IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 20, 2020, accepted November 2, 2020, date of publication November 5, 2020, date of current version November 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3036053

Local Foreground Removal Disocclusion Filling Method for View Synthesis

HAITAO LIANG, XIAODONG CHEN, HUAIYUAN XU, SIYU REN, HUAIYU CAI, AND YI WANG

Key Laboratory of Opto-electronics Information Technology, Ministry of Education, School of Precision Instrument and Opto-electronics Engineering, Tianjin University, Tianjin 300072, China

Corresponding author: Xiaodong Chen (xdchen@tju.edu.cn)

This work was supported in part by the National Major Project of Scientific and Technical Supporting Programs of China during the 13th Five-year Plan Period under Grant 2017YFC0109702, Grant 2017YFC0109901, and Grant 2018YFC0116202.

ABSTRACT View synthesis is an effective method to generate the contents of multiple views based on a limited number of reference views, which can be used in 2D to 3D conversion, free viewpoint video and multiview video rendering. Depth-image-based rendering (DIBR) is a practical technique to generate virtual view by using a 2D reference view and its depth image. However, a critical problem in DIBR process is that disocclusions might be produced in the synthesized image because the background occluded by the foreground objects in the reference view may be exposed in the virtual view. In this paper, a local foreground removal method is proposed for disocclusion filling. Morphology-based depth image preprocessing is performed before DIBR, aiming to correct the depth value of the ghosts and remove ghost artifacts. In the synthesized virtual image, pixels on the disocclusion edge are identified and classified. Then they are positioned in the reference image by inverse 3D warping. Local foreground regions that occlude the corresponding background are removed from both the reference image and its depth image based on the disocclusion edge pixels. Removed region is filled with surrounding background contents, and depth information is used in this process to prevent foreground penetration. The predicted background contents are warped to the disocclusion region, thereby achieving the hole filling. Experimental results show that the proposed method performs better than the other methods in disocclusion filling, and improves the subjective and objective quality of the synthesized view. In the evaluation results of PSNR, SSIM, FSIMc and VSI, our method improves by 0.32-2.43dB, 0.0036-0.0155, 0.0041-0.0198 and 0.0012-0.0057 respectively compared with competitive methods.

INDEX TERMS Free viewpoint video, depth-image-based rendering, disocclusion filling, local foreground removal, view synthesis.

I. INTRODUCTION

With the development of computer science and multimedia technology, 3D video and free viewpoint video (FVV) have drawn more attention. Compared to 2D video, 3D video introduces depth information, which can provide more immersive viewing experience to viewers [1], [2]. As the ultimate of 3D video, FVV allows the viewer to freely choose the viewpoint within a certain range [3]. Accordingly, it requires the data of multiple viewpoints. However, it is not practical to use a series of cameras to capture video from multiple viewpoints, and transmitting the multiple videos requires a lot of bandwidth [4]. Moreover, for 2D to 3D conversion, only a single viewpoint can be obtained [5]. In this case, a practical way is to use a depth-image-based rendering (DIBR) method to generate multiple virtual views, which only requires a single reference view and its associated depth image [6]. The core technology of DIBR is called 3D warping [7]. In this process, all pixels in the reference image are projected to the world coordinate based on the depth information, and then the resulting points are reprojected onto the imaging plane of the target view.

When using DIBR to synthesize the virtual view, a critical problem is that artifacts may appear in the virtual image [8]. The most serious one is disocclusion. It arises because the background occluded by the foreground objects in the reference view becomes visible in the virtual view. Since no

The associate editor coordinating the review of this manuscript and approving it for publication was Shaikh Anowarul Fattah.

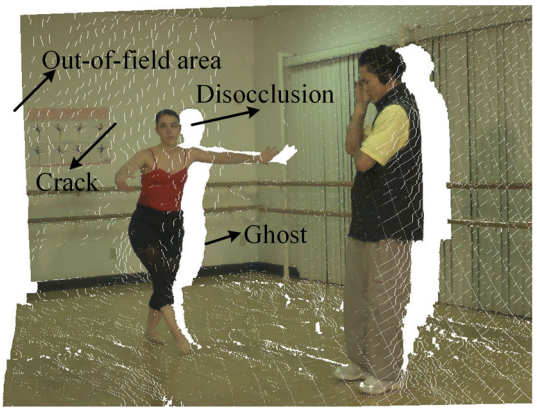


FIGURE 1. Artifacts in the virtual image.

pixels are warped to these regions, they appear as large holes, which seriously affect the visual quality of virtual view, as shown in Fig. 1. The disocclusion area is related to the baseline and depth discontinuity. As the baseline increases, the area of disocclusion gradually increases until the entire foreground object is projected onto the new background. In addition, artifacts in the virtual image include other types, such as cracks, ghosts, and out-of-field area (OOFA). Cracks are caused by rounding errors in 3D warping. Ghosts usually mean that the edges of foreground objects in the reference image are mismatched with those in the depth image. Some foreground edge pixels are given the depth values of background and projected to the background region in the virtual image. Since the virtual view exceeds the capture range of reference view, OOFA with no information appears on the edge of synthesized image. Therefore, reasonable handling of artifacts, especially disocclusion filling, is essential to improve the visual quality of virtual view.

To remove ghosts and fill the disocclusions with reliable contents, in this paper, a disocclusion filling method based on local foreground removal is proposed. Foreground edge is detected by the morphological approach in the depth image preprocessing. The depth value of ghost pixel is corrected so that the ghosts are projected to the correct position in 3D warping. Based on the disocclusion edge pixels, local foreground regions associated with the disocclusions are removed from the reference image and its depth image. The removed regions are filled with the texture and depth information of the surrounding background, which are then used to fill the disocclusions in the virtual image. In addition, a postprocessing approach is applied to deal with the remaining artifacts, including OOFAs and small holes introduced by depth errors. Our main contributions are as follows: 1) We introduce a morphology-based depth image preprocessing method to quickly detect foreground edges and correct ghost pixels. 2) We use the classification results of disocclusion edge to remove the local foreground, and then predict the content of the occlusion layer. 3) Disocclusion filling is performed in the reference image based on the modified inpainting method, which can prevent the artifacts generated in 3D warping.

The rest of the paper is organized as follows. Section 2 reviews the related work. The detailed description of the proposed method is given in Section 3. The experimental results and discussion are provided in Section 4. Finally, Section 5 concludes the paper and outlines the future work.

II. RELATED WORK

In General, disocclusion filling methods can be divided into two categories. The first one is to introduce a preprocessing process before 3D warping. Depth information is important for the calculation of coordinate position in 3D warping. Some depth image preprocessing methods are applied to prevent the generation of disocclusion. Due to the depth discontinuity between the foreground and background, symmetric or asymmetric low-pass filter is used to smooth the depth image, so that the area of disocclusion is reduced [9], [10]. However, the global filtering smoothes the regions that do not generate disocclusions, resulting in geometric distortion in the virtual image and reducing the 3D visual effect. To overcome this problem, Chen *et al.* [11] proposed an edge-based smoothing filter, focusing on the preprocessing of foreground edges. Zhu *et al.* [12] used morphological operators to detect foreground edges, and applied an asymmetric Gaussian filter to smooth the transition regions to avoid depth distortion. The above depth preprocessing methods are suitable for small baseline conditions. For the large baseline, due to the increase of disocclusion area, single smoothing is no longer competent. Therefore, additional postprocessing approaches are necessary. In addition, some other preprocessing methods are proposed to achieve ghost removal and hole decomposition. Lei *et al.* [13] proposed a divide-and-conquer hole filling method to decompose the disocclusion into several holes in the virtual image. In [14], depth image preprocessing is used to detect and remove ghosts. However, ghosts are caused by depth errors. The corresponding foreground content is effective, and should be projected to the correct position of the foreground edge.

The other type of method is to fill the disocclusion by using the texture correlation of surrounding pixels [15], [16]. Inpainting-based method is an alternative measure for hole filling [17], [18]. In Criminisi's method [17], the filling order is firstly calculated for pixels on the hole boundary. The priority consists of confidence term and data term. Then the beat matching patch is searched in the source region and copied to the hole. However, as disocclusions are originated from background region, they should be filled by background texture. The original inpainting algorithm gives foreground and background pixels the same weight, which allows the foreground texture to be sampled into the disocclusion region. Foreground blending is produced, which reduces the visual quality. To overcome this problem, some improved methods introduce depth information to fill the disocclusions. In Daribo's method [19], depth term is added to the priority calculation, so that the region with smaller depth variance is filled preferentially. But this method is performed under the assumption that the depth image of virtual view is known.

Ahn and Kim [14] generated the depth image of virtual view in 3D warping, and simultaneously filled disocclusions in the virtual image and its depth image. Kao [20] changed the priority calculation approach, and used the depth-based gray-level distance to calculate the matching cost of two patches. However, when the depth value of the foreground edge is incorrect, ghosts may appear and interfere with the extension of the background texture. Zhu *et al.* [12] introduced the mean square error when updating the priority to prevent the propagation of the wrong inpainted texture. To avoid the interference of foreground in disocclusion filling, some methods based on foreground-background segmentation are proposed. In Luo's method [21], edge detection algorithm is used to extract the foreground edge, and then the foreground object is extracted based on the depth information. Han *et al.* [22] used the multi-threshold Otsu method to segment the depth image into multiple layers and performed layered 3D warping. In [23], threshold segmentation is used to extract foreground object and the background layer is compensated. In [24], disocclusion edge pixels are divided into foreground and background based on the depth value. The confidence term and data term in the filling priority calculation are replaced by the depth term and background term. However, these methods are very dependent on the accuracy of foreground segmentation, which is a difficult task when there are several depth layers. Moreover, ghosts should also be considered in the segmentation process to avoid misclassification of foreground pixels.

