

Hole Filling in Image Conversion Using Weighted Local Gradients

Yung-Wen Lee,¹ Dong-Meau Chang,¹
Chun-Chi Chen² and Chyuan-Huei Thomas Yang^{3*}

¹School of Mathematics and Information Engineering, Longyan University, Longyan, Fujian 364012, China

²School of Life Sciences, Longyan University, Longyan, Fujian 364012, China

³Department of Information Management, Hsuan Chuang University, Hsinchu 30092, Taiwan

(Received July 20, 2020; accepted March 8, 2021)

Keywords: image conversion, weighted local gradients, hole filling, morphological operations

Three-dimensional displays have become an important medium. Most of the 3D content for these displays is converted from 2D content. The conversion requires appropriate depth information from 2D images, but the information is insufficient to create high-quality 3D images in many cases. Therefore, converting 2D images into 3D images has become an important issue in emerging 3D applications. As even sophisticated equipment produces noise, cracks, and holes in the converted images, the conversion requires an improved method for creating high-quality 3D images, especially a hole-filling method. This method is one of the techniques in depth image-based rendering (DIBR), which converts 2D single-depth images into 3D images by using a multiangle virtual view and stereoscopic image generation. We propose a new algorithm for hole filling in DIBR that converts 2D images into 3D images by maintaining the depth structure of occlusions and performing morphological operations on images and the background depth levels. This algorithm uses edge information, morphology, gradient vector flow, and weighted local gradients to render holes. Multiview images generated by the proposed algorithm have synthesized virtual views of better quality than those generated by previous methods.

1. Introduction

Video technology has rapidly developed from black and white to color, and to 3D televisions and movies. Nowadays, 3D technology can be integrated into virtual reality (VR) through the integration of computer graphics and vision technology. The future of stereoscopic displays is directed toward interactive free-viewpoint 3D TVs that show high-quality 3D images.

Three-dimensional technology requires the processing of stereoscopic images based on the principle of parallax. Applying the principle enables the conversion of a 2D image into a 3D image, which creates various depths of a picture and gives a sense of reality. Three-dimensional images can be seen with several different devices. A head-mounted VR display outputs two stereoscopic images at the same time to give a 3D reality. Red–blue glasses, polarizing glasses,

*Corresponding author: e-mail: chyang@hcu.edu.tw
<https://doi.org/10.18494/SAM.2021.3046>

and shutter glasses also produce 3D effects. All these devices apply the parallax principle to create the illusion of image depth. Creating sophisticated 3D stereoscopic images has attracted much research interest as the related technology requires the improvement of previous conversion methods.

A common way of converting 2D images into 3D images is to apply depth image-based rendering (DIBR), which simulates parallax to create 3D effects by using 2D plane images by synthesizing depths of the images and mapping multiview virtual maps. However, 2D images created from a low-resolution image sensor have noises, cracks, or holes. Although a high-resolution sensor creates less noise, it still generates holes in images with 3D warping. This is due to a lack of information in certain areas of images, which is unavoidable. Thus, it is still necessary to improve the conversion method for the developed image sensor technology.

DIBR is an essential technique for converting 2D images to 3D images.^(1–4) Different gradients or gradient vector flows (GVFs) of the depth in images need different filling techniques.^(5–10) Cheng *et al.* proposed an algorithm for a hypothesized depth gradient model and used a bilateral filter to generate visually comfortable depth maps and diminish block artifacts.⁽¹⁾ Cho *et al.* investigated the local and neighboring property of a depth map in which the background region was filled first by considering depth homogeneity.⁽²⁾ To reduce the number of artifacts, the boundary between the background and foreground was filled with background-relevant data. Lie *et al.* proposed a DIBR scheme with a background sprite model for disocclusion and hole filling for 3D TV.⁽³⁾ A hole-filling method for an edge-oriented morphological operation of DIBR was proposed by Yang *et al.*⁽⁴⁾ They applied a bilateral filter as a smooth filter in DIBR and found occlusion regions by edge detection in the depth map to maintain the depth structure. However, DIBR of a virtual image creates image displacement such as holes and distortions; thus, DIBR requires significant improvement.

