

Shape-Preserving Object Depth Control for Stereoscopic Images

Jianjun Lei, *Senior Member, IEEE*, Bo Peng, Changqing Zhang, Xuguang Mei,
Xiaochun Cao, *Senior Member, IEEE*, Xiaoting Fan, and Xuelong Li, *Fellow, IEEE*

Abstract—In the field of 3-D technology, it is interesting as well as meaningful issue to control object depth in 3-D space. Recently, some depth control methods for stereoscopic images have been proposed, which usually employ depth map or directly process color images to implement depth control. There are two main disadvantages for these methods. First, the results of these methods usually suffer from object deformation and holes. Second, these methods are prone to cause undesired object size changing in 3-D space. To address these issues, we propose a shape-preserving object depth control method for stereoscopic images. First, a novel depth mapping model is presented for calculating the ideal coordinates of the key points in depth control, so that the shape of the object can be well preserved. Afterward, the image content-based constraints are used to further preserve the structure of the object and its background. Finally, the warping technology is introduced to deal with images optimally as well as to avoid holes. Experimental results show that the proposed method can control object depth and preserve the shape of the object effectively without sensible background distortion.

Index Terms—Depth control, depth mapping model, shape preserving, stereoscopic image, warping technology, 3-D display.

I. INTRODUCTION

STEREOSCOPIC images and videos are becoming increasingly popular in recent years. Compared to 2D images, stereoscopic images provide depth information which enhances the visual experience greatly [1], [2]. Therefore, stereoscopic technology is widely used in a lot of applications such as 3D images and 3D movies. The popularity of the stereoscopic technology also boosted the development of 3D cameras and 3DTV [3], [11]. The prevalence of

Manuscript received February 8, 2017; revised July 11, 2017; accepted August 23, 2017. Date of publication September 4, 2017; date of current version December 4, 2018. This work was supported in part by the National Science Foundation of China under Grant 61722112, Grant 61602337, Grant 61761130079, and Grant 61520106002, in part by the Key Research Program of Frontier Sciences, CAS under Grant QYZDY-SSW-JSC044, and in part by the National Key Research and Development Program of China under Grant 2017YFB1002900. This paper was recommended by Associate Editor C. Zhang. (*Corresponding author: Changqing Zhang*.)

J. Lei, B. Peng, C. Zhang, X. Mei, and X. Fan are with Tianjin University, Tianjin 300072, China (e-mail: jjlei@tju.edu.cn; bpeng@tju.edu.cn; zhangchangqing@tju.edu.cn; meixg@tju.edu.cn; fanxiaoting1234@126.com).

X. Cao is with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China (e-mail: caoxiaochun@iie.ac.cn).

X. Li is with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China, and also with The University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: xuelong_li@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2017.2749146

stereoscopic technology brings a vigorous development period with a large number of researches focusing on stereoscopic technology. Among those researches, one interesting issue is the depth control for stereoscopic contents [5]–[7].

Since the new released phones and tablet PCs are with stereoscopic display functions, portable stereoscopic display devices become very common, which makes it convenient to watch stereoscopic images. However, when viewers watch 3D contents on portable devices with relatively small screens, the 3D effect is often inconspicuous [8]. Film producers may want to change the depth of the specific object to create more intensive 3D effects. In addition, viewers may feel fatigue if the objects in stereoscopic videos have long periods of large disparities [9]–[12]. For these reasons, it is desirable to well control the depth of a specific object in 3D space for stereoscopic images.

According to the usage of depth map, depth control methods for stereoscopic images can be divided into two categories, i.e., depth control methods using depth map and depth control methods without using depth map. For the first category, they need accurate disparity or depth map to implement depth control and often change depth map according to specific requirements, then use virtual view rendering technology to generate a depth controlled image [13]–[15]. However, high quality disparity map is often hard to achieve. For the second category, these methods implement depth control by processing stereoscopic images directly without using depth map [16], [17]. However, these methods usually cause object deformation or holes.

In this paper, we propose a novel shape-preserving object depth control method for stereoscopic images. The contributions of the paper mainly include the following: 1) We propose an object depth mapping model for 3D images which moves key points of an object in 3D space directly so that the shape of the object can be well preserved. 2) Efficiency constrains are proposed to keep the mesh size changing accordingly so that the structure of the selected object can be well preserved and avoid object deformation. 3) We combine our depth mapping model with warping technology to avoid holes when implementing depth control. 4) A novel buffer zone combined with edge detection is proposed as the solution to constrain the changes within the object region and preserve the contents of background.

The remainder of this paper is organized as follow. We review important related works in Section II, and present the proposed object depth control method in Section III.

After evaluating and making a comparison to related methods from literature in Section IV, this paper is concluded in Section V.

II. RELATED WORK

Depth map based depth control methods often change depth map according to specific requirements, and use virtual view rendering technology to generate a result image based on the changed depth map. For example, Lee *et al.* [18] proposed to analyze parallax histogram using GMM model first and separate parallax map into layers, then handle the position and size of each layer on the histogram level separately to achieve depth control. In [19], a depth sensation enhancement method was proposed. An energy function is designed according to the number of rendering views and the saliency analysis is used for optimally processing depth map. Then, depth control is achieved by using depth image based rendering. Kellnhofer *et al.* [20] expressed the move of object in 3D space by utilizing the depth map in the time field. By smoothing the depth change, the object is kept moving continuously in 3D space. Depth map based methods need an accurate depth map to achieve good results. However, depth estimation is still a challenging task, and it is difficult to achieve promising accuracy especially for the edge of objects.

To alleviate the difficulty of obtaining accurate depth map, researchers proposed to directly process images without using depth map [21], [22]. Conventional parallax shifting methods [23], [24] are widely used in commercial 3D display, such as 3DTV and 3D cameras. The methods shift the selected object with equal size horizontally to change parallax and then the depth. However, they usually change the actual size of the selected object in 3D space undesirably. Fig. 1 (a) shows how the viewer sees left and right images on the display screen to perceive the ball in 3D space. By shifting left image to the right and right image to the left, conventional parallax shifting methods can control the depth of the ball. However, it is shown in Fig. 1 (b) that the actual size of the ball, i.e., the scale of the ball in 3D space, will change at the same time. This undesired side effect may degrade the result of depth control. The scale contrast between the selected ball and other objects has significant effect on 3D result. Therefore, in order to change the depth of the selected object without changing its actual size in 3D space, the projections of the selected object on the images must be scaled accordingly as shown in Fig. 1 (c). Yan *et al.* [17] proposed a linear depth mapping method to adjust the depth range of a stereoscopic video according to the viewing configuration. They listed a series of energy terms which designed for depth control model and image content preservation. By minimizing the energy function, the optimal coordinates of mesh vertexes can be calculated. Then, the warping technology [25]–[27] is used to process images. Park *et al.* [28] proposed a method to control depth by utilizing a virtual fronto-parallel planar projection in 3D space. They first segment the object from the left and right images, and match object correspondence between them. Then, pixels on the selected object in the left and right images are projected to a virtual fronto-parallel plane. By moving the plane head to the viewer in 3D space and projecting the points back to the left

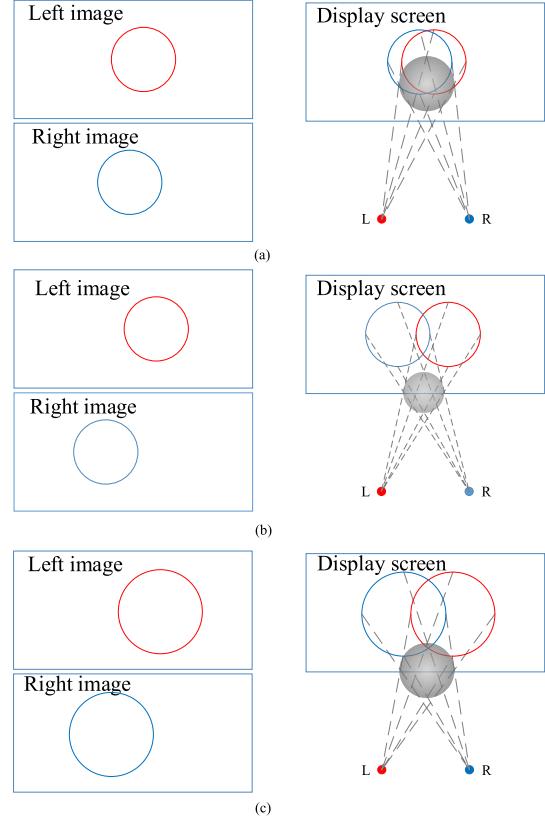


Fig. 1. Illustration of conventional parallax shifting method. (a) Original left and right images. (b) Result images after conventional parallax shifting adjustment, the actual size of the ball in 3D space becomes relatively small. (c) Result images after ideal depth control, the actual size of the ball in 3D space stays the same.

and right images, the depth controlled stereo 3D images can be obtained. However, this method does not consider moving object away from the viewer. It causes holes when the object is moved away.

