Depth-Included Curvature Inpainting for Disocclusion Filling in View Synthesis

Suryanarayana M. Muddala, Roger Olsson, and Mårten Sjöström

Dept. of Information and Communication Systems
Mid Sweden University
Sundsvall, Sweden 85170
Email: marten.sjostrom@miun.se

***Abstract*—Depth-image-based-rendering (DIBR) is the commonly used for generating additional views for 3DTV and FTV using 3D video formats such as video plus depth (V+D) and multiview-video-plus-depth (MVD). The synthesized views suffer from artifacts mainly with disocclusions when DIBR is used. Depth-based inpainting methods can solve these problems plausibly. In this paper, we analyze the influence of the depth information at various steps of the depth-included curvature inpainting method. The depth-based inpainting method relies on the depth information at every step of the inpainting process: boundary extraction for missing areas, data term computation for structure propagation and in the patch matching to find best data. The importance of depth at each step is evaluated using objective metrics and visual comparison. Our evaluation demonstrates that depth information in each step plays a key role. Moreover, to what degree depth can be used in each step of the inpainting process depends on the depth distribution.**

***Keywords*—3D; video plus depth; multiview video plus depth; 3D warping; depth-image-based rendering; image inpainting; disocclusion filling.**

I. INTRODUCTION

In recent years, Three Dimensional Television (3DTV) and Free Viewpoint Television (FTV) have become attractive in the 3D research area. The 3D video formats video-plus-depth (V+D) and multiview video-plus-depth (MVD) are efficient ways to send the 3D content to the end user. These data types provide backward compatibility with 2D displays, and at the same time enable rendering of virtual views corresponding to the requirement of current and future stereo and multiview 3D displays. Depth-image-based rendering (DIBR) is a fundamental method for producing these virtual views using the video or texture component and the depth per pixel information. A major concern when using DIBR on V+D content is the artifacts caused by disocclusions, i.e., areas that are occluded in the original view become visible in the rendered view when DIBR is used. When using MVD, DIBR methods can access multiple V+D components and thereby fill disocclusions in virtual views from several known views, which reduce the disocclusion problem yet does not solve it. Hence, the disocclusion problem inherent in the above 3D video formats is still an open question that needs to be addressed in order for applications such as 3DTV and FTV to be successful. The present work considered horizontal disparity for creating images on stereo or multiview displays. The depth-based inpainting [1] method does not explain the importance of depth at every step of the inpainting process, which raises the question how the depth information influence the total inpainting process in addressing the disocclusions.

There are several methods that aim to reduce the disocclusions by filling the holes they constitute, including preprocessing of depth map [2], linear interpolation, and inpainting. Preprocessing of the depth map before warping reduces the disocclusion but causes geometrical distortions in the resultant view. Using linear interpolation is a simple approach that gives acceptable results when the area of the disocclusion is smaller and less noticeable. However, when the disocclusions grow bigger, linear interpolation results a stark contrast to the structure of the texture outside the holes borders due to stretching of the border pixels into the holes. In general, inpainting methods aim to solve the missing areas by filling the unknown regions using neighborhood information. The disocclusion problems can be considered as missing texture information alone and it can be recovered by using texture synthesis methods [3]. Due to inpainting being an ill-posed problem the target of inpainting methods is to produce filled disocclusions that are as perceptually plausible as possible. In the following we will classify inpainting methods into the following categories: textural inpainting methods and structural inpaiting methods.

In texture inpainting methods, the missing regions are filled by replicating the repetitive patterns in the image, which surrounds the disocclusion. Many of these methods rely on Markov Random Field (MRF) to model the local patterns, and using small amount of known textures to generate new textures to fill the missing areas [4]. Texture inpainting methods perform well for uniform texture images but faces problems with real world images as they consist of a mixture of different textures and linear structures.

Structural inpainting methods, imitate the idea of manual inpainting that is propagating the linear structures present in the neighborhood of the missing regions using diffusion process [5]. Structural inpainting methods based on partial differential equations (PDE) fill the missing regions iteratively. These methods preserve and propagate linear structures in to missing regions. The disadvantages associated with these methods are blurring and structure discontinuity for large missing areas due to the diffusion process and not knowing the edge information. The total variation approach and curvature driven diffusion (CDD) model follow the anisotropic

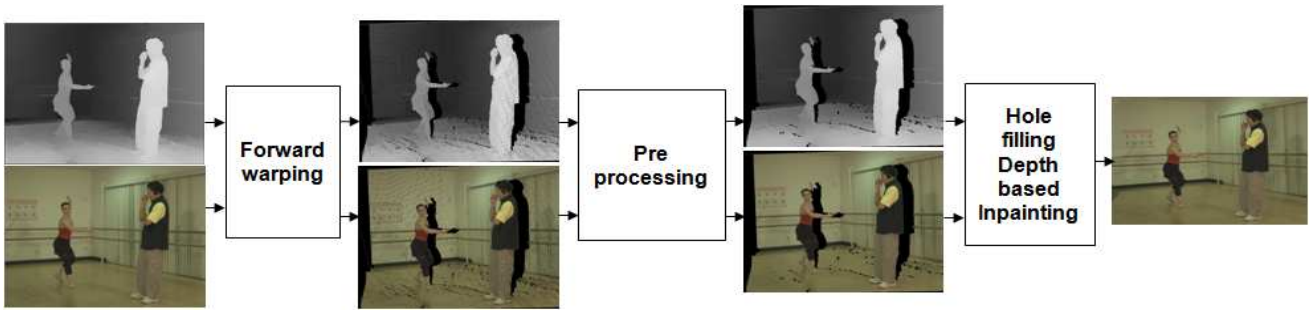


Figure 1. Block diagram of the rendering method using inpainting.

diffusion, which means diffusion strength reduces with respect to the structure variations [6] [7]. Total variation approach depends only on isotropic strength, whereas CDD depends on both isotropic strength and geometries. Although incorporating many constraints and higher order PDEs in the inpainting process, the resulted inpainted views still suffer from blurring artifacts for larger missing regions.

