DOI: 10.3217/978-3-85125-524-9-34

Depth-guided Disocclusion Inpainting for Novel View Synthesis*

Thomas Rittler[1,2], Matej Nezveda[1,2], Florian Seitner[2], and Margrit Gelautz[1]

Abstract— **The generation of novel views is a crucial processing step in 3D content generation, since it gives control over the amount of depth impression on (auto-)stereoscopic devices and enables free-viewpoint video viewing. A critical problem in novel view generation is the occurrence of disocclusions caused by a change in the viewing direction. Thus, areas in the novel views may become visible that were either covered by foreground objects or were located outside the borders in the original views. In this paper, we propose a depth-guided inpainting approach which relies on efficient patch matching to complete disocclusions along foreground objects and close to the image borders. Our method adapts its patch sizes depending on the disocclusion sizes and incorporates the depth information by focusing on the background scene content for patch selection. A subjective evaluation based on a user study demonstrates the effectiveness of the proposed approach in terms of quality of the 3D viewing experience.**

I. INTRODUCTION

The generation of novel views from an existing single view and its corresponding depth map is a crucial processing step for 3D content generation and processing. Such newly generated views enable the users to watch 3D content on different types of 3D displays, including multi-user autostereoscopic devices with a comfortable range of viewing perspectives, and navigate in 3D space for free-viewpoint video applications. The 2D input image and its associated depth map – known as 2D-plus-depth [11] – can be delivered by a variety of sources such as depth sensors based on time-of-flight or structured light (e.g., Microsoft's Kinect), stereo cameras, or 2D-to-3D conversion techniques.

A principal problem in novel view generation is the occurrence of disocclusions due to a change in the viewing direction. Some areas in the original views that were either covered by a foreground object or were located outside the image borders may become visible in the novel views. To deal with these disocclusions, one common approach is to pre-process the depth maps. In particular, filtering techniques are applied to the associated depth maps prior to the novel view generation [16]. Although this approach can reduce the appearance of disocclusions, it can also lead to spatial distortions in the scene geometry of the novel views.

Another approach is to use image inpainting techniques to fill in the disoccluded areas in the novel views with suitable estimates derived from the visible scene content. However, traditional inpainting algorithms (e.g., [5]) do not take into account additional knowledge provided by the depth data. For that reason, several inpainting strategies have been proposed that incorporate depth information during disocclusion filling [6], [8], [10], [13], [1], [15], [14]. While most related work aims at rendering photorealistic views, suitable inpainting approaches may also be required in the context of non-photorealistic rendering [9]. A few depth-induced inpainting strategies build upon PatchMatch (PM) [2], which is a randomized search algorithm that quickly finds correspondences between disjoint image patches. For example, He et al. [10] add the depth information to the PM algorithm by restricting the validity of patches used for inpainting. However, as their method was initially proposed for foreground object removal, the authors rely on a-priori depth information in the region to be filled which is not available when considering disocclusions. Morse et al. [13] extend PM from single image completion to stereo image pairs by not only incorporating depth information extracted from the stereo pairs but also allowing the matching of patches across the stereo pairs. However, the additional original view of a stereo image pair is not available in a 2D-plus-depth setup as considered in this work. Additionally, none of the aforementioned depth-guided inpainting approaches considers subjective quality assessment in the evaluation of their results. However, the results of Bosc et al. [3] indicate the need of subjective quality assessment in terms of novel views evaluation, as commonly used 2D quality metrics do not reflect the subjective quality of novel views. A very recent publication [4] gives an in-depth evaluation using the Middlebury ground truth data set, but does not incorporate user studies.

In this paper, we propose a depth-guided inpainting approach for disocclusion filling in novel views based on PM. Our approach incorporates the supplementary depth information to favor background patches during the disocclusion inpainting and uses adaptive patch sizes for efficient hole filling. We perform a paired comparison user study to evaluate our inpainting results in the context of stereoscopic viewing and present experimental results that show that our depth-guided inpainting approach yields better subjective quality compared to several earlier approaches.

The rest of the paper is organized as follows: Section 2 describes the proposed inpainting method. Section 3 provides details on our experimental setup. Section 4 presents the results of the user study along with some inpainting examples, and Section 5 concludes the paper.

*This work was supported by the Technology Agency of the City of Vienna (ZIT) under the project PAINT3D and finalized under the project Precise3D, funded by the Austrian Research Promotion Agency (FFG) and the Austrian Ministry BMVIT under the program ICT of the Future.

