts [56], and the use of *Laplacian pyramids* [60] instead of the gradient operator in a patch correspondence search framework due to the advantageous qualities of Laplacian pyramids, such as isotropy, rotation invariance, and lighter computation. Image completion has also been accomplished using mid-level scene understanding constraints [61] (Figure 6), semantically similar external databases of images [62, 63] (Figure 7), and deep convolutional neural networks aided by adversarial training [64, 65, 66] to plausibly complete color images.

The aforementioned methods certainly do not represent the entirety of the image completion literature and are not the focus of this study. As such, since wide-expanding surveys already exists on the issues of texture synthesis and inpainting within the context of color images [67, 68, 69, 70, 71], we will not delve any further into the subject and only refer to techniques directly pertinent to the issue at hand, depth completion.

4. Depth Hole Filling

Compared to the significant prior work in color image completion [32, 39, 36, 37, 42, 41, 68, 69, 70, 71], more limited literature exists on the removal of objects from scene depth [73, 74, 75] and filling the naturally occurring holes in depth images [7, 8, 9, 10, 11, 12] mainly because this is a relatively new area of research with significant challenges [40]. Inpainting methods provide excellent results when it comes to filling color images, but depth images have different attributes that can affect the results when a color image inpainting method is applied to them, and as such, other new or modified techniques are required for better results.

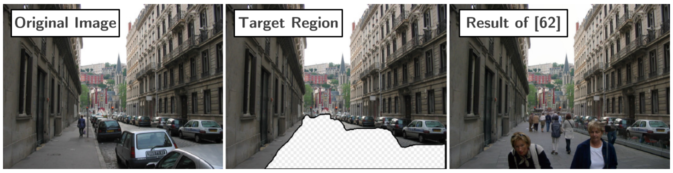

Fig. 7: Results of [62], achieved by extracting semantically similar images from a large database of photographs, and filling the target region by copying a region from a semantically valid image (reproduced from [62]).

Here, we will first focus on the different formulations of the depth inpainting problem, as numerous research works have attempted to solve this problem by concentrating on different challenges within the framework of depth filling. In later sections, we will attempt to provide a taxonomy of the current depth filling literature based on the information domain required for processing, input necessities, and the different aspects of the resulting output the completion process within individual techniques often focuses on.

4.1. Problem Formulation

Creatively reformulating an ill-posed problem such as scene depth completion and inpainting will lead to solutions that can fulfill particular required elements pertaining to certain situations, including time, computation, accuracy, and alike. In this section, we will discuss some of the most common ways in which depth filling has been posed and solved as a problem, and the effects each reformulation can have on the results.

4.1.1. Anisotropic Diffusion

Formulating the image completion and de-noising problem as anisotropic diffusion [76] has been a long-standing and successful technique in the field of color image inpainting [43, 39, 77, 59]. As such, diffusion-based solutions have also entered the realm of depth filling, since the smoothing and edge-preserving qualities of the diffusion-based depth filling output is desirable in certain downstream applications such as localization and mapping [78, 79].

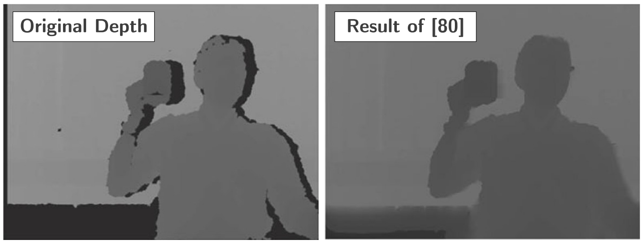

Fig. 8: The results of the method in [80]. The approach is an anisotropic diffusion-based method with real-time capabilities (reproduced from [80]).

Anisotropic diffusion is a non-linear partial differential equation scheme [76] with edge-preserving smoothing qualities. As a space-variant transformation of an input image, it generates a family of smoothed parametrized images, each of which corresponds with a filter that depends on the local statistics of the input image.

More formally put, let $I(\cdot, t)$ be a family of parametrized images, then the anisotropic diffusion is:

$$I_t = div(c(x, y, t)\nabla I) = c(x, y, t)\Delta I = \nabla c \cdot \nabla I, \qquad (1)$$

where div is the divergence operator, ∇ and Δ denote the gradient and Laplacian operators respectively, and $c(x, y, t)$ is the diffusion coefficient, which can be a constant or a function of the image gradient.

Equation 1 can be discretized using a 4-neighborhood scheme, as in [80] where the color image is used to guide the diffusion in an iterative process. In this approach [80], the depth image is completed at a low resolution, and the ensuing iterative color-guided anisotropic diffusion within the upsampling steps corrects the depth image (see an example of the results of the approach [80] in Figure 8).

Another example of the use of diffusion in depth completion can be seen in [81]. The approach attempts to fill depth holes by extracting the edges from the accompanying color image captured from a structured-light device. Subsequently, different diffusion algorithms are applied to smooth and edge regions. The separation of these regions before the diffusion process is performed based on their observation that surfaces which need to be smooth in the depth may be textured in the color image, and object boundaries within the depth image can be missed during the color edge extraction process due to the low contrast in the color image.

Using diffusion methods, the resulting completed depth image can be smooth in the presence of flat planes with sharp edges. While smooth surfaces and strong edges and object boundaries can be very desirable traits in a depth image, the implementation requires discretization and will bring forth numerical stability issues and is computationally expensive. The longer run-time of diffusion-based methods make them intractable for real-time requirements within applications.

4.1.2. Energy Minimization

Following the successes of energy minimization used within the color image completion framework [52, 53, 54], the technique has been used in various depth filling approaches.

The foundations of an energy minimization approach stem from certain assumptions made about the color and/or depth image, based on which an energy function is designed. The function is subsequently optimized, completing and enhancing the original image based on the criteria set within the different terms added to said energy function.

Depth filling approaches using energy minimization are mostly accurate and produce plausible results, but more importantly the capability of these approaches to focus on specific features within the image based on the terms added to the energy function is highly advantageous.

For instance, the energy function in [82] incorporates the characteristics of a depth image acquired via a structured light device (Kinect) into the filling process. The noise model of the capture device and structure information are taken into account using terms added to the energy function, performing the regularization during the minimization process.

The approach in [83] assumes a linear correlation between depth and color values within small local neighborhoods. An additional regularization term based on [84] enforces sparsity in vertical and horizontal gradients of the depth image, resulting in more crisp object boundaries with less noise (Figure 23). The work of [85] includes a data term that favors pixels surrounding hole boundaries and a smoothing prior that encourage flat and smooth surfaces within the depth image. This is very advantageous in terms of geometry and structure of the scene following

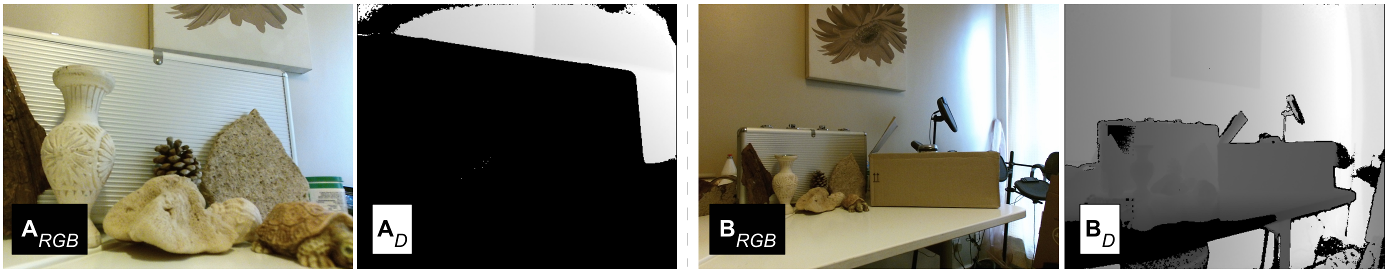

Fig. 9: Examples representing challenges involving depth textures (captured using Microsoft Kinect v2). When captured from close proximity, depth values of highly textured objects are missing due to the short camera distance (left). The same objects captured from a distance are smooth with little granular relief (right).

the design of the energy function, even though important information such as relief and texture is lost in the output.

Designing an energy function based on the characteristics of the input and the requirements of the output can be very beneficial, as the function can be modified or regularized to produce desirable outputs based on the specific application of the resulting depth image. However, the optimization process can come with implementation difficulties, numerical instabilities, and computationally intensive necessities.

4.1.3. *Exemplar-based Filling*

One of the most important challenges involving the depth filling problem is related to texture, which not unlike the related work in color image completion can be solved by copying and pasting textured patches from the known regions of the image (exemplar-based image completion). However, there can be major pitfalls with using an exemplar-based technique used for color image completion for a depth image.

While texture and relief are very important in many modern computer vision tasks, most active depth sensing devices do not cope well with texture, which is why missing and invalid depth values in close views of highly textured regions is commonplace. In Figure 9 (A), we can see that an attempt to capture the depth of a highly textured object from a close distance fails, due to the fact that most active depth sensors (in this case a time-of-flight camera) cannot deal with objects *too close to the camera*. On the other hand, if more distance is allowed between the camera and the textured object to resolve this issue, the resulting depth image is more smooth and shape-based than the equivalent color image (Figure 9 - B). We can see in Figure 9(A) that the color (RGB) image contains highly textured objects close to the capture device, while all the depth values in these objects are missing in the depth (D) image, because of the close proximity of the camera to the objects in the scene. Figure 9(B) indicates that the same objects captured from a distance, while still visibly textured in the color image (representing the human perception of relief), are far smoother in the depth image. Passively obtained depth is normally better textured than depth information acquired through active devices, but the amount of captured texture and fine relief is still not comparable to the relief perceived from the scene by a human observer (Figure 1).

Furthermore, simply assuming that a depth image is just a gray-scale image with no texture [86] is a significant oversight and ignores the many potentials an accurate and textured depth image can have.

Even though there have been many attempts to directly use structure-based or exemplar-based *color image completion* approaches for depth hole filling [47, 48, 41], particular factors create obstacles. As mentioned earlier, a depth image is not as visibly textured as a color image of the same scene. Therefore, when a structural inpainting technique is being used to propagate the shapes and structures into the target regions, identifying the points at which the propagation must be terminated is challenging. There is little texture present, and in many cases object boundaries lie within or adjacent to the holes, which makes detecting them extremely challenging.

Formulating depth filling as an exemplar-based completion problem based on specific depth image characteristics can rectify many of these issues but is not without its own challenges. Lack of color texture on a smooth surface which leads to unified depth can confuse an exemplar-based approach to a great degree. As seen in Figure 5, the notable exemplar-based in-

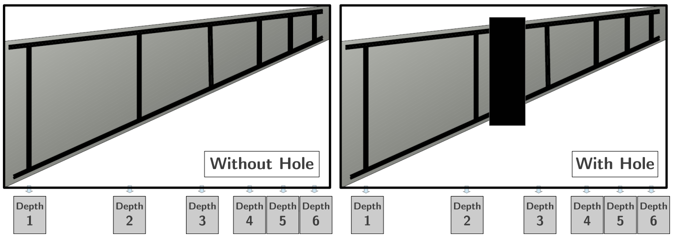

Fig. 10: A simple virtual image used to describe the issues of inpainting methods applied to depth images.

painting method of [36] is capable of filling the target region post object removal from the color image in a reasonably plausible way due to existence of color texture in the background (Figure 5 - left), but within the context of depth, when color texture is removed and uniform depth of a flat plane is all that is left, results are not nearly as impressive (Figure 5 - right). Please note that the goal is to remove an object (the baby) from both the color and depth images and plausibly complete the remaining hole post removal and at the same time fill the existing holes in the depth image (represented by black markings on the depth map).

Moreover, attempting to replicate texture usually requires copying pixels or entire patches from the known regions of the image and using them to fill the holes. One drawback stems from the fact that there may not be enough useful information in the known region to sample from, which is a very common problem in filling depth images if they are not of a fronto-parallel view, which does occur with color image completion as well. However, as mentioned before, the problem can be solved for color images by including the transformed version of the patches by varying rotation, scale, shear, aspect ratio, keystone corrections, gain and bias color adjustments, and other photometric transformations in the search space when trying to find similar patches to sample from. This will exponentially increases the size of the search space and effect the efficiency and accuracy, but is still a solution to the problem, nonetheless. However, with depth images, there may be scenes in which no suitable patch can be found to fill a specific part of the hole even if the search space contains all possible transformed patches from the input.

Imagine Figure 10 (left) contains the outline of a color image. As we go deeper into the image, the intensity values in the color channels of the image may change. However, the hue remains the same, while the illumination changes. Therefore, by transforming patches sampled from the known region (outside the black rectangle in Figure 10-right) using homographic and photometric transformations like illumination, suitable samples can be found that can fill the target region.

On the other hand, assume Figure 10 represents an *ideally* accurate depth image where the depth continuously varies from pixel to pixel as we move deeper into the image (i.e. in a row of pixels on the fence, no two pixels have the same depth value). In a scenario like this, neither homography transformation nor any commonly used photometric variation can guarantee that patches exist in the resulting search space that can be used to accurately fill the target region. Essentially, the 3D depth variation of the scene is captured within the 2D topology of the depth image, but exemplar-based completion following a 2D paradigm will inherently fail in such an *ideal* depth image. Instead, a full 3D transformation of a given patch may be required in terms of rotation, translation, and scale.

It should be noted that the depth images captured using current 3D sensing technology are not *ideal*, and in reality patches that fit the criteria required to fill the target are often found in the depth images obtained through the currently existing technology, but this does not guarantee that an exemplar-based image completion solution will always fill depth images successfully as it does within color images. That said, there are specific depth filling techniques that still take advantage of classic inpainting approaches such as [36] and [41], which have been commonly employed, with or without additional improvements, for depth value filling [87, 86, 58, 41, 8].