Criminisi et al. [8] proposed an efficient exemplar-based image inpainting technique, a kind of hybrid inpainting that combines the benefits of structural and textural inpainting methods to fill the missing regions. However, this method introduces inconsistent structures in to disocclusions areas due to not considering what is foreground (objects closer to the camera) and background (objects away from the camera) in the rendered view. Daribo et al. [9] extended the exemplar-based inpainting to address this limitation by introducing the depth constraint. However, hole filling with this method causes noticeable artifacts around foreground objects in the virtual view. Gautier et al. [10] extended the Criminisi et al. method by taking into account the structure tensor as a data term that identifies the strongest structure in the neighborhood, and the depth information to fill the holes. Moreover, filling with these methods require true depth information. Usually, the depth at the required camera view position needs to be estimated. By having access to the true depth at the virtual camera view position disocclusions can be filled according to depth level.

We proposed depth-included curvature inpainting method that consider the scene depth in all parts of the inpainting process [1]. This paper presents the method in more detail and also performs an analysis of how the depth information affects the quality of the inpainting result. The proposed method also relies on the fundamental method introduced in [8], but extends and improves upon it by using the available depth information at different stages of the inpainting process. In contrast to [9], [10], the proposed depth-included curvature inpainting method does not rely on having access to a true depth map but instead considers a more general and realistic case of having access to a warped depth when performing the inpainting.

The outline of the paper is as follows: A more general review of depth-image-based rendering is given in Section II. A selected summary of related work within the field of inpainting is presented in Section III and the proposed depth-included curvature inpainting method is described in Section IV. The methodology used in this work, including test arrangement and evaluation criteria, are described in Section V, followed by results and in Section VI. Finally, Section VII concludes the work.

II. Depth-image based rendering

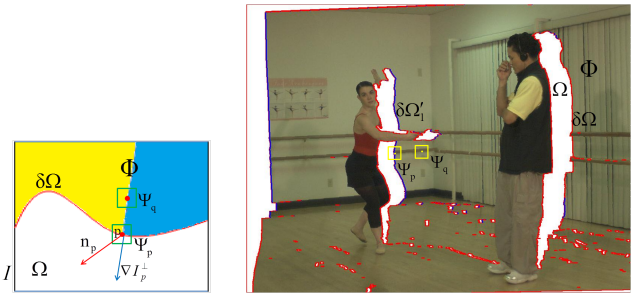
With the development of many rendering algorithms, the rendering methods are categorized into image based rendering and model based rendering methods. Image based rendering methods, use the available image data to render new views whereas model based methods require information about 3D models to compute synthesized views. Image based rendering methods can be further classified depending on the amount of geometry used. DIBR is an approximation of rendering with partial geometry, i.e., depth data. The depth data can be obtained from multiview images or range sensors [11], [12]. Using this depth data and texture, new views can be synthesized using basic principle 3D warping [13]. The warping can be done in two ways either forward warping or backward warping. In forward warping both depth and texture are warped to the virtual view position, whereas in backward warping first the depth map is warped and then pixel locations of the virtual view are located in the original image. The inherent problems with the DIBR are cracks, translucent cracks, missing areas and corona like artifacts when views are synthesized from multi view format. The causes of artifacts and different solutions are presented in [14].

In this work, we present the total DIBR using image inpainting technique to fill the disocclusion problems. Initially, the texture and depth map are warped to the virtual view positions and then the crack filling is applied by using averages of neighboring pixels. In the next step, the ghosting-like artifacts are removed on the borders. They appear as a thin layer of foreground which is projected on the background at the boundary of the disocclusion, due to depth and texture misalignments. These ghosting problems can be removed by simply extending the hole area on the background side. Otherwise, the filling will be affected due to the mixed colour pixels presented at the boundary area and the filling depends on the boundary data. Thus these problems need to be addressed before starting the inpainting process. The total DIBR method is illustrated in the Fig. 1.

III. Related work

The exemplar-based texture synthesis introduced by Criminisi et al. effectively replicates both structure and texture by using the advantages of both partial differential equations (PDE) based inpainting method and non-parametric texture

synthesis. The quality of the inpainted image highly depends on the order of the filling direction.



(a) Notation diagram (b) Warped image with notations

Figure 2. Schematic illustration.