[2] emotion3D GmbH, Gartengasse 21/3, 1050 Vienna, Austria; {nezveda, seitner}@emotion3d.tv

[1] Institute of Software Technology and Interactive Systems, Vienna University of Technology, Favoritenstrasse 9-11/188-2, 1040 Vienna, Austria; margrit.gelautz@tuwien.ac.at

II. Proposed Approach

We suggest an inpainting technique that builds upon PM as an efficient strategy for finding patch correspondences based on color differences. The proposed approach incorporates adaptive patch sizes and search space restrictions based on depth information, as explained in the following subsections. First, the formalism of the general inpainting problem is recapped [5]: Let I be an input image and $\Omega \subseteq I$ a "hole" to be filled, called the *target* region. That is, Ω denotes all the missing pixels within I. Additionally, the *source* region Φ provides samples used in the infilling process. The goal is now to complete the missing region Ω with data from Φ so that the resulting image will be visually coherent. While conventionally $\Phi = I \setminus \Omega$, we restrict Φ to candidates from the image background as part of our approach.

A. Adaptive patch sizes

As opposed to iterative inpainting approaches that shrink the holes by successively copying patches of constant size, we perform the inpainting step only once at the end of the image completion chain, with the goal to avoid propagating erroneous inpainting results from one iteration step to the next. Our non-iterative approach is enabled by the usage of adaptive patch sizes. If fixed-size patches are used and the patch size is smaller than the size of Ω, there are some target patches containing no valid image information (see blue rectangle in Fig. 1a) that is required to compute the patch similarities.

For that purpose, a threshold τ_1 is specified to ensure a minimum percentage of valid pixels in each target patch. The corresponding patch size for each target pixel is determined by successively incrementing the patch dimensions until the percentage of the valid source pixels exceeds τ_1. Hence, the selected patches are smaller near the borders and are growing as the patch's central pixel is moving towards the hole's centroid, as illustrated in Fig. 1b. As a side effect, fewer patches are involved in the color synthesis of an individual pixel (based on weighted color averaging of overlapping patches) near the boundaries of Ω, which helps avoid blurring artifacts in these regions.

By introducing adaptive patch sizes it is guaranteed that the majority of the target patches contain a certain percentage of valid pixels. However, there may arise situations where the combination of target and source patches becomes impractical, as schematically illustrated in Fig. 1c. Hence, a second threshold τ_2 (equal to or smaller than τ_1) is specified to maintain the majority of valid pixels in the matching step and to ensure a minimal overlap between valid pixels of the target patch and the corresponding source patch.

B. Depth

There are two major reasons for disocclusions that cause blank areas in novel views: (a) areas that had been covered by a foreground object in the original view, and (b) areas along the image borders that had been outside the field of view in the original image. While scene depth is not taken into account when dealing with case (b), it is reasonable to fill occlusions of group (a) with image data obtained from background regions. As these holes emerge due to sharp depth transitions (i.e., depth discontinuities) at object boundaries, a target patch may comprise pixels that belong to foreground objects as well as pixels that are part of the background, as illustrated in Fig. 1d. Consequently, inpainting artifacts occur – hereinafter also referred to as *foreground color blur* – which are caused by color bleeding from the foreground. Therefore, depth information is incorporated in the matching stage to find appropriate patch correspondences and prevent foreground regions from being used for filling disoccluded regions.

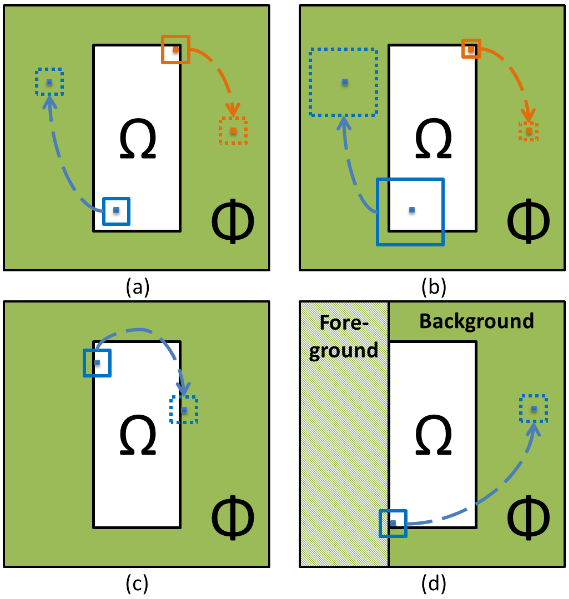

Fig. 1. Schematical overview of the basic concepts of our inpainting approach: (a) constant versus (b) adaptive patch size; (c) problem of non-overlapping valid pixels between target and source patch; (d) target patch comprising foreground and background pixels. Further details are given in the text.