For instance, Atapour-Abarghouei et al. [75] performs the challenging task of object removal and depth hole filling in RGB-D images by decomposing the image into high/low spatial frequency components by filtering in Fourier space. The high frequency information (object boundaries and texture relief) is filled using a classic texture synthesis method [32] reformulated as a pixel-by-pixel exemplar-based filling approach

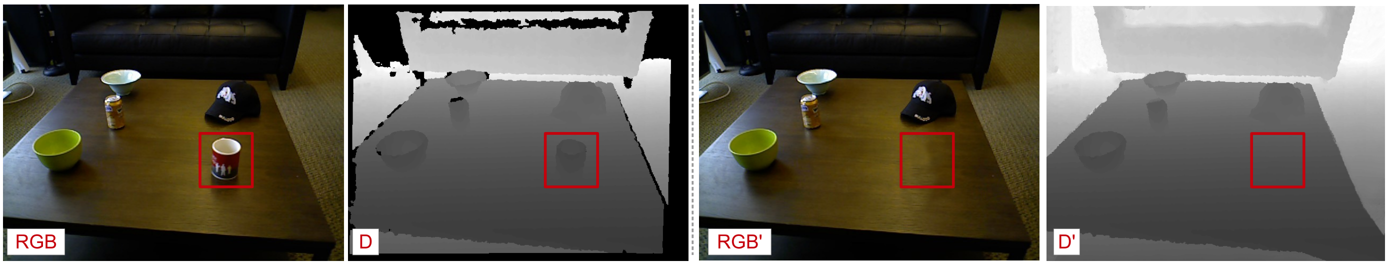

Fig. 11: Object removed from RGB-D image with holes filled using [75]. High and low spatial frequency components are filled independently resulting in sharper and more crisp outputs.

and enhanced by means of query expansion within the search space, and the low frequency component (underlying shape geometry) is completed via [37]. The results are then recombined in the frequency domain to generate the final output. As seen in Figure 11, the produced images are sharp, crisp, and without additional artefacts, although the reliance of the approach on [32] limits its overall computational efficiency.

On the other hand, the work in [88] attempts to perform object removal in multi-view images with an extracted depth image, and uses both structure propagation and structure-guided completion to fill the images, which results in better geometric and structural coherence. The target region is completed in one of a set of multi-view photographs casually taken in a scene. The obtained images are first used to estimate depth via Structure from Motion (SfM). Structure propagation and structure-guided completion are employed to create the final results after an initial color and depth filling step. The individual steps of this algorithm use previously developed color image completion method of [53], and the patch based exemplar approach of [50] to generate results. The approach is relatively costly due to the fact that each of the three steps require an independent form of image completion.

Linag et al. [73] proposes a color and depth inpainting method using a segmentation based approach in stereo images. They make use of the fact that parts of the removed region in one stereo image may still be visible in the other, and try to complete both images via 3D warping. They later fill in both color and depth via depth-assisted texture synthesis, a modified version of the well-known exemplar-based filling technique in [36]. However, in cases where stereo or multi-camera views are not available, as in many active 3D sensing devices such as time-of-flight (ToF) cameras, but missing depth data is abundant (e.g. Figure 9), other filling approaches not dependent on stereo or multi-view images have to be used to fill the naturally occurring holes in depth images. Furthermore, this method has no built-in mechanism to handle large structures, and geometric structures are not accounted for.

Another occasion where exemplar-based depth filling is often used is in Depth Image-Based Rendering techniques (DIBR). This is an extension to Image-Based Rendering (IBR) that tries to create a novel “virtual” view from a set of “real” views, with the added benefit of having depth information available. The images are normally warped and combined to create new synthetic views [92], but the greatest part of the challenge is to deal with the newly exposed holes that are created after the warping. There have been attempts to solve the depth image issues using exemplar-based image completion techniques such as the one proposed in [36]. Daribo and Saito [47] directly utilize this method in their approach to DIBR. Hervieu et al. [48] has modified said method to complete stereo-vision generated disparity maps, where the information from the complementary disparity is used to fill the missing information.

Solving the depth filling problem using an exemplar-based framework has the potential to produce outputs in which structural continuity within the scene is preserved and granular relief texture is accurately and consistently replicated in the missing depth regions. However, if the scene depth is not captured from a fronto-parallel view, there is no guarantee that correct depth values can be predicted for the missing regions even if the patches sampled within the exemplar-based filling approach undergo different transformations.

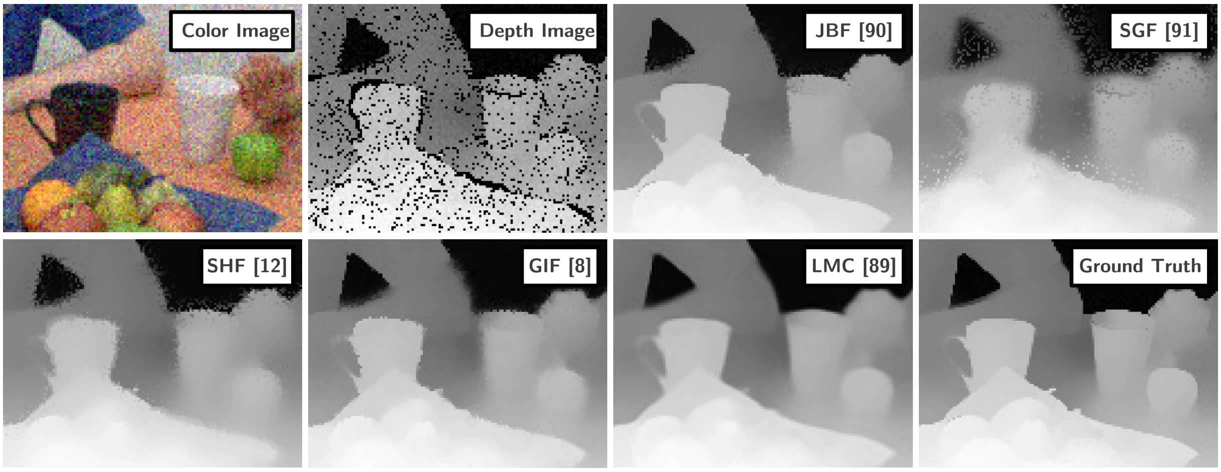

Fig. 12: Example of the results of [89] using low-rank matrix operations (denoted by LMC) compared to joint bilateral filtering method (JBF) [90], structure guided fusion (SGF) [91], spatio-temporal hole filling (SHF) [12], and the guided inpainting and filtering approach (GIF) [8] (reproduced from [89]).

4.1.4. Matrix Completion

Even though completing images using matrices is not conventionally done, it has been observed [89] that similar patches in an RGB-D image lie in a low-dimensional subspace and can be approximated by a matrix with a low rank.

Lu et al. [89] presents a linear algebraic method for low-rank matrix completion-based depth image enhancement to simultaneously remove noise and complete the depth image using the accompanying color image that might be noisy. In order to accomplish simultaneous denoising and hole filling, the low-rank subspace constraint is enforced on a matrix with RGB-D patches via incomplete factorization, which results in capturing the potentially scene-dependent image structures both in the depth and color space.

The rank differs from patch to patch depending on the image structures, so a method is proposed to automatically estimate a rank number based on the data. Figure 12 illustrates how this method can outperform some of the other methods previously referred to in the literature as state-of-the-art approaches, such as the joint bilateral filtering method (JBF) [90], structure guided fusion (SGF) [91], spatio-temporal hole filling (SHF) [12], and the guided inpainting and filtering approach (GIF) [8]. These methods will be explained in the upcoming parts. It is worth mentioning that the approach [89] generates particularly impressive results in that the color image used as the input is noisy (Figure 12 - Color Image). Before the comparisons, a state-of-the-art denoising method [93] was applied to the noisy color image used as an input for the comparators.

Discussion: The problem of depth image completion, being an inherently ill-posed one, can of course be formulated in a variety of ways, including but not limited to diffusion, energy minimization, exemplar-based completion, and alike. Reformulating the depth filling problem results in a variety of solutions that generate completed depth maps with different qualities appropriate for the application for which the depth information is intended. Additionally, there is great potential in attempting to complete an image using a learning-based approach that is capable of understanding the scene intricacies, objects, and their spatial relationships. In recent years, deep neural network approaches have made advances in predicting depth from a single monocular color image [94, 95, 96, 97, 98, 99] and depth super-resolution and upscaling [100, 101, 102], many of which actually learn spatial and/or temporal information within the scene to accomplish their tasks. However, to the best of our knowledge, no attempts have been made so far to complete depth images using deep neural networks, but any approach capable of learning about scene context and content can potentially produce promising results compared to the conventional methods we focus on here.

Our goal is to facilitate the comprehension of the many approaches functioning in or around the field of depth filling [73, 88, 89, 90, 91, 8, 83, 103]. Although highly varied and multifarious, we have made significant strides to divide depth image completion strategies into specific groups for a better and deeper understanding of their functionalities, which would lead to an easier choice of the right approach for researchers based on their requirements and desired effects.

In the upcoming sections, we will categorize depth filling strategies based on three different characterizations: their dependence on the accompanying color image (which may not always be available), the main objective focus that an approach attempts to fulfill (associated with the principles of inpainting outlined by [39] explained in Section 3), and the type of information used within the scene to complete missing depth.

4.2. Information Domain Used for Depth Filling

There are three general types of approaches commonly used to deal with holes in depth images obtained using active or passive 3D capture methods based on the domain of information used to carry out the filling process. An approach may only use the spatial information locally contained within the depth map and potentially the accompanying color image, temporal information extracted from a sequence used to complete or homogenize the scene depth, or even a combination of both in various ways. A brief overview of the approaches utilizing these types of input information is presented in this section. Furthermore, Table 1 provides a short summery of the advantages and the disadvantages of all the categories. Please note that the listed advantages and disadvantages for a given class of approaches in the table obviously vary in degree and strength for different methods in that category, and are generalized to be more comprehensive. Figure 13 provides a general overview of depth filling techniques categorized based on their input dependencies and required information domain.

4.2.1. Spatial-Based Depth Hole Filling

The methods in the first group of depth hole-filling approaches use the neighboring pixel values and other information available in a single depth image to complete any missing or invalid data in the depth image. There are also several approaches that take advantage of the information available in the color image of the same scene to fill the missing data in the depth image.

Even though there are clear limitations to using this type of approach, such as a possible lack of specific information that can be construed as useful to a particular hole region in the current image, there are many important advantages. For instance, when temporal and motion information is taken into consideration in depth completion, filling one frame in a video requires processing multiple consecutive frames around it, and so either the processing has to be done off-line or if real-time results are needed, the results of each frame will appear with a delay. However, if there is no dependence on other frames, with an efficient spatial-based method, real-time results can be generated without any delay.

Being the most-widely-studied depth hole filling approach in the literature, several spatial-based methods have been proposed to complete depth images, the majority of which can be categorized in three different classes: methods that rely upon filtering, interpolation, and extrapolation techniques, inpainting-based methods, and finally, reconstruction-based methods. Examples of the seminal works in these areas are presented in Table 1.

4.2.1.1. Filtering, Interpolation, and Extrapolation

The easiest, yet not always the best, solution to the depth hole filling problem is applying a filter to the depth data. Some common filters of choice would be the median filter [125] or the Gaussian filter [126], but with their use comes significant blurring and loss of texture and edge detail. As mentioned earlier, there are specific filters that have edge-preserving qualities, such as the bilateral filter [127] and non-local filter [128]. However, these filters will not only preserve edges at object boundaries, but the undesirable depth discontinuities caused by depth sensing issues as well. There is also a possibility of distortion in non-hole regions.

As with most depth images obtained via structured light devices, stereo correspondence, and so forth, there is a secondary color or gray-scale image. The visual information present in this accompanying image can be employed to improve the accuracy of the depth image within or near object boundaries (Table 3). It has also been utilized to reduce the

Categories	Subcategories	Advantages	Disadvantages	Examples
Spatial-Based Methods	Filtering, Interpolation, Extrapolation	• Simple implementation. • Potential to be fast and efficient. • Potential to provide a clean image.	• Edges and boundaries may be smoothed. • Undesirable depth discontinuities may stay. • May fail in filling large holes.	[104, 105, 106, 107, 108, 109, 7, 9, 110, 111, 112, 113]
	Inpainting-based	• Effective in smooth regions. • Edges are not smoothed by mistake.	• Artefacts may be added around boundaries and discontinuities. • Not very fast and efficient.	[47, 48, 91, 8, 86, 81, 80]
	Reconstruction-based	• Higher levels of accuracy in both edge and smooth regions. • No artefacts around boundaries and edges.	• Mostly require guidance from color image. • Complicated implementation process.	[114, 82, 83, 115, 103, 116]
Temporal-Based Methods		• Not limited by sampling constraints. • Texture can be preserved more accurately. • Can maintain depth consistency in a sequence.	• Delays are noticeable in presenting results. • Mostly suited for off-line applications. • Incapable of completing a single image.	[11, 117, 118, 119, 120]
Spatio-temporal-Based Methods		• Can avoid jagging and blurring usual in spatial-based methods. • Potential to be more efficient than temporal-based methods.	• Delays still exist in on-line applications. • Cannot complete a single depth effectively.	[121, 12, 122, 123, 124, 90]

Table 1: Advantages and disadvantages of different categories of depth hole filling methods.