To address this problem, Huang *et al.*⁽⁵⁾ proposed a multiframe super-resolution (SR) method that applied a gradient vector flow hybrid field algorithm with an anisotropic diffusion shock filter (GVFHF-ADSF) and effectively achieved image denoising and enhancement. Wu *et al.*⁽⁶⁾ adopted masks of four neighborhoods to replace the original masks and calculated the GVF for the extended neighborhood. Zhou *et al.* proposed a mean-shift-based GVF (MSGVF) segmentation algorithm that correctly located borders.⁽⁷⁾ The MSGVF kept the smoothness constraint of image pixels until the contour reached an equilibrium and reduced under-segmentation. Kim and Ro proposed an optimized hole-filling method by using the information from the adjacent and previous filled frame to maintain the spatiotemporal consistency and binocular symmetry in synthesized 3D videos from multiple VR viewpoints.⁽⁸⁾ Luo and Zhu proposed a foreground removal approach for hole filling that removed the foreground objects from the corresponding depth map to eliminate the holes in the synthesized video.⁽⁹⁾ Seiler *et al.* proposed an optimized processing order for improving the extrapolation quality of 3D frequency-selective extrapolation.⁽¹⁰⁾

Despite the above research results, a new method is still necessary to solve the problems of DIBR to obtain a more natural image quality. Therefore, we propose a method with a new algorithm to fill holes by applying DIBR to image warping with a weighted local GVF. The new method is expected to provide a basis for creating better 3D stereoscopic images than those obtained with previous technologies.

2. Materials and Methods

2.1 Image warping

Three-dimensional image synthesis techniques are divided into image-based rendering (IBR) and DIBR. DIBR is based on a stereoscopic system for objects, and its process consists of three parts: (1) preprocessing, (2) 3D image warping, and (3) hole filling. Preprocessing uses a depth map with noise reduced by a smooth filter and sets a zero plane (left and right eye images of parallax as the coordinates of the origin). General views of the zero plane for the depth value have 128 or 255 points. Figure 1 shows a diagram of a camera and a stereoscopic image, where C_l , C_r , and C_c are the optical centers of the left eye, the right eye, and the camera, respectively.

The image is taken from two cameras on the left and right sides and synthesized by considering the displacements on both sides, which are expressed by

$$x_l = x_c + \frac{t_x f}{2 Z}, \quad (1)$$

$$x_r = x_c + \frac{t_x f}{2 Z}, \quad (2)$$

where x_l , x_r , and x_c are the coordinates of the pixels in the left eye, right eye, and the center image, respectively, f is the focal length (a constant), t_x is the baseline length, and Z is the depth value at location p (the smaller the Z , the greater the depth grayscale value). A larger parallax causes the images on the left and right sides to shift further. When synthesizing a virtual view with a single color and depth image, occlusion regions appear. An occlusion is a visible region created as a result of a different viewpoint as shown in Fig. 2. Since the occlusion regions have no input data, they are shown as holes in the synthesized image.

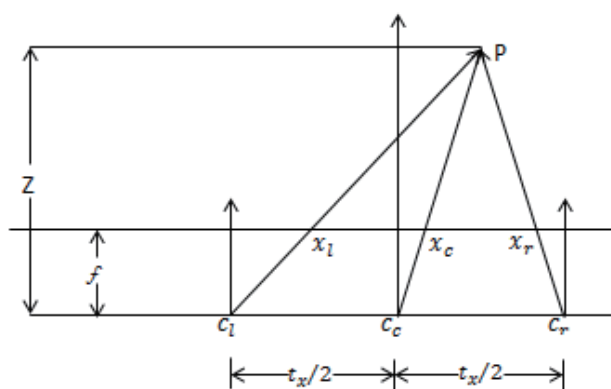


Fig. 1. Diagram of a camera and a stereoscopic image.

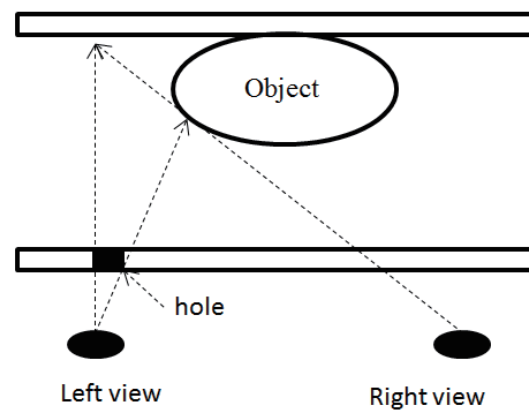


Fig. 2. Hole appearing in a synthesized image.