Since depth information is not available in the target region, the depth values have to be synthesized first from the warped depth values in the surrounding. For every hole in Ω, each scanline is first filled by a constant value determined as the maximum depth value of the left and right pixel located at the hole boundary. Then, the minimum of the newly filled in depth values is selected as a lower bound of permissible depth levels in the nearest-neighbor search for target patches of the respective hole. An additional outlier removal based on the statistics of the depth histogram is applied to make the procedure more robust to depth map inaccuracies.

III. Experimental Setup

In order to investigate the effectiveness of our proposed inpainting algorithm on the perceived quality of stereoscopic images, a pair-wise comparison study was conducted. The stereo pairs used for evaluation were formed by the original left views and novel right views, i.e., synthesized views derived from the left views and the corresponding depth maps with disocclusions filled by inpainting. This section

TABLE I

This table lists the name, number (percentage) of disoccluded pixels and the characteristics of the images used in the subjective study.

Name	Disocclusions	Characteristics
Arm	54050 (2.6%)	low-textured background
Bird	29790 (1.4%)	moderately textured background
Crowd	57711 (2.7%)	cluttered repetitive background
Edge	51173 (2.5%)	highly textured background
Flower	50483 (2.4%)	repetitive background

describes the test material, the inpainting techniques used for comparison and the selected subjective methodology including a description of the test environment and subjects.

A. Dataset

All inpainting methods are evaluated on footage from a movie sequence. Five still images – termed as *Arm*, *Bird*, *Crowd*, *Edge* and *Flower* – have been chosen as test images, with a resolution of 1920 × 1080 pixels. The selected images cover different image characteristics including varying densities of background texture and diverse amounts of disoccluded pixels, as summarized in Table I.

B. Algorithms

We compare our depth-guided PM inpainting approach (DPM), which was described in Section 2, against our implementation of PM [2] with constant patch sizes of 51×51 pixels, the image completion function content-aware fill (CAF) of Adobe's Photoshop CS5[3], which does not use depth information, and horizontal background replication (HBR) [7]. We use the following, same constant parameter settings to generate the results: $\{\tau_1, \tau_2\} = \{10\%, 10\%\}$. The thresholds have been chosen to provide a small but reasonable amount of valid pixels to be used for patch matching while preventing target patches from becoming too large, which would lead to blurrier inpainting results and increase the overall runtime of the algorithm.

C. Subjective assessment procedure

The Pair Comparison (PC) method has been chosen to quantify the subjective ratings [12]. In the PC method, a pair of stimuli is compared and the subjects are asked to rate the quality of the stimuli in terms of preferences using a ternary scale (i.e., stimulus A is preferred, stimulus B is preferred, or stimuli A and B are equally preferred).

Particularly, using 4 inpainting approaches and 5 images, a total number of 30 pair comparisons had to be performed by each subject. Each pair was presented successively in random order. The subjects were allowed to switch interactively between the two stimuli of a pair. Moreover, each subject performed a trial run in which the test methodology was introduced.

We compute the quality score for each method by increasing its respective counter by 1 in case of a preference and 0.5 in case of an equal valuation. The accumulated value is then divided by the number of comparisons per method and by the total number of participants. Hence, the final score shows the percentage of comparisons "won", e.g., a value of 100 indicates that this method has always been preferred over any other approach.

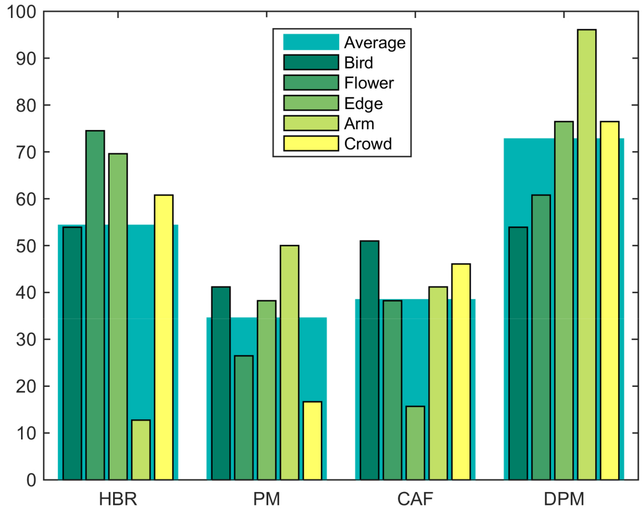

Fig. 2. Pair comparison scores of the subjective study.