This paper proposes a shape-preserving depth control method for stereoscopic image. A depth mapping model is proposed for 3D images, which moves key points of an object in 3D space directly so that the shape of the object can be well preserved. In addition, inspired by the widely applied warping energy function, our depth mapping model is implemented with warping energy function to avoid holes. Finally, we combine a novel buffer zone with edge detection as the solution to constrain the changes within the object region and preserve the contents of background. Experimental results show that the proposed method can effectively control the depth of selected object when moving close and away without sensible distortions.

III. PROPOSED DEPTH CONTROL METHOD

A. Framework Overview

In this section, we give an overview of the proposed method. There are mainly four components in our method as shown in Fig. 2. In the first component, SIFT key points [29] are extracted from left and right images separately and the correspondences between the two sets of extracted key points are

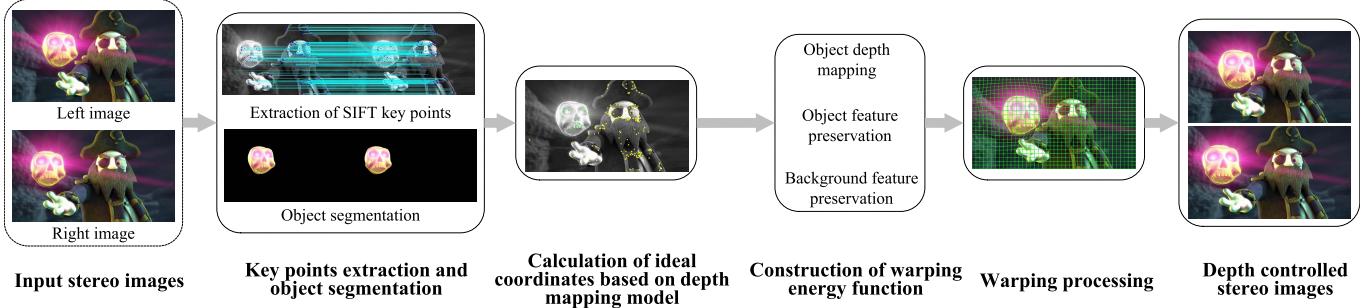


Fig. 2. The framework of the proposed method.

TABLE I
SUMMARY OF NOTATIONS

$P_{1L} = (X_{1L}, Y_{1L}, Z_{1L})^T$	Point in the left image
$P_{1R} = (X_{1R}, Y_{1R}, Z_{1R})^T$	Point corresponding to P_{1L} in the right image
$P_1 = (X_1, Y_1, Z_1)^T$	Point in 3D space perceived by viewer from P_{1L} and P_{1R}
$P'_1 = (X'_1, Y'_1, Z'_1)^T$	Moved point of P_1
$P'_{1L} = (X'_{1L}, Y'_{1L}, Z'_{1L})^T$	Depth-adjusted point in the left image
$P'_{1R} = (X'_{1R}, Y'_{1R}, Z'_{1R})^T$	Depth-adjusted point that corresponds to P'_{1L} in the right image
$P_c = (X_c, Y_c, Z_c)^T$	Proximate center of the selected object
B	Distance between two eyes
t	Vertical distance from viewer to the display screen
D	Parallax between a pair of corresponding pixels
d_z	Difference between the original and the target depth values
d_x	X-direction offset value
d_y	Y-direction offset value
c_l	Set of key points coordinates in the left image
c'_r	Set of key points coordinates in the right image
c'_v	Set of mesh vertex coordinates
c_{el}	Set of pixel coordinates on the detected edges in the left image
c'_{er}	Set of pixel coordinates on the detected edges in the right image

matched to represent image features. The object whose depth needs to be adjusted should be selected in the stereoscopic images. By using existing object segmentation methods [30]–[32], the selected object can be segmented from the given stereoscopic images. Second, all the key points on the selected object are handled by our depth mapping model to obtain the ideal coordinates of key points. Then, warping energy function is constructed based on the constraints of object depth mapping, object feature preservation, and background feature preservation. Finally, the energy function is used in the warping procedure to process the images and get the final result. To facilitate understanding, the summary of notations is presented in Table I.

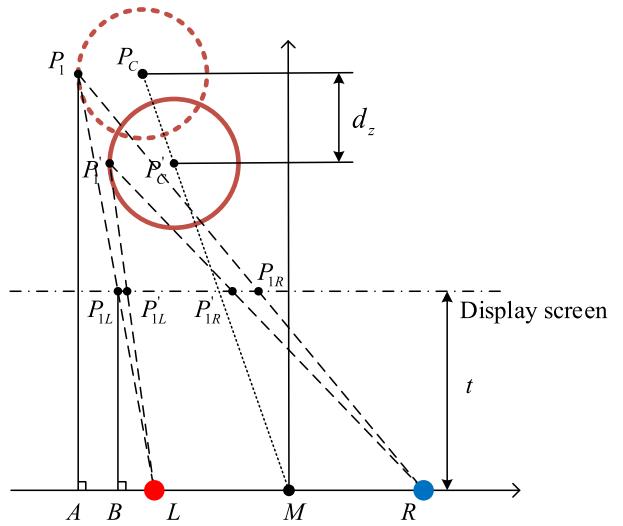


Fig. 3. Illustration of our depth mapping model.

B. Our Depth Mapping Model

In our depth mapping model, all the key points on the selected object are moved with equal size in 3D space according to specific configuration. Since we move all the key points on the selected object in 3D space directly, the actual size of the selected object can be preserved. Fig. 3 illustrates the proposed method of control object depth in 3D space. The dotted circle represents the original selected object in 3D space. The origin M of 3D space is set as the midpoint between two eyes of the viewer. We assume that $P_1(X_1, Y_1, Z_1)$ is one of the key points which represent features of the selected object. Since all the matched key point pairs on the selected object are supposed to move with equal size, we use P_1 as an example to show the computational process of coordinates. The depth Z of one point is defined as the vertical distance between the viewer and the point, and it can be calculated by [33]:

$$Z = \frac{Bt}{B - D}, \quad (1)$$

where B is the interaxial distance between two eyes of the viewer, t is the vertical distance between the viewer and the screen, and D is the parallax between the corresponding pixels

which can be calculated as

$$D = X_{1R} - X_{1L}. \quad (2)$$

The point $P_1(X_1, Y_1, Z_1)$ in 3D space can be calculated by key point pair $P_{1R}(X_{1R}, Y_{1R}, Z_{1R})$ and $P_{1L}(X_{1L}, Y_{1L}, Z_{1L})$ using (1) and two similar triangles as follows:

$$\begin{aligned} X_1 &= \frac{B}{B-D} \left(X_{1L} + \frac{B}{2} \right) - \frac{B}{2}, \\ Y_1 &= \frac{B}{B-D} Y_{1L}, \\ Z_1 &= \frac{B}{B-D} t. \end{aligned} \quad (3)$$

After calculating the position of every key point in 3D space, the proximate center $P_c(X_c, Y_c, Z_c)$ of the selected object can be obtained by the mean value of all the key points, and can be formulated as follows:

$$K_c = \frac{1}{n} \sum_{i=1}^n K_i, \quad K \in \{X, Y, Z\} \quad (4)$$

where n is the number of key points on the selected object.