For view synthesis of video sequences, the correlation of time information can be used to deal with artifacts. In video content analysis, some shot boundary detection algorithms are used to segment a video into clips [25]–[27]. They can detect changes in the scene and are used to update the background model. In the time domain, for moving foreground objects, the occluded region in the current frame may be exposed in other frames. In this case, some methods have been proposed to build background models, such as Gaussian Mixture Model-based methods [28]–[30]. However, these methods cannot obtain the background content occluded by the still foreground object or the background model of a still image. Therefore, disocclusion handling in the spatial domain is still worth studying. In this paper, we propose a spatial domain disocclusion filling method. The occlusion layer is predicted based on the local foreground removal approach, which is used to fill the disocclusions in the virtual image.

III. PROPOSED METHOD

The flowchart of the proposed method is shown in Fig. 2. Our framework mainly contains five parts: morphology-based preprocessing, classification of disocclusion edge, local foreground removal, removed region filling, and disocclusion filling and postprocessing. In the following, these steps will be described in detail.

A. MORPHOLOGY-BASED PREPROCESSING

Depth value reflects the distance between the object and the camera, which plays an important role in 3D warping. It can be obtained by depth camera, structured light, and stereo matching. Due to the limitation of algorithm and equipment accuracy, the depth value of foreground edge might be coarse. The foreground edge pixels with wrong depth values are projected to the background region and mixed with background texture, which are called ghosts. As the disocclusion is caused by depth discontinuity, ghosts are located at the edge of disocclusion, which disturb the disocclusion filling, resulting in the propagation of foreground texture and reducing the visual quality of virtual view. Some methods detect and remove ghosts in the horizontal direction [8], [13]. But the same problem exists in the vertical direction. In addition, the contents of ghosts in the reference image are reliable, thus a more reasonable way is to modify their depth values and project them to the correct position. In this paper, a morphology-based preprocessing method is proposed to detect foreground edges that may produce ghosts, and correct their depth values instead of smoothing. The preprocessing is performed in the region around the foreground edge, which would not introduce blur results and geometric distortion.

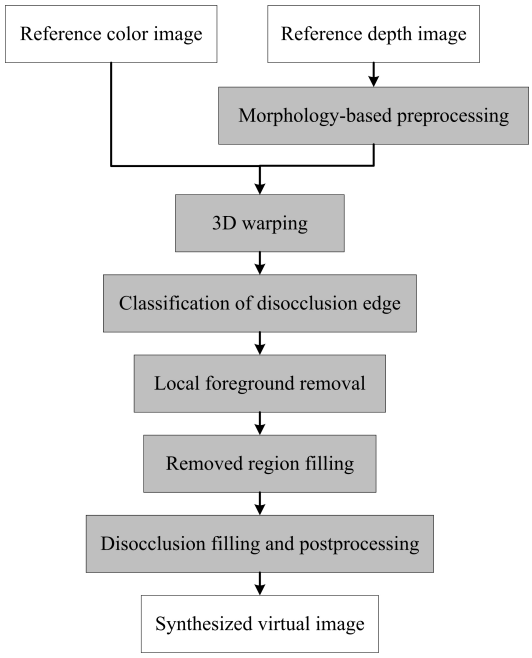
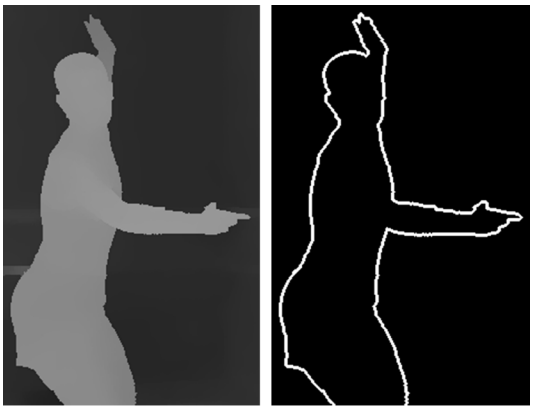


FIGURE 2. Flowchart of the proposed method.

Depth image preprocessing contains two steps: foreground edge detection and depth correction. Since the depth image may contain multiple depth layers, single threshold edge detection algorithm cannot extract accurate foreground edge. In this section, morphology-based edge detection is introduced. The foreground edge pixels that may contain ghosts are defined as:

$$E_r(u,v) = \begin{cases} 1, & d(u,v) \oplus L - d(u,v) > th \\ 0, & \text{otherwise}, \end{cases} \quad (1)$$

where d denotes the value of pixel (u, v) in the depth image or disparity image. As shown in Fig. 3(a), the depth value of the foreground is higher than the background it



(a) (b)

FIGURE 3. Foreground edge extraction for ghost removal. (a) Depth image. (b) Extracted foreground edge.

occluded [21]. $\oplus$ represents the morphological dilation operation, and L is the structural element, which is defined as:

$$L = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \tag{2}$$

The extracted foreground edge is shown in Fig. 3(b). These pixels with background depth values contain the foreground texture. Therefore, the depth correction process is to replace the depth value of the marked pixel with the four-neighborhood foreground depth value. Since only the depth values of some pixels are modified, the 3D visual effect of the synthesized result would not be reduced. Considering that the ghosts are usually 1-2 pixels wide [31], the above process is performed twice to ensure that all the ghost pixels are corrected. We use four-neighborhood operator instead of eight-neighborhood operator in preprocessing. This is because eight-neighborhood operator cannot remove all the ghost pixels in one execution process, and executing twice would cause too many background pixels to be included. It is noted that some background pixels may be given the foreground depth value in this process and still be adjacent to the foreground edge in the virtual image. In this case, the pixels on both sides of the disocclusion in virtual image belong to the background, that is, FG-BG-disocclusion-BG. Since the disocclusion also belongs to the background, these background pixels would not affect the disocclusion handling, and after the disocclusion filling, the texture distribution can be expressed as FG-BG, which is consistent with the real scene.

B. CLASSIFICATION OF DISOCCLUSION EDGE

Reference image and the preprocessed depth image are used as the input of modified 3D warping [32] to synthesize the virtual image. All pixels in the reference image are projected to the world coordinate based on their depth values. In the process of reprojecting these pixels onto the virtual view imaging plane, cracks are filled by the surrounding valid pixels, and the pixel overlap is resolved based on the depth value to correctly maintain the occlusion relationship.

(a) (b)

FIGURE 4. Synthesized result of modified 3D warping. (a) Right synthesized virtual image. (b) Left synthesized virtual image.

In the synthesized result, ghosts are warped to the correct foreground edge. Therefore, disocclusion filling becomes the main task. Traditional methods perform the inpainting process in the virtual image. The error introduced by 3D warping would affect the filling accuracy. Therefore, in this paper, we extract the foreground covering the disocclusion in the reference image and use the background content to fill the removed region. Since the content in the reference image is reliable, the filling result is used to fill the disocclusion in the virtual image.

Local foreground removal in the reference image is achieved based on the information of the disocclusion edge. The generation of disocclusion is related to depth discontinuity between the foreground and background. Different depth values make adjacent pixels separate in the virtual image. In generally, for the right synthesized virtual view, disocclusion appears on the right side of foreground [12]. Therefore, the pixels on the right side usually belong to the background, while the pixels on the left side may belong to the foreground or background as shown in Fig. 4(a), which is related to the width of the foreground and the baseline. There is a vice versa for the left synthesized virtual view, as shown in Fig. 4(b). To identify disocclusion edge pixels, we apply the Laplacian operator to depth image because of its sensitivity and direction invariance to depth discontinuity [33]. The foreground pixels on the disocclusion edge are marked as follows:

$$F(u,v) = \begin{cases} 1, & (\Delta d)_w(u,v) < 0 \\ 0, & (\Delta d)_w(u,v) > 0, \end{cases} \qquad \text{for } (u,v) \in \delta\Omega_d, \tag{3}$$

where Δd represents the Laplacian of the depth image and $(\Delta d)_w$ is the warped Laplacian image. $\delta\Omega_d$ represents the contour of disocclusion. After the foreground pixels are marked, the remaining pixels on the disocclusion edge belong to the local background. It is noted that Laplacian value equal to zero means that there is no depth discontinuity. Since there must be a depth discontinuity when the disocclusion is generated, this situation is not discussed in (3). The result of disocclusion edge classification is shown in Fig. 5. Based on their depth values, the marked pixels are projected to the reference image by inverse 3D warping. Then the corresponding pixels are located for the local foreground removal.