Consider 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. 2). The basic steps of Criminisi's algorithm are as follows: (i) Identify the boundary and compute the priorities on the boundary region and (ii) Find the patch with the maximum priority and find the best patch that matches the selected patch using patch matching and filling (iii) Update the confidence values. 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:

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

$$C(\mathbf{p}) = \frac{1}{|\Psi_{\mathbf{p}}|} \sum_{\mathbf{q} \in \Psi_{\mathbf{p}} \cap \Phi} C(\mathbf{q}), \tag{2}$$

$$D(\mathbf{p}) = \frac{\left\langle \nabla I_{\mathbf{p}}^{\perp}, \mathbf{n}_{\mathbf{p}} \right\rangle}{\alpha}, \tag{3}$$

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. $|\Psi_{\mathbf{p}}|$ is number of pixels in $\Psi_{\mathbf{p}}$, α is normalization factor (e.g., 255 for gray scale image), $\mathbf{n}_{\mathbf{p}}$ is a unit vector orthogonal to $\delta\Omega$ at a point $\mathbf{p}$, and $\nabla I_{\mathbf{p}}^{\perp}$ is the direction and of the isophote.

After the 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. Next, a block matching algorithm is used to find the best similar source patch $\Psi_{\hat{\mathbf{q}}}$ in order to fill-in the missing pixels in the target patch:

$$\Psi_{\hat{\mathbf{q}}} = \arg \min_{\Psi_{\mathbf{q}} \in \Phi} \left\{ d(\Psi_{\hat{\mathbf{p}}}, \Psi_{\mathbf{q}}) \right\}, \tag{4}$$

where d is the distance between two patches defined as sum of squared difference (SSD). After the most similar source patch $\Psi_{\hat{\mathbf{q}}}$ is found, the values of the missing 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 source patch $\Psi_{\hat{\mathbf{q}}}$. Once the target patch $\Psi_{\hat{\mathbf{p}}}$ is filled, the update of the confidence term $C(\mathbf{p})$ is as follows:

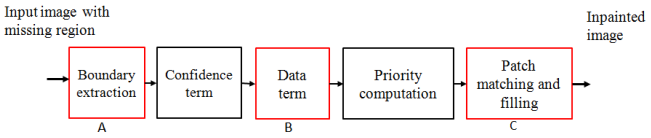


Figure 3. Block diagram of the exemplar-inpainting method; The highlighted blocks indicates the depth-included curvature inpainting method enhancements.

$$C(\mathbf{q}) = C(\hat{\mathbf{p}}), \forall \mathbf{q} \in \Psi_{\hat{\mathbf{p}}} \cap \Omega. \tag{5}$$

However, this method is not aimed at 3DV formats to handle disocclusions and thereby could not recognize the differences between foreground and background in a virtual view. As a result, missing areas are sometimes filled with foreground information instead of background.

Daribo et al. extended the Criminisi et al. method by using depth information, first by introducing a depth regularity term in the priority term calculation $P(\mathbf{p}) = C(\mathbf{p}) \cdot D(\mathbf{p}) \cdot L(\mathbf{p})$ and is given as:

$$L(\mathbf{p}) = \frac{|Z_{\mathbf{p}}|}{|Z_{\mathbf{p}}| + \sum_{\mathbf{q} \in, Z_{\mathbf{p}} \cap \Phi} \left(Z_{\mathbf{p}}(\mathbf{q}) - \overline{Z_{\mathbf{p}}} \right)^2}, \tag{6}$$

The depth regularity term $L(\mathbf{p})$ is approximated as the inverse variance of the depth patch $Z_{\mathbf{p}}$ centered at $\mathbf{p}$. The depth regularity term presented in their inpainting method controls the inpainting process by favoring the filling order that comes from the background. In addition, the patch matching step is modified by adding the depth into the search process to find a best patch in both the texture and the depth domain. Although this method uses the depth in the inpainting process reduces the problem to a degree as it still partly fills the disocclusion regions with the foreground information and wrong textures.

Gautier et al. followed the [9] method in considering depth map and extending the exemplar approach to help the inpainting process. They introduced a 3D tensor to calculate the 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.

$$J = \sum_{l=R,G,B,Z} \nabla I_l \nabla I_l^T, \tag{7}$$

$$D(\mathbf{p}) = a + (1-a) \exp\left(\frac{-C_1}{(\lambda_1 - \lambda_2)^2} \right), \tag{8}$$

where ∇I_l is local spatial gradient over a 3x3 window. J is the 3D structure tensor and λ_1,λ_2 are the eigen values of J, which gives amount of structure variation, C_1 is a constant positive value and $a \in [0, 1]$.

Moreover, these previous work both Daribo et al. and Gautier et al. rely on having true depth information available at the rendered view position. In general, this assumption is not feasible or realistic since the depth information of the virtual view also must be estimated.

(a) Total boundary (b) One sided boundary (c) Depth-guided one sided boundary

Figure 4. Boundary extraction.

IV. Proposed Depth-included curvature inpainting method

The depth-included curvature inpainting method also followed the Criminisi method by introducing the new data term and depth based constraints in all stages of the inpainting process. The depth information is added in the following steps of the depth-included curvature inpainting process:

A. Depth guided directional priority
B. Depth included curvature data term
C. Depth-based source region selection

Fig. 3 illustrates how these steps relate to the general inpainting process. Step A consists of defining a depth guided directional priority, which helps the selection of background patches to be filled first. In Step B, we used the Curvature Driven Diffusion (CDD) model [6] similarly to [15] as the data term $D(\mathbf{p})$ in the priority computation and extend the data term by incorporating depth to give the importance to the isophote curvature and strength. Finally, Step C prevents the foreground data from the source region, using depth constraints derived from the warped depth to favor the background filling. In addition, a weighted combination of - N best patches is used to define the target patch in the patch matching.