The proposed method controls the position of the selected object in 3D space by moving all the key points with equal size on the basis of its center. In order to avoid the selected object shifting out of the image, it is moved along the line from the object to the midpoint of two eyes of the viewer, which means P_1 is moved to P'_1 and its new coordinate can be calculated as follows:

$$\begin{aligned} X'_1 &= \frac{B}{B-D} \left(X_{1L} + \frac{B}{2} \right) - \frac{B}{2} - d_x, \\ Y'_1 &= \frac{B}{B-D} Y_{1L} - d_y, \\ Z'_1 &= \frac{B}{B-D} t - d_z, \end{aligned} \quad (5)$$

where d_z is the difference between the original and the target depth values, and d_x , d_y are offset values which can be calculated as:

$$\begin{aligned} d_x &= d_z \frac{X_c}{Z_c}, \\ d_y &= d_z \frac{Y_c}{Z_c}. \end{aligned} \quad (6)$$

Similarly, all the key points on the selected object can be calculated by using the same offset value, which keeps the scale of the selected object in 3D space unchanged.

Once the coordinate of the depth controlled point P'_1 is calculated, the ideal coordinates of P'_{1L} and P'_{1R} on the image can be obtained. To calculate ideal coordinates of the depth controlled points on the image, the relationships among P'_1 , P'_{1L} and P'_{1R} are needed, which can be calculated using two similar triangles. The relationship between P'_1 and P'_{1L} can be calculated as follows:

$$\begin{aligned} \frac{t}{Z'_1} &= \frac{X'_{1L} + \frac{B}{2}}{X'_1 + \frac{B}{2}}, \\ \frac{t}{Z'_1} &= \frac{Y'_{1L}}{Y'_1}. \end{aligned} \quad (7)$$

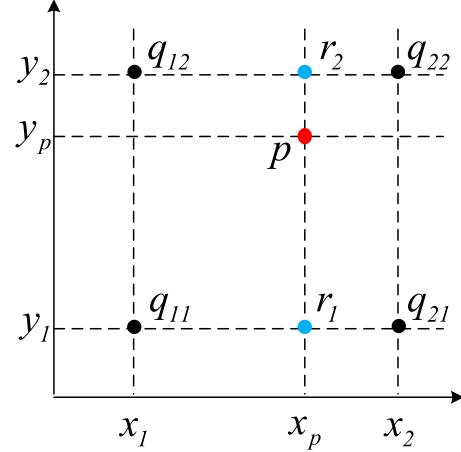


Fig. 4. Illustration of representing key point by using mesh vertexes.

Similarly, the relationship between P'_1 and P'_{1R} is:

$$\begin{aligned} \frac{t}{Z'_1} &= \frac{\frac{B}{2} - X'_{1R}}{\frac{B}{2} - X'_1}, \\ \frac{t}{Z'_1} &= \frac{Y'_{1R}}{Y'_1}. \end{aligned} \quad (8)$$

Finally, the ideal coordinates of key point pairs in the images can be calculated using (5), (7) and (8). For the left image, $P'_{1L}(X'_{1L}, Y'_{1L})$ can be calculated as follows:

$$\begin{aligned} X'_{1L} &= \frac{\frac{B}{B-D} \left(X_{1L} + \frac{B}{2} \right) - d_x - \frac{B}{2}}{\frac{B}{B-D} - \frac{d_z}{t}}, \\ Y'_{1L} &= \frac{\frac{B}{B-D} Y_{1L} - d_y}{\frac{B}{B-D} - \frac{d_z}{t}}. \end{aligned} \quad (9)$$

For the right image, $P'_{1R}(X'_{1R}, Y'_{1R})$ can be calculated similarly:

$$\begin{aligned} X'_{1R} &= \frac{\frac{B}{2} - \frac{B - \frac{B}{B-D} \left(X_{1L} + \frac{B}{2} \right) + d_x}{\frac{B}{B-D} - \frac{d_z}{t}}}{\frac{B}{B-D} - \frac{d_z}{t}}, \\ Y'_{1R} &= \frac{\frac{B}{B-D} Y_{1R} - d_y}{\frac{B}{B-D} - \frac{d_z}{t}}. \end{aligned} \quad (10)$$

According to (9) and (10), the depth of the selected object can be controlled by changing a single parameter d_z intuitively, which is the difference between the original depth and the target depth. The ideal coordinates of controlled key points on the left and right images can be easily obtained by using (9) and (10).

C. Construction of Warping Energy Function

Since the proposed method utilizes SIFT key points to represent image feature, sparse region may not have enough key points. To solve this problem, grid meshes are constructed on the left and right images after calculating the ideal coordinates of depth controlled key point pairs. The method uses mesh vertices to represent key points as shown in Fig. 4.

$q_{11}, q_{12}, q_{21}, q_{22}$ are mesh vertexes around the key point p . The line from r_1 to r_2 passes key point p and parallel to the line $q_{11}q_{12}$ and the line $q_{21}q_{22}$. Coordinate c_{r_2} can be represented by coordinate $c_{q_{12}}$ and $c_{q_{22}}$, and coordinate c_{r_1} represented by $c_{q_{11}}$ and $c_{q_{21}}$ similarly as follows:

$$\begin{aligned} c_{r_1} &\approx \frac{x_2 - x_p}{x_2 - x_1} c_{q_{11}} + \frac{x_p - x_1}{x_2 - x_1} c_{q_{21}}, \\ c_{r_2} &\approx \frac{x_2 - x_p}{x_2 - x_1} c_{q_{12}} + \frac{x_p - x_1}{x_2 - x_1} c_{q_{22}}. \end{aligned} \quad (11)$$

Then, the key point coordinate c_p can be represented by c_{r_1} and c_{r_2} using (11) as:

$$c_p \approx \frac{y_2 - y_p}{y_2 - y_1} c_{r_1} + \frac{y_p - y_1}{y_2 - y_1} c_{r_2}. \quad (12)$$

1) *Object Depth Mapping*: All the key points on the selected object are used to calculate its central position. For each key point, the ideal coordinate can be calculated through our depth mapping model with factors d_z . Assume $c'_{i,l}$ and $c'_{i,r}$ are the ideal coordinates, $\hat{c}_{i,l}$ and $\hat{c}_{i,r}$ are the adjusted coordinates. The following energy terms are used for the warping process to make the adjusted coordinates optimally approach to the ideal coordinates.

$$\begin{aligned} E_{zl}(\hat{c}_l) &= \sum_{i \in n} \|\hat{c}_{i,l} - c'_{i,l}\|^2, \\ E_{zr}(\hat{c}_r) &= \sum_{i \in n} \|\hat{c}_{i,r} - c'_{i,r}\|^2, \end{aligned} \quad (13)$$

where n stands for all key point pairs on the selected object. The two energy terms in (13) are combined to achieve the total depth control energy term as follows:

$$E_z(\hat{c}_l, \hat{c}_r) = E_{zl}(\hat{c}_l) + E_{zr}(\hat{c}_r). \quad (14)$$

2) *Object Shape and Feature Preservation*: Since our method uses SIFT key points [29] to represent image features and to control the selected object, the shape of the selected object may deform while implementing depth control. This problem is solved by constraining the size of meshes that cover the selected object.

According to our depth mapping model, the size of the selected object in the left and right images will change along with depth. We attempt to make the border of the meshes changing synchronously to preserve the structure of the selected object. The object size changing factor k_i can be set as:

$$k_{jk} = \frac{X'_j - X'_k}{X_j - X_k}. \quad (15)$$

where j and k represent two key points on the target object with the same depth.

Bring (9) and (10) into (15), the relation ship between the size and the depth of the selected object can be calculated as

$$k_{jk} = \frac{Z_{jk}}{Z'_{jk}}, \quad (16)$$

where

$$Z'_{jk} = Z_{jk} - d_z. \quad (17)$$

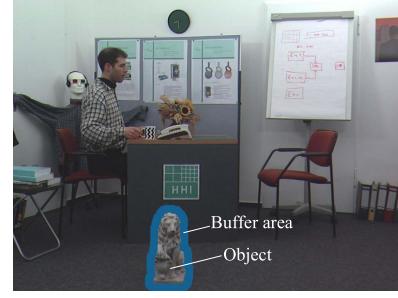


Fig. 5. The buffer area between object and background.