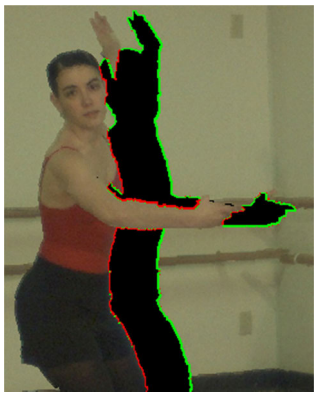


FIGURE 5. Classification result of disocclusion edge (foreground is marked in red and background is marked in green).

C. LOCAL FOREGROUND REMOVAL

In this section, local foreground covering the disocclusion is removed from the reference image and its depth image based on the obtained disocclusion edge pixels and the hole width. After inverse 3D warping, the relative position of the left and right edge pixels is still maintained. Therefore, the foreground removal starts from the background side. Taking the right synthesized view as an example, the foreground extraction process is performed from the right edge to the left because disocclusion appears on the right side of the foreground.

Since the left edge of the disocclusion may contain foreground and background pixels, local foreground extraction includes two cases. If the left edge pixel belongs to the foreground, it means that only part of the foreground region is projected on the new background in the virtual view, while the remaining part still occludes the original background. Therefore, the local foreground region is removed based on the width of disocclusion. In the case where the left edge pixel belongs to background, the entire foreground region is warped to the new background in the virtual view, so that the entire background that is occluded in the reference view is exposed. Therefore, all foreground pixels between the two sides are removed. In addition, the extracted foreground mask is processed by morphological dilation to accommodate to some possible depth changes and slight view rotation, which may affect the location of edge pixels in inverse 3D warping. The extraction result of the local foreground is shown in Fig. 6. It can be seen that the occluded regions visible in the virtual view are exposed. Since they belong to the background, the depth value and texture information of background should be used for removed region filling.

D. REMOVED REGION FILLING

In this section, the removed region in the depth image is predicted first. The depth image will then be used to assist the removed region filling in the reference image. Using depth information to guide the inpainting process helps to select more reasonable background contents to fill the removed region and prevent the incorrect propagation of foreground texture.

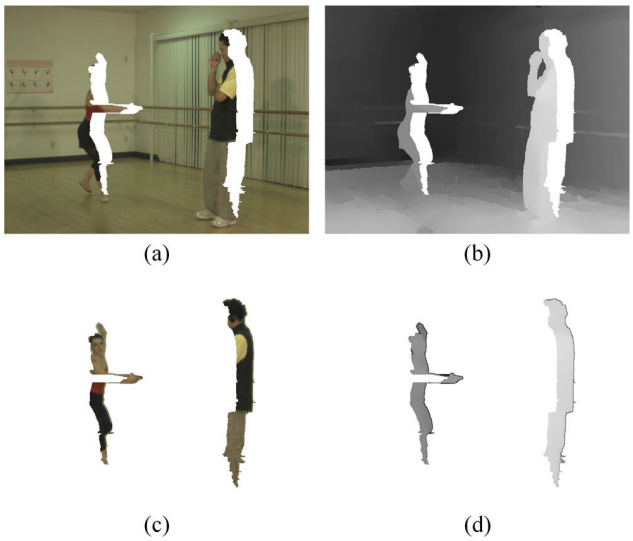


FIGURE 6. Local foreground removal result. (a) Local foreground is removed from the reference image. (b) Local foreground is removed from the depth image. (c) Local foreground regions in the reference image. (d) Local foreground regions in the depth image.

1) DEPTH PREDICTION FOR REMOVED REGION

For the depth prediction of the removed region, it is assumed that the removed region and its surrounding background content belong to the same physical surface, so they should have similar depth values [21]. The classification of pixels on both sides of the removed region includes two cases: foreground-background and background-background. For the former, the depth value of the background edge pixel is used to predict the depth value of the removed region on the horizontal line. For the latter, we apply a fast linear interpolation approach to achieve smooth transition of depth values. Therefore, the depth value of the removed region can be expressed as follows:

$$d(u,v) = \begin{cases} d(u_l,v) + s(u-u_l), (u_l,v) \text{ and } (u_r,v) \in \text{BG} \\ d(u_l,v), (u_l,v) \in \text{BG} \text{ and } (u_r,v) \in \text{FG} \\ d(u_r,v), (u_l,v) \in \text{FG} \text{ and } (u_r,v) \in \text{BG}, \end{cases} \quad (4)$$

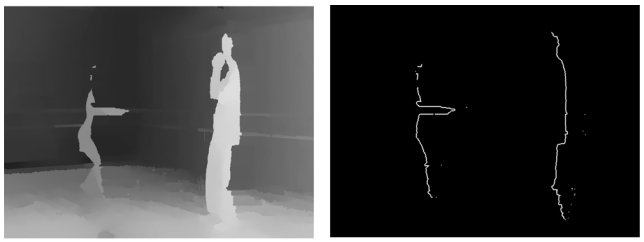
where (u_l, v) and (u_r, v) are the coordinates of the left and right edge pixels of the removed region, respectively. s represents the slope of the depth value and is defined as:

$$s = \frac{d(u_r,v) - d(u_l,v)}{u_r - u_l}, \quad (5)$$

The filling result of the removed region in the depth image is shown in Fig. 7(a). The filled depth image is used to predict the texture of the removed region in the reference image.

2) TEXTURE PREDICTION FOR REMOVED REGION

In the reference image, since the local foreground region is removed, the removed region should be filled with reasonable background textures. Criminisi's method [17] is an effective method for hole filling, which can simultaneously transmit texture and structural information. However, for removed



(a) (b)

FIGURE 7. Depth prediction result for removed region. (a) Predicted result of depth image. (b) Foreground detection result of the removed region edge.

region filling, since the foreground may still exist around the hole, directly applying the Criminisi's method would cause foreground texture to be sampled into the hole, reducing the visual quality of the filling result. Therefore, we propose a modified inpainting method based on the depth image to recover the texture of the removed region. Our contributions mainly include the optimization of priority calculation and patch matching cost.

For the input image I, Ω is the removed region in I.The remaining valid region is the source region Φ. In order to determine the filling order, the priority of each pixel on the edge of the removed region $\delta\Omega$ is calculated first. For pixel $p \in \delta\Omega$, Ψ_p denotes the square template centered at p. The priority of pixel p is defined as:

$$P(p) = [C(p) \cdot D(p) \cdot Z(p)] B(p), \tag{6}$$

where $C(p)$ and $D(p)$ are the confidence term and data term as defined in [17]. Confidence term represents the percentage of valid pixels in the patch, which tends to fill the patch with more valid pixels. Data term reflects the intensity of the isophote, which encourages linear structures to be preferentially synthesized and propagated to the hole. These two terms do not consider that the removed region belongs to the background and should be filled by background texture. Therefore, we add depth term and background term to the priority calculation to improve the filling order. The newly added terms are defined as follows:

$$Z(p) = \frac{d_{\max} - \frac{\sum_{q \in \Psi_p \cap \Phi} d(q)}{|\Psi_p \cap \Phi|}}{d_{\max} - d_{\min}}, \tag{7}$$

$$B(p) = \begin{cases} 0, & \text{for } p \in FG \\ 1, & \text{for } p \in BG, \end{cases} \tag{8}$$

where $d_{\max}$ and $d_{\min}$ are the highest and lowest nonzero depth values in the depth image. The depth term considers the weight of the depth information and gives higher priority to the background pixel with lower depth value. The background term is introduced to identify the foreground pixels on the edge of the removed region. It ensures that the hole filling starts from the background side and keeps distinct foreground boundary. In our method, as the depth value of the removed region is predicted, the Laplacian operator is applied to achieve the classification of the pixels on $\delta\Omega$. This method is practical for images with multiple depth layers, and has better robustness than the threshold-based method. The foreground detection result of the removed region edge is shown in Fig. 7(b), and the priority of the relevant foreground pixel is set to 0 to prevent the propagation of the foreground.

After all priorities on $\delta\Omega$ are computed, the patch $\Psi_{\hat{p}}$ with the highest priority would be filled first. The search for the best matching patch is performed in the source region to find the most similar patch $\Psi_{\hat{q}}$ to fill the hole in $\Psi_{\hat{p}}$. In Criminisi's algorithm, the matching cost is defined as the sum of the squared differences (SSD) of the valid pixels in the two patches. It is noted that directly applying this principle may cause some foreground textures to be sampled into hole, resulting some artifacts. Therefore, it is necessary to select candidate patches located in the same depth layer as the patch to be inpainted. To solve this problem, in our modified method, depth information is introduced to limit the search range of candidate patches. The best matching patch $\Psi_{\hat{q}}$ is searched by:

$$\Psi_{\hat{q}} = \left\{ \Psi_{\hat{q}} \left| \arg \min_{\Psi_q \in \Phi'} \mathrm{SSD}_{\mathrm{color}}\left(\Psi_{\hat{p}}, \Psi_q\right) \cap \mathrm{DD}\left(Z_{\hat{p}}, Z_q\right) \le 0.2 \right. \right\}, \tag{9}$$

where SSD between two patches is calculated in RGB color space. $Z_{\hat{p}}$ and Z_q represent the average depth value of valid pixels in $\Psi_{\hat{p}}$ and Ψ_q respectively. DD represents the depth difference between the two patches, which is defined as:

$$\mathrm{DD} = \frac{\mathrm{abs}\left(Z_q - Z_{\hat{p}}\right)}{Z_{\hat{p}}}, \tag{10}$$

For the candidate patch in the source region, the matching cost is calculated only if it is at the same depth level as $\Psi_{\hat{p}}$. The improved matching equation ensures the depth consistency between the candidate patch and the target patch. Since the depth value of the hole edge pixel is dynamic, the depth limit for the candidate patch also changes dynamically. This can guide the best matching patch to be selected in source region with similar depth value. The combination of SSD and DD makes the filling result have similar properties with the surrounding background in both texture and depth information. In addition, the search process is changed from global search to local search, which reduces the computational complexity and allows the spatial locality of textures to be explored. After the best matching patch searching, the content in $\Psi_{\hat{q}}$ is copied to the hole region in $\Psi_{\hat{p}}$. Then the terms in the priority are updated and the next iteration is performed until the entire removal region is filled. The filling result of the removed region is shown in Fig. 8. The predicted occlusion layer is used to fill the disocclusions in the virtual image.