A. Depth guided direction priority

The boundary extraction block of Fig. 3 is improved by using depth information to guide the filling such that it starts from the background. This is because disocclusions result from depth discontinuities between foreground and background, and the disocclusion regions belong to the background which makes filling the disocclusion from the horizontal background side reasonable. Moreover, when the disocclusion appears between two foreground objects, selection of one sided boundary results foreground propagation into the holes. Hence, the one-sided boundary is constrained by using the depth, which controls the foreground scattering problem. 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 $\delta\Omega_1 = DM * H$. Disocclusion map is a mask, which represents the hole regions with 1s and remaining regions with 0s so the convolution gives one sided edges. Second, the directional priority selection is constrained by using a depth 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. The one-sided boundary and depth guided boundaries are shown in the Fig. 4.

$$\delta\Omega_1^{'} = \delta\Omega_1(\mathbf{q})|_{\mathbf{q}\in\delta\Omega_1\cap(Z(\mathbf{q})<M\cdot max(Z))}, \tag{9}$$

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 $\mathbf{q}$. Convolution kernel H is defined as follows according to the warped view:

$$H = \begin{cases} [1 \quad -1 \quad 0] & \text{if left warped view ;} \\ [0 \quad -1 \quad 1] & \text{if right warped view.} \end{cases} \tag{10}$$

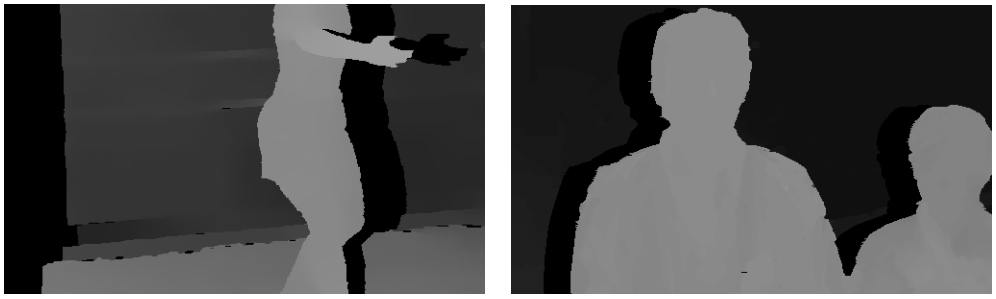
Once the hole boundary is obtained, using (9), priorities are calculated according to (1) utilizing the data term (14). Then the holes in the background regions are filled using the selected depth guided direction priority. However, the filling with the depth-guided directions handles holes to a certain depth level, the remaining holes are filled with one sided boundary priority. Moreover, when the virtual view camera is not horizontal, the holes do not appear according to the assumption that a right virtual view has holes on right side. In that case, hole filling is processed with the total boundary extraction.

B. Depth included curvature 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 is a variational approach, which solves the PDE to fill the missing regions using diffusion [6]. It mainly aims at satisfying the connectivity principle, i.e., broken lines should connect over missing areas. The CDD introduces the curvature along with the gradient strength to achieve that goal by using diffusion process.

Although the present approach is not a diffusion-based inpainting, the CDD model computes the structures information using the curvature as it gives the geometry of the isophote. Hence, we adopt and add the depth while calculating the data term to propagate the structure details into the missing areas.

$$g(s) = s^{\alpha}, s > 0, \alpha \geq 1 \tag{11}$$



(a) "Ballet" depth map (b) "Lovebird" depth map

Figure 5. Different depth distributions in the background.

$$k_{\mathbf{p}} = \nabla \cdot \left(\frac{\nabla I_{\mathbf{p}}}{|\nabla I_{\mathbf{p}}|} \right) \tag{12}$$

$$\frac{\partial I_{\mathbf{p}}}{\partial t} = \nabla \cdot \left(\frac{g(|k_{\mathbf{p}}|)}{|\nabla I_{\mathbf{p}}|} \nabla I_{\mathbf{p}} \right), \tag{13}$$

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. By incorporating the CDD model as a data term and setting $\alpha = 1$ in (11), the data term becomes:

$$D(\mathbf{p}) = \left| \nabla \cdot \left(\frac{k_{\mathbf{p}}}{|\nabla I_{\mathbf{p}}|} \nabla I_{\mathbf{p}} \right) \right|. \tag{14}$$

The data term is calculated using the depth gradient. Therefore, less variation in depth reduces the influence of the data term. Fig. 5 shows two examples of depth variations, where Fig. 5(a) exhibits more changes in depth than Fig. 5(b).

