Disocclusion Handling Using Depth-Based Inpainting

Suryanarayana M. Muddala, Roger Olsson and Mårten Sjöström
Dept. of Information Technology and Media
Mid Sweden University
Sundsvall, Sweden
suryanarayana.muddala@miun.se, roger.olsson@miun.se and marten.sjostrom@miun.se

Abstract—**Depth image based rendering (DIBR) plays an important role in producing virtual views using 3D-video formats such as video plus depth (V+D) and multi view-video-plus-depth (MVD). Pixel regions with non-defined values (due to disoccluded areas) are exposed when DIBR is used. In this paper, we propose a depth-based inpainting method aimed to handle disocclusions in DIBR from V+D and MVD. Our proposed method adopts the curvature driven diffusion (CDD) model as a data term, to which we add a depth constraint. In addition, we add depth to further guide a directional priority term in the exemplar based texture synthesis. Finally, we add depth in the patch-matching step to prioritize background texture when inpainting. The proposed method is evaluated by comparing inpainted virtual views with corresponding views produced by three state-of-the-art inpainting methods as references. The evaluation shows the proposed method yielding an increased objective quality compared to the reference methods, and visual inspection further indicate an improved visual quality.**

Keywords-3D; video plus depth; warping; depth-image-based rendering; inpainting;

I. Introduction

In recent years, Three Dimensional Television (3DTV) and Free Viewpoint Television (FTV) have become hot topics in the 3D research area. A common way to transmit the 3D content required for these applications is to use video-plus-depth (V+D) and multi view-plus-depth (MVD) formats, as these ensure a display agnostic rendering of virtual views for both stereoscopic and autostereoscopic multiview displays. A required tool for V+D and MVD formats is view synthesis, which creates content suitable for each specific display type. A fundamental view synthesis method is depth-image-based rendering (DIBR), which produces virtual views using pixel dense texture and depth information. Unfortunately DIBR brings inherent artifacts, mainly caused by disocclusions [1]. Disocclusions are areas that are occluded in an original view that is stored in the format, which become visible in rendered virtual views. Although MVD permits virtual views to be rendered using information from not one but two or more V+D data sets, there still exists a disocclusion problem that needs to be addressed. Mainly due to content with a baseline that significantly differs from that required by a specific display.

Inpainting methods aim to solve the disocclusion problem by filling the unknown regions using neighborhood informa-

tion. Disoccluded areas can be considered as missing texture information alone, as is being done by texture synthesis methods [2]. Criminisi et al. proposed an efficient image inpainting technique that combines the structural and textural propagation into the missing regions [3]. However, this method was not aimed at V+D or MVD formats and thereby could not recognize the differences between foreground (objects closer to the camera) and background parts (objects away from the camera) in a virtual view. As a result it propagates foreground information into the disoccluded areas, which should only contain background information. Daribo et al. extended the exemplar based inpainting to address this limitation by introducing the depth constraint. However, this method only reduces the problem to a degree as it still partly propagates the foreground information into disoccluded regions [4]. Gautier et al. extended the Criminisi method by considering the 3D structure tensor as a data term that identifies the strongest structure in the neighborhood, and added the depth information to calculate the required inpainting priorities [5]. Worth noting with these previous work is that both Daribo et al. and Gautier et. al relies on having true depth map available at the rendered virtual view position. This assumption is in general not feasible or realistic since the depth map of the virtual view also must be estimated.

This paper proposes a novel method to inpainting for V+D and MVD based DIBR. The proposed method relies on the fundamental method introduced in [3] but enhanced using the available depth information. In contrast to [4], [5], we have not relied on having access to a true depth map but instead considered a more general case with having a warped depth map available in our inpainting process.

The outline of the paper is as follows: The related work is briefly reviewed in Section II and the proposed inpainting method is presented in Section III. The test arrangement and evaluation criteria are described in Section IV. The results and analysis are given in Section V and finally we conclude the work in Section VI.