E. DISOCCLUSION FILLING AND POSTPROCESSING

After the texture and depth value of the removed region are predicted, the occlusion layer is warped to the virtual view, as shown in Fig. 9(b). The corresponding region in the virtual

FIGURE 8. Texture prediction result for removed region.

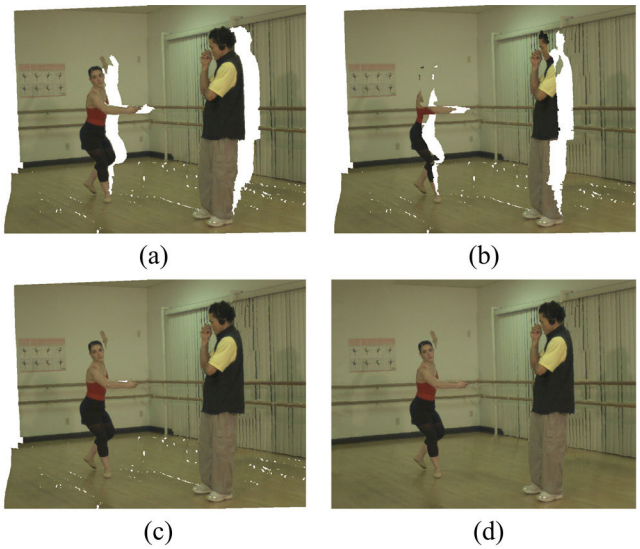


FIGURE 9. Results of disocclusion filling and postprocessing. (a) Warped virtual image. (b) Rendered image with local foreground removal. (c) Disocclusion filling result. (d) Postprocessing result.

image belongs to the disocclusion. In this case, the disocclusion in the virtual image is filled by the warped occlusion layer, as shown in Fig. 9(c). In addition, small holes caused by the error of depth value and OOFAs still exist. We introduce a postprocessing approach to deal with these artifacts. Since they are not caused by occlusion, the inpainting process uses the method mentioned above with minor adjustment. The restriction of background term is redundant. In the postprocessing process, we also use depth information as an auxiliary to keep the consistency of texture and depth information. The result of postprocessing is shown in Fig. 9(d).

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

The proposed algorithm is implemented in MATLAB R2014a. In our experiment, Five public multiview video-plus-depth (MVD) sequences (Ballet, Breakdancers [34], PoznanHall2, PoznanStreet and UndoDancer [35]) and seven public image-plus-depth sequences from the Middlebury Stereo Data Sets [36] are used to evaluate the performance of the proposed method. Video sequence provided by Microsoft Research contains 8 viewpoints, which have a resolution of

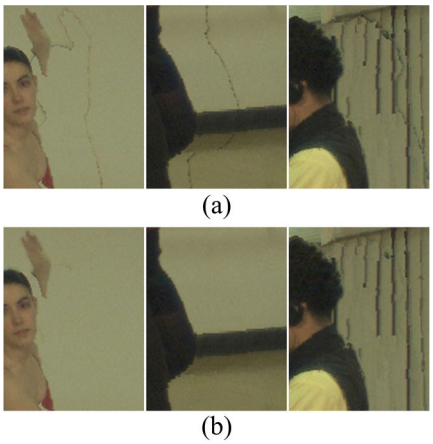


FIGURE 10. Comparison results of ghost removal. (a) Virtual image without ghost removal. (b) Virtual image with ghost removal.

1024 × 768 pixels and 100 frames long. The depth image is generated using stereo vision algorithm. The Ballet sequence contains two moving foregrounds which have large depth discontinuity with the background. The Breakdancers sequence contains multiple moving foregrounds with similar depth values, and the foreground overlaps in some frames. The resolution of the remaining three sequences is 1920 × 1088. They include complex backgrounds, camera motion, and changed illumination. The image sequences contain high-quality multiview images in parallel camera configuration, representing different types of natural scenes. Full-size images are used in our experiments. The ground truth of depth image is calculated using high-precision structured light. The camera parameters of the test sequences are known, including internal and external parameters. The experiments are performed in five scenarios, including the performance of ghost removal, ablation study on FG-BG detection, subjective visual quality evaluation, quantitative objective quality evaluation and computational cost analysis. The parameters used in the proposed method are set as follows. The threshold th is set to 20 to accurately obtain ghost pixels with smaller depth discontinuity. In the removed region filling, according to [17], the patch size is expected to be larger than the largest texel or the thickest structure. Therefore, the patch size is set to 9×9. The search window is experimentally set to 160×120 based on [14] and the aspect ratio of the data set. If the search window is too large, the search time for the best matching patch will increase. Conversely, a smaller window makes the algorithm propagate repeated content successively because it cannot explore the spatial locality of the texture.

B. RESULT OF GHOST REMOVAL

In the sampling and quantization of depth image, some errors may occur due to the limitations of equipment and algorithm accuracy. In the original DIBR, foreground pixels with incorrect depth values are warped to the background in the virtual view, and ghosts are generated. The proposed method detects pixels that may produce ghosts based on morphology and corrects their depth values. Fig. 10 shows the comparison

(a) (b)

FIGURE 11. Ablation study on FG-BG detection. (a) Filling result without FG-BG detection. (b) Filling result with FG-BG detection.

results of ghost removal. It can be seen that the proposed method can effectively handle the ghost pixels and warp them to the correct position instead of directly deleting the relevant content. The blending of foreground and background textures can be avoided in the virtual image, thereby improving the visual quality, and preventing the propagation of foreground texture during the disocclusion filling.

C. ABLATION STUDY ON FG-BG DETECTION

To evaluate the performance of the proposed inpainting method with FG-BG detection. The ablation study is performed in this section. The comparison result for removed region filling is shown in Fig. 11. Fig. 11(a) shows the filling result without FG-BG detection. In this case, the filling process starts at the same time from the foreground and background sides. Even with the guidance of depth information, the best matching patch is still selected from the foreground region, leading to the propagation of foreground texture. After adding FG-BG detection, the filling result is shown in Fig. 11(b). The proposed method effectively maintains the foreground edge and ensure that the prediction of the occlusion layer starts from the background side.

D. VISUAL QUALITY EVALUATION OF SYNTHESIZED VIEW

In our experiment, six competitive schemes are selected for comparison, including Criminisi's exemplar-based inpainting method [17], Daribo's inpainting method [19], Ahn's depth-based inpainting method [14], Kao's synthesis method [20], Zhu's approach [12] and Oliveira's method [24]. Among them, Criminisi's inpainting method and Ahn's inpainting method are implemented based on the codes provided by the authors, and the other methods are implemented based on published papers. For the parameters not indicated in the papers, we used the same default parameters as the proposed method. In the evaluation of subjective visual quality, for video sequence rendering, the comparison results of the synthesized view are shown in Fig. 12 and 13. It is noted that the synthesized view is named after the sequence name and rendering information. For example, BA54 represents the synthesized view of Ballet sequence from view 5 to view 4. BA54 and BR41 represent the view synthesis in the small baseline and the large baseline configuration, respectively. The comparison results in Fig. 12 show that the proposed method outperforms others in disocclusion filling, and provides plausible results, while other methods contain some artifacts and unrealistic textures. In Criminisi's method [17], depth information is not considered in hole filling. The inpainting process is performed simultaneously from the foreground and background edges, and some foreground textures are sampled into the disocclusion region, as shown in Fig. 12(b). In Daribo's method [19], depth variance is introduced into the computation of priority and patch distance, but the presence of ghosts makes some artifacts appear in the disocclusion region, as shown in Fig. 12(c). In Ahn's method [14], due to the mismatch of the foreground edge in reference image and depth image, ghosts disturb the hole filling process. Some artifacts appear at the edges of foreground objects, including the penetration of foreground texture and incorrect inpainting results, as shown in Fig. 12(d). In Kao's method [20], the depth image preprocessing is introduced before 3D warping, but the depth expansion process is only performed in the horizontal direction. The priority based on inverse variance of depth is not ideal in distinguishing foreground and background, especially for scenes with multiple depth layers. Some unexpected results are produced in the hole region as shown in Fig. 12(e), which reduce the visual quality. In Zhu's method [12], preprocessing approach based on asymmetric filter can reduce the area of holes to some extent and prevent depth distortion. But smoothing without ghost removal may cause foreground contents to propagate to disocclusions and produce some unrealistic results, as shown in Fig. 12(f). In Oliveira's method [24], depth information dominates the disocclusion filling process. The confidence and data terms in the priority calculation are replaced by depth and background terms. This method encourages the filling to be performed from the background side, but the texture and structure information is ignored. During the filling process, regions with high confidence and linear structures cannot be propagated preferentially, resulting in some artifacts and distorted textures, as shown in Fig. 12(g). The proposed method applies morphology-based preprocessing to eliminate the interference of ghosts and warps them to the correct place in 3D warping. The contents for disocclusion filling come from the reliable texture of the reference image. From the experimental results in Fig. 12(h), the proposed method can keep the sharp foreground edge and use the background texture to fill the disocclusions. The overall synthesized results look the most likely the ground truth. Fig. 13 shows the global synthesized results. Our method performs better than the other methods in artifact handling and provides a realistic virtual view.