C. Depth-based source region selection

The patch-matching step is an improvement to the method of [9] and [10], where they add depth information in the patch matching step to find the best texture according to the depth range, and moreover, they assume the original depth map is available. If the original depth information is available, hole filling will be straight forward to find the source patch according to the right depth level. The improvement to the reference methods consists of classifying the source region using warped depth information, in order to select similar patches from the nearest depth range. The source region is divided into background region and foreground regions using depth information in the target patch. By considering Φ to be the known source region, which contains both foreground and background regions, the best source patch selection from foreground region is avoided by sub-dividing Φ using depth threshold Z_c according to:

$$\Phi_b = \Phi - \Phi_f, \tag{15}$$

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 selected adaptively from the variance of the known pixel values of the target depth patch. When the depth patch lies near to foreground (See Fig. 6(a)), the variance of the target depth patch is greater than a threshold γ, and the patch might contain unwanted foreground values (See Fig. 6(b)). The average value of the depth patch is then chosen instead as the depth threshold in order to deduct the foreground parts. Otherwise, the patch contains the uniform or continuous depth values, so the maximum value in the depth patch is used as the depth threshold in order to get the best patch according to the depth level. The depth threshold Z_c is defined as follows:

$$Z_c = \begin{cases} \overline{Z_{\hat{\mathbf{p}}}} & \text{if } var(Z_{\hat{\mathbf{p}}}(\mathbf{q})|_{\mathbf{q} \in \Psi_{\hat{\mathbf{p}}} \cap \Phi}) > \gamma; \\ \max(Z_{\hat{\mathbf{p}}}) & \text{otherwise.} \end{cases} \tag{16}$$

$\Psi_{\hat{\mathbf{p}}}$ 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 $\Psi_{\hat{\mathbf{p}}}$ from the priority term and depth-based source region Φ_b defined in (15) are computed, the target patch is filled with the best N number of patches within the source region.

$$\Psi_{\hat{\mathbf{q}}} = \arg \min_{\Psi_{\mathbf{q}} \in \Phi_b} \left\{ d(\Psi_{\hat{\mathbf{p}}}, \Psi_{\mathbf{q}}) + \beta \cdot d(Z_{\hat{\mathbf{p}}}, Z_{\mathbf{q}}) \right\}, \tag{17}$$

where d is SSD, and β is a value to give weight to the depth as important as the texture. In contrast to the reference methods, the depth-included curvature method used the warped depth information in the inpainting process. In order to help the inpainting process using the depth information, the holes in the depth image should be filled simultaneously, or the hole free depth image should be available for depth-guided inpainting process. In depth-included curvature method, we considered that the depth map should be filled simultaneously along with the texture.

From the idea of [16], we used a weighted average of N patches from the patch matching to fill the missing information in the disocclusion. Weighted average minimizes the noise in the selected patch and helps the smooth continuation filling process. The depth information helps the inpainting process to fill with the background textures otherwise when the depth information in the patch-matching is not considered, we put the filling in risk, i.e., the holes being filled with foreground textures.

(a) Warped depth-information (b) Source region before separation (c) Depth-based background source region

Figure 6. Depth-based source region selection.

In [1], the data term is calculated iteratively after every target patch is filled because the new copied texture to fill the hole region is a combination of the N best patches. In addition, the source region is updated such that the filled area is also available as source region for the next target patch.

V. Test arrangement and evaluation criteria

A. Compared to related work

Initially, the depth-included curvature inpainting method results are evaluated by using objective measurements as well as using visual comparison. 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 former two sequences are captured with 8 cameras and a baseline of 300 mm and 400 mm respectively [12]. The latter sequence is captured with 12 cameras and a baseline of 35 mm [17]. The test sequences have different depth and texture characteristics that make them suitable for testing different disocclusion filling attributes of inpainting methods as the method relies on depth. The "Ballet" sequence has large depth discontinuities at different depth ranges, which results in large disocclusions at different depth levels. The "Break dancers" sequence has a large number of objects located in almost the same depth level and smaller holes due to gradual depth discontinuities. The "Lovebird1" sequence has complicated scene with complex texture, structured background and large depth discontinuities.

B. Sensitivity analysis

The sensitivity of the depth at various steps of the proposed inpainting process is analyzed by restricting the depth at those stages. The depth information is used in three steps of the proposed inpainting process. In the first step, sensitivity of the depth in boundary extraction (SZB) is analyzed by performing the inpainting method without considering the depth information, i.e., using one sided boundary. In the next step, sensitivity of depth in data term (SZD) is analyzed by computing the data term using R, G, and B channels without incorporating the depth information as an additional channel. In the last step, the depth is used in source region selection and in block matching. Sensitivity of depth in patch matching (SZP) is analyzed separately without using depth. The sensitivity of depth in source region selection (SZSR) and sensitivity of depth in block matching (SZBM) are analyzed by using the Φ as a source region and finding SSD for only R, G, and B channels. While measuring the sensitivity at one step, the depth is not changed in other steps.

All test sequences are used in a DIBR of V+D scenario with possibility of full reference evaluation, i.e., access to ground truth texture and available depth at the required camera view positions. For the first two sequences, camera view 4 is rendered from camera view 5 and compared with ground truth image at camera view 4. In the third "Lovebird1" sequence, camera view 4 is rendered from camera view 6 and compared with ground truth image at camera view 4. The virtual views are first rendered and small holes and ghosting artifacts are removed using preprocessing step (see Fig. 1). Thereafter, the processed warped views are used as inputs to the inpainting method presented in Section IV. Important parameter and values of the proposed inpainting method is given in [1]. The best exemplars are searched in the warped depth and texture images, where as in [9] and [10] methods exemplars are searched in the warped texture and original depth at the virtual camera position. The implementation of inpainting method [8] is available on [18]. We have also implemented the method [9] for the comparison. The following two objective evaluation metrics are considered for assessing the results: peak signal to noise ratio of the luminance component (Y-PSNR) and mean structural similarity index (MSSIM). Both of these metrics measure the image quality where as MSSIM mostly corresponds to the perceptual visual quality [19]. Both metrics are applied to the full image, although the disocclusion areas only corresponds to $3-15\%$ of the total number of pixels. This results in the quality change shown in the sensitivity analysis to only manifest in the fractional part of the metric values. Another approach that would emphasize the changes would be to evaluate disocclusion pixels only. However, that would produce results that are not comparable with previous work.