II. Related work

Criminisi et al. introduced the exemplar based texture synthesis, which effectively replicates both structure and
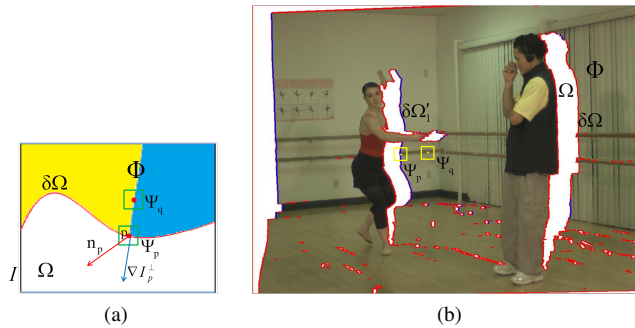
Figure 1. Schematic illustration: (a) notation diagram; (b) warped view with notations.

texture by using the advantages of partial differential equations (PDE) based inpainting method and non-parametric texture synthesis. The quality of the inpainted image is highly dependent on the order in which filling is performed.

For an input image I with an empty region Ω, also known as hole, the source region Φ (the remaining part of the image except the empty region) is defined as $\Phi = I - \Omega$. The boundary between Φ and Ω is denoted as $\delta\Omega$ (see Fig. 1). The basic steps of Criminisi's algorithm are (i) Computing the priorities on the boundary region and (ii) Finding the best match using patch matching. Suppose a patch $\Psi_{\mathbf{p}}$ centered at a pixel $\mathbf{p}$ for some $\mathbf{p} \in \delta\Omega$ and the priority is computed as the product of two terms:

$$P(\mathbf{p}) = C(\mathbf{p}) \cdot D(\mathbf{p}), \quad (1)$$

where $C(\mathbf{p})$ is the confidence term indicating the amount of non-missing pixels in a patch and the data term $D(\mathbf{p})$ gives importance to the isophote direction.

Once all priorities on boundary $\delta\Omega$ are computed, the highest priority patch $\Psi_{\hat{\mathbf{p}}}$ centered at $\hat{\mathbf{p}}$ is selected to be filled first. A block matching algorithm is used to find the best similar patch $\Psi_{\hat{\mathbf{q}}}$ from which to fill-in the missing pixels:

$$\Psi_{\hat{\mathbf{q}}} = \arg \min_{\Psi_{\mathbf{q}} \in \Phi} \{d(\Psi_{\hat{\mathbf{p}}}, \Psi_{\mathbf{q}})\}, \quad (2)$$

where d is the distance between two patches defined as sum of squared difference (SSD). After the most similar patch $\Psi_{\hat{\mathbf{q}}}$ is found, the values of the hole pixels in the target patch $\acute{\mathbf{p}}|\acute{\mathbf{p}} \in \Psi_{\hat{\mathbf{p}}} \cap \Omega$ are copied from their corresponding pixels inside $\Psi_{\hat{\mathbf{q}}}$. Once the patch $\Psi_{\hat{\mathbf{p}}}$ is filled, the confidence term $C(\mathbf{p})$ is updated as follows:

$$C(\mathbf{q}) = C(\hat{\mathbf{p}}), \forall \mathbf{q} \in \Psi_{\hat{\mathbf{p}}} \cap \Omega. \quad (3)$$

Daribo et al. extended the Criminisi method first by introducing a depth regularity term in the priority term calculation in (1). The depth regularity term is defined as the inverse variance of the depth patch centered at $\mathbf{p}$. Their depth regularity term is described as controlling the inpainting process such that the filling order favors the background. Furthermore, the patch matching step is modified by searching for a best patch in both the texture and the depth domain.

Gautier et al. followed the Darios method in considering depth map to help the inpainting process, but introduced a 3D tensor as a data term in the priority calculation of (1) and a one-sided priority to restrict the filling direction. In the patch matching step they also used a weighted combination of the best patches as the final selected patch.