Z and Z' represent the depths of target object before and after depth control. It should be noted that different parts on the same object may have different depths. But since the depth range of an object is relatively small, our method uses mean value to set total k as:

$$k = \frac{\frac{1}{n} \sum_{i \in \beta} Z_i}{\left(\frac{1}{n} \sum_{i \in \beta} Z_i\right) - d_z}, \quad (18)$$

where β means the set of key points on the selected object, and the number of the set β is n .

Then the mesh constraint energy term can be set by using (18) as follows:

$$E_{mesh}(\hat{c}_o) = \sum_{o} \sum_{i,j \in o} \|\hat{c}_i - \hat{c}_j - k(c_i - c_j)\|^2, \quad (19)$$

where o means all the meshes on the selected object and v means every vertex of the meshes. c_i, c_j are the original vertex coordinates of the meshes and \hat{c}_i, \hat{c}_j are the adjusted vertex coordinates, \hat{c}_o is the set of all the adjusted vertex coordinates.

3) *Background Feature Preservation*: By using the depth mapping energy term, the depth of the selected object can be controlled. But the background and other objects may distort while changing the depth of the selected object, and the shape of the selected object may also change a little. So there should be some energy terms to keep the background and other objects approximately unchanged.

For the key points not on the selected object, our method applies the following energy terms to make them remain the same as far as possible:

$$E_{bl}(\hat{c}_l) = \sum_{j \in \gamma} \|\hat{c}_{j,l} - c_{j,l}\|^2$$

and

$$E_{br}(\hat{c}_r) = \sum_{j \in \gamma} \|\hat{c}_{j,r} - c_{j,r}\|^2, \quad (20)$$

where γ means all the key points on the background. Ideal coordinates $c_{j,l}$ and $c_{j,r}$ in this case are their original coordinates. The total background content preservation energy term is obtained by the two energy terms in (20) as follows:

$$E_b(\hat{c}_l, \hat{c}_r) = E_{bl}(\hat{c}_l) + E_{br}(\hat{c}_r). \quad (21)$$

The energy terms (14) and (21) are set different weight values so that the background can keep unchanged when

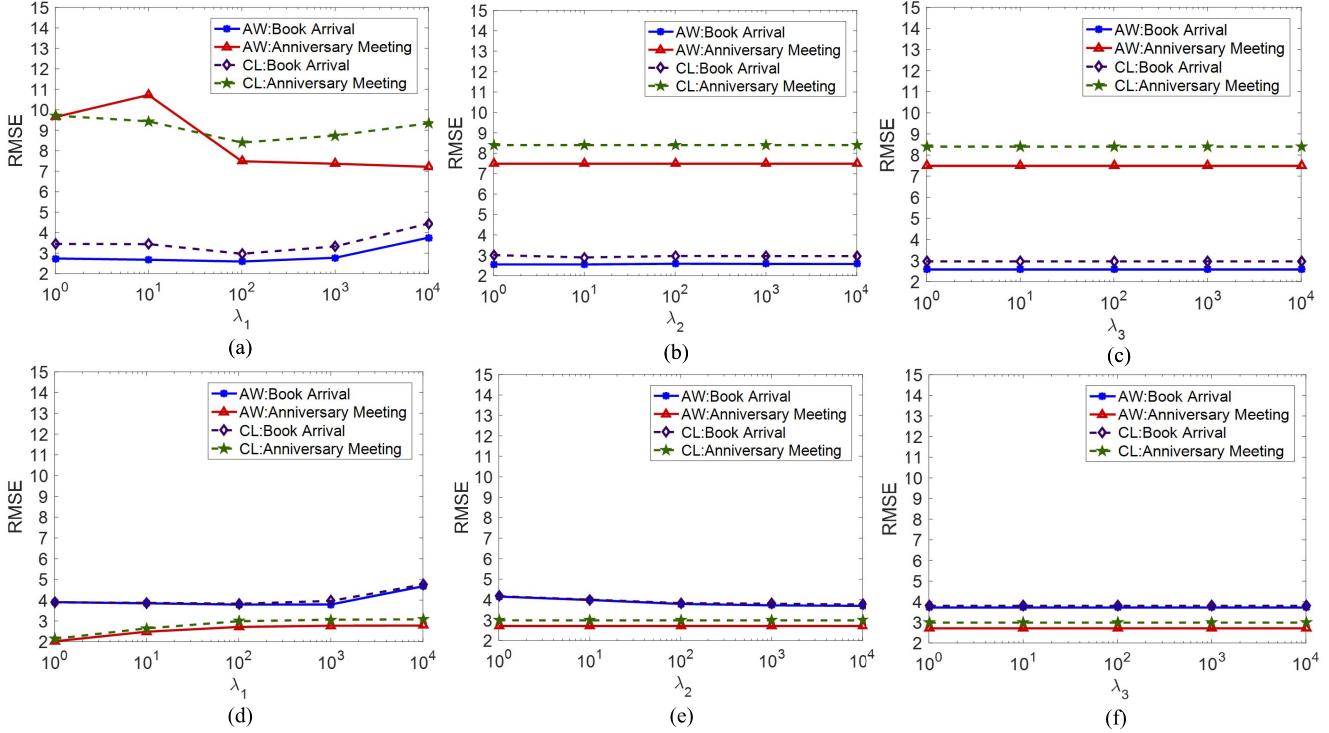


Fig. 6. RMSE results that pushing the selected object away from (AW) and close to (CL) the viewer. Top: RMSE of selected object; Bottom: RMSE of background. (a-c) RMSE results of selected object with different λ_1 , λ_2 and λ_3 ; (d-f) RMSE results of background with different λ_1 , λ_2 and λ_3

controlling the depth of the selected object. While using (14) and (21), some background key points close to the selected object may cause shape distortion. That is because the size changing of the selected object may make it cover some background key points, which cause conflict between (14) and (21). To solve this problem, the region within α pixels to the selected object is set as the buffer area, and the background key points on the buffer area are removed. Fig. 5 shows the buffer area can solve the conflict between (14) and (21), and keep most of the distortions within this area.

Other objects in the background may have some parts close and some parts far from the selected object, the closer these parts are to the selected object the more distorted these objects tend to be. So our method also detects background edges to preserve background contents. The Canny operator is employed to detect all the edges in the image, and then the edges on the selected object and buffer area are removed to get the background edges.

For each pixel on the background edges, the following energy term is defined to constraint them not move too far from their original position:

$$\begin{aligned} E_{el}(\hat{\mathbf{c}}_{el}) &= \sum_{i \in edge} \|\hat{c}_{i,l} - c_{i,l}\|^2, \\ E_{er}(\hat{\mathbf{c}}_{er}) &= \sum_{i \in edge} \|\hat{c}_{i,r} - c_{i,r}\|^2, \end{aligned} \quad (22)$$

where *edge* means all the pixels on the detected background edges, $\hat{\mathbf{c}}_{el}$ and $\hat{\mathbf{c}}_{er}$ are the set of all the pixels on the detected

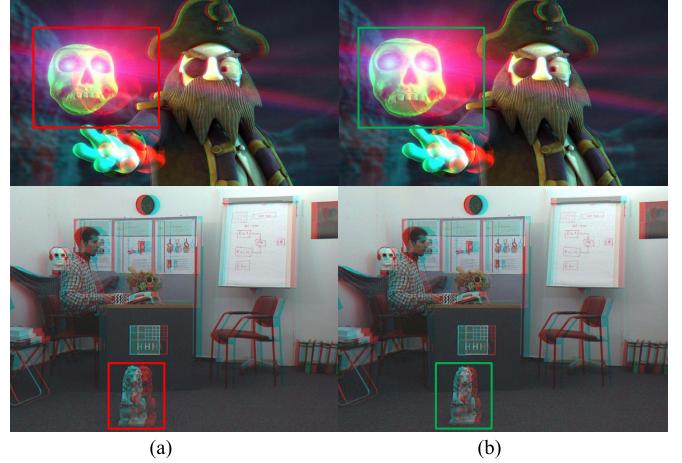


Fig. 7. Result of our method. Moving the selected object close and away are shown in the top row and the bottom row respectively. (a) Original stereoscopic images; (b) The result images from our method.