2.2 Relevant algorithm

Figure 3 shows a flowchart of a hole-filling process. DIBR creates images on the left and right sides with cracks and holes by 3D warping, which generates two distorted images on both sides when creating a single image. To remove them, methods based on edge detection, the GVF, median filter, bilateral filter, and mathematical morphology are applied.

Sobel edge detection was used in this study for detecting edges. It contains image convolution operations, mainly in 3×3 masks. The vertical and horizontal directions (G_x and G_y , respectively) of the computing image satisfy the following.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (3)$$

Mathematical morphology contains a set of algebraic operators for analyzing and processing image shapes and structures through collecting image pixels. The operation of the morphology is divided into four main steps.

Dilation: Image A and a collection of pixels, B, of structural elements are used to extend A-origin pixels with pixel point Q to extend the edge lines of objects.

$$D(A, B) = A \oplus B = \bigcup_{Q \in B} (A)_Q \quad (4)$$

Erosion: Image A and the structure element pixel collection B are used with pixel point Q to find the echoing B-origin pixel, which is used to cut the edge line of the object.

$$E(A, B) = A \ominus B = \{Q : B_Q \subseteq A\} \quad (5)$$

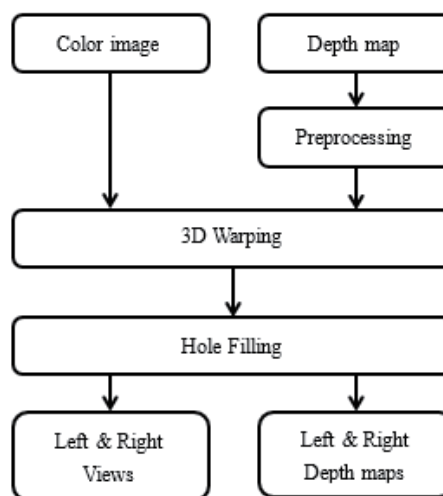


Fig. 3. Flowchart of hole-filling process.

Opening: First, image erosion is performed, and then image dilation is performed to remove small points of noise from the image.

$$O(AB) = A \circ B = (A \ominus B) \oplus B \quad (6)$$

Closing: An image dilation operation is first performed, followed by an image erosion operation to repair the connection of small holes and broken wires in the image.

$$C(AB) = A \bullet B = (A \oplus B) \ominus B \quad (7)$$

The GVF modifies the characteristics of the original gradient vector field and expands the edge image in the vector field with high homogeneity. Thus, the recursive secondary vector field increases with the expansion of the range of the GVF and the direction of the edge vector field. The GVF is

$$G(x, y) = [u(x, y), v(x, y)]. \quad (8)$$

The vector flow that minimizes the energy correspondence is defined as

$$\varepsilon = \iint \mu (u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |v - \nabla f|^2 dx dy, \quad (9)$$

where μ is the weight of the recursive parameters.

The median filter technique is an example of nonlinear filtering. It sorts pixel values in the filter range and replaces the current pixel values with the median value to effectively preserve the edge characteristics. The median filter is often used to reduce image noises such as salt-and-pepper and speckle noises.

The bilateral filter also performs nonlinear filtering, and uses two functions to process the geometric distance, brightness, and color difference between pixels. It effectively handles the noise and saves the edge information of the image, which takes a longer time than when using the median filter technique. The bilateral filter satisfies

$$I^{filtered}(x) = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) f_r(\|I(x_i) - I(x)\|) g_s(\|x_i - x\|). \quad (10)$$

2.3 Proposed algorithm

The proposed algorithm is executed by the following steps.

Step 1: Input a color image and depth.

Step 2: The bilateral filter is applied to the color image and depth, and acts on the adjacent color changes to eliminate the distortion by noise. Three-dimensional warping is executed using the color image and depth after bilateral filtering.

Step 3: The median filtering and morphological operation fill the gaps and holes in the previous color view and depth maps on the left and right sides. Then, the view and depth maps are processed after the holes are filled.