For still image sequence rendering, the comparison results of the synthesized image are shown in Fig. 14. As the depth image of the virtual view in the Middlebury data sets is not provided, the performance of Daribo's method [19] is not evaluated. Compared with the stereo matching algorithm, the per-pixel depth image generated by structured light is more accurate. However, there are still some depth errors in the foreground edge because the texture in the color image

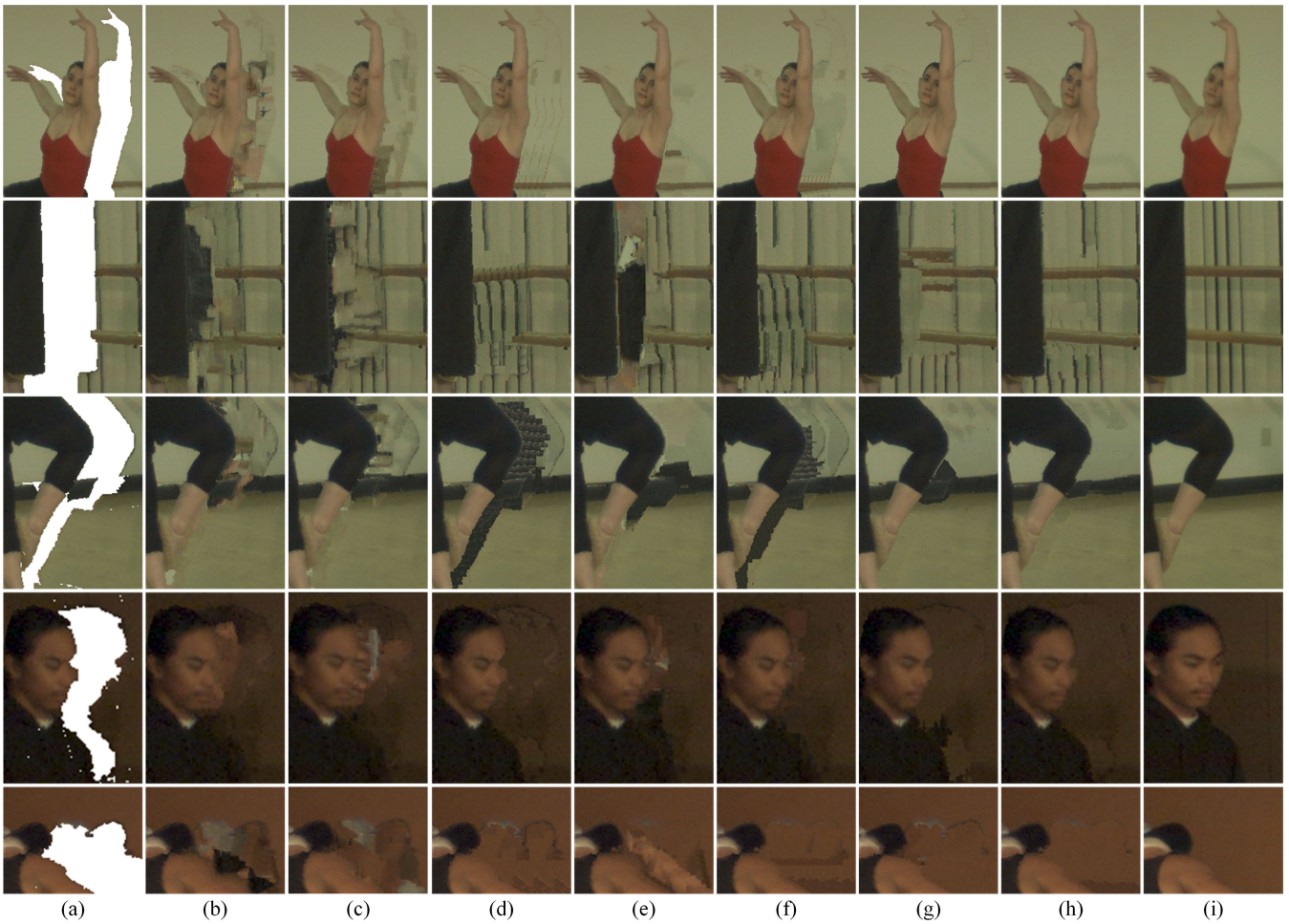


FIGURE 12. Visual quality comparison results of disocclusion filling for MVD sequences. (a) Hole regions. (b) Criminisi's method. (c) Daribo's method. (d) Ahn's method. (e) Kao's method. (f) Zhu's method. (g) Oliveira's method. (h) Proposed method. (i) Ground truth.

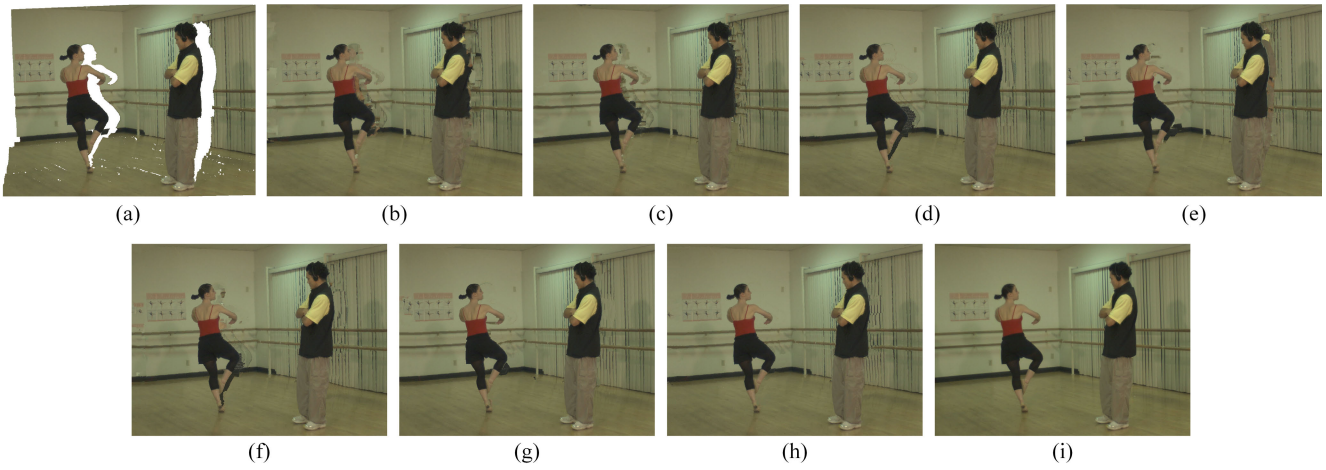


FIGURE 13. Visual quality comparison results of view synthesis for MVD sequences. (a) Warped virtual image. (b) Criminisi's method. (c) Daribo's method. (d) Ahn's method. (e) Kao's method. (f) Zhu's method. (g) Oliveira's method. (h) Proposed method. (i) Ground truth.

is gradual, but depth discontinuity occurs at the foreground edge in the depth image. Therefore, the ghost correction is still necessary. From the visual quality comparison results, the proposed algorithm performs better than other methods, although the data sets contain multiple simple and complex scenes, while other methods contain some defects, including