background edges. By combining the two energy terms in (22), the total edge constraint can be set as:

$$E_e(\hat{\mathbf{c}}_{el}, \hat{\mathbf{c}}_{er}) = E_{el}(\hat{\mathbf{c}}_{el}) + E_{er}(\hat{\mathbf{c}}_{er}). \quad (23)$$

4) Energy Function: Since $\hat{\mathbf{c}}_l$, $\hat{\mathbf{c}}_r$, $\hat{\mathbf{c}}_{el}$, $\hat{\mathbf{c}}_{er}$ can be represented by $\hat{\mathbf{c}}_v$ using (12), the total energy can be calculated as follows:

$$\begin{aligned} E(\hat{\mathbf{c}}_v) &= E_Z(\hat{\mathbf{c}}_l, \hat{\mathbf{c}}_r) + \lambda_1 E_{mesh}(\hat{\mathbf{c}}_v) \\ &\quad + \lambda_2 E_b(\hat{\mathbf{c}}_l, \hat{\mathbf{c}}_r) + \lambda_3 E_e(\hat{\mathbf{c}}_{el}, \hat{\mathbf{c}}_{er}). \end{aligned} \quad (24)$$

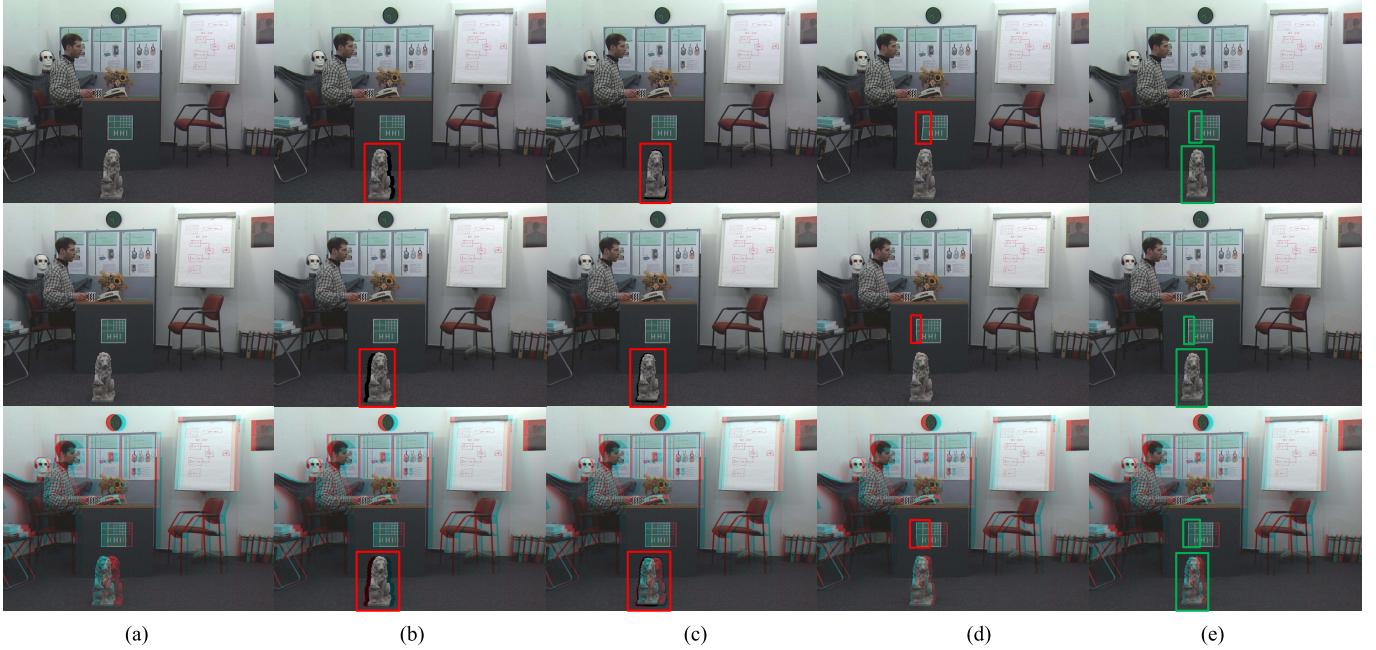


Fig. 8. Result Book Arrival images that pushing the selected object away from the viewer. From top to bottom: the left image, right image, and red/cyan color anaglyph image. (a) Original image; (b) Parallax shifting method [24]; (c) Park's method [28]; (d) Lang's method [25]; (e) Our method.

Parameter λ_1 , λ_2 and λ_3 are set as weight values. Then, the optimized $\hat{\mathbf{c}}_v$ can be calculated by:

$$\hat{\mathbf{c}}_v = \arg \min_{\hat{\mathbf{c}}_v} E(\hat{\mathbf{c}}_v). \quad (25)$$

The energy term (25) is a least square fitting problem, it can be easily solved by:

$$\frac{\partial E(\hat{\mathbf{c}}_v)}{\partial \hat{\mathbf{c}}_v} = 0. \quad (26)$$

By solving (26), the optimized coordinates of mesh vertices are calculated to get the new geometry for the grid mesh. By using warping technology [34], the original left and right images are warped based on the new grid mesh and the depth controlled images can be obtained.

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

The proposed method is evaluated with four images, which including Book Arrival, Curse of Skull Rock, Skydiving, and Anniversary Meeting [17], [35]. First, the depth-controlled stereoscopic images processed by our method, parallax shifting method [24], Park's method [28], and Lang's method [25] are shown for comparison. Then, a subjective user study and complexity analysis are performed to evaluate the performance. We conducted experiments under the simulation environments of Intel(R) Core(TM) i7-6700 CPU @3.40GHz with 8GB memory and 64 bit Windows 7 operating system. In our method, the selected object is moved by $d_z = 4\text{cm}$ and $d_z = -4\text{cm}$, the buffer size parameter α is set as 30. We set the horizontal disparity in the parallax sifting method [24] as 4% of the image width, set ρ as 0.09 and -0.09 in Park's method [28], and η as 1° and -1° in Lang's method [25] to

move the selected object with equal size and obtain similar depth control.

There are three parameters, λ_1 , λ_2 and λ_3 in Eq. (24), which correspond to object feature preservation term, background content preservation term, and edge constraint term, respectively. Initially, the parameters λ_1 , λ_2 and λ_3 are all set as 100. Then, we tune one parameter by fixing the others. The parameters are tuned from 1 to 10000 with an interval of 10 times. The relationships between RMSE (Root-Mean-Square Error) and the values of parameters are shown in Fig. 6. In this paper, the RMSE of selected object is computed based on the SIFT key points, since the size of the selected object in left and right images will change along with depth. It can be observed from Fig. 6 (a) and Fig. 6 (d) that, when λ_1 is set as 100, the minimum RMSE of the selected object and background can be obtained. We further tune the second parameter λ_2 . As shown in Fig. 6 (e), when λ_2 becomes larger, the RMSE decreases gradually and reaches the minimum value at 1000 for Book Arrival. Our method gives a relatively good performance when the second parameter λ_2 set as 1000. As shown in Fig. 6 (c) and Fig. 6 (f), RMSE is robust with respect to the third parameter λ_3 , therefore, we set it to the initial value 100. Based on the above analyses, λ_1 , λ_2 and λ_3 are set as 100, 1000 and 100 by considering the trade-off between the selected object and background.