Step 4: The edge detection finds the edges of large holes by the image-labeling method. The direction of the GVF is used to find color or depth information of the edges of the large holes.

Step 5: An n by n weighted GVF is created for the labeled large holes of the images and the maps, and the weighted GVF is used to process the images and the maps.

Step 6: The median filter fills the holes.

Step 7: Color images and depth maps are output.

The algorithm and flowchart of the process in this proposed method are respectively shown in Figs. 4 and 5. We applied a GVF of 3 by 3 or 5 by 5 masks to find the maximum magnitude of the gradient of the local area. Then, the position with the maximum magnitude becomes the candidate direction and the candidate values of color or depth. The masks used for the local weights of the above algorithm are given in Fig. 6. On the basis of the algorithm and the local GVF, we established the flowchart of processing images in the proposed method as shown in Fig. 5.

```

Input  $A_i$ , where  $A_1$ : RGB left view;  $A_2$ : RGB right view;  $A_3$ : left depth map;  $A_4$ : right depth map
for  $i = 1$  to 4
     $I = A_i$ ; %preprocessing(step 4)
     $J = \text{magnitude}(\text{gradient}(I))$ ; % magnitude of gradient vectors
     $L = \text{label}(A_i)$ ; % labelling the large holes
     $\bar{A}_i = A_i$ ; % initialization
    If( $p \neq 0, p \in L$ )
         $\bar{A}_i(p) = \omega \times I(p) + (1 - \omega) \left[ \left( \frac{J_{n \times n(p)} \times K_l}{\sum J_{n \times n(p)} \times K_l} \right) \times I_{n \times n}(p) \right]$ ;
        where  $0 \leq \omega \leq 1$ 
        where  $K_{1(n \times n)} = K_{3(n \times n)}$  (left view, left depth)
             $K_{1(i,j)} = K_{3(i,j)} = \begin{cases} 1, j \leq \left(\frac{n+1}{2}\right), 1 \leq i \cdot j \leq n, n \text{ is an odd}, n \geq 3 \\ 0, \text{otherwise.} \end{cases}$ 
        where  $K_{2(n \times n)} = K_{4(n \times n)}$  (right view, right depth)
            where  $K_{2(i,j)} = K_{4(i,j)} = \begin{cases} 1, j \geq \left(\frac{n+1}{2}\right), 1 \leq i \cdot j \leq n, n \text{ is an odd}, n \geq 3 \\ 0, \text{otherwise.} \end{cases}$ 
        end
         $\bar{A}_i(p) = \text{median filter}(\bar{A}_i(p))$ 
    end
end
Output  $\bar{A}_i$ , where  $\bar{A}_i$  are final results of RGB images or depth maps,  $1 \leq i \leq 4$ 

```

Fig. 4. Algorithm of the proposed method.

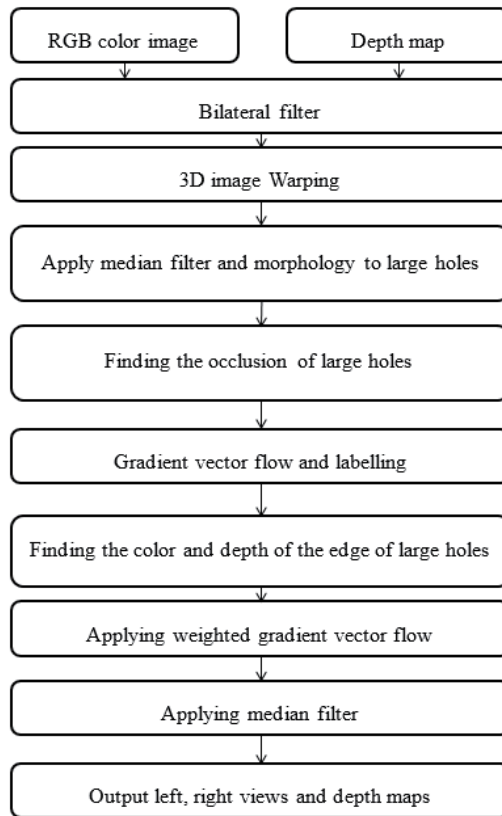


Fig. 5. Flowchart of the proposed method.