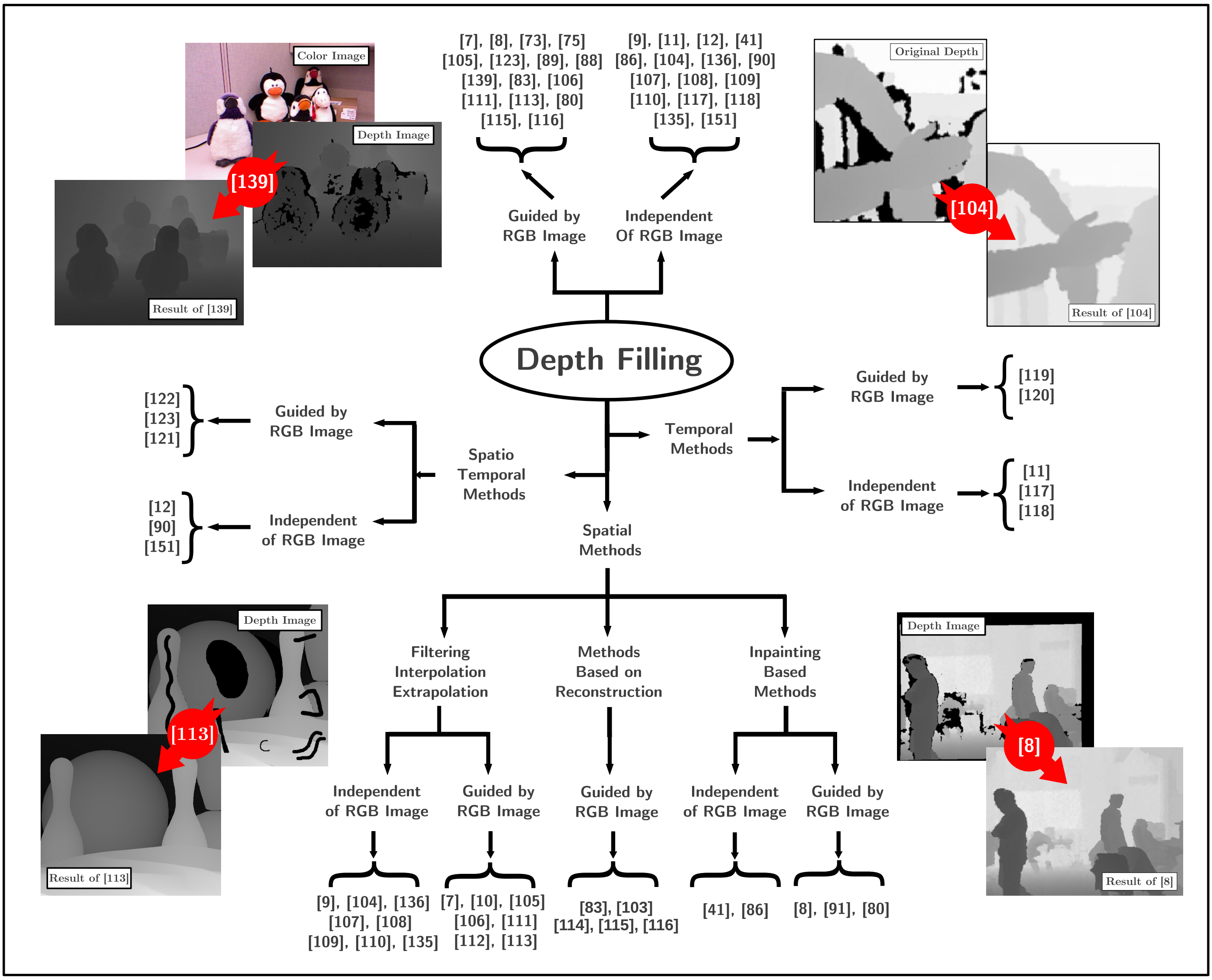

Fig. 13: A diagrammatic taxonomy of filling approaches drawn based on their inputs and information domain used during the filling process.

noise in depth images that is generated by upsampling procedures [129, 130, 10, 131, 100], where the goal is to increase the sharpness, accuracy, and the resolution of the depth image. Moreover, it can also be used to assist filtering approaches, as seen in methods such as joint-bilateral filtering [132], joint trilateral filtering [133], and alike.

He et al. [134] even proposed a fast and non-approximate linear time guided filtering method, the output of which is generated based on the contents of a guidance image. It can transfer the structures of the guidance image into the output and has edge-preserving qualities like the bilateral filter, but can perform even better near object boundaries and edges by avoiding reversal artefacts. Due to its efficiency and performance, it has been used as the basis for several depth completion methods [8, 115].

Yang et al. [104] fills the depth holes based on the depth distribution of its neighboring pixels after labeling each hole and dilating each labeled hole to get the value of the surrounding pixels. Cross-bilateral filtering is subsequently used to refine the results. In Figure 14, the results are compared with the temporal based method in [11], which will be reviewed subsequently.

Chen et al. [105] attempted to detect and fill the invalid and missing depth information using a region growing technique based on the accompanying color image. To increase the accuracy of the values used to fill holes, joint bilateral filtering

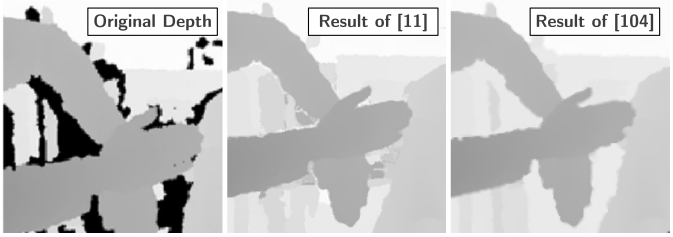

Fig. 14: Example of the result of [104] compared to [11]. Each hole is filled based on the distribution of its neighboring pixels [104] (reproduced from [11]).

is utilized. Once again, since the detection and filling of invalid depth values depends on the color image, in regions where the color values do not match the depth values, validity of the filled hole is questionable even though the results seem plausible and without visible defects. Figure 15 demonstrates how the method can fill depth holes without adding artefacts or blurring.

In the method proposed by Min et al. [106], an approach based on weighted mode filtering and a joint histogram of the color image and the depth image is used. A weight value is calculated according to the color similarity between the target and neighboring pixels on the color image and used for counting each bin on the joint histogram of the depth image. Subsequently, they expand their method to include temporal information for a temporally consistent estimate on the depth video. This method is effective against depth values being blurred on the boundaries.

With regards to improving depth images after a novel virtual viewpoint has been created in DIBR (Depth Image-Based Rendering), Chen et al. [107] utilize a simple average filter to fill depth holes. However, to avoid smoothing and blurring the textured regions and edges, an adaptive method that considers edges and directions is used to enhance the accuracy of object boundaries. Daribo et al. [108] make use of simple filtering but based on a weighted Gaussian filter taking into account the distance to the contours, so as to apply smoothing close to object boundaries but avoid filtering the smooth areas in the depth image. However, in both of these methods, the novel virtual viewpoint is on the same axis as the real view point, which restricts the applicability of the approach.

Mueller et al. [135] uses adaptive cross-trilateral median filtering to reduce the noise and inaccuracies commonly found in depth estimates obtained via stereo correspondence. Parameters of the filter are adapted to the local structures, and a confidence kernel is employed in selecting the filter weights to reduce the number of mismatches.

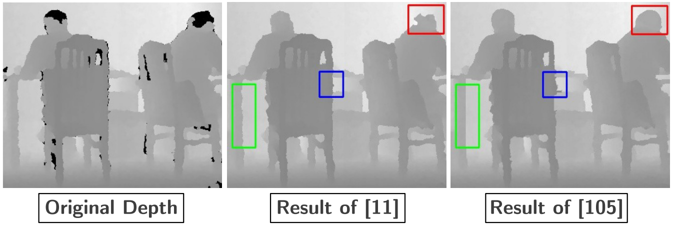

Fig. 15: Result of [105] compared to [11]. The depth is filled using region growing based on the accompanying color image [105] (reproduced from [105]).

In [109], object boundaries are first extracted, and then a discontinuity-adaptive smoothing filter is applied based on the distance of the object boundary and the amount of depth discontinuities. Nguyen et al. [136] proposes a propagation method, inspired by [137], that makes use of a cross bilateral filter to fill the holes in the warped image. Directional depth information is propagated based on camera calibration to fill the holes caused by disocclusion from 3D warping. Whilst the method produces good results (Figure 16), it only accounts for holes caused by transformation and warping.

Lai et al. [7] attempted to handle the false contours and noisy artefacts that exist in the depth information estimated through stereo correspondence methods. They used a joint multilateral filter that consists of kernels measuring proximity of depth samples, similarity between the values of said samples, and similarity between the corresponding color values. Shape of the filter is adaptive to brightness variations. Although the results are promising, there are instances of blurring in the resulting depth images (Figure 18 - left).

A non-parametric interpolation method was recently proposed in [113]. This grammar-inspired approach utilizes a segmentation step [138] and redefines and identifies holes within a set of 12 completion cases with each hole existing in a single row of a single object. The depth pattern is propagated into hole regions according to the individual cases. The approach requires a segmentation step and only works well if enough depth information is available within the object where the hole lies, which means large holes cannot be filled accurately. However,

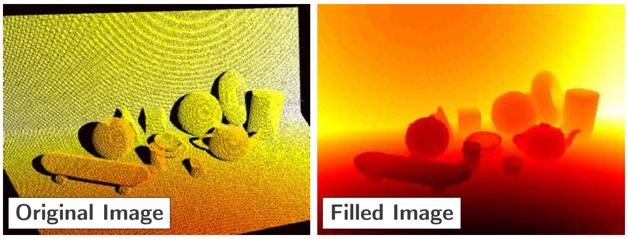

Fig. 16: Depth completion after view rendering [136]. The method uses cross bilateral filtering to fill the holes (reproduced from [136]).

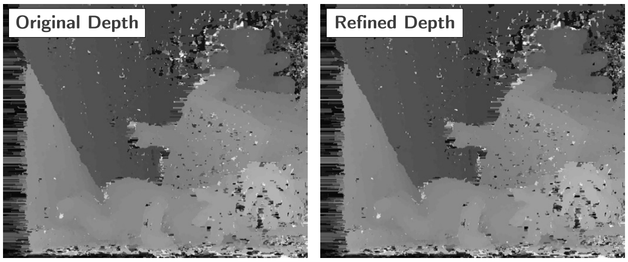

Fig. 17: Depth enhancement via [112]. Noise is removed across object boundaries via a slope depth compensation filter (reproduced from [112]).

for reasonably sized holes the approach works effectively and very efficiently. Figure 19 demonstrates the efficacy of the approach [113] compared to [8, 85, 41, 37, 75] when tested on a synthetic image simulating exaggerated texture. The results are clearly in favor of [113]. The method in [75] provides comparable results qualitatively, but [113] functions in a manner of milliseconds, while [75] can take hours. This can be seen in Table 2, which demonstrates that the approach (signified by “DC” [113] in Table 2) is highly efficient compared to widely-used comparators such as guided inpainting and filtering (GIF) [8], second-order smoothing inpainting (SSI) [85], the fast marching based inpainting method (FMM) [41], exemplar-based inpainting (EBI) [36], Fourier-based inpainting (FBI) [75], and diffusion-based exemplar filling (DEF) [37].

There are interpolation techniques that fill the holes horizontally or vertically within the boundary of the hole by calculating a normalized distance between opposite points of the border (horizontally or vertically) and interpolating the pixels accordingly [9]. These types of approaches will face obvious problems when the target covers parts of certain structures that are neither horizontal nor vertical. Po et al. [9] proposes a multi-directional extrapolation technique for hole filling that uses the neighboring pixel texture features to estimate the direction in which extrapolation is to take place, rather than using the classic horizontal or vertical directions that create obvious deficiencies in the completed image. They propose sets of nine directions to fill the holes so that there is a higher possibility for the completed holes to match the texture or structure of the background and the surrounding objects (Figure 18 - right).

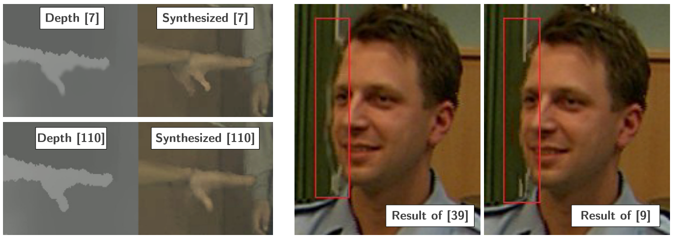

Fig. 18: Result of [7] (reproduced from [7]) compared to [110] (left) and result of [9] (reproduced from [9]) compared to [39] (right).

Shen and Cheung [139] separate the scene into a static background and a number of dynamic foreground objects by assuming different depth layers. As a result, they propose a stochastic framework that combines various RGB-D noise models to determine the label of each depth layer. In order to fill the missing depth values, joint bilateral filter is used, considering the fact that only the neighboring pixels that are on the same depth layer contribute to filling the central pixel. Furthermore, not only are missing depth pixels filled, erroneous depth values are corrected by identifying pixels whose values significantly differ from other neighboring values and refilling them as if they were holes. Figure 20 demonstrates the effectiveness of this method compared to the method proposed in [12].

Xu et al. [110] criticizes the use of bilateral and trilateral filters as the major solution used in completing, enhancing, and refining depth images in DIBR (Depth Image-Based Rendering) [7, 136] by pointing out that artefacts around edges and object boundaries still exist due to the fact that color and depth edges are characteristically different. While the method by Xu et al. [110] does not focus on actually filling depth holes, it does attempt to remove and refine the artefacts that can usually be seen in and around filled areas after the holes have been filled using other methods such as [140, 41, 36, 9]. They use watershed color segmentation [141] to correct any misalignments, and enhance disoccluded regions and sharp depth edges within or without object boundaries by extending the object bound-

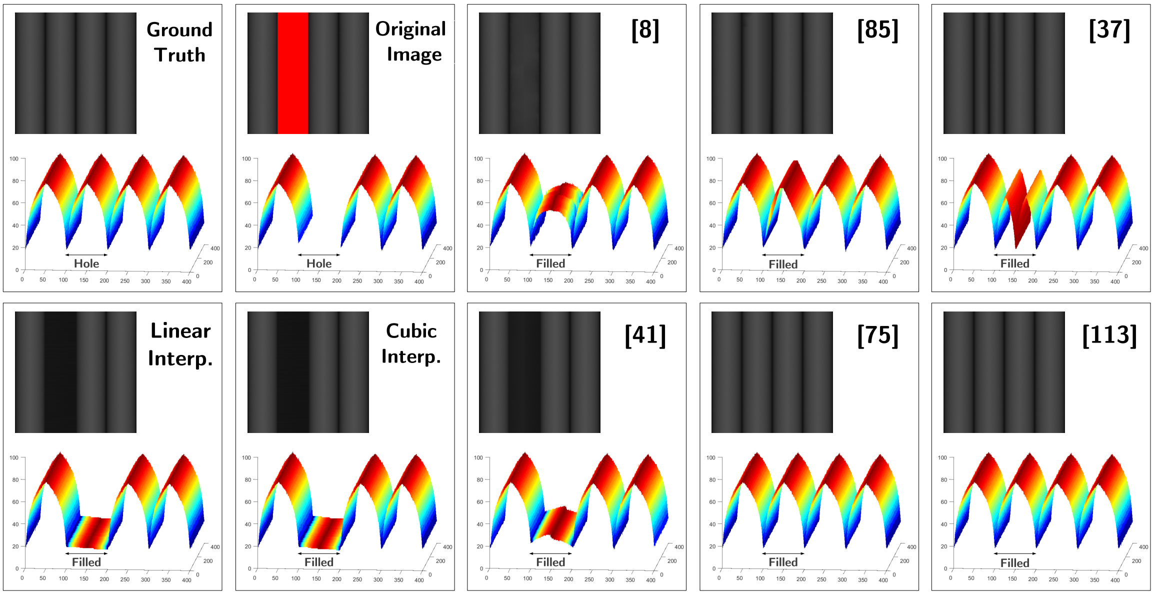

Fig. 19: Example [113] of comparing [113], [8], [85], [41], [37], [75] and linear and cubic interpolation (synthetic test image with available ground truth depth).

aries in depth images to cover the transitional edge regions of color images (Figure 18 - left). Although the resulting depth images are without any burring, the segmentation adds to the computational cost of the approach.