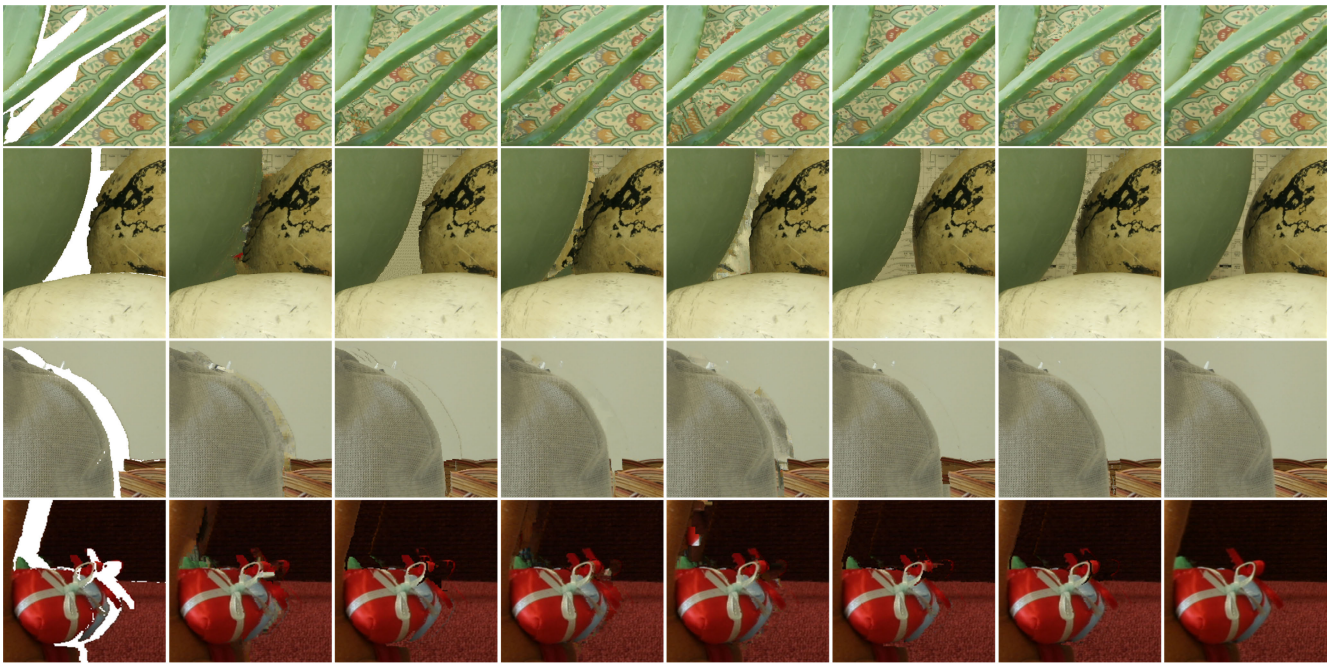


FIGURE 14. Visual quality comparison results of disocclusion filling for still image sequences. (a) Hole regions. (b) Criminisi's method. (c) Ahn's method. (d) Kao's method. (e) Zhu's method. (f) Oliveira's method. (g) Proposed method. (h) Ground truth.

the penetration of foreground texture and the unrealistic filling results.

E. OBJECTIVE QUALITY EVALUATION OF SYNTHESIZED VIRTUAL VIEW

In order to quantitatively evaluate the performance of the proposed method, in our experiment, peak signal to noise ratio (PSNR), structural similarity (SSIM) [37], feature similarity index (color) (FSIMc) [38], and visual saliency-induced index (VSI) [39] are used to evaluate the objective quality of the synthesized view. The calculation formulas for these metrics are as follows:

$$PSNR = 10 \log \frac{255^2}{MSE}, \tag{11}$$

$$MSE = \frac{1}{w \times h} \sum_{u=1}^{w} \sum_{v=1}^{h} \left[I_v(u,v) - I_g(u,v) \right]^2, \tag{12}$$

$$SSIM = \frac{\left(2\mu_{I_v}\mu_{I_g} + c_1\right)\left(2\sigma_{I_v,I_g} + c_2\right)}{\left(\mu_{I_v}^2 + \mu_{I_g}^2 + c_1\right)\left(\sigma_{I_v}^2 + \sigma_{I_g}^2 + c_2\right)}, \tag{13}$$

$$FSIMc = \frac{\sum_{u=1}^{w} \sum_{v=1}^{h} S_L(u,v) \cdot [S_C(u,v)]^{\lambda} \cdot PC_m(u,v)}{\sum_{u=1}^{w} \sum_{v=1}^{h} PC_m(u,v)}, \tag{14}$$

$$VSI = \frac{\sum_{u=1}^{w} \sum_{v=1}^{h} S(u,v) \cdot VS_m(u,v)}{\sum_{u=1}^{w} \sum_{v=1}^{h} VS_m(u,v)}, \tag{15}$$

where I_v and I_g represent the virtual image and the ground truth respectively. w and h represent the width and height of the image. Other parameters are defined in [37]–[39] and we use the default parameter values recommended by the authors. PSNR measures the difference in pixels between two images. SSIM compares the structural similarity between the synthesized image and the ground truth. FSIMc and VSI consider the similarity of the two images in terms of gradient and color features. The evaluation results of SSIM, FSIMc, and VSI are normalized to 0-1. A higher metric value indicates that the synthesized result is closer to the ground truth. The average objective evaluation results for video sequence rendering are shown in Table 1 and 2. The proposed method achieves the best overall results and shows the highest metric value in most scenarios. For the scenario BA54, the proposed method surpasses the previous methods by 0.62-5.62dB in terms of PSNR. It also shows higher metric value in SSIM, FSIMc and VSI. Likewise, there is evident promotion in terms of other scenarios. It is noted that the objective evaluation metrics of BA56 are lower than those of BA54. This is because the disocclusion region exposed in the view 6 contains texture that does not exist in view 5. Our method cannot create new textures to fill the disocclusion. For the image sequence rendering, the objective comparison results are shown in Table 3 and 4. Under the parallel camera configuration and high-precision per-pixel depth image, the result of disocclusion filling is improved. The comparison of the evaluation results shows that the proposed method has better performance than the competitive methods and shows the most realistic filling results. Overall, the virtual view

TABLE 1. PSNR and SSIM comparison results for MVD sequences.

Test	PSNR (dB)							SSIM						
seq.	[17]	[19]	[14]	[20]	[12]	[24]	Ours	[17]	[19]	[14]	[20]	[12]	[24]	Ours
BA54	26.16	29.13	29.57	28.56	30.37	31.16	**31.78**	0.7850	0.7937	0.8081	0.7912	0.8117	0.8159	**0.8290**
BA52	23.95	24.82	24.89	23.97	25.64	25.71	**26.39**	0.7306	0.7415	0.7448	0.7341	0.7422	0.7410	**0.7587**
BA56	25.85	27.54	27.69	27.79	28.13	28.19	**28.81**	0.7855	0.7913	0.7930	0.7894	0.8031	0.8043	**0.8082**
BR43	28.27	28.81	29.63	29.68	30.16	30.58	**30.62**	0.7828	0.7840	0.7884	0.7864	0.7923	0.7946	**0.7955**
BR41	25.67	26.40	26.59	26.43	27.41	27.47	**27.71**	0.7424	0.7455	0.7487	0.7499	0.7541	0.7546	**0.7589**
BR45	28.19	29.55	29.57	29.74	30.11	30.36	**30.88**	0.7831	0.7857	0.7899	0.7892	0.7904	0.7922	**0.7968**
PH67	32.73	32.78	32.80	32.71	32.85	32.94	**33.11**	0.8716	0.8725	0.8728	0.8713	0.8741	0.8749	**0.8756**
PS34	29.79	29.91	29.88	29.83	30.15	30.24	**30.29**	0.8538	0.8565	0.8559	0.8543	0.8582	0.8618	**0.8632**
UD15	26.70	27.43	27.61	27.18	27.75	28.23	**28.36**	0.9243	0.9269	0.9284	0.9254	0.9299	0.9306	**0.9328**

TABLE 2. FSIMc and VSI comparison results for MVD sequences.

Test	FSIMc							VSI						
seq.	[17]	[19]	[14]	[20]	[12]	[24]	Ours	[17]	[19]	[14]	[20]	[12]	[24]	Ours
BA54	0.9269	0.9338	0.9409	0.9292	0.9504	0.9516	**0.9565**	0.9775	0.9842	0.9879	0.9833	0.9886	0.9908	**0.9925**
BA52	0.8238	0.8317	0.8345	0.8235	0.8354	0.8359	**0.8530**	0.9577	0.9607	0.9611	0.9604	0.9625	0.9640	**0.9685**
BA56	0.9206	0.9273	0.9367	0.9337	0.9402	0.9395	**0.9456**	0.9791	0.9823	0.9851	0.9855	0.9871	0.9870	**0.9893**
BR43	0.9492	0.9535	0.9528	0.9530	0.9551	0.9567	**0.9572**	0.9913	0.9918	0.9916	0.9915	0.9922	0.9926	**0.9929**
BR41	0.9011	0.9092	0.9123	0.9009	0.9045	0.9096	**0.9205**	0.9819	0.9839	0.9840	0.9825	0.9836	0.9839	**0.9871**
BR45	0.9576	0.9618	0.9612	0.9607	0.9615	0.9623	**0.9638**	0.9924	0.9933	0.9935	0.9934	0.9938	0.9939	**0.9943**
PH67	0.9651	0.9678	0.9713	0.9684	0.9741	0.9748	**0.9754**	0.9950	0.9955	0.9962	0.9957	0.9965	0.9967	**0.9968**
PS34	0.9711	0.9723	0.9736	0.9729	0.9744	0.9746	**0.9752**	0.9919	0.9927	0.9931	0.9929	0.9933	0.9937	**0.9943**
UD15	0.9636	0.9659	0.9697	0.9691	0.9718	0.9748	**0.9775**	0.9931	0.9938	0.9946	0.9943	0.9947	0.9951	**0.9956**

TABLE 3. PSNR and SSIM comparison results for still image data sets.