1	1	0
1	1	0
1	1	0

(a)

0	1	1
0	1	1
0	1	1

(b)

1	1	1	0	0
1	1	1	0	0
1	1	1	0	0
1	1	1	0	0
1	1	1	0	0

(c)

0	0	1	1	1
0	0	1	1	1
0	0	1	1	1
0	0	1	1	1
0	0	1	1	1

(d)

Fig. 6. Masks used for a local weighted GVF. (a) $K_{1(3 \times 3)} = K_{3(3 \times 3)}$. (b) $K_{2(3 \times 3)} = K_{2(3 \times 3)}$. (c) $K_{1(5 \times 5)} = K_{3(5 \times 5)}$. (d) $K_{2(5 \times 5)} = K_{2(5 \times 5)}$.

3. Experimental Results

We used images and depth maps of a ballet dancer for the experiment (Fig. 7). The original image and depth map were preprocessed using a bilateral filter, and then 3D warping was performed to obtain Fig. 8. After 3D warping, cracks or holes were found in the views and the depth maps. To get rid of the cracks and fill the holes, the GVF and the magnitude of the gradient were applied (Figs. 9 and 10).



Fig. 7. (Color online) Image and depth map used for the experiment. (a) RGB image of a ballet dancer. (b) Depth map.

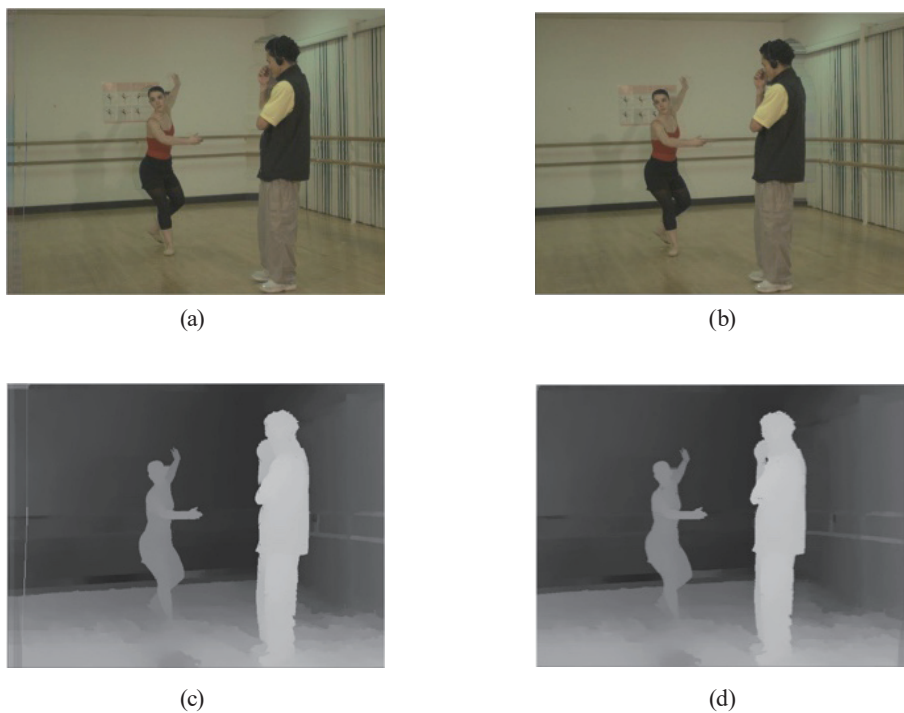


Fig. 8. (Color online) Preprocessed images and depth maps of the original image. (a) Left view of RGB image. (b) Right view of RGB image. (c) Left depth map. (d) Right depth map.



Fig. 9. (Color online) GVFs in (a) left depth map and (b) right depth map.



Fig. 10. Magnitude of gradients in (a) left depth map and (b) right depth map.



Fig. 11. Labels of large holes in (a) left depth map and (b) right depth map.

The edge detection identified the edges of the occlusion areas and found the color information of the edges by using the direction of the GVF. Then, the large holes were labeled. Figure 11 shows the large holes of the left and right depth maps.

Then, mathematical morphology was used to fill the large holes, which were then modified by the locally weighted gradient with the median filter for the rendering regions. Figure 12 shows the results.