VI. Results and analysis

A. Compared to related work

The objective evaluation results from the related work and proposed methods are shown in Fig. 7. The PSNR and MSSIM graphs consistently demonstrate that the depth-included curvature inpainting method performs better than the

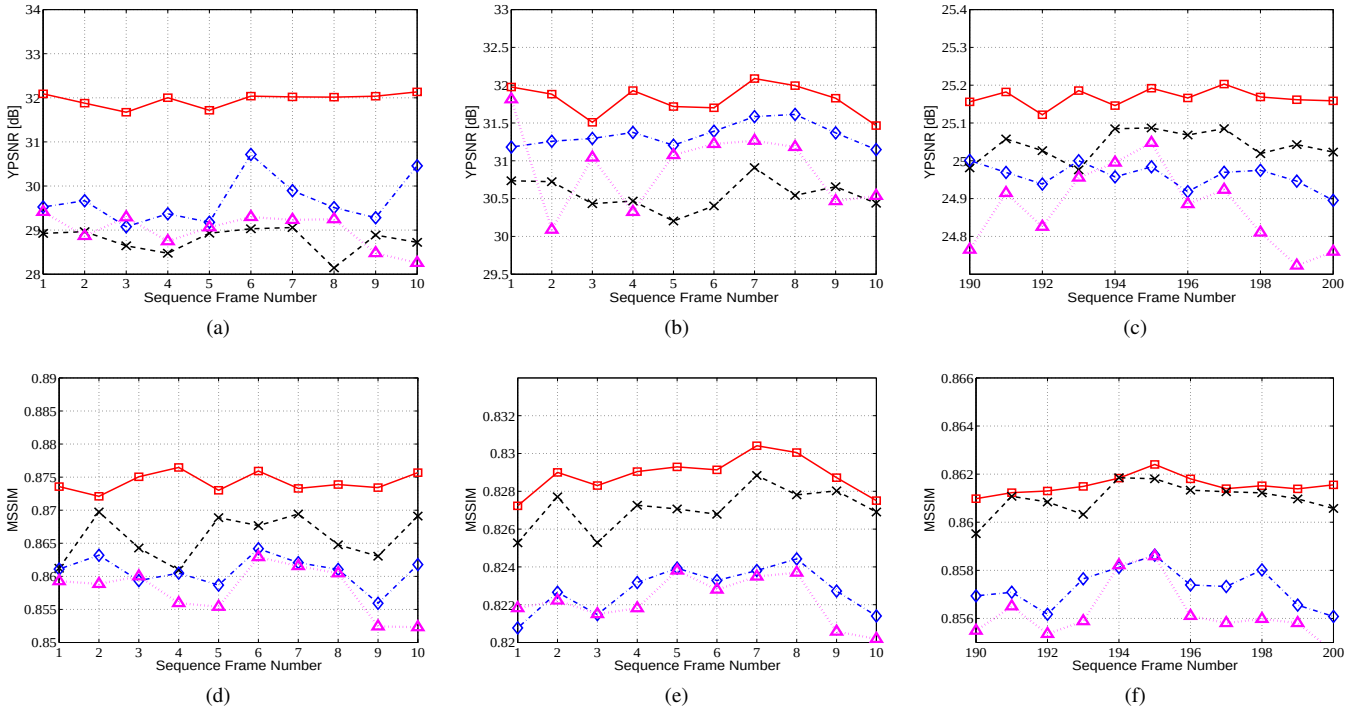


Figure 7. Compared to related work: Objective metrics PSNR and MSSIM of the investigated sequences; Proposed method (−□−), Gautier et al. method (− × −), Daribo et al. method (− · ◊ · −) and Criminisi method (· · △ · ·), PSNR for each rendered frame at view position 4 of "Ballet" (a), at view position 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).

reference Criminisi, Daribo and Gautier methods. In addition to the objective results, Fig. 8 shows the synthesized views of the "Ballet" and "Lovebird1" images with the missing areas and inpainted images from different inpainting methods for visual assessment. Missing regions in Fig. 8(b) are filled with foreground information since no information about the depth is used to assist the filling process. Although the Daribo and Gautier methods are aided with true depth information, the missing areas are filled with the unwanted textures due to the lack of depth constraints and filling order.

The depth-included curvature inpainting method operates in a more realistic setting that only depends on warped depth and it shows visual improvements compared to the reference methods. The results from Fig. 8(e) show that the depth-included curvature inpainting method propagates the necessary neighboring information into the missing areas, by retaining both smooth areas (at the left side of the "Ballet" image) and propagating neighborhood structure (on the curtain in the "Ballet" image and at the head of the women in the "Lovebird1" image). The inpainting method might not reproduce the exact structure as in ground truth images due to the lack of knowledge about the scene contents, but it almost replicates the main structure.