The test sequences were displayed on a 23.6″ stereoscopic display (i.e., Acer GD245HQ) with a native resolution of 1920 × 1080 pixels and the NVIDIA 3D vision controller. To provide an ideal test setup, the room was darkened to avoid external visual disturbances and the viewing distance was set to one and a half times the screen size.

Seventeen non-expert observers (six female and eleven male observers aged between 17 and 49) participated in the study. All of the subjects were screened for visual acuity, color vision and stereo vision according to ITU-R BT.1438 recommendation [12].

IV. Results and Discussion

In Fig. 2, the PC scores obtained for the five test images are presented, grouped by the evaluated inpainting methods. Our proposed approach DPM performs best and is preferred on average in 72.75% of all comparisons. In contrast, the other PatchMatch-based inpainting methods PM and CAF attain significantly lower average PC scores of 34.51% and 38.43%, respectively.

Fig. 3 offers a closer look at some examples of inpainted regions. Regarding our approach, the study participants remarked a clear delineation of the foreground objects. A possible explanation is the reduction of artifacts caused by foreground color blur, which are mainly perceived as unnatural shadows of the objects (cf. DPM and PM in the second and third row of Fig. 3). Additionally, it can be seen that for holes at the image border, it is possible to inpaint coherent information by using adaptive instead of fixed-size patches.

The lower score of our approach (60.78%) compared to HBR (74.51%) for the image *Flower* may be caused by significant inaccuracies of the corresponding depth map. In

[3] http://www.adobe.com/technology/projects/content-aware-fill.html

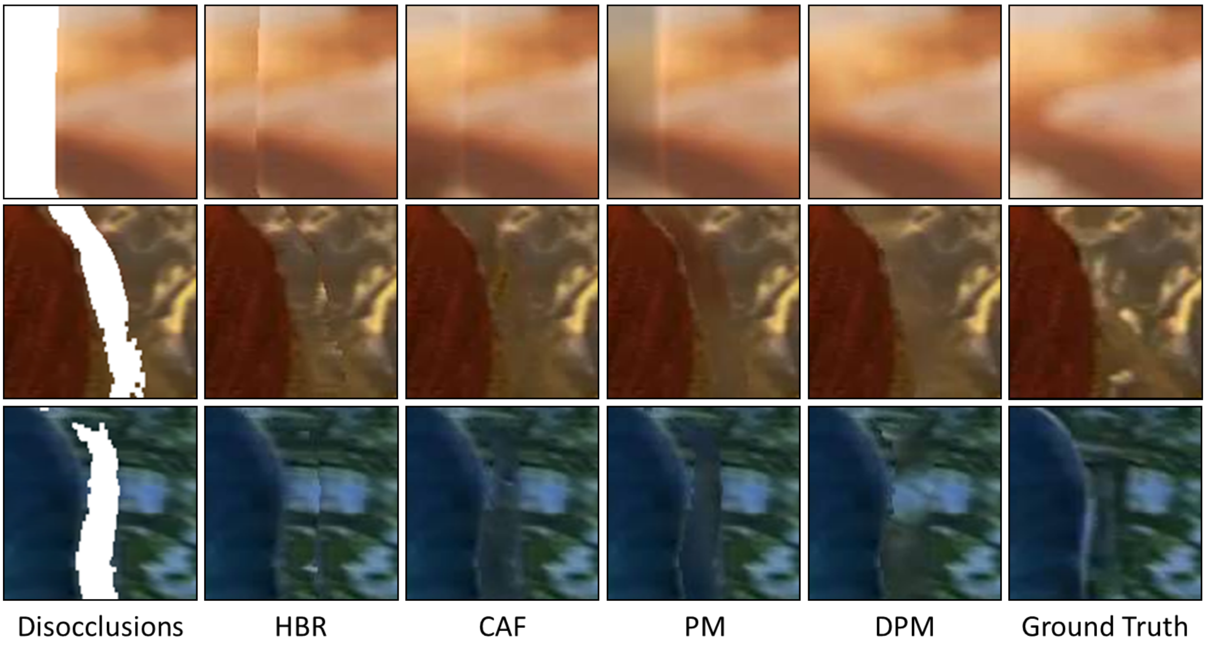

Fig. 3. Visual comparison of inpainting results. The first row shows a snippet including a hole at the border of image *Flower*. The second and third row show snippets including holes caused by depth discontinuities for images *Crowd* and *Edge*, respectively. Best viewed in color.

particular, parts of the background area have been erroneously labeled as foreground and thus are not taken into account in the patch matching step according to the predefined depth constraints. Consequently, artifacts are present in the inpainted region, which however could be avoided by adjusting the depth-based outlier removal.