The large holes in the image were filled with the locally weighted gradient vector; then, the holes were processed with a median filter. Figure 13 shows the result of the filling with the median filter of the locally weighted gradient vector.

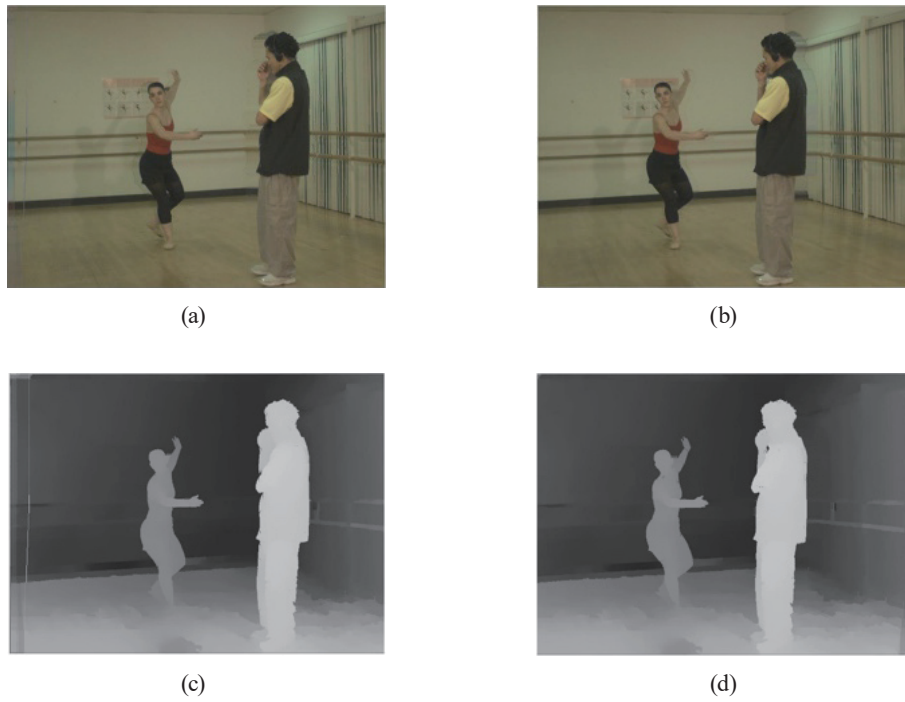


Fig. 12. (Color online) Hole filling modified by the locally weighted gradient vector. (a) Left view of RGB image. (b) Right view of RGB image. (c) Left depth map. (d) Right depth map.