III. PROPOSED INPAINTING METHOD

The novelty of our proposed depth-based inpainting method can be described in three steps:
- A. Depth guided directional priority
- B. Depth included curvature data term
- C. Depth-based source region selection

Fig. 2 shows how these steps relate to the general inpainting process. Step A, consists of defining a depth guided directional priority that selects background patches to be filled first. In Step B, we adopt the Curvature Driven Diffusion (CDD) model similarly to [6] as data term $D(\mathbf{p})$, and extend the CDD model by incorporating depth information. Finally, Step C excludes foreground information from the source region, using depth constraints derived from the warped depth. In the patch matching, a weighted combination of - N best patches is used to define the target patch.

A. Depth guided direction priority

In this step, the boundary extraction block of Fig. 2 is improved by using depth information to guide the filling such that it starts from the background. This because disocclusions result from depth discontinuities between foreground and background, which makes filling the disocclusion from the horizontal background side reasonable. The background side of the disocclusion is obtained as follows. First, a one sided boundary $\delta\Omega_1$ of the disocclusion area is obtained by applying the convolution operation on a disocclusion map (DM) as given in (4). Second, the directional priority selection is further improved by using a depth constraint on $\delta\Omega_1$, such that pixels whose depth values are less than M percent of the maximum depth value in the warped depth map are selected (see the blue colored border in Fig. 1(b)):

$$\delta\Omega_1 = DM * H \quad (4)$$

$$\delta\Omega_1' = \delta\Omega_1(\mathbf{q})|_{\mathbf{q} \in \delta\Omega_1 \cap (Z(\mathbf{q}) < M \cdot max(Z))}, \quad (5)$$

where $\delta\Omega_1'$ is the depth guided boundary, Z is the depth map and $Z(\mathbf{q})$ is the depth value at pixel location $\mathbf{q}$. The convolution kernel H is defined as follows, depending of from which direction the warp is performed:

$$H = \begin{cases} \begin{bmatrix} 0 & 0 & 0 \\ 1 & -8 & 0 \\ 0 & 0 & 0 \end{bmatrix} & \text{if left warped view;} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & -8 & 1 \\ 0 & 0 & 0 \end{bmatrix} & \text{if right warped view.} \end{cases} \quad (6)$$

Once the hole boundary is obtained, using (4) and (5), priorities are calculated according to (1) utilizing the proposed data term (10). Then the holes in the background regions are filled using the depth guided direction priority. The filling process continues with the one sided boundary priority and finally the holes which are not filled using the one sided boundary priority are processed with total boundary extraction.

B. Depth included data term

As the data term in the general inpainting process we adopt, and add depth to, the CDD model in order to consider the depth curvature along with the texture. The CDD model uses the strength and geometry of an isophote [7], where the latter obtained using scalar curvature. The CDD model is defined as follows:

$$g(s) = s^\alpha, s > 0, \alpha \geq 1 \quad (7)$$

$$k_{\mathbf{p}} = \nabla \cdot \left(\frac{\nabla I_{\mathbf{p}}}{|\nabla I_{\mathbf{p}}|} \right) \quad (8)$$

$$\frac{\partial I_{\mathbf{p}}}{\partial t} = \nabla \cdot \left(\frac{g(|k_{\mathbf{p}}|)}{|\nabla I_{\mathbf{p}}|} \nabla I_{\mathbf{p}} \right), \quad (9)$$

where $k_{\mathbf{p}}$ is the curvature of the isophote through some pixel $\mathbf{p}$, $\nabla \cdot$ is the divergence at $\mathbf{p}$, and g is the control function to adjust the curvature. The conductive coefficient of CDD model is influenced by the isophote strength and curvature. By incorporating the CDD model as a data term in the proposed method and setting $\alpha = 1$ in (7), the data term becomes:

$$D(\mathbf{p}) = \left| \nabla \cdot \left(\frac{k_{\mathbf{p}}}{|\nabla I_{\mathbf{p}}|} \nabla I_{\mathbf{p}} \right) \right|, \quad (10)$$

The depth information is considered as an additional channel along with R, G, and B when calculating the curvature and isophote values.