The approach in [111] uses a joint trilateral filtering method made up of domain, range, and depth filters. In this approach, local patch pattern matching is first performed between the image and the depth image, and the results are used to tune the parameters of the filter. The range and depth filters are thus adjusted in a way that the edges in the depth image that accurately correspond with the image edges are rewarded, and therefore, sharper object boundaries are produced.

Matsuo et al. [112] proposed a depth refinement technique that is not meant for hole filling but its elements can certainly be used in filling missing depth data. Their filter attempts to reduce noise by matching the boundary of an object in the color image with the boundary of the object in the depth image. They subsequently remove blurring and ringing across the boundary of the object using an additional slope depth compensation filter. The method is not very computationally costly, but they do note that there is always a trade-off between efficiency and accuracy. An example of the results of the method when applied to depth images with large quantities of noise and holes can be seen in Figure 17. It is important to note once again that, as seen in Figure 17, the approach is not created to fill holes, but to improve and enhance depth images.

Garro et al. [10] presented a segmentation-based interpolation technique to upsample, refine, and enhance depth images. The strategy uses segmentation methods that combine depth and color information [142, 143] in the presence of texture or segmentation techniques based on graph cuts [144] when the image is not particularly highly textured to identify the surfaces and objects in the color image, which are assumed to align with those in the depth image. The low-resolution depth image is later projected on the segmented color image and interpolation is subsequently performed on the output. This method is highly dependent on the precision of the registration between the color image and the depth image and the accuracy of the standard segmentation step.

Discussion: Among spatial-based depth filling strategies, filtering, interpolation, and extrapolation approaches are of the most used and most efficient methods. Filtering methods are widely used in filling depth holes, but most of them have a tendency to blur the image, introduce artefacts around boundaries, and produce invalid edges. As seen in many of the aforementioned examples, many researchers try to overcome these issues by combining the filtering techniques with other methods, constraining their filtering elements, or adding post-processing

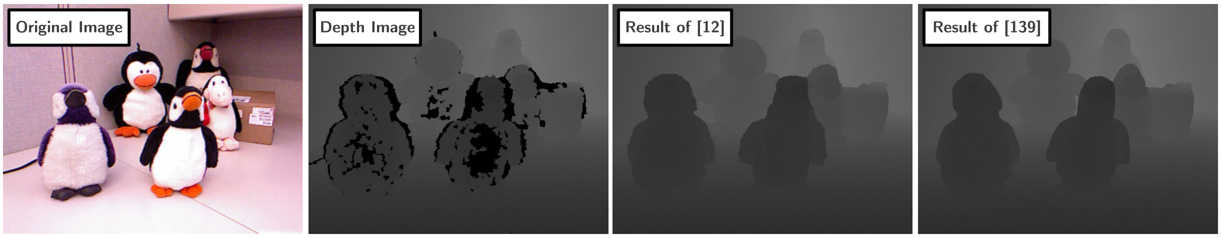

Fig. 20: Results of [139] and [12]. Depth is completed in [139] by assuming different depth layers for foreground and background objects (reproduced from [139]).

stages to refine the filled data. Although many of these solutions are effective, they tend to diminish one of the most valuable aspects of this type of hole filling- the potential for high computational efficiency.

Interpolation and extrapolation techniques are certainly the most efficient strategies due to their low computational cost and are applicable where real-time results are needed. However, simplistic interpolation methods (linear, bilinear, and alike) can cause streaking effects and are only capable of filling small holes on flat planes, just as seen in Figures 19 and 26. There have, however, been methods that take advantage of semantic object boundaries and implicit or explicit diagonal associations to plausibly fill depth holes without significant artefacts (Figures 19 and 26 - result of [113]).

4.2.1.2. Inpainting-Based Approaches

The next category of depth hole filling approaches are heavily based on traditional inpainting techniques (normally used for color images, Section 3). Although many inpainting-based methods yield promising results (at least more so than many filtering techniques), a majority of them are computationally expensive and can only be used in off-line applications of depth hole filling.

Structure-guided inpainting [44] is used in depth hole-filling, but the diffusion, which is used to propagate the structures, results in blurring and hence the loss of detail and texture. The method proposed by Criminisi et al. [36] has been widely used in depth completion. It has been utilized in depth image-based rendering [47] and modified to recover missing data in depth estimates acquired via stereo correspondence [48]. Telea's method [41] is another popular approach, but it does not perform well on depth images, and cannot fill large holes plausibly (Figure 26).

Qi et al. [91] attempt to recover the missing depth information using a fusion-based method integrated with a non-local filtering strategy. They note that the object boundaries and other stopping points that mark the termination of structure continuation process are not easy to locate in depth images which generally have little or no texture, or the boundaries or stopping points might be in the hole region of the depth image. Therefore, the color image is used to assist with spotting the boundaries, and their corresponding positions in the depth image are estimated according to calibration parameters. Their depth inpainting framework follows the work of Bugeau et al. [59] that takes advantage of a scheme similar to the non-local means scheme to make more accurate predictions for pixel values based on image textures. To solve the issue of structure propagation termination, a weight function is proposed in the inpainting framework that takes geometric distance, depth similarity, and the structure information within the color image into account.

Liu et al. [8] improve upon the fast marching method-based inpainting proposed by Telea [41] for depth value in-filling. They essentially use the color image to guide the depth inpainting process. By assuming that the adjacent pixels that have similar color values have a higher probability of having similar depth values as well, they introduce an additional *color term* into the weighting function to increase the contribution of the pixels with the same color. They also change the order of filling, so that the pixels near edges and object boundaries are filled later, in order to produce sharper edges. However, even with all the improvements, this guided depth inpainting method is still not immune to noise and added artefacts around object boundaries (Figure 21 and Figure 26); therefore, the guided filter pro-

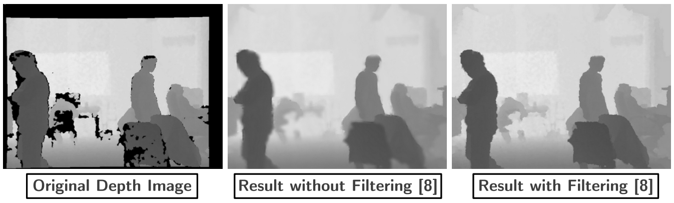

Fig. 21: Results of depth inpainting [8]. The approach is an improved fast marching-based [41] method guided by the color image (reproduced from [8]).

posed by He et al. [134] is used in the post-processing stage to refine the depth image. An example of the results of this widely-acclaimed method with and without the final filtering stage is seen in Figure 21.

Xu et al. [86] introduced an exemplar-based inpainting method to avoid blurring while filling holes in novel views synthesized through depth image-based rendering. In the two separate stages of warped depth image hole filling and warped color image hole filling, the focus is mainly on depth-assisted color image completion with texture. The depth image is assumed to be only a gray-scale image with no texture, and is therefore filled using any available background information (i.e. depth pixels are filled by being assigned the minimum of the neighboring values). The assumptions that depth images have no texture, that texture and relief are not of any significant importance in depth images, and depth holes can be plausibly filled using neighboring background depth are obviously not true, and lead to ignoring the utter importance of accurate 3D information in the state of the art. As a result, although the inpainting method proposed here to complete newly synthesized views based on depth is reasonable, the depth filling itself is lacking.

Miao et al. [81] proposed a color-assisted depth inpainting method that uses diffusion approaches with different rules for two separated components of a depth image: the edge regions and the smooth regions. They note that the depth edges shrinking or fattening is a common problem seen in the results of depth image inpainting methods. To combat the issue, they introduce the concept of a *fluctuating edge region*, which has an adaptive size and is used in the inpainting process. The big issue is that the mean of the depth values in the fluctuating edge region is used to determine the missing pixels near the boundaries, which does not result in a very accurate representation.

Vijayanagar et al. [80] introduced an anisotropic diffusion-based method that can have real-time capabilities by means of a GPU. The color image is used to guide the diffusion in the depth image, which saves computation in the multi-scale pyramid scheme since the color image does not change. In order to guarantee the alignment of the object boundaries in the color image and the depth image, anisotropic diffusion is also applied to object boundaries (see results in Figure 8).

Discussion: There is certainly a greater literature supporting inpainting-based depth filling methods as they are mostly inspired by color image completion techniques, which have a longer history in image processing and computer vision. This class of filling approaches are capable of generating plausible outputs, yet not without their own flaws.

Many inpainting based approaches utilize diffusion techniques and partial differential equations that inherently carry with themselves numerical instabilities and implementation issues. Moreover, efficiency is always a concern when depth filling is needed as preprocessing facet of other applications. As seen in Table 2, inpainting based methods ([41, 8]) need in the order of seconds to process a single image. Although modern hardware and GPUs can facilitate a faster performance with such methods, an independent cross-platform application is still more desirable in the real-world.

Figures 19 and 26 demonstrate the efficacy of the inpainting-based methods in [41], [8], and [37]. While in general these approaches perform better than simple interpolation techniques (linear or cubic interpolation) and even more complex methods such as [85], they are still behind [75] and the very efficient method of [113].

4.2.1.3. Reconstruction-Based Methods

Although filtering and inpainting based depth filling techniques can produce reasonable and efficient results, there is a higher possibility of blurring, ringing, and added artefacts especially around object boundaries, sharp discontinuities and highly textured regions. In reconstruction-based methods, missing depth values are predicted using common synthesis approaches. Since a closed-loop strategy is mostly used to resolve the recon-

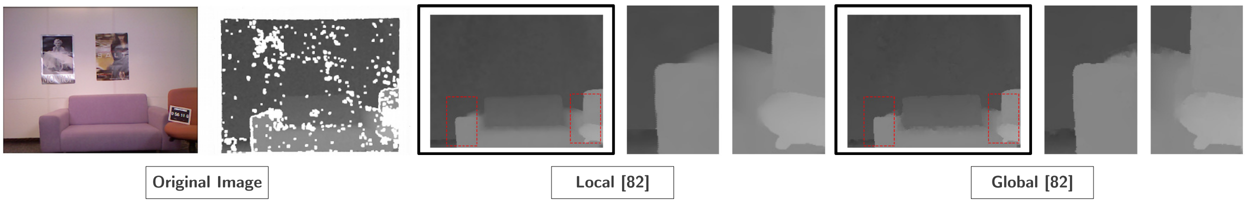

Fig. 22: Local and global framework of [82]. The energy function is made up of a fidelity term (generated depth data characteristics) and a regularization term (joint-bilateral and joint-trilateral kernels). Local filtering can be used instead of global filtering to make parallelization possible (reproduced from [82]).

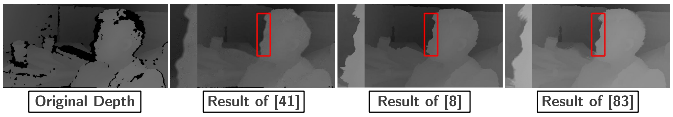

Fig. 23: Example of the results of [83] compared to [41] and [8]. The method's energy function [83] assumes that in small local neighborhoods, depth and color values are linearly correlated (reproduced from [83]).

struction coefficients in terms of the minimization of residuals, higher levels of accuracy can be accomplished in depth hole filling. There are numerous different models found in the literature that are used to represent the hole filling problem.

Chen et al. [114, 82] defines the depth hole filling problem, specifically generated by consumer depth sensors such as Microsoft Kinect, as an energy minimization problem, the function of which is made up of a fidelity term that considers the characteristics of consumer device generated depth data and a regularization term that incorporates the joint-bilateral kernel and the joint-trilateral kernel. The joint-bilateral filter is tuned to incorporate the structure information and the joint-trilateral kernel is adapted to the noise model of consumer device generated depth data. Since the approach is relatively computationally-expensive, local filtering is used to approximate the global optimization framework in order to make parallelization possible, which brings forth the long-pondered question of accuracy versus efficiency. A comparison between examples of the results generated through both local and global frameworks is seen in Figure 22.

Liu et al. [83] proposed a method mainly inspired by the work of Levin et al. in image matting [145]. They designed an energy function based on their assumption that in small local neighborhoods, there is a linear correlation between depth and color values. In order to remove noise and create sharper object boundaries and edges, a regularization term originally proposed by Barbero and Sra [84] is added to the energy function. This added term makes the gradient of the depth image be both horizontally and vertically sparse, which results in less noise and sharper edges. A comparison between the results of this method and inpainting methods in [41] and [8], considered to be very powerful within the literature, is shown in Figure 23.

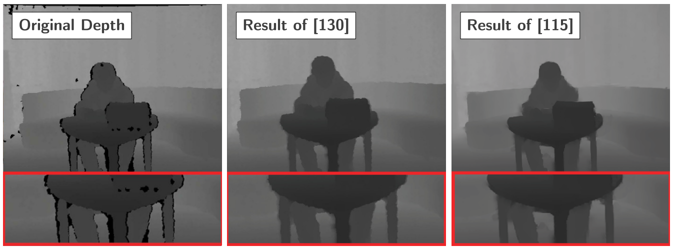

Fig. 24: Result of [115] compared to [130]. [115] follows an adaptive color-guided auto-regressive model for depth recovery (reproduced from [115]).