Test seq.	PSNR (dB)						SSIM					
	[17]	[14]	[20]	[12]	[24]	Ours	[17]	[14]	[20]	[12]	[24]	Ours
Aloe	28.13	27.91	29.28	29.37	29.41	**29.76**	0.9007	0.8943	0.9069	0.9065	0.9063	**0.9083**
Art	26.81	27.41	27.78	29.65	29.77	**30.25**	0.8979	0.9049	0.9094	0.9137	0.9160	**0.9169**
Baby1	32.65	33.07	33.42	33.51	33.70	**33.87**	0.9306	0.9289	0.9322	0.9325	0.9336	**0.9347**
Bowling2	27.11	26.35	27.92	28.29	30.43	**30.51**	0.9098	0.8991	0.9161	0.9278	0.9286	**0.9293**
Lampshade2	31.36	33.78	35.49	35.99	36.42	**36.85**	0.9426	0.9456	0.9563	0.9564	0.9578	**0.9583**
Midd2	29.03	28.34	29.97	30.15	30.08	**30.41**	0.9204	0.9185	0.9302	0.9311	0.9337	**0.9345**
Reindeer	32.17	33.06	32.93	33.28	33.57	**33.76**	0.9230	0.9251	0.9256	0.9274	0.9294	**0.9322**

TABLE 4. FSIMc and VSI comparison results for still image data sets.

Test seq.	FSIMc						VSI					
	[17]	[14]	[20]	[12]	[24]	Ours	[17]	[14]	[20]	[12]	[24]	Ours
Aloe	0.9554	0.9596	0.9736	0.9734	0.9742	**0.9768**	0.9887	0.9891	0.9933	0.9916	0.9924	**0.9931**
Art	0.9394	0.9567	0.9671	0.9711	0.9738	**0.9742**	0.9809	0.9901	0.9907	0.9914	0.9930	**0.9938**
Baby1	0.9716	0.9723	0.9793	0.9771	0.9787	**0.9805**	0.9955	0.9953	0.9963	0.9962	0.9966	**0.9972**
Bowling2	0.9369	0.9316	0.9509	0.9694	0.9696	**0.9726**	0.9881	0.9857	0.9889	0.9912	0.9942	**0.9947**
Lampshade2	0.9534	0.9726	0.9799	0.9809	0.9826	**0.9867**	0.9914	0.9932	0.9975	0.9956	0.9977	**0.9984**
Midd2	0.9616	0.9606	0.9752	0.9806	0.9779	**0.9827**	0.9939	0.9926	0.9966	0.9964	0.9952	**0.9968**
Reindeer	0.9747	0.9819	0.9780	0.9838	0.9863	**0.9909**	0.9941	0.9966	0.9957	0.9966	0.9972	**0.9979**

generated by the proposed method has better objective quality and shows robustness for scenes containing moving objects and complex textures to some extent.

F. COMPUTATIONAL COST ANALYSIS

Compared to directly applying image inpainting algorithm to fill the disocclusions. the computational cost of our method increases because we add the morphology-based preprocessing and local foreground removal process. Table 5 shows the comparison of running time among our method and competitive methods on two MVD sequences. Criminisi's method [17] has the lowest running time because it only requires color images as input. Daribo's method [19] introduces depth information, thus increasing the computational cost. In Ahn's

TABLE 5. Comparison of running time (Unit: s).

Test seq.	Running time						
	[17]	[19]	[14]	[20]	[12]	[24]	Ours
BA54	13.5	16.1	18.3	18.9	19.1	20.3	21.9
UD15	26.2	32.5	33.8	33.2	33.5	35.4	37.1

method [14], Kao's method [20], and Zhu's method [12], the depth image of virtual view is filled synchronously without using the ground truth. Oliveira's method [24] adds the reverse warping process and searches for the best matching patch in the reference image. The computational cost of our method is the greatest because the corresponding foreground is removed to predict the occlusion layer. In addition, searching in the reference image can obtain more effective background contents and prevent foreground interference. Our work mainly focuses on the improvement of the virtual view quality and achieves higher evaluation metrics than the other methods. By using parallel computing, the running time can be effectively reduced.

G. DISCUSSION

Experiments on public video and image sequences show that the proposed method can effectively fill the disocclusion and generate high-quality virtual view. Morphology-based depth image preprocessing detects foreground edge pixels in all directions and corrects ghosts. These pixels are still located at the foreground edge in the virtual image, thus maintaining the boundary of the foreground object without deleting the valid pixels. For disocclusion filling, we remove the corresponding local foreground in the reference image and use the surrounding background contents to predict the occlusion layer, instead of directly applying the inpainting method in the virtual image. This operation can prevent the errors generated in 3D warping from being sampled into the disocclusion. Compared to entire foreground removal, removing the local foreground region can decrease computational cost. In addition, if edge detection is directly applied to extract foreground contours in the reference image, such as the cross-bilateral filtering combined with the canny operator proposed by Luo and Kim [21], some foreground pixels are also ignored because the depth discontinuity is too small. In our method, based on the disocclusion edge pixels, it can be ensured that the associated foreground pixels are removed in the reference image. The inpainting process uses depth information to encourage the background texture to be preferentially propagated and search for target patch that has similar depth and texture information to fill the removed region. The subjective and objective evaluation results show that the proposed method has an improvement in the disocclusion handling compared with the other methods. For video sequence rendering, some methods use background modeling to prevent the flicker between adjacent frames. The background model obtained in the current frame can be used to fill the disocclusions in other frames [4], [40], [41]. This is effective for still background scene. For scenes where background texture changes or still foreground objects, the currently established background model is no longer applicable in subsequent frames. In this case, frame-by-frame disocclusion filling is more advantageous. The proposed method performs occlusion layer prediction in the reference image. For common disocclusions in adjacent frames, the prediction result of the occlusion layer is similar because the removed regions are the same. This helps maintain the temporal consistency of synthesized image. In terms of computational complexity, the proposed method adds a series of approaches to deal with artifacts based on 3D warping. While improving the subjective and objective quality, the computational complexity is increased. The proposed method includes depth image preprocessing, forward and inverse 3D warping, and the most time-consuming inpainting process. To improve rendering efficiency, GPU-based parallel computing is a measure that can be considered. In some steps of our method, such as the classification of disocclusion edge and the priority calculation, each pixel is processed independently. Therefore, the use of parallel computing in these processes can greatly reduce the computation cost.

Recently, deep learning has been widely used in the field of image processing, such as depth prediction [42], moving object detection [43], and image inpainting [44]. This can provide some help for DIBR-based methods. Traditional image inpainting algorithms predict texture based on existing information. But deep learning technology can generate new textures from large amounts of data. In addition, some end-to-end view synthesis techniques based on deep learning have been proposed [45], [46]. They allow for the generation of new views of a scene given a single input image. In our follow-up work, combining the proposed method with deep learning is a feasible measure that can effectively improve the visual quality of the virtual view.

V. CONCLUSION

This paper presents an effective disocclusion handling method for virtual view synthesis.We perform morphological-based ghost removal before 3D warping, therefore the ghosts can be warped to the correct place. Disocclusion filling is achieved based on the local foreground removal approach. By locating and classifying the disocclusion edge pixels, we remove the corresponding local foreground in the reference image. The predicted occlusion layer is projected to the virtual view and completes the disocclusion filling. Experimental results demonstrate that the proposed method can effectively remove ghosts, and keep sharp foreground boundaries with reasonable texture in disocclusion filling. Compared with the other methods, the proposed method has better performance in visual quality and objective evaluation. Our current work focuses on improving the quality of virtual view and the computational complexity is increased correspondingly. In the future, we will investigate to improve the rendering efficiency while maintaining the visual quality. Moreover, the temporal correlation of adjacent frames will

also be considered [47]. As the foreground object moves, the removed foreground region also changes, so the background content visible in different frames can be used to build a background model in the time domain. Since the content of the background model is reliable instead of predicted, the combination of local foreground removal and background modeling helps to further improve the quality of the virtual view and minimize the flicker between frames. In addition to the general quality assessment metrics, some assessment metrics for view synthesis have been proposed [48], [49]. In the future, we will study these methods to better evaluate the quality of the synthesized image.

ACKNOWLEDGMENT

The authors would like to thank the Interactive Visual Media Group at Microsoft Research for making the MSR 3D Video Dataset publicly available.