B. Comparison of Resulting Images

First, the result images from our method are shown in Fig. 7 to show its performance generally. The disparities of the selected object become larger when moving close and smaller when moving away, respectively. The size of the selected object on the image changes accordingly. The background

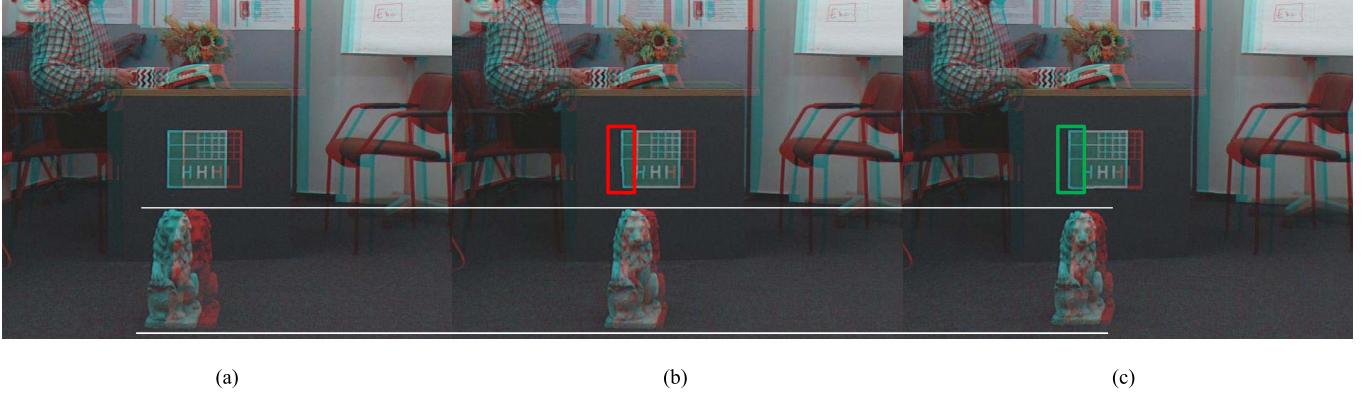


Fig. 9. Enlarged red/cyan color anaglyph images. The statue in (b) and (c) is moved close to the viewer. (a) Original image; (b) Lang's method [25]; (c) Our method.

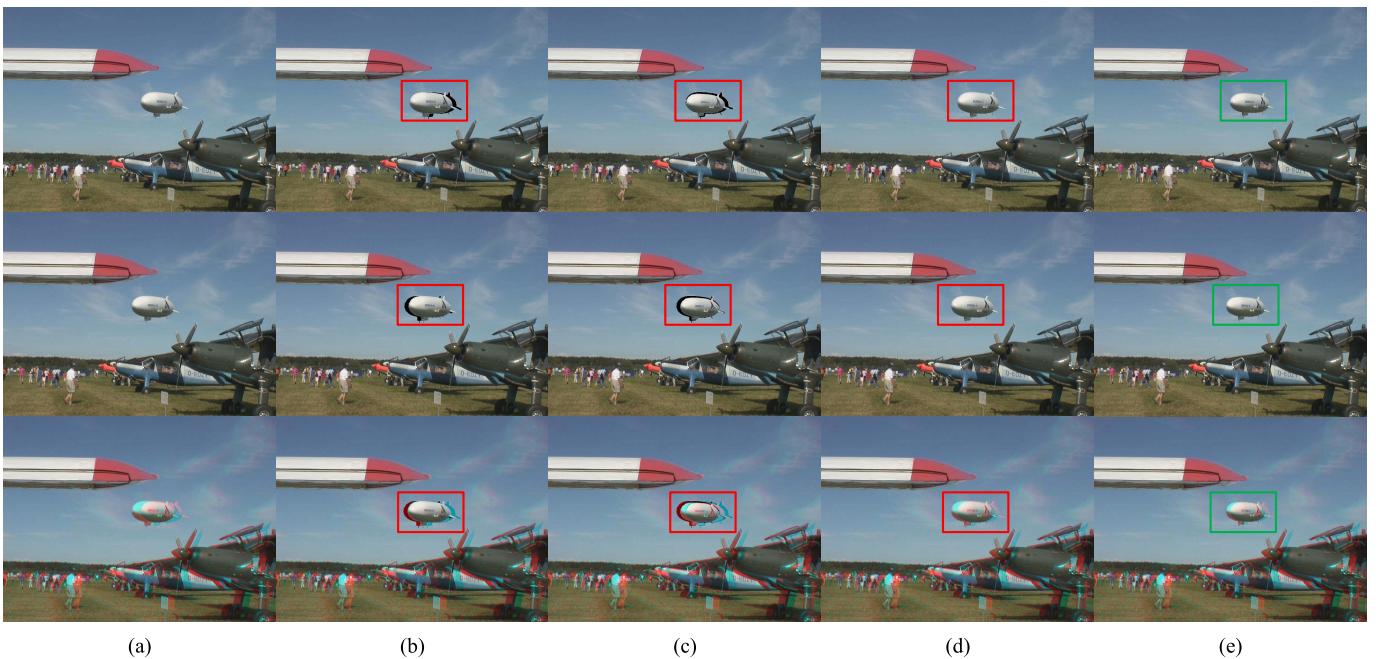


Fig. 10. Result Anniversary Meeting images that pushing the selected object away from the viewer. From top to bottom: the left image, right image, and red/cyan color anaglyph image. (a) Original image; (b) Parallax shifting method [24]; (c) Park's method [28]; (d) Lang's method [25]; (e) Our method.

contents do not have visible distortion, and the shape of the selected object is well preserved.

Next, we compare our method with the parallax shifting method [24], Park's method [28] and Lang's method [25] under different situations including the selected objects close to and away from the viewer. The results of moving the selected object away are shown in Fig. 8 and Fig. 10, the partial enlarged detail of Fig. 8 is shown in Fig. 9. The results of moving close are shown in Fig. 11 and Fig. 12. Mainly two aspects are evaluated: image quality and visual perception.

1) Image Quality: It can be observed from Fig. 8, Fig. 10, Fig. 11 and Fig. 12 that the parallax shifting method [24] causes holes when controlling the depth of the object. Park's method [28] always causes holes when moving the selected object away as shown in Fig. 8 and Fig. 10. When moving the selected object close, Fig. 11 shows that Park's method [28] also causes holes in some specific condition. Lang's method [25] causes deformation when moving the

selected object away, the edge of graph and the table in Fig. 9 is curved. On the contrast, the proposed method does not have prominent distortion or holes.

Following is the analysis of the detailed reasons behind such performance. The parallax shifting method [24] causes holes since this method merely shifts the selected object horizontally with equal size. Park's method [28] causes holes since it does not consider the situation of moving the selected object away. Meanwhile, according to the shape of the selected object and the viewing configuration, the selected object may not be able to cover the holes even when moving close. This happens especially when the screen is close to the viewers. For Lang's method [25], since it is proposed to control the depth range of the image wholly, its shape preservation constraint does not suit for object depth control. The proposed method uses warping technology and has shape constraints designed for object depth control, thus can avoid holes and deformation.



Fig. 11. Result Skydiving images that moving the selected object close to the viewer. From top to bottom: the left image, right image, and red/cyan color anaglyph image. (a) Original image; (b) Parallax shifting method [24]; (c) Park's method [28]; (d) Lang's method [25]; (e) Our method.