Yang et al. [115] suggests an adaptive color-guided Auto-regressive (AR) model for depth image recovery. Upon verifying the idea that the AR model fits depth images of generic scenes, they formulate the problem as a minimization of AR prediction errors subject to measurement consistency. Both the local correlation in the original depth image and the non-local similarity in the color image play a role in creating the AR predictor for each pixel. In order to accomplish more accuracy, a parameter adaptation strategy was designed to increase stability. An example of the results is seen in Figure 24.

Wang et al. [103] builds upon their previous work [116] that used a locality regularized representation (LRR) guided by the color image to determine the weights from juxtaposed patches to increase the contribution of the most relevant pixels. However, to mend the shortcomings of their previous method, which ignores the effects of geometric distance and position and only concentrates on the impact of locality on coefficient learning, they suggest using a trilateral constrained sparse representation (SR) which takes intensity similarity and spatial distance between reference patches and the target on sparsity penalty term,

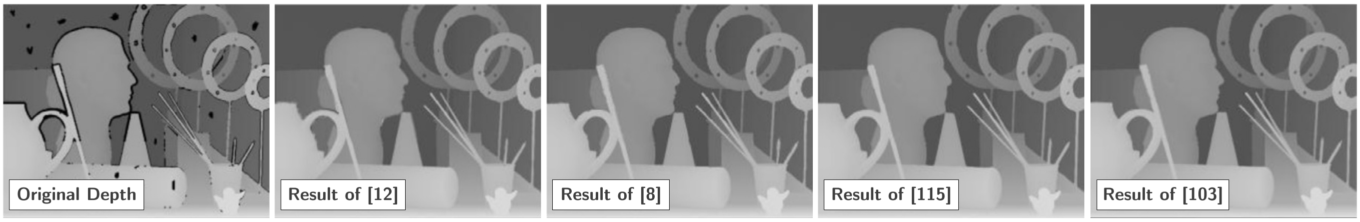

Fig. 25: Result of [103] compared to [12], [8], and [115]. The approach [103] uses a trilateral constrained sparse representation (reproduced from [103]).

and position constraint of central pixel in the target patch on data-fidelity term into account. It should be noted that SR models have been successfully used in stereo vision applications [146, 147, 148] for depth estimation, noise removal, and reconstruction. However, in hole filling, where the depth values in the target region are unavailable, reconstruction coefficient learning has to be performed via the accompanying color image. Figure 25 contains a comparison between [103] and some of the more commonly used depth filling methods of [12, 8, 115].

Discussion: Reconstruction-based methods may be of high complexity, difficult to implement and somewhat computationally expensive, but as seen with the aforementioned approaches, they generate more desirable results, without too much blurring or added artefacts. The object boundaries are also estimated more accurately than most other approaches.

In Figure 26, we can see a comparison of some of the spatial-based depth filling methods [113, 8, 85, 113], image completion techniques [41, 37, 36], and bilinear interpolation over examples from the Middlebury dataset [149]. Table 2 presents the numerical evaluation of the same approaches by comparing their Root Mean Square Error (RMSE), Percentage of Bad Matching Pixels (PBMP), and their run-time. As you can see, even though spatial-based methods are certainly capable of achieving real-time results (unlike temporal-based methods), the current literature epitomizes the long-standing trade-off between accuracy and efficiency. Many of these methods are capable of filling only small holes and others are extremely inefficient. Any future work will need to work towards achieving higher standards of accuracy and plausibility in shorter periods of time. Recent machine learning techniques capable of learning the context and content of a scene [64, 66, 65], may be the next leap forward.

4.2.2. Temporal-Based Depth Hole Filling

In this section, we discuss a group of algorithms that use motion and temporal information in a stream of depth images and perhaps additionally the accompanying color images to fill holes and refine the depth images [11, 117].

One of the techniques most commonly used as a comparator in the literature is the method proposed by Matyunin et al. [11] that uses motion information and the difference between the depth values in the current image and those in the consecutive frames to fill the holes by giving the pixels the weighted average values of the corresponding pixels in other frames. Although the results are mostly plausible, one drawback is that the value of the edges of objects cannot be accurately estimated to an acceptable level (Figures 14 and 15), other than the fact that there is a need for a sequence of depth images, and therefore, the holes in a single depth image cannot be filled. Moreover, this is designed to be an off-line approach and cannot be utilized in real-time applications. Also, when the color information does not correspond with the depth data, the results often contain invalid depth values.

The KinectFusion approach proposed by Izadi et al. [118] takes advantage of the depth images of the neighboring frames to complete the missing information during real-time 3D reconstruction. However, camera motion and a static scene are of utmost importance and although the approach is robust, it cannot be utilized for a static view of a scene without any camera motion.

In [117], holes are grouped into one of two categories: the ones created as a result of occlusion by foreground objects which are assumed to be in motion, and the holes created by reflective surfaces and other random factors. Subsequently, they use the deepest neighboring values to fill pixels according to

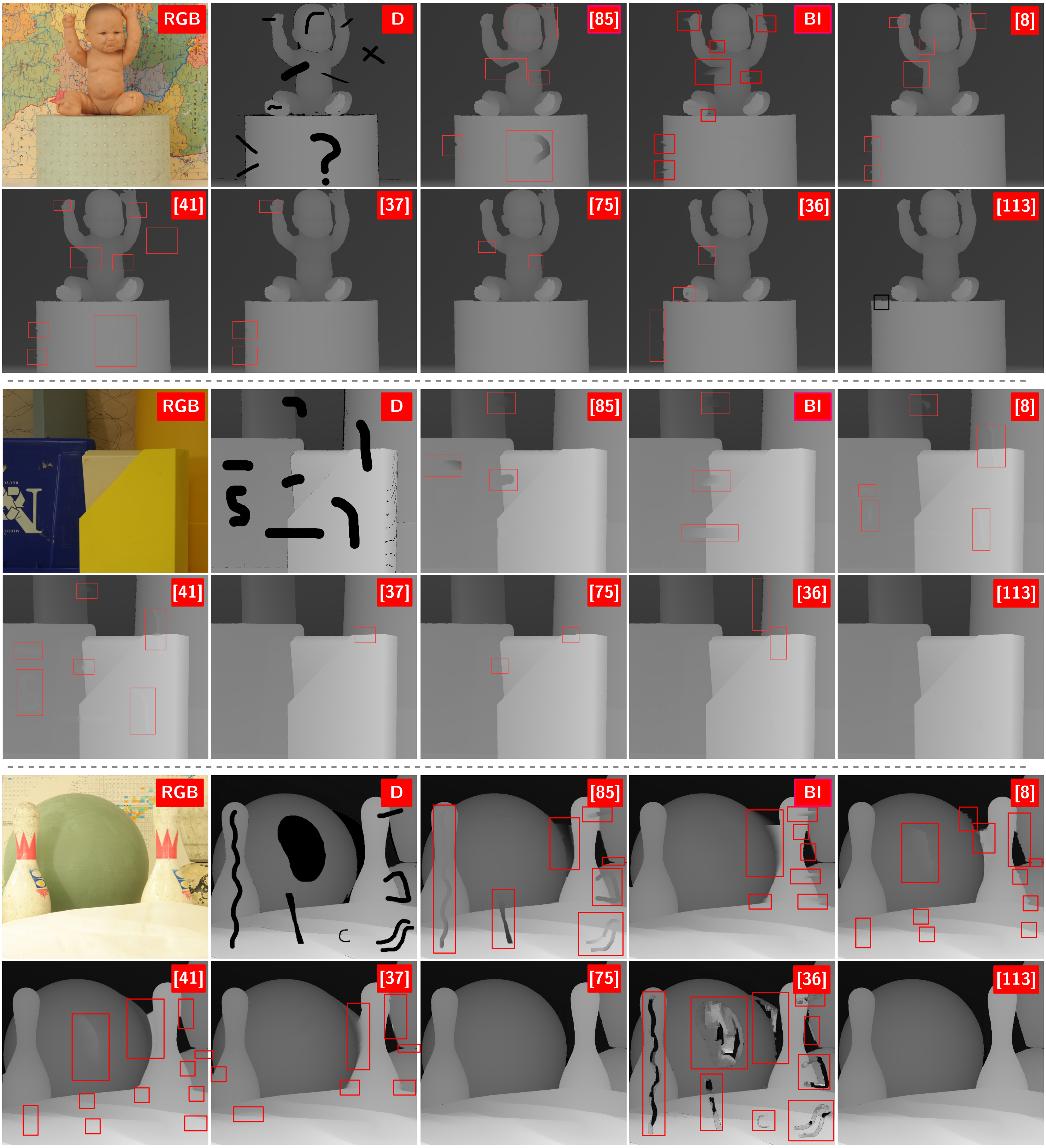

Fig. 26: Comparing the results of [113], [8], [85], [41], [37], [75], [36] and bilinear interpolation (BI) over examples from the Middlebury dataset [149].

the groups they are placed in. Even though their assumptions might be true in many real-life scenarios, they are not universal, and static objects can be the cause of missing or invalid data in depth images captured via many consumer depth sensors.

Fu et al. [119] focuses on repairing the inconsistencies in depth image videos. Depth values of certain objects in one frame sometimes vary from the values of the same objects in a neighboring frame, while the planar existence of the object has not changed. They proposed an adaptive temporal filtering based on the correspondence between depth and color se-

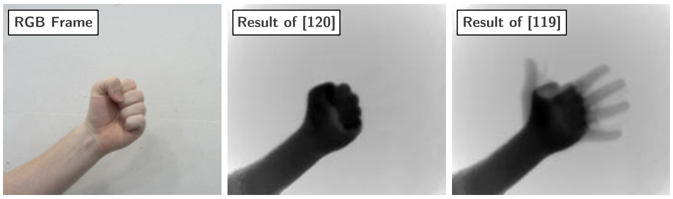

Fig. 27: An example of the results of [120] compared to [119]. The approach [120] repairs depth inconsistencies in videos (reproduced from [120]).

quences. Sheng et al. [120] notes that the challenge in detecting and mending temporal inconsistencies in depth videos is due to the dynamic content and outliers. Consequently, they propose using the intrinsic static structure, which is initialized by taking the first frame and refined as more frames are available. The depth values are then enhanced by combining the input depth and the intrinsic static structure, the weight of which depends on the probability of the input value belonging to the structure. As seen in Figure 27, the method proposed by Sheng et al. [120] does not introduce artefacts into the results due to motion delay because temporal consistency is only enforced on static regions, as opposed to Fu et al.'s method [119], which applies temporal filtering to all regions.

Discussion: Temporal-based methods generate reasonable results even when spatial-based approaches are unable to, and are necessary when depth consistency and homogeneity is important in a depth sequence, which it often is. On the other hand, the dependency on other frames is a hindrance that causes delays or renders the method only applicable as an off-line approach. Moreover, there are many scenarios where a depth sequence is simply not available, but a single depth image still needs to be completed.

4.2.3. Spatio-Temporal Depth Hole Filling

The third class of algorithms combines the elements of the spatial and temporal based methods and attempt to fill holes using *spatio-temporal* information in depth images [121, 12].

In the method proposed by Wang et al. [121], hole filling is attempted in two stages. First, a "*deepest depth image*" is generated by combining the spatio-temporal information in the depth image and the color image, and used to fill the holes. Subsequently, the filled depth image is enhanced based on the joint information of geometry and color. To preserve local features of the depth image, filters that are adapted to the features of the color images are utilized.

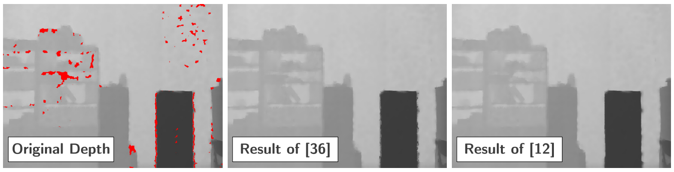

Fig. 28: The results of [12] and [36]. Joint-bilateral filtering is applied to neighboring pixels, and a temporal consistency map is created to track the reliability of the depth values near the holes [12] (reproduced from [12]).

In another widely-used method, Camplani and Salgado [122] use an adaptive spatio-temporal approach to fill depth holes utilizing bilateral and Kalman filters. Their approach is made up of three blocks: an adaptive joint bilateral filter that combines the depth and color information is used, and then random fluctuations of pixel values are subsequently handled by applying an adaptive Kalman filter on each pixel. Finally, an interpolation system uses the stable values in the regions neighboring the holes provided by the previous blocks, and by means of a 2D Gaussian kernel, fills the missing depth values.

In another method [12], the depth holes are filled using a joint-bilateral filter applied to neighboring pixels, the weights of which are determined based on visual data, depth information, and a temporal consistency map that is created to track the reliability of the depth values near the hole regions. The resulting values are taken into account when filtering successive frames, and iterative filtering can ensure increasing accuracy as new samples are acquired and filtered. As seen in Figure 28, the results are superior to the ones produced by the inpainting algorithm proposed by Criminisi et al. [36], which is one of the most commonly-used inpainting methods when it comes to depth hole filling.

Kim et al. [123] once again uses a joint bilateral filter taking both the color and depth information into account for spatial enhancement. For temporal enhancement, they take advantage of block matching applied to the previous and current frame in the color video to detect stationary objects. Therefore by using block matching, they can predict the movement of objects by estimating the similarity between the blocks, measured by mean absolute difference, from frame to frame. The method generates a sharper and clearer depth image, as seen in Figure

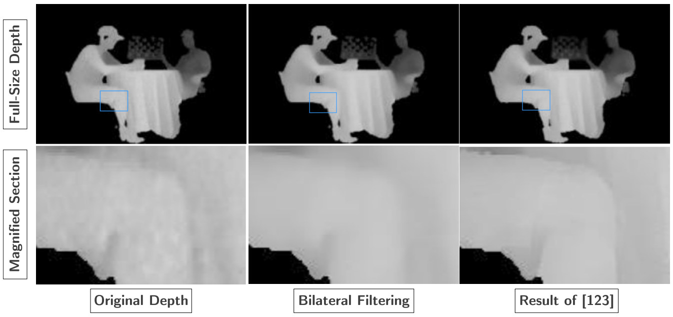

Fig. 29: The results of [123], in which block matching applied to previous and current color frames provides temporal enhancement (reproduced from [123]).