B. Sensitivity analysis

The sensitivity of the depth in various stages of the depth-included curvature inpainting method is analyzed by using both the objective metrics and visual comparison. The results for the objective measurements are presented in Table I and Table II. Fig. 9 and Fig. 10 show the synthesized views of the "Ballet" and "Lovebird1" images with the missing areas and sensitivity of the depth in depth-included curvature inpainting method for visual assessment. The average PSNR and MSSIM values and visual comparison consistently demonstrate that the influence of the depth depends on the scene content and available depth information.

1) Sensitivity of depth in boundary extraction: The results from the sensitivity of depth in boundary extraction show that the depth information is important to handle the disocclusions, plausibly when they occur between foregrounds (see Fig. 9(b)). Although depth information is used in source region selection and block matching in order to fill holes from the background, still the holes are filled with the foreground due to the lack of knowledge about the depth on boundary, and as a result foreground boundary is selected in the boundary extraction. Thus adding the depth constraint on the selection of the boundary improves the hole filling process.

2) Sensitivity of depth in data term: The results from the sensitivity of depth in data term show that the depth information is less important for filling holes in the inpainted view (see Fig. 9(c) and Fig. 10(c)). However, the depth information gives the priority to structures when the depth contains several layers. For example, Fig. 5(a) contains several depth layers, whereas the other depth image does not contain less depth layers (see Fig. 5(b)). Moreover, the depth characteristics depend on the depth acquisition method. Thus, if there are some layers in the depth map enriching the depth information, it favors the inpainting process.

TABLE I. Sensitivity analysis: Average Y-PSNR.

Test sequence	SZB	SZD	SZSR	SZBM	Proposed
Ballet	31.91	31.88	31.76	**31.99**	31.96
Break Dancer	**31.84**	31.82	31.82	31.80	31.80
Love bird1	**25.16**	25.15	24.91	25.13	**25.16**

TABLE II. Sensitivity analysis: Average MSSIM.

Test sequence	SZB	SZD	SZSR	SZBM	Proposed
Ballet	0.8745	0.8745	0.8741	**0.8748**	0.8742
Break Dancer	0.8289	0.8289	**0.8298**	0.8296	0.8289
Love bird1	0.8614	0.8613	0.8607	**0.8615**	**0.8615**

3) Sensitivity of depth in patch matching: The results from the sensitivity of depth in source region selection (SZSR) in the patch matching step demonstrate that the depth information is necessary in order to avoid the selection of the similar texture regions from the foreground. Without using the depth information, regions between two foreground objects are filled with the foreground texture (see Fig. 9(d) and Fig. 10(d)). Other results from the sensitivity of depth in source region selection (SZBM) in the patch matching step demonstrate that the use of depth in the source region is not so important when the depth data contains no layers (see Fig. 9(e) and Fig. 10(e)). In contrast, the depth data is essential for filling holes when the scene contains multiple depth layers in order to propagate the similar texture according to the depth level.

In summary, the quality of the virtual view is highly dependent on the available depth information. Moreover, the depth-information plays a crucial role in filling the missing regions in the synthesized views by guiding the filling process to proceed from the background direction and copying the best texture from the background data. It is important that the depth map is filled with the background information otherwise errors will propagate because all stages of the proposed inpainting method depend on the depth information.

VII. Conclusion

We have analyzed the importance of the depth information in the proposed depth-based inpainting method to fill disocclusions in a virtual view. The depth information guide the direction from which the filling should proceed, helping the data term calculation and, moreover, in the patch matching to select the best texture from the background. The influence of the depth information at each stage of the inpaiting process is analyzed by using objective measurements and visual inspection and the results are compared with the depth-included curvature inpainting method results. The evaluation demonstrates that the proposed method performs better than related work.

To what degree depth can be used in each step of the inpainting process depends on the depth distribution, which is presented in our visual analysis and objective evaluation. More elaborate knowledge about the depth distribution allows for tradeoffs that may reduce computational requirements without sacrificing quality. One such example is in the case where the scene has no disocclusions between foreground, and depth may be excluded from the boundary extraction step. Moreover, when the depth map contains only foreground and background layers, the depth information in the data term and block matching demonstrate less impact on visual quality.

For the future work, we will focus on reducing the computation time as the current method uses third order PDEs and is iteratively calculating the data term. Another focus in the future work will be on temporal coherency by using information from neighboring frames with valid tests and analysis.

Acknowledgment

This work has been supported by grant 00156702 of the EU European Regional Development Fund, Mellersta Norrland, Sweden, and by grant 00155148 of Länsstyrelsen Västernorrland, Sweden. We would like to acknowledge our colleagues Yun Li and Mitra Damghanian for their help. We would also like to acknowledge reviewers for their valuable suggestions.

References

[1] S. M. Muddala, R. Olsson, and M. Sjöström, "Disocclusion handling using depth-based inpainting," in *The Fifth International Conferences on Advances in Multimedia (MMEDIA)*, April 2013, pp. 136–141.

[2] W. J. Tam, G. Alain, L. Zhang, T. Martin, and R. Renaud, "Smoothing depth maps for improved steroscopic image quality," *Three-Dimensional TV, Video, and Display III*, vol. 5599, pp. 162–172, 2004.