C. Depth-based source region selection

The patch-matching step in the proposed inpainting method is an improvement of the method of [4] and [5]. The improvement consists of classifying the source region using depth information, in order to select similar patches from the
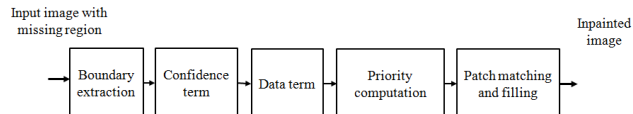

Figure 2. Block diagram of the inpainting method.

nearest depth range. The idea of separating the background region has been previously employed by [8] using patch averages. However, here we classify the source region to enhance the patch-matching step. By considering Φ to be the known source region, which contains both foreground and background regions we avoid patch selection from foreground region by sub-dividing Φ using depth threshold Z_c according to:

$$\Phi_b = \Phi - \Phi_f, \quad (11)$$

where Φ_f is the source region whose depth values are higher than the depth threshold Z_c.

The depth threshold has two different values depending on the variance of the depth patch. If the variance of the depth patch is greater than the threshold γ, the patch might contain unwanted foreground values. The average value of the depth patch is then instead chosen to deduct the foreground parts. Otherwise, the patch contains the constant or continuous area values and so the maximum value in the depth patch is used as the depth threshold to get the best patch according to the depth level. So the depth threshold Z_c is defined as follows:

$$Z_c = \begin{cases} \overline{Z_{\hat{\mathbf{p}}}} & \text{if } var(Z_{\hat{\mathbf{p}}}(\mathbf{q})|_{\mathbf{q} \in \Psi_{\hat{\mathbf{p}}} \cap \Phi}) > \gamma; \\ \max(Z_{\hat{\mathbf{p}}}) & \text{otherwise.} \end{cases} \quad (12)$$

$\Psi_{\hat{\mathbf{p}}}$ is the highest priority patch, $Z_{\hat{\mathbf{p}}}$ is the depth patch centered at $\hat{\mathbf{p}}$; and $\overline{Z_{\hat{\mathbf{p}}}}$ is the average value of the depth patch. $Z_{\hat{\mathbf{p}}}(\mathbf{q})$ is the depth value at pixel $\mathbf{q}$ and γ is the depth variance threshold.

Once the highest priority patch $\Psi_{\hat{\mathbf{p}}}$ and depth-based source region Φ_b defined in (11) are computed, we search for the best N number of patches within the source region.

$$\Psi_{\hat{\mathbf{q}}} = \arg \min_{\Psi_{\mathbf{q}} \in \Phi_b} \{d(\Psi_{\hat{\mathbf{p}}}, \Psi_{\mathbf{q}}) + \beta \cdot d(Z_{\hat{\mathbf{p}}}, Z_{\mathbf{q}})\}, \quad (13)$$

where d is SSD, and β is a parameter to emphasize the depth. The depth map is considered in the patch matching process to find the similar patches in the depth domain and simultaneously fill the disocclusion in the depth map along with the texture.