Method	RMSE	PBMP	Run-time
Linear Inter.	1.3082	0.0246	25.12 *ms*
Cubic Inter.	1.3501	0.0236	27.85 *ms*
GIF [8]	0.7797	0.0383	3.521*e*3 *ms*
SSI [85]	3.7382	0.0245	51.56*e*3 *ms*
FMM [41]	1.0117	0.0365	4.31*e*3 *ms*
DEF [37]	0.6188	0.0030	8.25*e*5 *ms*
EBI [36]	0.6541	0.0062	9.68*e*5 *ms*
FBI [75]	0.6944	0.0058	3.84*e*6 *ms*
DC [113]	0.4869	0.0016	99.09 *ms*

Table 2: Average RMSE, PBMP, & run-time (images from Middlebury [149]).

29. However, this method only accounts for the existence of motion, and not the length of motion vectors. Therefore, the depth image is stabilized only for stationary objects.

Xu et al. [124] uses the temporal sequence and motion to create a moving body detection strategy for occlusion filling. Background differentials and the original images are used to extract the moving bodies, and then a 4-neighbor interpolation technique is utilized over the background areas before filling the body areas. The edge can be reasonably preserved, but for an interpolation method, the approach is time-consuming.

Richardt et al. [90] discussed the improvements they made to what can be obtained from a regular video camera alongside a time-of-flight camera. They focus on depth upsampling, color and depth alignment, etc. One of the issues they address is filling holes, which is performed via multi-scale completion technique following the works in [150] and [130]. The output undergoes joint bilateral filtering and spatio-temporal processing to remove noise by averaging values from several consecutive frames. A comparison of their results obtained via spatial filtering only and using spatio-temporal filtering is presented in Figure 30.

More recently, an approach is presented in [151] which uses a sequence of frames to locate outliers with respect to depth consistency within the frame, and utilizes an improved and more efficient regression technique using least median of squares (LMedS) [152] to fill holes and replace outliers with valid depth values. The approach is capable of hole filling and sharp depth refinement within a sequence of frames, but can fail in the presence of invalid depth shared between frames and sudden changes in depth due to fast moving dynamic objects within the scene.

Discussion: Spatio-temporal methods certainly take advantage of the best elements of both spatial-based methods and temporal-based methods, but they also inherit the negatives along with the positives. Temporal and motion information can play a part in helping with the blurring, jagging, and mismatched object contours that are sometimes created by spatial-based methods. However, they also bring forth the issues of off-line applicability and delay in real-time generation of results.

4.3. Use of Secondary Guidance Image

Many modern 3D sensing technologies can provide the user with a depth map and a color image of the same scene. While the filling process on its own is focused on the depth map, there is valuable information contained within the accompanying color image that can significantly improve the quality of the results. There are approaches that take advantage of the object boundaries and edges of the color image to preserve and align the structures in the depth [8, 115, 111]. Even so, it has been pointed out that this can still lead to undesirable artefacts around edges and object boundaries since color and depth edges are characteristically different [110]. Some other approaches have taken to using the color image as a means to segment the scene before depth filling takes place [73, 10, 110, 113], which can provide the filling process with semantically valid scene objects to sample homogeneous depth information from.

However, despite the advantages the color information can offer, not all depth acquisition technologies produce an aligned or easily alignable color image, and requirements of the application may not always allow for the additional computation that comes with the color image processing. In these situations, a depth filling approach that is fully dependent on the color image as a secondary guidance image may not be desirable.

As such, we provide a simple overview of depth filling approaches by categorizing them based on the their use of the color image to provide guidance for the depth completion process. Table 3 presents the aforementioned split over the filling approaches commonly used in the literature. Moreover, Figure 13 provides a taxonomy of the literature based on the requirements of the approaches in terms of their dependence on a secondary input images and the information domain used for the filling process.

Discussion: Among depth filling approaches, some heavily rely on the view of the scene in color to guide the depth completion process. While this can positively affect the outcome in terms of quality and consistency, certain limitations ensue. Aside from the color image not being available at all times, computational requirements can create issues when the application demands light and real-time processing. As seen in Table 3 and Figure 13, a variety of approaches operate in both spaces, giving researchers the opportunity to select a desirable depth filling technique.

4.4. Texture, Boundaries and Smoothing

Four simple rules were proposed in [39] to provide a set of guidelines for generating more plausible and realistic results when attempting to solve the problem of color image completion (Section 3). While not all of these rules apply to depth images (depth maps obviously do not contain any color information), preserving texture, relief and clear object boundaries or smoothing can be important factors in selecting a suitable depth filling approach.

In certain downstream applications, fine-grained texture and relief over surfaces and a clear separation between objects within the depth map is of utmost importance [113, 106], whereas smooth and consistent scene depth [85, 108] can satisfy the requirements of other systems.

It is important to note that preserving fine relief within the depth information of a scene object is a difficult task. Additionally, depth filling is an inherently ill-posed problem. As a result, if texture and relief generation is unnecessarily carried out based on insufficient information, the resulting output can contain more outliers and invalid depth information, which is a hindrance on its own.

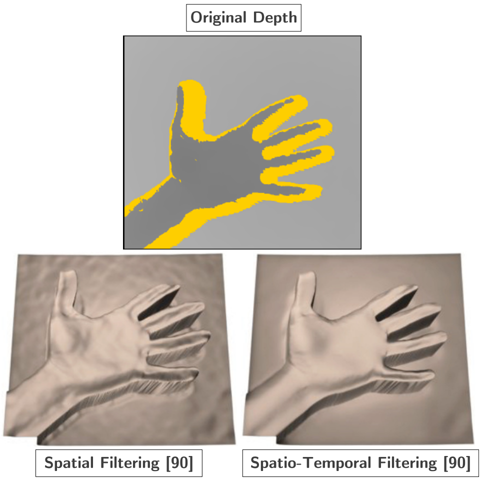

Fig. 30: Result of [90] with spatial and spatio-temporal filtering [90].

Hence, it is important for researchers to identify what is expected of the 3D information gathered from the scene in terms of purpose and functionality to decide what filling approach can produce the ideal results for their specific task.

Table 4 presents a list of depth filling approaches categorized according to their main objectives. Some techniques concentrate on providing very accurate texture and object boundaries, while others generate overly smooth depth in the output with their focus on the structural integrity of the scene depth.

Discussion: The exact characteristics of a depth map depends on its purpose. In certain applications such as object recognition [153, 154] or detection [155], accurate boundaries and relief of an object in the depth map can play an important role in the semantic value of that object within the scene. However, other applications such as localization and mapping [78, 79] do not require fine texture and relief for each individual scene object and accurate structure within the scene depth is sufficient. As seen in Table 4, different filling techniques exist that can generate complete depth either with fine relief or smoothed object surfaces.

Input Image Required	Advantages	Disadvantages	Examples of Filling Techniques
Depth and Color Images	• more processing information • more accurate results	• possible lack of color input • more computationally intensive	[75], [73], [88], [89], [139], [115], [105], [106] [8], [7], [91], [80], [83], [116], [123], [113]
Depth Image Only	• no dependence on extra inputs • more efficient processing	• less information for processing • lower quality outputs	[104], [107], [108], [135], [136], [109], [9], [110] [86], [11], [117], [118], [12], [90], [151], [41]

Table 3: Examples of filling approaches categorized according to the type of images required as their input.

5. Conclusion

In this survey, we focused on reviewing the various techniques that have been developed to complete, enhance, and refine depth images. Although significant efforts are under way with regards to improving scene depth capture technologies, there are still several issues blocking the path to a perfect depth image such as missing data, invalid depth values, low resolution, and noise.

Although this is still an area of increasing interest and importance, numerous approaches have already been proposed to deal with the aforementioned issues. Most of the methods are unique and propose creative solutions to the problem, but we have attempted to categorize the existing algorithms according to the nature of their formulation, information domain needed for the filling process, input requirements, and the focus of the approach, in order to provide a better means of analysis and understanding.

The problem of depth hole filling has been formulated in a variety of different ways, as has the related problem of color image completion, which offers creative solutions appropriate for different application facets. Diffusion-based and energy minimization solutions to the problem are accurate with respect to structural continuity within the scene depth and can produce smooth surfaces within object boundaries, which can be a desirable trait for certain applications. However, these solutions are often inefficient, computationally expensive, and fraught with implementation issues. Similar to some of the most successful image completion approaches, the depth filling problem can be solved using an exemplar-based paradigm, which can accurately replicate object texture and relief as well as preserve the necessary geometric structures within the scene. There are, of course, a variety of other problem formulations, such as matrix completion, labelling, and alike, each focusing on certain aspects of the completed depth output.

As for the input requirements for a depth filling approach, depending on the acquisition method, depth images are sometimes obtained along with an aligned or easily alignable color image of the same scene. The information contained within this color image can be used to better guide the filling approach applied to the depth image. However, not all depth images are accompanied by a color image and processing the color information intensifies the computation that may not be necessary depending on the requirements of the application.

Additionally, some approaches produce completed depth images with fine-grain texture, relief, and accurate object boundaries in mind, which is an outcome that is very desirable for certain applications. On the other hand, some systems only require accurate structure and scene geometry within the depth information and smooth object surfaces with no granular texture and relief whatsoever are sufficient.

Regarding the information domain used to carry out the filling process, there are spatial-based methods that limit themselves to the information in the neighboring regions of the depth image and possibly the accompanying color image. Some of these algorithms make use of filtering techniques, while some utilize interpolation and extrapolation approaches. The filtering, interpolation, and extrapolation methods can provide fast and clean results but suffer from issues like smoothed boundaries and blurred edges. Many researchers have proposed using inpainting-based techniques, which have been proven successful in completing color images, for filling depth holes. Although the results are satisfactory, these methods are not all efficient and can generate additional artefacts near target and object boundaries. Reconstruction methods provide very accurate results by using techniques inspired by scene synthesis methods. However, they are difficult to implement and mostly have

Main Focus of Filling Approach	Examples of Filling Techniques
Relief and Object Boundary Preservation	[75], [139], [113], [8], [83], [90], [89], [106], [111], [107], [123]
Accurate Structure and Smooth Surfaces	[86], [48], [47], [12], [104], [105], [41], [91], [109], [85], [112]

Table 4: Examples of filling approaches categorized according to the main focus of the filling approach (structure vs. texture and accurate boundaries).

a strict dependency on the accompanying color image.

Temporal-based hole filling techniques take advantage of the motion information and the values in the neighboring frames in a sequence to complete depth images. Sometimes the information in a single depth image is not enough to complete that image, which is where spatial-based methods fall short. Temporal-based approaches, however, do not suffer from this issue and have a larger supply of information at their disposal. This class of methods is still not perfect and the need to process other frames to complete a depth image makes them more suited for off-line applications rather than real-time systems.

Finally, various spatio-temporal-based methods have been proposed that use both the spatial information in the depth image and the motion and temporal information extracted from a depth sequence to complete a depth image. Although these methods can be more accurate than spatial-based methods and more efficient than temporal-based methods, they still suffer from the issues of both these categories.

Based on our careful examination of all the approaches discussed in this study and our own experimental comparisons, we observe general trends with respect to output quality, input requirements, algorithm complexity and speed. The needs of an individual user based on these considerations determines their choice of approach.

In terms of speed, spatial-based methods are essentially the only group of techniques potentially capable of processing images in a real-time fashion, even though they may not always live up to this potential. Within this category, filtering, interpolation, and extrapolation techniques are the most efficient with the least amount of complexity while constrained when it comes to output quality. Inpainting-based approaches are more complex and in certain cases numerically unstable, yet they offer a better trade-off between efficiency and output quality. On the other hand, reconstruction-based methods, though complex, difficult to implement, and somewhat dependent on scene conditions (e.g. scene object sizes, static objects, dynamic viewpoint, and alike), can produce higher quality outputs with acceptable efficiency.

As for input requirements, an accompanying color image may or may not be indispensable to specific spatial-based methods, but there is no need for temporal depth information received from others adjacent frames, which is indeed a necessity for temporal-based and spatio-temporal-based methods, rendering their use less than satisfactory in applications that demand real-time depth processing. Moreover, although these approaches can suffer from issues stemming from complexity, they can provide homogeneous and consistent depth within a stream, which is an important quality for certain applications.

Although the taxonomies drawn in this work may create an illusion of certain unsolvable constraints on the problem in general, new and innovative techniques that focus on efficiently generating more accurate depth images with higher qualities can be developed by considering semantic aspects such as scene analysis, object recognition, and constrained reconstruction.

Furthermore, whilst future avenues of research need to explicitly consider computational efficiency, within the contemporary application domains of consumer depth cameras and stereo-based depth recovery, it is also highly likely they will be able to exploit temporal aspects of a live “depth stream”. It is thus possible that both temporal and spatio-temporal genres within our taxonomy will become the primary areas of growth within this domain over the coming years. This trend will be heavily supported by aspects of machine learning and potentially on-line machine learning as depth streams become increasingly widespread, of which we see limited leverage in depth completion and enhancement to date [100], and that of depth-driven odometry [156] and related scene mapping techniques.

6. Acknowledgments

Figures 3, 4, 6, 7, 11, 12, 20, 14, 15, 16, 17, 18, 19, 21, 8, 22, 23, 24, 25, 27, 28, 29, and 30 were reproduced from the original works (as referenced) with the kind permission of the original authors.

References

[1] Tippetts, B, Lee, DJ, Lillywhite, K, Archibald, J. Review of stereo vision algorithms and their suitability for resource-limited systems. Real-Time Image Processing 2016;11(1):5–25.

[2] Zhang, Z. Microsoft kinect sensor and its effect. IEEE Multimedia 2012;19(2):4–10.

[3] Cong, P, Xiong, Z, Zhang, Y, Zhao, S, Wu, F. Accurate dynamic 3d sensing with fourier-assisted phase shifting. Selected Topics in Signal Processing 2015;9(3):396–408.

[4] Yang, Q, Tan, KH, Culbertson, B, Apostolopoulos, J. Fusion of active and passive sensors for fast 3d capture. In: Int. Workshop on Multimedia Signal Processing. IEEE; 2010, p. 69–74.

[5] Gudmundsson, SA, Aanaes, H, Larsen, R. Fusion of stereo vision and time-of-flight imaging for improved 3d estimation. Intelligent Systems Technologies and Applications 2008;5(3-4):425–433.