[3] 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, pp. 527–540, 2007.

[4] A. Efros and T. Leung, "Texture Synthesis by Non-parametric Sampling," in *International Conference on Computer Vision*, 1999, pp. 1033–1038.

[5] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of ACM Conf. Comp. Graphics (SIGGRAPH)*, 2000, pp. 417–424.

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

[7] T. Chan and J. Shen, "Mathematical Models for Local Nontexture Inpaintings," CAM TR 00-11, March 2000.

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

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

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

[11] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE Journal of Quantum Electronics*, vol. 37, pp. 390–397, March 2001.

[12] 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, pp. 600–608, Aug. 2004.

[13] 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*, pp. 93–104, Jan. 2004.

[14] S. M. Muddala, M. Sjöström, and R. Olsson, "Edge-preserving depth-image-based rendering method," in *International Conference on 3D Imaging 2012 (IC3D)*, December 2012.

[15] 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)*, March 2012, vol. 2, pp. 326–329.

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

[17] 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.

[18] Sooraj Bhat, "Matlab implementaion of inpainting," http://www.cc.gatech.edu/~sooraj/inpainting/.

[19] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.

(a) Warped images of Ballet and Love bird1

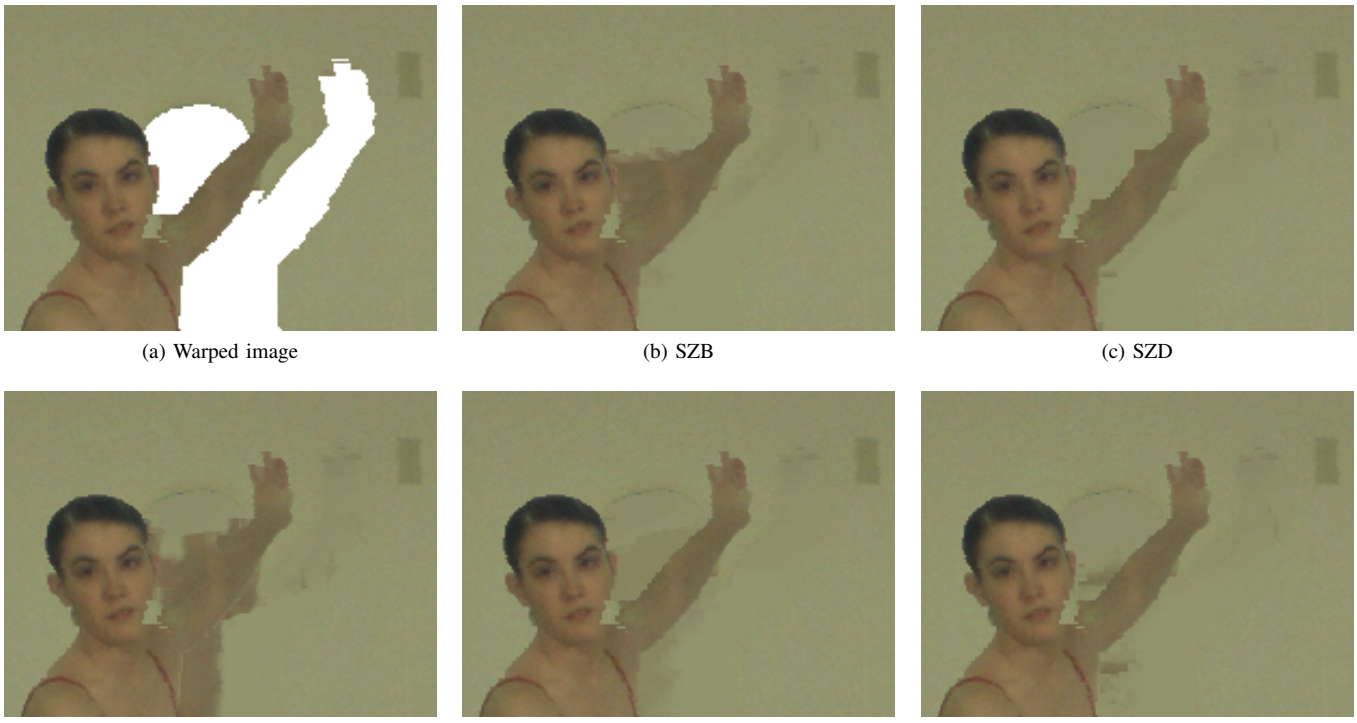
(b) Inpainted with Criminisi et al. method

(c) Inpainted with Daribo et al. method

(d) Inpainted with Gautier et al. method

(e) Inpainted with Proposed method

Figure 8. Compared to related work: inpainting method results for the investigated sequence frames “Ballet” first frame in the coulmn1 and “Lovebird1” 190th frame in column 2.



(a) Warped image (b) SZB (c) SZD

(d) SZSR (e) SZBM (f) Proposed method (depth in all steps)

Figure 9. Sensitivity analysis results for the investigated “Ballet” sequence second frame.

(a) Warped image (b) SZB (c) SZD

(d) SZSR (e) SZBM (f) Proposed method (depth in all steps)

Figure 10. Sensitivity analysis results for the investigated “Lovebird1” sequence frame198.