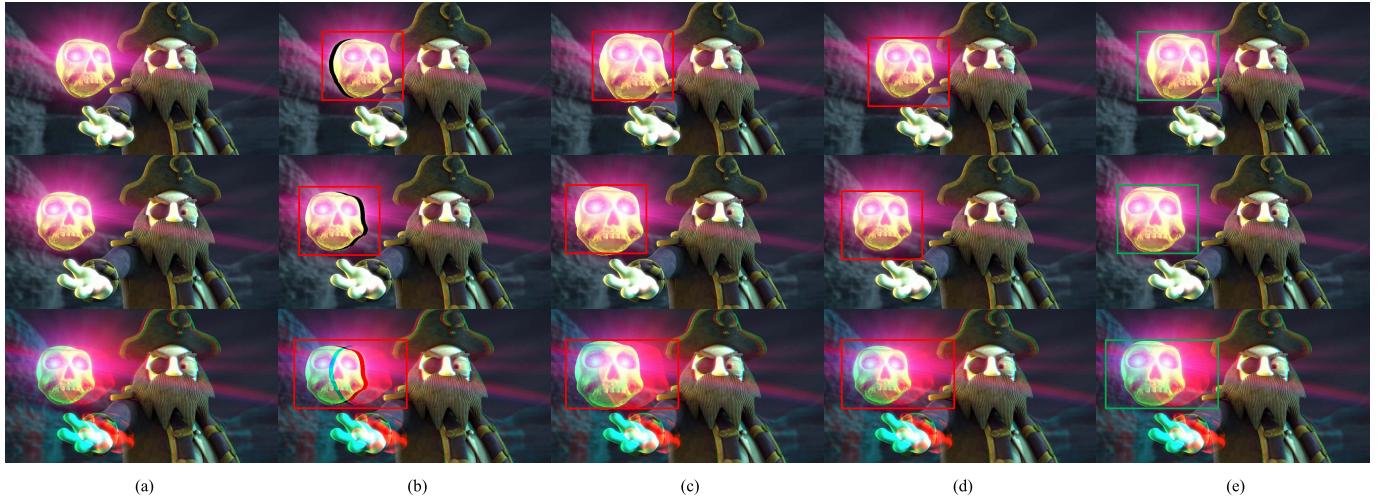


Fig. 12. Result Curse of Skull Rock images that moving the selected object close to the viewer. From top to bottom: the left image, right image, and red/cyan color anaglyph image. (a) Original image; (b) Parallax shifting method [24]; (c) Park's method [28]; (d) Lang's method [25]; (e) Our method.

2) Visual Perception: It can be observed from Fig. 8, Fig. 10, Fig. 11 and Fig. 12 that the parallax shifting method [24] does not change the size of the selected object on the image accordingly when implementing depth control. Lang's method [25] uses the depth mapping model similar to the parallax shifting method [24], the size of the selected object on the image does not change accordingly as shown in Fig. 9 and Fig. 12. In our result, the size of the selected object on the image changes correspondingly to the depth control. The selected object becomes relatively smaller when moving away and larger when moving close as shown in Fig. 9 and Fig. 12.

The scale contrast between the selected object and other objects has significant effect on depth perception. Viewers are usually sensitive to the size of the object, so changing the size in left and right images while controlling depth not only preserves the actual size of the selected object in 3D space, but also makes the depth change more noticeable.

C. User Study

To evaluate the effectiveness of our method subjectively, we used DSCQS (Double Stimulus Continuous Quality Scale) to compare our method with the parallax shifting method [24], Park's method [28], and Lang's method [25]. We conducted the subjective evaluation with 20 people. The participants included 10 females and 10 males with an average age of 25. We employed four images for user study as shown in Fig. 13. Images were displayed on a 17-inch screen and viewers were about 0.5m from the screen. Experimental results were divided into two categories. In the first category, the selected object was moved close, while in the other the selected object was moved away. The images were displayed in pairs. All the results were evaluated by two aspects, i.e., depth perception and visual comfort [36], [37]. Depth perception means to evaluate the effectiveness of depth control. The visual comfort is set to evaluate the preservation of image contents, the result images should not have distortion or holes. The paired t-test



Fig. 13. Four stereo 3D images used in subjective evaluation.

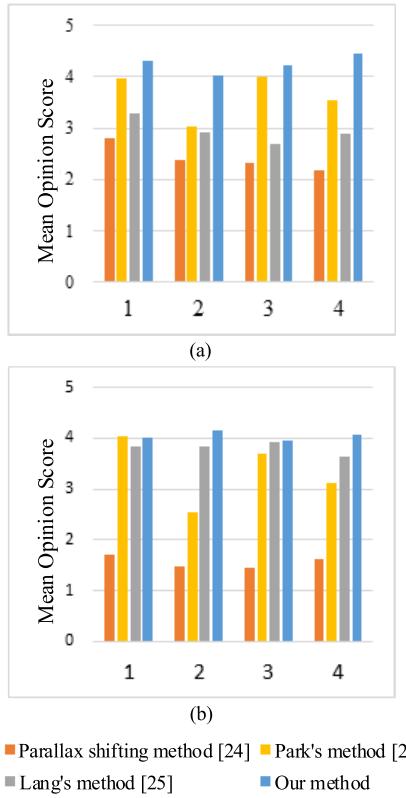


Fig. 14. Subjective evaluation on the result images when moving object close. (a) Scores of the depth perception; (b) Scores of the visual comfort.

was performed to evaluate the statistical significance of the subjective evaluation results. When a universal threshold of 0.05 is used, the comparison result is statistically significant.

Experimental results in the condition of moving the selected object close to the viewer are shown in Fig. 14. Fig. 14 (a) shows that our method significantly outperforms parallax shifting method [24] and Lang's method [25], and gets better scores than Park's method [28] in depth perception performance. As for the visual comfort, our method is significantly better than parallax shifting method [24], and outperforms Park's method [28] and Lang's method [25] as shown in Fig. 14 (b). Parallax shifting method [24] and Lang's method [25] do not change the size of the selected object on the image along with the depth which affect their depth perception performance. Park's method [28] causes holes in the Skydiving image even when moving the selected object close, while our method is better in such condition.

As for moving the selected object away from the viewer, the experimental results are shown in Fig. 15. It can be observed from the figure that the scores of our method are higher than other three methods both in depth perception and visual comfort. Parallax shifting method [24] and Lang's method [25]

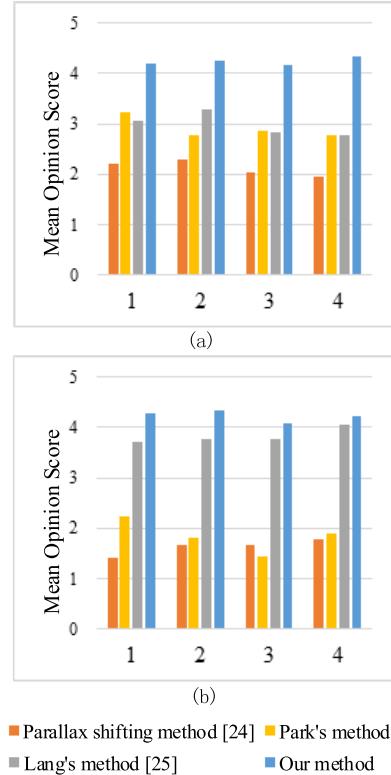


Fig. 15. Subjective evaluation on the result images when moving object away. (a) Scores of the depth perception; (b) Scores of the visual comfort.

have the same problem as moving the selected object close, the size of the selected object on the image does not change along with the depth. Since Park's method [28] does not consider moving the selected object away from the viewer, it causes holes that significantly affect their performance.

D. Computing Complexity

In order to investigate the computation complexity of the proposed method, we conducted experiments on a PC with Intel(R) Core(TM) i7-6700 CPU @3.40GHz and 8GB memory. The proposed method and Lang's method [25] are implemented in Matlab, and the parallax shifting method [24] and Park's method [28] are implemented in C++. For an image with resolution of 1280×720, it takes 11.78s to generate the adjusted result by the proposed method. As a comparison, the parallax shifting method [24] takes 0.22s, Park's method [28] takes 0.26s, and Lang's method [25] takes 12.44s to adjust the image. The parallax shifting method [24] and Park's method [28] do not need the energy based optimization process, and thus have low computing complexities. However, the proposed method and Lang's method [25] have similar computational complexities and take more times to adjust the image, since the adoption of optimization process based on warping energy function.

V. CONCLUSION

This paper presented an object depth control method for stereoscopic images, which can effectively change the selected

object to the target depth position. Our object depth control model for stereoscopic images can change the depth of the selected object as well as preserve its actual size in 3D space. Combining with the warping technology, the method can also avoid holes. Meanwhile, our proposed method well preserves the visual property of images via properly handling the background and edges. Experimental results show that our method can well control the depth of the selected object when moving close and away without causing distortions or holes. In the future, we will extend the proposed method to process stereoscopic videos.