[6] Cruz, L, Lucio, D, Velho, L. Kinect and rgb-d images: Challenges and applications. In: Conf. Graphics, Patterns and Images Tutorials. IEEE; 2012, p. 36–49.

[7] Lai, P, Tian, D, Lopez, P. Depth map processing with iterative joint multilateral filtering. In: Picture Coding Symposium. IEEE; 2010, p. 9–12.

[8] Liu, J, Gong, X, Liu, J. Guided inpainting and filtering for kinect depth maps. In: Int. Conf. Pattern Recognition. IEEE; 2012, p. 2055–2058.

[9] Po, LM, Zhang, S, Xu, X, Zhu, Y. A new multi-directional extrapolation hole-filling method for depth-image-based rendering. In: Int. Conf. Image Processing. IEEE; 2011, p. 2589–2592.

[10] Garro, V, Mutto, CD, Zanuttigh, P, Cortelazzo, GM. A novel interpolation scheme for range data with side information. In: Conf. Visual Media Production. IEEE; 2009, p. 52–60.

[11] Matyunin, S, Vatolin, D, Berdnikov, Y, Smirnov, M. Temporal filtering for depth maps generated by kinect depth camera. In: 3DTV Conference. IEEE; 2011, p. 1–4.

[12] Camplani, M, Salgado, L. Efficient spatiotemporal hole filling strategy for kinect depth maps. In: IS&T/SPIE Electronic Imaging. 2012, p. 82900E–82900E.

[13] Scharstein, D, Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Computer Vision 2002;47(1-3):7–42.

[14] Seitz, SM, Curless, B, Diebel, J, Scharstein, D, Szeliski, R. A comparison and evaluation of multi-view stereo reconstruction algorithms. In: IEEE Conf. Computer Vision and Pattern Recognition; vol. 1. 2006, p. 519–528.

[15] Brown, LG. A survey of image registration techniques. Computing Surveys 1992;24(4):325–376.

[16] Pages, J, Salvi, J, Garcia, R, Matabosch, C. Overview of coded light projection techniques for automatic 3d profiling. In: Int. Conf. Robotics and Automation; vol. 1. IEEE; 2003, p. 133–138.

[17] Han, J, Shao, L, Xu, D, Shotton, J. Enhanced computer vision with microsoft kinect sensor: A review. IEEE Trans Cybernetics 2013;43(5):1318–1334.

[18] Khoshelham, K. Accuracy analysis of kinect depth data. In: ISPRS Workshop on Laser Scanning; vol. 38. 2011, p. W12.

[19] Kolb, A, Barth, E, Koch, R, Larsen, R. Time-of-flight cameras in computer graphics. In: Computer Graphics Forum; vol. 29. Wiley Online Library; 2010, p. 141–159.

[20] Sell, J, O'Connor, P. The xbox one system on a chip and kinect sensor. IEEE Micro 2014;34(2):44–53.

[21] Gokturk, SB, Yalcin, H, Bamji, C. A time-of-flight depth sensor-system description, issues and solutions. In: Conf. Workshop on Computer Vision and Pattern Recognition. IEEE; 2004, p. 35–35.

[22] El-laithy, RA, Huang, J, Yeh, M. Study on the use of microsoft kinect for robotics applications. In: Position Location and Navigation Symposium. IEEE; 2012, p. 1280–1288.

[23] Berger, K, Ruhl, K, Schroeder, Y, Bruemmer, C, Scholz, A, Magnor, MA. Markerless motion capture using multiple color-depth sensors. In: Vision Modeling and Visualization. 2011, p. 317–324.

[24] Butler, A, Izadi, S, Hilliges, O, Molyneaux, D, Hodges, S, Kim, D. Shake'n'sense: Reducing interference for overlapping structured light depth cameras. In: Conf. Human Factors in Computing Systems. 2012, p. 19331936.

[25] Sabov, A, Krüger, J. Identification and correction of flying pixels in range camera data. In: Proc. Conf. Computer Graphics. ACM; 2008, p. 135–142.

[26] Sarbolandi, H, Lefloch, D, Kolb, A. Kinect range sensing: Structured-light versus time-of-flight kinect. Computer Vision and Image Understanding 2015;139:1–20.

[27] Hansard, M, Lee, S, Choi, O, Horaud, RP. Time-of-Flight Cameras: Principles, Methods and Applications. Springer Science & Business Media; 2012.

[28] Ihrke, I, Kutulakos, KN, Lensch, H, Magnor, M, Heidrich, W. Transparent and specular object reconstruction. In: Computer Graphics Forum; vol. 29. Wiley Online Library; 2010, p. 2400–2426.

[29] Ringbeck, T, Möller, T, Hagebeuker, B. Multidimensional measurement by using 3d pmd sensors. Advances in Radio Science 2007;5:135.

[30] Lindner, M, Schiller, I, Kolb, A, Koch, R. Time-of-flight sensor calibration for accurate range sensing. Computer Vision and Image Understanding 2010;114(12):1318–1328.

[31] Popat, K, Picard, RW. Novel cluster-based probability model for texture synthesis, classification, and compression. In: Visual Communications. 1993, p. 756–768.

[32] Efros, AA, Leung, TK. Texture synthesis by non-parametric sampling. In: Int. Conf. Computer Vision; vol. 2. IEEE; 1999, p. 1033–1038.

[33] Liang, L, Liu, C, Xu, YQ, Guo, B, Shum, HY. Real-time texture synthesis by patch-based sampling. ACM Trans Graphics 2001;20(3):127–150.

[34] Efros, AA, Freeman, WT. Image quilting for texture synthesis and transfer. In: Conf. Computer Graphics and Interactive Techniques. ACM; 2001, p. 341–346.

[35] Nealen, A, Alexa, M. Hybrid Texture Synthesis. Techn. Univ., Fachbereich Informatik, Fachgebiet Graphisch-Interaktive Systeme; 2003.

[36] Criminisi, A, Pérez, P, Toyama, K. Region filling and object removal by exemplar-based image inpainting. IEEE Trans Image Processing 2004;13(9):1200–1212.

[37] Arias, P, Facciolo, G, Caselles, V, Sapiro, G. A variational framework for exemplar-based image inpainting. Computer Vision 2011;93(3):319–347.

[38] Jia, J, Tang, CK. Image repairing: Robust image synthesis by adaptive n-d tensor voting. In: IEEE Conf. Computer Vision and Pattern Recognition; vol. 1. 2003, p. I–643.

[39] Bertalmio, M, Sapiro, G, Caselles, V, Ballester, C. Image inpainting. In: Int. Conf. Computer Graphics and Interactive Techniques. 2000, p. 417–424.

[40] Breckon, TP, Fisher, RB. Amodal volume completion: 3d visual completion. Computer Vision and Image Understanding 2005;99(3):499–526.

[41] Telea, A. An image inpainting technique based on the fast marching method. Graphics Tools 2004;9(1):23–34.

[42] Chan, T, Shen, J. Mathematical models for local deterministic inpaintings. Tech. Rep.; Technical Report CAM TR 00-11, UCLA; 2000.

[43] Chan, TF, Shen, J. Non-texture inpainting by curvature-driven diffusions. Visual Communication and Image Representation 2001;12(4):436–449.

[44] Richard, MMOBB, Chang, MYS. Fast digital image inpainting. In: Int. Conf. Visualization, Imaging and Image Processing. 2001, p. 106–107.

[45] Harrison, P. A non-hierarchical procedure for resynthesis of complex textures 2001;.

[46] Bertalmio, M, Vese, L, Sapiro, G, Osher, S. Simultaneous

structure and texture image inpainting. IEEE Trans Image Processing 2003;12(8):882–889.

[47] Daribo, I, Saito, H. A novel inpainting-based layered depth video for 3dtv. IEEE Trans Broadcasting 2011;57(2):533–541.

[48] Hervieu, A, Papadakis, N, Bugeau, A, Gargallo, P, Caselles, V. Stereoscopic image inpainting: Distinct depth maps and images inpainting. In: Int. Conf. Pattern Recognition. IEEE; 2010, p. 4101–4104.

[49] Mansfield, A, Prasad, M, Rother, C, Sharp, T, Kohli, P, Van Gool, LJ. Transforming image completion. In: Proc. British Machine Vision Conference. 2011, p. 1–11.

[50] Darabi, S, Shechtman, E, Barnes, C, Goldman, DB, Sen, P. Image melding: Combining inconsistent images using patch-based synthesis. ACM Trans Graphics 2012;31(4):82–1.

[51] Kumar, V, Mukherjee, J, Mandal, SKD. Image inpainting through metric labeling via guided patch mixing. IEEE Trans Image Processing 2016;25(11):5212–5226.

[52] Kwatra, V, Essa, I, Bobick, A, Kwatra, N. Texture optimization for example-based synthesis. In: ACM Trans. Graphics; vol. 24. 2005, p. 795–802.

[53] Wexler, Y, Shechtman, E, Irani, M. Space-time completion of video. IEEE Trans Pattern Analysis and Machine Intelligence 2007;29(3):463–476.

[54] Barnes, C, Shechtman, E, Finkelstein, A, Goldman, D. Patchmatch: A randomized correspondence algorithm for structural image editing. ACM Trans Graphics 2009;28(3):24.

[55] Komodakis, N, Tziritas, G. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. IEEE Trans Image Processing 2007;16(11):2649–2661.

[56] Pritch, Y, Kav-Venaki, E, Peleg, S. Shift-map image editing. In: Int. Conf. Computer Vision; vol. 9. 2009, p. 151–158.

[57] Sun, J, Yuan, L, Jia, J, Shum, HY. Image completion with structure propagation. In: ACM Trans. Graphics; vol. 24. 2005, p. 861–868.

[58] Liu, Y, Caselles, V. Exemplar-based image inpainting using multiscale graph cuts. IEEE trans Image Processing 2013;22(5):1699–1711.

[59] Bugeau, A, Bertalmío, M, Caselles, V, Sapiro, G. A comprehensive framework for image inpainting. IEEE Trans Image Processing 2010;19(10):2634–2645.

[60] Lee, JH, Choi, I, Kim, MH. Laplacian patch-based image synthesis. In: IEEE Conf. Computer Vision and Pattern Recognition. 2016, p. 2727–2735.

[61] Huang, JB, Kang, SB, Ahuja, N, Kopf, J. Image completion using planar structure guidance. ACM Trans Graphics 2014;33(4):129.

[62] Hays, J, Efros, AA. Scene completion using millions of photographs. ACM Trans Graphics 2007;26(3):4.

[63] Whyte, O, Sivic, J, Zisserman, A. Get out of my picture! internet-based inpainting. In: British Machine Vision Conference. 2009, p. 1–11.

[64] Pathak, D, Krahenbuhl, P, Donahue, J, Darrell, T, Efros, AA. Context encoders: Feature learning by inpainting. In: IEEE Conf. Computer Vision and Pattern Recognition. 2016, p. 2536–2544.

[65] Yeh, R, Chen, C, Lim, TY, Hasegawa-Johnson, M, Do, MN. Semantic image inpainting with perceptual and contextual losses. arXiv preprint arXiv:160707539 2016;.

[66] Yang, C, Lu, X, Lin, Z, Shechtman, E, Wang, O, Li, H. High-resolution image inpainting using multi-scale neural patch synthesis. arXiv preprint arXiv:161109969 2016;.

[67] Wei, LY, Lefebvre, S, Kwatra, V, Turk, G. State of the art in example-based texture synthesis. In: Eurographics State of the Art Report. 2009, p. 93–117.

[68] Guillemot, C, Le Meur, O. Image inpainting: Overview and recent advances. Signal Processing Magazine 2014;31(1):127–144.

[69] Fidaner, IB. A survey on variational image inpainting, texture synthesis and image completion. Bogazici University 2008;.

[70] Zhang, Hy, Peng, Qc. A survey on digital image inpainting. Image and Graphics 2007;12(1):1–10.

[71] Janarthanan, V, Jananii, G. A detailed survey on various image inpainting techniques. Advances in Image Processing 2012;2(2):1.

[72] Lowe, DG. Distinctive image features from scale-invariant keypoints. Computer Vision 2004;60(2):91–110.

[73] Wang, L, Jin, H, Yang, R, Gong, M. Stereoscopic inpainting: Joint color and depth completion from stereo images. In: IEEE Conf. Computer Vision and Pattern Recognition. 2008, p. 1–8.

[74] Hamilton, O, Breckon, T. Generalized dynamic object removal for dense stereo vision based scene mapping using synthesised optical flow. In: Int. Conf. Image Processing. IEEE; 2016, p. 3439–3443.

[75] Atapour-Abarghouei, A, Payen de La Garanderie, G, Breckon, TP. Back to butterworth - a fourier basis for 3d surface relief hole filling within rgb-d imagery. In: Int. Conf. Pattern Recognition. IEEE; 2016, p. 2813–2818.

[76] Perona, P, Malik, J. Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Analysis and Machine Intelligence 1990;12(7):629–639.

[77] Ballester, C, Caselles, V, Verdera, J, Bertalmio, M, Sapiro, G. A variational model for filling-in gray level and color images. In: Int. Conf. Computer Vision; vol. 1. IEEE; 2001, p. 10–16.

[78] Hu, G, Huang, S, Zhao, L, Alempijevic, A, Dissanayake, G. A robust rgb-d slam algorithm. In: Int. Conf. Intelligent Robots and Systems. IEEE; 2012, p. 1714–1719.

[79] Mur-Artal, R, Tardós, JD. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. IEEE Trans Robotics 2017;33(5):1255–1262.

[80] Vijayanagar, KR, Loghman, M, Kim, J. Real-time refinement of kinect depth maps using multi-resolution anisotropic diffusion. Mobile Networks and Applications 2014;19(3):414–425.

[81] Miao, D, Fu, J, Lu, Y, Li, S, Chen, CW. Texture-assisted kinect depth inpainting. In: Int. Symp. Circuits and Systems. IEEE; 2012, p. 604–607.

[82] Chen, C, Cai, J, Zheng, J, Cham, TJ, Shi, G. Kinect depth recovery using a color-guided, region-adaptive, and depth-selective framework. ACM Trans Intelligent Systems and Technology 2015;6(2):12.