The best N number of patches obtained from the patch matching step are not equally reliable [9]. Therefore, we adopt a weighted average of N patches when fill the missing information of the disocclusion.
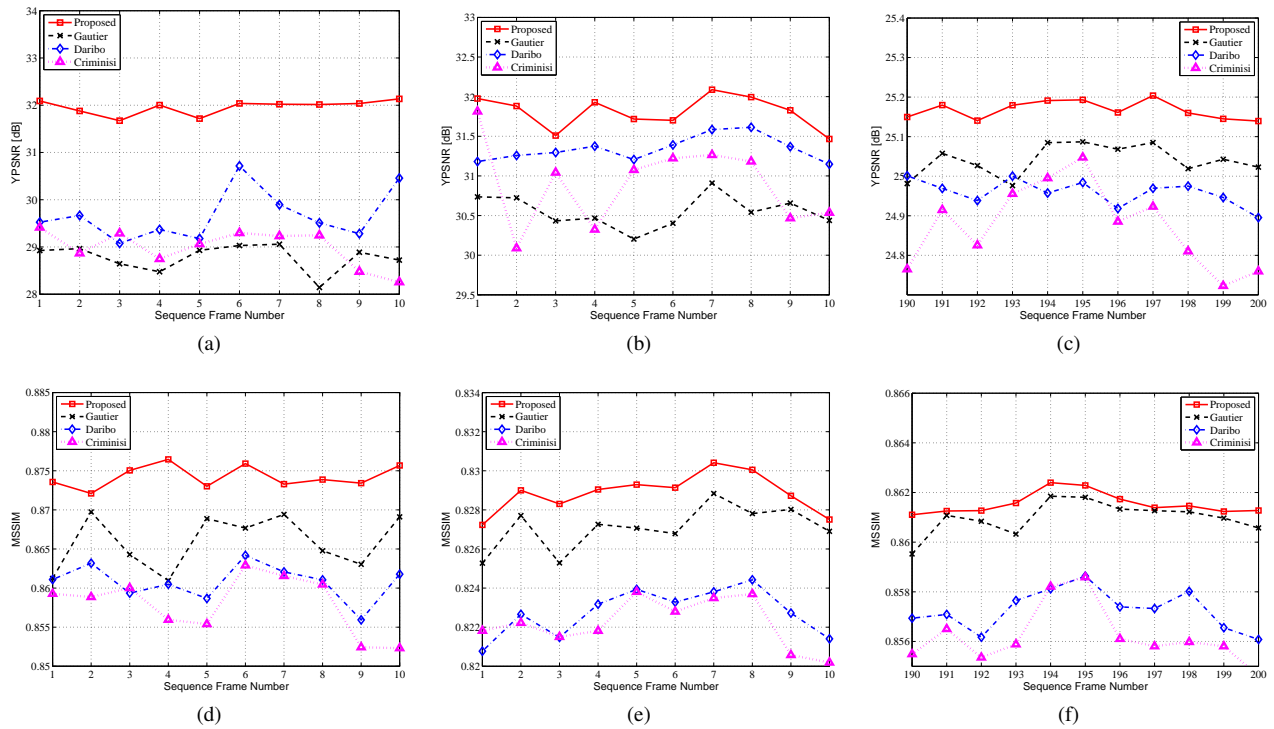
Figure 3. Objective metrics PSNR and MSSIM of the investigated sequences; PSNR for each rendered frame at view position 4 of "Ballet" (a), at view position 4 of "Break dancers" (b), at view position 4 of "Lovebird1" (c); MSSIM for each rendered frame at view position 4 of "Ballet" (d), at view position 4 of "Break dancers" (e) and at view position 4 of "Lovebird1" (f).

IV. TEST ARRANGEMENT AND EVALUATION CRITERIA

Results from the proposed method are evaluated by objective measurements as well as visual inspection. A set of 10 frames are selected from the three MVD sequences "Ballet", "Break dancers" and "Lovebird1" for objective evaluation. All three sequences have a spatial resolution of 1024x768 pixels. The two first sequences are captured with 8 cameras and a baseline of 300 mm and 400 mm respectively [10]. The third sequence is captured with 12 cameras and a baseline of 35 mm [11]. The chosen MVD sequences have characteristics that make them suitable for testing different disocclusion filling attributes of inpainting methods. The "Ballet" sequence has large depth discontinuities at two different depth ranges, which results in big disocclusion areas at different depth levels. The "Break dancers" sequence has a large number of objects located in almost the same depth level. The "Lovebird1" sequence has complex texture and more structured background, with larger depth discontinuities.

All sequences are used in a DIBR of V+D scenario with full reference evaluation possible, i.e. access to ground truth texture and depth is available. More specifically, the first two sequences renders view 4 from view 5 and in the third "Lovebird1" sequence, view 4 is rendered from view 6. Post processing is applied on the rendered view and the depth

to remove the cracks and ghosting artifacts before starting the inpainting process. Important parameters of the proposed inpainting method is a patch matching window size of 120 pixels, $M = 0.4$, $\gamma = 80$ in (12), $\beta = 3$ in (13), and $N = 5$. For evaluation purposes, two objective evaluation metrics are considered: peak signal to noise ratio of the luminance component (Y-PSNR) and mean structural similarity index (MSSIM).