Another interesting finding is the approximately uniform distribution of PC scores among the investigated inpainting methods for the image *Bird*. The observers declared that they found it hard to detect any differences, which might be due to the fact that *Bird* exhibits the smallest number of disoccluded pixels (see Table I). Additionally, these disoccluded pixels are located in primarily low textured areas outside the main focus of the observer's attention. Similarly, the better result of the relatively straightforward inpainting method HBR (54.31% on average) compared to PM (34.51% on average) and CAF (38.43% on average) may lie in the fact that in our test images the inconsistencies caused by HBR inpainting become mainly noticeable in highly textured background regions near the image margin, whereas observers tend to pay more attention to the central image area covered by the foreground object.

V. Conclusion

We have introduced a depth-guided inpainting approach that addresses the filling of disocclusions in novel views. Our method is based on efficient patch matching and produces visually very satisfying results for both disocclusions at image borders and disocclusions along the boundaries of foreground objects. Our method adapts its patch sizes to the disocclusion sizes. For disocclusions along objects, we additionally incorporate the depth information by focusing on the background scene content for patch selection. A subjective evaluation of the stereoscopically perceived quality of the synthesized novel views showed the effectiveness of our proposed approach. For future work, we plan to extend our technique to disocclusion inpainting of video sources.

References

[1] I. Ahn and C. Kim, "Depth-based disocclusion filling for virtual view synthesis," in *IEEE International Conference on Multimedia and Expo*, 7 2012, pp. 109–114.

[2] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 28, pp. 24:1–24:11, 2009.

[3] E. Bosc, R. Pepion, P. L. Callet, M. Pressigout, and L. Morin, "Reliability of 2D quality assessment methods for synthesized views evaluation in stereoscopic viewing conditions," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, 2012, pp. 1–4.

[4] P. Buyssens, O. Le Meur, M. Daisy, D. Tschumperlé, and O. Lézoray, "Depth-guided disocclusion inpainting of synthesized RGB-D images," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 525–538, 2017.

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

[6] I. Daribo and B. Pesquet-Popescu, "Depth-aided image inpainting for novel view synthesis," in *IEEE International Workshop on Multimedia Signal Processing*, 2010, pp. 167–170.

[7] M. Eisenbarth, F. Seitner, and M. Gelautz, "Quality analysis of virtual views on stereoscopic video content," in *19th International Conference on Systems, Signals and Image Processing*, 2012, pp. 333–336.

[8] J. Gautier, L. M. Josselin, and C. Guillemot, "Depth-based image completion for view synthesis," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, 2011, pp. 1–4.

[9] M. Gelautz, E. Stavrakis, and M. Bleyer, "Stereo-based image and video analysis for multimedia applications," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Systems (XXth ISPRS Congress)*, 2004, pp. 998–1003.

[10] L. He, M. Bleyer, and M. Gelautz, "Object removal by depth-guided inpainting," in *Austrian Association for Pattern Recognition Workshop*, vol. 2, 2011, pp. 1–8.

[11] ISO/IEC 23002-3, "Information technology – MPEG video technologies – Part 3: Representation of auxiliary video and supplemental information," 2007.

[12] ITU-R Recommendation BT.1438, "Subjective assessment of stereoscopic television pictures," 2000.

[13] B. Morse, J. Howard, S. Cohen, and B. Price, "Patchmatch-based content completion of stereo image pairs," in *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, 2012, pp. 555–562.

[14] S. M. Muddala, R. Olsson, and M. Sjöström, "Spatio-temporal consistent depth-image-based rendering using layered depth image and inpainting," *EURASIP Journal on Image and Video Processing*, vol. 2016, no. 1, pp. 1–19, 2016.

[15] S. M. Muddala, M. Sjostrom, and R. Olsson, "Depth-based inpainting for disocclusion filling," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, 2014, pp. 1–4.

[16] M. Nezveda, N. Brosch, F. Seitner, and M. Gelautz, "Depth map post-processing for depth-image-based rendering: A user study," in *IS&T/SPIE Electronic Imaging*, 2014, pp. 90 110K–90 110K.