[83] Liu, S, Wang, Y, Wang, J, Wang, H, Zhang, J, Pan, C. Kinect depth restoration via energy minimization with tv 21 regularization. In: Int. Conf. Image Processing. IEEE; 2013, p. 724–724.

[84] Barbero, A, Sra, S. Fast newton-type methods for total variation regularization. In: Int. Conf. Machine Learning. 2011, p. 313–320.

[85] Herrera, D, Kannala, J, Heikkilä, J, et al. Depth map inpainting under a second-order smoothness prior. In: Scandinavian Conf. Image Analysis. Springer; 2013, p. 555–566.

[86] Xu, X, Po, LM, Cheung, CH, Feng, L, Ng, KH, Cheung, KW. Depth-aided exemplar-based hole filling for dibr view synthesis. In: Int. Symp. Circuits and Systems. IEEE; 2013, p. 2840–2843.

[87] Zhang, L, Shen, P, Zhang, S, Song, J, Zhu, G. Depth enhancement with improved exemplar-based inpainting and joint trilateral guided filtering. In: Int. Conf. Image Processing. IEEE; 2016, p. 4102–4106.

[88] Baek, SH, Choi, I, Kim, MH. Multiview image completion with space structure propagation. In: IEEE Conf. Computer Vision and Pattern Recognition. 2016, p. 488–496.

[89] Lu, S, Ren, X, Liu, F. Depth enhancement via low-rank matrix completion. In: IEEE Conf. Computer Vision and Pattern Recognition. 2014, p. 3390–3397.

[90] Richardt, C, Stoll, C, Dodgson, NA, Seidel, HP, Theobalt, C. Coherent spatiotemporal filtering, upsampling and rendering of rgbz videos. In: Computer Graphics Forum; vol. 31. Wiley Online Library; 2012, p. 247–256.

[91] Qi, F, Han, J, Wang, P, Shi, G, Li, F. Structure guided fusion for depth map inpainting. Pattern Recognition Letters 2013;34(1):70–76.

[92] McMillan Jr, L. An image-based approach to three-dimensional computer graphics. Ph.D. thesis; Citeseer; 1997.

[93] Dabov, K, Foi, A, Katkovnik, V, Egiazarian, K. Image denoising by sparse 3d transform-domain collaborative filtering. IEEE Trans Image Processing 2007;16(8):2080–2095.

[94] Eigen, D, Fergus, R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: Int. Conf. Computer Vision. 2015, p. 2650–2658.

[95] Kuznietsov, Y, Stückler, J, Leibe, B. Semi-supervised deep learning for monocular depth map prediction. arXiv preprint arXiv:170202706 2017;.

[96] Garg, R, Carneiro, G, Reid, I. Unsupervised cnn for single view depth estimation: Geometry to the rescue. arXiv preprint arXiv:160304992 2016;.

[97] Tatarchenko, M, Dosovitskiy, A, Brox, T. Multi-view 3d models from single images with a convolutional network. arXiv preprint arXiv:151106702 2015;.

[98] Chakrabarti, A, Shao, J, Shakhnarovich, G. Depth from a single image by harmonizing overcomplete local network predictions. In: Advances

in Neural Information Processing Systems. 2016, p. 2658–2666.

[99] Zhao, XR, Wang, X, Chen, QC. Temporally consistent depth map prediction using deep convolutional neural network and spatial-temporal conditional random field. Computer Science and Technology 2017;32(3):443–456.

[100] Riegler, G, Ferstl, D, Rüther, M, Bischof, H. A deep primal-dual network for guided depth super-resolution. arXiv preprint arXiv:160708569 2016;.

[101] Lei, J, Li, L, Yue, H, Wu, F, Ling, N, Hou, C. Depth map super-resolution considering view synthesis quality. IEEE Trans Image Processing 2017;26(4):1732–1745.

[102] Hui, TW, Loy, CC, Tang, X. Depth map super-resolution by deep multi-scale guidance. In: European Conf. Computer Vision. Springer; 2016, p. 353–369.

[103] Wang, Z, Hu, J, Wang, S, Lu, T. Trilateral constrained sparse representation for kinect depth hole filling. Pattern Recognition Letters 2015;65:95–102.

[104] Yang, NE, Kim, YG, Park, RH. Depth hole filling using the depth distribution of neighboring regions of depth holes in the kinect sensor. In: Int. Conf. Signal Processing, Communication and Computing. IEEE; 2012, p. 658–661.

[105] Chen, L, Lin, H, Li, S. Depth image enhancement for kinect using region growing and bilateral filter. In: Int. Conf. Pattern Recognition. IEEE; 2012, p. 3070–3073.

[106] Min, D, Lu, J, Do, MN. Depth video enhancement based on weighted mode filtering. IEEE Trans Image Processing 2012;21(3):1176–1190.

[107] Chen, WY, Chang, YL, Lin, SF, Ding, LF, Chen, LG. Efficient depth image based rendering with edge dependent depth filter and interpolation. In: Int. Conf. Multimedia and Expo. IEEE; 2005, p. 1314–1317.

[108] Daribo, I, Tillier, C, Pesquet-Popescu, B. Distance dependent depth filtering in 3d warping for 3dtv. In: Workshop on Multimedia Signal Processing. IEEE; 2007, p. 312–315.

[109] Lee, SB, Ho, YS. Discontinuity-adaptive depth map filtering for 3d view generation. In: Int. Conf. Immersive Telecommunications. ICST; 2009, p. 8.

[110] Xu, X, Po, LM, Ng, KH, Feng, L, Cheung, KW, Cheung, CH, et al. Depth map misalignment correction and dilation for dibr view synthesis. Signal Processing: Image Communication 2013;28(9):1023–1045.

[111] Jung, SW. Enhancement of image and depth map using adaptive joint trilateral filter. IEEE Trans Circuits and Systems for Video Technology 2013;23(2):258–269.

[112] Matsuo, T, Fukushima, N, Ishibashi, Y. Weighted joint bilateral filter with slope depth compensation filter for depth map refinement. In: Int. Conf. Computer Vision Theory and Applications. 2013, p. 300–309.

[113] Atapour-Abarghouei, A, Breckon, T. Depthcomp: Real-time depth image completion based on prior semantic scene segmentation. In: British Machine Vision Conference. BMVA; 2017,.

[114] Chen, C, Cai, J, Zheng, J, Cham, TJ, Shi, G. A color-guided, region-adaptive and depth-selective unified framework for kinect depth recovery. In: Int. Workshop on Multimedia Signal Processing. IEEE; 2013, p. 007–012.

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

[116] Hu, J, Hu, R, Wang, Z, Gong, Y, Duan, M. Color image guided locality regularized representation for kinect depth holes filling. In: Visual Communications and Image Processing. IEEE; 2013, p. 1–6.

[117] Berdnikov, Y, Vatolin, D. Real-time depth map occlusion filling and scene background restoration for projected-pattern based depth cameras. In: Graphic Conf. IETP. 2011,.

[118] Izadi, S, Kim, D, Hilliges, O, Molyneaux, D, Newcombe, R, Kohli, P, et al. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In: ACM Symp. User Interface Software and Technology. 2011, p. 559–568.

[119] Fu, D, Zhao, Y, Yu, L. Temporal consistency enhancement on depth sequences. In: Picture Coding Symposium. IEEE; 2010, p. 342–345.

[120] Sheng, L, Ngan, KN, Li, S. Temporal depth video enhancement based on intrinsic static structure. In: Int. Conf. Image Processing. IEEE; 2014, p. 2893–2897.

[121] Wang, J, An, P, Zuo, Y, You, Z, Zhang, Z. High accuracy hole filling for kinect depth maps. In: SPIE/COS Photonics Asia. 2014, p. 92732L–92732L.

[122] Camplani, M, Salgado, L. Adaptive spatiotemporal filter for low-cost camera depth maps. In: Int. Conf. Emerging Signal Processing Applications. IEEE; 2012, p. 33–36.

[123] Kim, SY, Cho, JH, Koschan, A, Abidi, MA. Spatial and temporal enhancement of depth images captured by a time-of-flight depth sensor. In: Int. Conf. Pattern Recognition. IEEE; 2010, p. 2358–2361.

[124] Xu, K, Zhou, J, Wang, Z. A method of hole-filling for the depth map generated by kinect with moving objects detection. In: Int. Symp. Broadband Multimedia Systems and Broadcasting. IEEE; 2012, p. 1–5.

[125] Lai, K, Bo, L, Ren, X, Fox, D. A large-scale hierarchical multi-view rgb-d object dataset. In: Int. Conf. Robotics and Automation. IEEE; 2011, p. 1817–1824.

[126] Zhang, L, Tam, WJ, Wang, D. Stereoscopic image generation based on depth images. In: Int. Conf. Image Processing; vol. 5. IEEE; 2004, p. 2993–2996.

[127] Tomasi, C, Manduchi, R. Bilateral filtering for gray and color images. In: Int. Conf. Computer Vision. IEEE; 1998, p. 839–846.

[128] Buades, A, Coll, B, Morel, JM. A non-local algorithm for image denoising. In: Int. Conf. Computer Vision and Pattern Recognition; vol. 2. IEEE; 2005, p. 60–65.

[129] Gangwal, OP, Djapic, B. Real-time implementation of depth map post-processing for 3d-tv in dedicated hardware. In: Int. Conf. Consumer Electronics. IEEE; 2010, p. 173–174.

[130] Kopf, J, Cohen, MF, Lischinski, D, Uyttendaele, M. Joint bilateral upsampling. ACM Trans Graphics 2007;26(3):96.

[131] Kim, Y, Ham, B, Oh, C, Sohn, K. Structure selective depth superresolution for rgb-d cameras. IEEE Trans Image Processing 2016;25(11):5227–5238.

[132] Petschnigg, G, Szeliski, R, Agrawala, M, Cohen, M, Hoppe, H, Toyama, K. Digital photography with flash and no-flash image pairs. In: ACM Trans. Graphics; vol. 23. 2004, p. 664–672.

[133] Liu, S, Lai, P, Tian, D, Gomila, C, Chen, CW. Joint trilateral filtering for depth map compression. In: Visual Communications and Image Processing. International Society for Optics and Photonics; 2010, p. 77440F–77440F.

[134] He, K, Sun, J, Tang, X. Guided image filtering. In: European Conf. Computer Vision. Springer; 2010, p. 1–14.

[135] Mueller, M, Zilly, F, Kauff, P. Adaptive cross-trilateral depth map filtering. In: 3DTV Conference. IEEE; 2010, p. 1–4.

[136] Nguyen, QH, Do, MN, Patel, SJ. Depth image-based rendering from multiple cameras with 3d propagation algorithm. In: Int. Conf. Immersive Telecommunications. ICST; 2009, p. 6.

[137] Nguyen, HT, Do, MN. Image-based rendering with depth information using the propagation algorithm. In: Int. Conf. Acoustics, Speech, and Signal Processing. 2005, p. 589–592.

[138] Badrinarayanan, V, Kendall, A, Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:151100561 2015;.

[139] Shen, J, Cheung, SC. Layer depth de-noising and completion for structured-light rgb-d cameras. In: IEEE Conf. Computer Vision and Pattern Recognition. 2013, p. 1187–1194.

[140] Vázquez, C, Tam, WJ, Speranza, F. Stereoscopic imaging: Filling disoccluded areas in depth image-based rendering. In: Optics East. 2006, p. 63920D–63920D.

[141] Meyer, F. Color image segmentation. In: Int. Conf. Image Processing and its Applications. IET; 1992, p. 303–306.

[142] Crabb, R, Tracey, C, Puranik, A, Davis, J. Real-time foreground segmentation via range and color imaging. In: Computer Vision and Pattern Recognition Workshops. 2008, p. 1–5.

[143] Ma, Y, Worrall, S, Kondoz, AM. Automatic video object segmentation using depth information and an active contour model. In: Workshop on Multimedia Signal Processing. IEEE; 2008, p. 910–914.

[144] Felzenszwalb, PF, Huttenlocher, DP. Efficient graph-based image segmentation. Computer Vision 2004;59(2):167–181.

[145] Levin, A, Lischinski, D, Weiss, Y. A closed-form solution to natural image matting. IEEE Trans Pattern Analysis and Machine Intelligence 2008;30(2):228–242.

[146] Tošić, I, Olshausen, BA, Culpepper, BJ. Learning sparse representations of depth. Selected Topics in Signal Processing 2011;5(5):941–952.

[147] Tosic, I, Drewes, S. Learning joint intensity-depth sparse representations. IEEE Trans Image Processing 2014;23(5):2122–2132.

[148] Harsha, GN, Majumdar, A, Ward, R. Disparity map computation for

stereo images using compressive sampling. Signal and Image Processing 2013;:804–809.

[149] Hirschmuller, H, Scharstein, D. Evaluation of cost functions for stereo matching. In: IEEE Conf. Computer Vision and Pattern Recognition. 2007, p. 1–8.

[150] Gortler, SJ, Grzeszczuk, R, Szeliski, R, Cohen, MF. The lumigraph. In: Conf. Computer Graphics and Interactive Techniques. ACM; 1996, p. 43–54.

[151] Islam, AT, Scheel, C, Pajarola, R, Staadt, O. Robust enhancement of depth images from depth sensors. Computers & Graphics 2017;.

[152] Rousseeuw, PJ. Least median of squares regression. American Statistical Association 1984;79(388):871–880.

[153] Gupta, S, Girshick, R, Arbeláez, P, Malik, J. Learning rich features from rgb-d images for object detection and segmentation. In: European Conf. Computer Vision. Springer; 2014, p. 345–360.

[154] Bo, L, Ren, X, Fox, D. Unsupervised feature learning for rgb-d based object recognition. In: Experimental Robotics. Springer; 2013, p. 387–402.

[155] Spinello, L, Arras, KO. People detection in rgb-d data. In: Int. Conf. Intelligent Robots and Systems. IEEE; 2011, p. 3838–3843.

[156] Domínguez, S, Zalama, E, García-Bermejo, JG, Worst, R, Behnke, S. Fast 6d odometry based on visual features and depth. In: Frontiers of Intelligent Autonomous Systems. Springer; 2013, p. 5–16.