V. RESULTS AND ANALYSIS

The rendered and inpainted virtual views were generated and compared for disocclusion handling using methodology presented in the previous. Results from the objective evaluation are shown in Fig. 3. The PSNR and MSSIM graphs consistently demonstrate that the proposed depth-based inpainting method performs better than the Criminisi, Daribo and Gautier methods. Fig. 4 shows the rendered views with disoccluded areas (denoted with white color) and inpainting methods results of the "Ballet" and "Lovebird1" images for visual comparison. Note that the disocclusion regions in Fig. 4(c) and (d) are filled with foreground information since no depth is available to assist the filling process. Although the Daribo and Gautier methods are aided with true depth information, there still exists artifacts in the virtual views disocclusions. The proposed inpainting method shows visual improvements with respect to all the

reference methods, although it is operating in a more realistic setting where only warped depth information is available. The results from Fig. 4(i) and (j) show that the proposed method propagates the required neighboring information into the disocclusions region, retaining both smooth areas (at the left side of the "Ballet" image) and continuing neighborhood structure (on the curtain in the "Ballet" image and at the head of the women in the "Lovebird1" image). The proposed method still show some jaggedness effects at object boundaries, which is due to constraints on the source region selection and patch matching. In summary, the proposed method performs better than the reference methods both objectively and visually, which is a result of utilizing the depth-based direction priority, the depth included data term and search constraints incorporating depth information.

VI. CONCLUSION

We have proposed a new depth-based inpainting method to fill disocclusions in a virtual view by employing a depth guided directional term, a depth enhanced curvature driven diffusion model and depth searching constraints in the exemplar based texture synthesis. The results of the proposed method have been compared with the inpainting method of Criminisi, Daribo and Gautier using objective quality metrics and visual inspection. Both ways of evaluating consistently demonstrates that the proposed method offers an improved quality. In future work, we will focus on reducing the computational time that is inherent with processing large disocclusions, temporal consistency, and more elaborate subjective tests to further validate our results.

ACKNOWLEDGMENT

REFERENCES

[1] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XI, Jan. 2004, pp. 93–104.

[2] Z. Tauber, Z. N. Li, and M. S. Drew, "Review and preview: Disocclusion by inpainting for image-based rendering," IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, vol. 37, no. 4, 2007, pp. 527–540.

[3] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," IEEE Transactions on Image Processing, vol. 13, 2004, pp. 1200–1212.

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

[5] J. Gautier, O. L. Meur, and C. Guillemot, "Depth-based image completion for view synthesis," in 3DTV conference, 2011, pp. 1–4.

[6] S. Li, R. Wang, J. Xie, and Y. Dong, "Exemplar image inpainting by means of curvature-driven method," in Computer Science and Electronics Engineering (ICCSEE), vol. 2, march 2012, pp. 326 –329.

[7] T. F. Chan and J. Shen, "Non-texture inpainting by curvature-driven diffusions (cdd)," J. Visual Comm. Image Rep, vol. 12, 2001, pp. 436–449.

[8] I. Ahn and C. Kim, "Depth-based disocclusion filling for virtual view synthesis," in ICME, 2012, pp. 109–114.

[9] Y. Wexler, E. Shechtman, and M. Irani, "Space-time completion of video," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 3, 2007, pp. 463–476.

[10] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," ACM Trans. Graph., vol. 23, no. 3, Aug. 2004, pp. 600–608.

[11] G. M. Um, G. Bang, N. Hur, J. Kim, and Y. S. Ho, "3d video test material of outdoor scene," ISO/IEC JTC1/SC29/WG11/M15371, April 2008.

Figure 4. Illustration of the different inpainting method results for the investigated sequence frames "Ballet" first frame in the coulmn1 and "Lovebird1" 190th frame in column 2; (a)(b) rendered view images (disocclusions are represented with white regions); (c)(d) The results of Criminisi method; (e)(f) The results of Daribo method; (g)(h) The results of Gautiers method; (i)(j) The results of Proposed method.