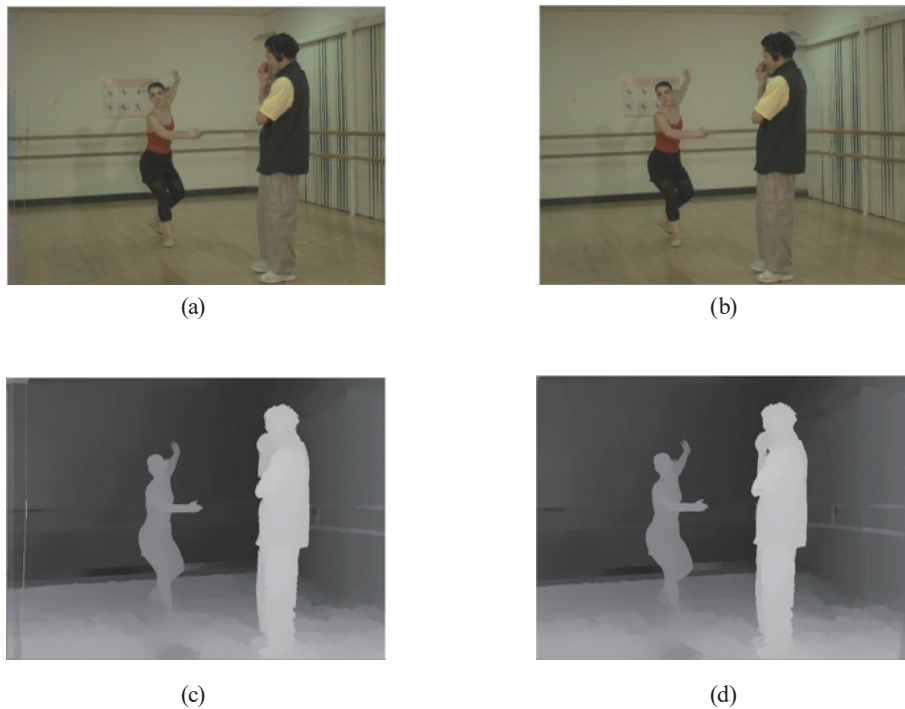


Fig. 13. (Color online) Hole filling by the proposed method using locally weighted gradient vector. (a) Left view of RGB image. (b) Right view of RGB image. (c) Left depth map. (d) Right depth map.

4. Conclusions

A new algorithm was proposed to fill holes of different sizes in 3D stereoscopic images. This technique included the use of bilateral filtering, 3D warping, median filtering, morphological operation, edge detection, a labeling method, the GVF, and the gradient. The bilateral filter was used to handle depth and adjacent color changes in the image and depth maps and eliminated the distortion of the noise of the image. After 3D warping followed by median filtering, a morphological operation mended the gaps and holes in the views and maps on the left- and right-side views. Then, edge detection found the edges of the large holes. The direction of the GVF was used to obtain the color information of the edge of the holes, and the labeling method labeled the holes. The locally weighted gradient vector and flow were applied to fill the holes. The experimental results show that the proposed method effectively filled the holes and produced high-quality images. The new method is expected to lead to a future study on converting 2D images to 3D images in better and more efficient ways and to provide more real and natural images than the previous methods.

Acknowledgments

The works are supported by the Longyan University's Qi Mai Science and Technology Innovation Fund Project of Longyan City (2017QM0201 and 2018LYQM0202), Longyan University's Qi Mai Science and Technology Innovation Fund Project of Liancheng ([2018]132) and Shanghang (2019SHQM05) County, Longyan and Longyan University's Research and Development Team Fund ((2018)8), Great project of production, teaching, research of Fujian provincial science and Technology Department (2019H6023).

References

- 1 C.-C. Cheng, C.-T. Li, and L.-G. Chen: IEEE Trans. Consum. Electron. **56** (2010) 1739. <https://doi.org/10.1109/TCE.2010.5606320>
- 2 J.-H. Cho, W. Song, H. Choi, and T. Kim: IEEE Signal Process. Lett. **24** (2017) 329. <https://doi.org/10.1109/LSP.2017.2661319>
- 3 W.-N. Lie, C.-Y. Hsieh, and G.-S. Lin: IEEE T. Multimedia. **20** (2018) 1075. <https://doi.org/10.1109/TMM.2017.2763319>
- 4 C.-H. Yang, C.-C. Wu, and P.-S. Chen: Proc. 2017 Int. Workshop on Marketing Research in the Digital Age (MR2017) (NETs, 2017) 2–14
- 5 S. Huang, J. Sun, Y. Yang, Y. Fang, and P. Lin: IEEE Access **5** (2017) 21669. <https://doi.org/10.1109/ACCESS.2017.2757239>
- 6 Q. Wu, Z. Yong, and Z. Zhang: Proc. 2015 6th Int. Conf. Intelligent Systems Design and Engineering Applications (IEEE, 2015) 460–462. <https://doi.org/10.1109/ISDEA.2015.121>
- 7 H. Zhou, X. Li, G. Schaefer, M. Celebi, and P. Miller: Comput. Vis. Image Understanding **117** (2013) 1004. <https://doi.org/10.1016/j.cviu.2012.11.015>
- 8 H. G. Kim and Y. M. Ro: IEEE Trans. Circuits Syst. Video Technol. **27** (2017) 1435. <https://doi.org/10.1109/TCSVT.2016.2515360>
- 9 G. Luo and Y. Zhu: IEEE Trans. Circuits Syst. Video Technol. **27** (2017) 2118. <https://doi.org/10.1109/TCSVT.2016.2583978>
- 10 J. Seiler, S. Scholl, W. Schnurrer, and A. Kaup: Proc. 2016 Picture Coding Symp. (IEEE, 2016) 1–5. <https://doi.org/10.1109/PCS.2016.7906329>