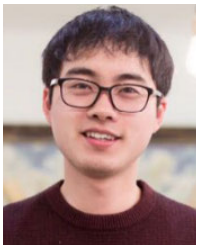
REFERENCES

[1] G. Tech, Y. Chen, K. Muller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, Jan. 2016.
[2] H. Xu, X. Chen, H. Liang, S. Ren, Y. Wang, and H. Cai, "Crosspatch-based rolling label expansion for dense stereo matching," *IEEE Access*, vol. 8, pp. 63470–63481, 2020.
[3] M. Tanimoto, "FTV: Free-viewpoint television," *Signal Process., Image Commun.*, vol. 27, no. 6, pp. 555–570, Jul. 2012.
[4] G. Luo, Y. Zhu, Z. Li, and L. Zhang, "A hole filling approach based on background reconstruction for view synthesis in 3D video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1781–1789.
[5] S. Yin, H. Dong, G. Jiang, L. Liu, and S. Wei, "A novel 2D-to-3D video conversion method using time-coherent depth maps," *Sensors*, vol. 15, no. 7, pp. 15246–15264, Jun. 2015.
[6] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," *Proc. SPIE*, vol. 5291, pp. 93–104, May 2004.
[7] W. R. Mark, L. Mcmillan, and G. Bishop, "Post-rendering 3D warping," in *Proc. Symp. Interact. 3D Graph. (SI3D)*, 1997, pp. 7–16.
[8] X. D. Chen, H. T. Liang, H. Y. Xu, S. Y. Ren, H. Y. Cai, and Y. Wang, "Virtual view synthesis based on asymmetric bidirectional DIBR for 3D video and free viewpoint video," *Appl. Sci.-Basel*, vol. 10, no. 5, p. 19, Mar. 2020.
[9] W. J. Tam, G. Alain, L. Zhang, T. Martin, and R. Renaud, "Smoothing depth maps for improved steroscopic image quality," *Proc. SPIE*, vol. 5599, Oct. 2004, pp. 162–172.
[10] Y.-R. Horng, Y.-C. Tseng, and T.-S. Chang, "Stereoscopic images generation with directional Gaussian filter," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2010, pp. 2650–2653.
[11] W.-Y. Chen, Y.-L. Chang, S.-F. Lin, L.-F. Ding, and L.-G. Chen, "Efficient depth image based rendering with edge dependent depth filter and interpolation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2005, Amsterdam, The Netherlands, pp. 1314–1317.
[12] S. Zhu, H. Xu, and L. Yan, "An improved depth image based virtual view synthesis method for interactive 3D video," *IEEE Access*, vol. 7, pp. 115171–115180, 2019.
[13] J. Lei, C. Zhang, M. Wu, L. You, K. Fan, and C. Hou, "A divide-and-conquer hole-filling method for handling disocclusion in single-view rendering," *Multimedia Tools Appl.*, vol. 76, no. 6, pp. 7661–7676, Mar. 2017.
[14] I. Ahn and C. Kim, "A novel depth-based virtual view synthesis method for free viewpoint video," *IEEE Trans. Broadcast.*, vol. 59, no. 4, pp. 614–626, Dec. 2013.
[15] M. S. Farid, M. Lucenteforte, and M. Grangetto, "Depth image based rendering with inverse mapping," in *Proc. IEEE 15th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2013, pp. 135–140.
[16] S. Zinger, L. Do, and P. H. N. de With, "Free-viewpoint depth image based rendering," *J. Vis. Commun. Image Represent.*, vol. 21, nos. 5–6, pp. 533–541, Jul. 2010.
[17] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.
[18] G. Luo, Y. Zhu, and B. Guo, "Fast MRF-based hole filling for view synthesis," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 75–79, Jan. 2018.
[19] I. Daribo and H. Saito, "A novel inpainting-based layered depth video for 3DTV," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 533–541, Jun. 2011.
[20] C.-C. Kao, "Stereoscopic image generation with depth image based rendering," *Multimedia Tools Appl.*, vol. 76, no. 11, pp. 12981–12999, Jun. 2017.
[21] G. Luo and Y. Zhu, "Foreground removal approach for hole filling in 3D video and FVV synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2118–2131, Oct. 2017.
[22] D. Han, H. Chen, C. Tu, and Y. Xu, "View synthesis using foreground object extraction for disparity control and image inpainting," *J. Vis. Commun. Image Represent.*, vol. 56, pp. 287–295, Oct. 2018.
[23] S. Smirnov, F. Battisti, and A. Gotchev, "Layered approach for improving the quality of free-viewpoint depth-image-based rendering images," *J. Electron. Imag.*, vol. 28, no. 1, p. 17, Jan. 2019.
[24] A. Q. de Oliveira, M. Walter, and C. R. Jung, "An artifact-type aware DIBR method for view synthesis," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1705–1709, Nov. 2018.
[25] S. H. Abdulhussain, B. M. Mahmmod, M. I. Saripan, S. A. R. Al-Haddad, T. Baker, W. N. Flayyih, and W. A. Jassim, "A fast feature extraction algorithm for image and video processing," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2019, pp. 1–8.
[26] S. H. Abdulhussain, S. A. R. Al-Haddad, M. I. Saripan, B. M. Mahmmod, and A. Hussien, "Fast temporal video segmentation based on Krawtchouk-Tchebichef moments," *IEEE Access*, vol. 8, pp. 72347–72359, 2020.
[27] A. Sasithradevi and S. M. M. Roomi, "A new pyramidal opponent color-shape model based video shot boundary detection," *J. Vis. Commun. Image Represent.*, vol. 67, p. 12, Feb. 2020.
[28] C. Yao, T. Tillo, Y. Zhao, J. Xiao, H. Bai, and C. Lin, "Depth map driven hole filling algorithm exploiting temporal correlation information," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 394–404, Jun. 2014.
[29] D. M. M. Rahaman and M. Paul, "Hole-filling for single-view plus-depth based rendering with temporal texture synthesis," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2016, pp. 1–6.
[30] D. M. M. Rahaman and M. Paul, "Virtual view synthesis for free viewpoint video and multiview video compression using Gaussian mixture modelling," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1190–1201, Mar. 2018.
[31] S. M. Muddala, M. Sjöström, and R. Olsson, "Virtual view synthesis using layered depth image generation and depth-based inpainting for filling disocclusions and translucent disocclusions," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 351–366, Jul. 2016.
[32] X. D. Chen, H. T. Liang, H. Y. Xu, S. Y. Ren, H. Y. Cai, and Y. Wang, "Artifact handling based on depth image for view synthesis," *Appl. Sci.-Basel*, vol. 9, no. 9, p. 19, May 2019.
[33] H. Lim, Y. S. Kim, S. Lee, O. Choi, J. D. K. Kim, and C. Kim, "Bi-layer inpainting for novel view synthesis," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 1089–1092.
[34] 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.
[35] H. Schwarz, D. Marpe, and T. Wiegand, "Description of explorationexperiments in 3D video coding," Int. Org. Standardization, Dresden, Germany, Tech. Rep. MPEG2010/N11274, Apr. 2010.
[36] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[38] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[39] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014.
[40] Z. M. Deng and M. J. Wang, "Reliability-based view synthesis for free viewpoint video," *Appl. Sci.-Basel*, vol. 8, no. 5, p. 15, May 2018.
[41] G. Luo and Y. Zhu, "Hole filling for view synthesis using depth guided global optimization," *IEEE Access*, vol. 6, pp. 32874–32889, 2018.

[42] Z. Li and N. Snavely, "MegaDepth: Learning single-view depth prediction from Internet photos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2041–2050.
[43] H. Zhu, X. Yan, H. Tang, Y. Chang, B. Li, and X. Yuan, "Moving object detection with deep CNNs," *IEEE Access*, vol. 8, pp. 29729–29741, 2020.
[44] Y. Jiang, J. Xu, B. Yang, J. Xu, and J. Zhu, "Image inpainting based on generative adversarial networks," *IEEE Access*, vol. 8, pp. 22884–22892, 2020.
[45] H. Zhou, J. Liu, Z. Liu, Y. Liu, and X. Wang, "Rotate-and-render: Unsupervised photorealistic face rotation from single-view images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5910–5919.
[46] O. Wiles, G. Gkioxari, R. Szeliski, and J. Johnson, "SynSin: End-to-end view synthesis from a single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7465–7475.
[47] G. Luo, Y. Zhu, Z. Weng, and Z. Li, "A disocclusion inpainting framework for depth-based view synthesis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 6, pp. 1289–1302, Jun. 2020.
[48] D. Sandic-Stankovic, D. Kukolj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological pyramids," in *Proc. 3DTV-Conf., True Vis. Capture, Transmiss. Display 3D Video (3DTV-CON)*, Jul. 2015, pp. 1–4.
[49] X. Wang, F. Shao, Q. Jiang, R. Fu, and Y.-S. Ho, "Quality assessment of 3D synthesized images via measuring local feature similarity and global sharpness," *IEEE Access*, vol. 7, pp. 10242–10253, 2019.

HAITAO LIANG received the B.S. degree from Tianjin University, Tianjin, China, in 2016, where he is currently pursuing the Ph.D. degree in optical engineering. His research interests include image processing, computer vision, and multimedia technology.

XIAODONG CHEN received the Ph.D. degree in optical engineering from Tianjin University. He is currently a Professor with the School of Precision Instruments and Opto-Electronic Engineering, Tianjin University. He is the author of two books, more than 180 articles, and more than seven inventions. His research interests include photoelectric detection technology and instrument, image processing, and machine vision detection.



HUAIYUAN XU received the M.S. degree from Tianjin University, Tianjin, China, in 2017, where he is currently pursuing the Ph.D. degree in optical engineering. His research interests include stereo matching and image processing.

SIYU REN received the B.S. degree from Tianjin University, China, in 2018, where he is currently pursuing the M.S. degree. His research interests include computer vision and image processing.

HUAIYU CAI received the Ph.D. degree in optical engineering from Tianjin University. She is currently a Professor with the School of Precision Instruments and Opto-Electronic Engineering, Tianjin University. She is the author of one book and more than 70 articles. Her research interests include photoelectric imaging and detection technology, information optics, and image processing technology.

YI WANG received the Ph.D. degree in optical engineering from Tianjin University. She currently works with the Key Laboratory of Opto-Electronics Information Technology, Ministry of Education, Tianjin University. Her research interests include photoelectric detection technology and instrument, optical coherence tomography, and medical image processing.