REFERENCES

- [1] G. Tech, Y. Chen, K. Müller, 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] C. Zhu, Y. Zhao, L. Yu, and M. Tanimoto, Eds., *3D-TV System With Depth-Image-Based Rendering: Architecture, Techniques and Challenges*. New York, NY, USA: Springer, 2013.
- [3] W. Yan, C. Hou, J. Lei, Y. Fang, Z. Gu, and N. Ling, "Stereoscopic image stitching based on a hybrid warping model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 9, pp. 1934–1946, Sep. 2017, doi: 10.1109/TCSVT.2016.2564838.
- [4] Y. Fang, J. Wang, M. Narwaria, P. Le Callet, and W. Lin, "Saliency detection for stereoscopic images," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2625–2636, Jun. 2014.
- [5] F. Shao, Z. Li, Q. Jiang, G. Jiang, M. Yu, and Z. Peng, "Visual discomfort relaxation for stereoscopic 3D images by adjusting zero-disparity plane for projection," *Displays*, vol. 39, pp. 125–132, Oct. 2015.
- [6] H. Park, H. Lee, and S. Sull, "Adaptive parallax control based on 3D scene change detection," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 1589–1592.
- [7] C. Wu, C.-T. Li, C.-H. Chung, C.-Y. Ko, and L.-G. Chen, "A viewer centric depth adjustment for stereoscopic images," in *Proc. IEEE Int. Conf. Consum. Electron.*, Berlin, Germany, Sep. 2012, pp. 10–14.
- [8] S. Mangiat and J. Gibson, "Disparity remapping for handheld 3D video communications," in *Proc. IEEE Int. Conf. Emerg. Signal Process. Appl.*, Jan. 2012, pp. 147–150.
- [9] Y. J. Jung, H. Sohn, S.-I. Lee, F. Speranza, and Y. M. Ro, "Visual importance- and discomfort region-selective low-pass filtering for reducing visual discomfort in stereoscopic displays," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 8, pp. 1408–1421, Aug. 2013.
- [10] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940–1953, May 2013.
- [11] Y. Fang, W. Lin, Z. Chen, C.-M. Tsai, and C.-W. Lin, "A video saliency detection model in compressed domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 1, pp. 27–38, Jan. 2014.
- [12] J. Shen, D. Wang, and X. Li, "Depth-aware image seam carving," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1453–1461, Oct. 2013.
- [13] 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.
- [14] J. Jin, A. Wang, Y. Zhao, C. Lin, and B. Zeng, "Region-aware 3-D warping for DIBR," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 953–966, Jun. 2016.
- [15] M. Guttmann, L. Wolf, and D. Cohen-Or, "Semi-automatic stereo extraction from video footage," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2009, pp. 136–142.
- [16] C.-H. Chang, C.-K. Liang, and Y.-Y. Chuang, "Content-aware display adaptation and interactive editing for stereoscopic images," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 589–601, Aug. 2011.
- [17] T. Yan, R. W. Lau, Y. Xu, and L. Huang, "Depth mapping for stereoscopic videos," *Int. J. Comput. Vis.*, vol. 102, nos. 1–3, pp. 293–307, 2013.
- [18] S. Lee, Y. Kim, J. Lee, K. Kim, K. Lee, and J. Noh, "Depth manipulation using disparity histogram analysis for stereoscopic 3D," *Vis. Comput.*, vol. 30, no. 4, pp. 455–465, 2014.
- [19] J. Lei, C. Zhang, Y. Fang, Z. Gu, N. Ling, and C. Hou, "Depth sensation enhancement for multiple virtual view rendering," *IEEE Trans. Multimedia*, vol. 17, no. 4, pp. 457–469, Apr. 2015.
- [20] P. Kellnhofer, T. Ritschel, K. Myszkowski, and H.-P. Seidel, "Optimizing disparity for motion in depth," *Comput. Graph. Forum*, vol. 32, no. 4, pp. 143–152, 2013.
- [21] C. Oh, B. Ham, S. Choi, and K. Sohn, "Visual fatigue relaxation for stereoscopic video via nonlinear disparity remapping," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 142–153, Jun. 2015.
- [22] Y. Jung, H. Sohn, S.-I. Lee, and Y. Ro, "Visual comfort improvement in stereoscopic 3D displays using perceptually plausible assessment metric of visual comfort," *IEEE Trans. Consum. Electron.*, vol. 60, no. 1, pp. 1–9, Feb. 2014.
- [23] H. Deng, Q.-H. Wang, D.-H. Li, W.-X. Zhao, Y.-H. Tao, and A.-H. Wang, "Parallax images acquired by parallel camera array with shift," *Acta Photon. Sinica*, vol. 38, no. 11, pp. 2985–2988, 2009.
- [24] J. Lei, S. Li, B. Wang, K. Fan, and C. Hou, "Stereoscopic visual attention guided disparity control for multiview images," *J. Display Technol.*, vol. 10, no. 5, pp. 373–379, May 2014.
- [25] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3D," *ACM Trans. Graph.*, vol. 29, no. 4, 2010, Art. no. 75.
- [26] Y. Niu, F. Liu, X. Li, and M. Gleicher, "Warp propagation for video resizing," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 537–544.
- [27] Y. Niu, W.-C. Feng, and F. Liu, "Enabling warping on stereoscopic images," *ACM Trans. Graph.*, vol. 31, no. 6, 2012, Art. no. 183.
- [28] H. Park, H. Lee, and S. Sull, "Efficient viewer-centric depth adjustment based on virtual fronto-parallel planar projection in stereo 3D images," *IEEE Trans. Multimedia*, vol. 16, no. 2, pp. 326–336, Feb. 2014.
- [29] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [30] C. Rother, A. Blake, and V. Kolmogorov, "'GrabCut': Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.
- [31] K. Zhang, Q. Liu, H. Song, and X. Li, "A variational approach to simultaneous image segmentation and bias correction," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1426–1437, Aug. 2015.
- [32] F. Liu and H. Li, "Joint sparsity and fidelity regularization for segmentation-driven CT image preprocessing," *Sci. China Inf. Sci.*, vol. 59, no. 3, pp. 1–7, 2016.
- [33] R. Cormack and R. Fox, "The computation of disparity and depth in stereograms," *Attention, Perception, Psychophys.*, vol. 38, no. 4, pp. 375–380, 1985.
- [34] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, 2008, Art. no. 118.
- [35] I. Feldmann, *HHI Test Material for 3D Video*, ISO/IEC JTC1/SC29/WG11, document MPEG M15413, ISO, Archamps, France, Apr. 2008.
- [36] Y. Fang, C. Zhang, J. Li, J. Lei, M. P. Da Silva, and P. Le Callet, "Visual attention modeling for stereoscopic video: A benchmark and computational model," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4684–4696, Oct. 2017.
- [37] Y. Pang, H. Zhu, X. Li, and J. Pan, "Motion blur detection with an indicator function for surveillance machines," *IEEE Trans. Ind. Electron.*, vol. 63, no. 9, pp. 5592–5601, Sep. 2016.



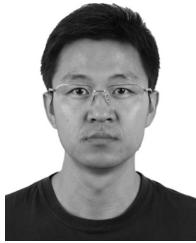
Jianjun Lei (M'11–SM'17) received the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications, Beijing, China, in 2007.

He was a Visiting Researcher with the Department of Electrical Engineering, University of Washington, Seattle, WA, from 2012 to 2013. He is currently a Professor with Tianjin University, Tianjin, China. His research interests include 3-D video processing, 3-D display, and computer vision.



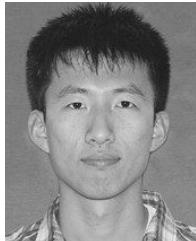
Bo Peng received the M.S. degree in communication and information systems from Xidian University, Xi'an, Shaanxi, China, in 2016.

She is currently pursuing the Ph.D. degree with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. Her research interests include computer vision, image processing, and video analysis.



Changqing Zhang received the B.S. and M.S. degrees from the College of Computer Science, Sichuan University, Sichuan, China, in 2005 and 2008, respectively, and the Ph.D. degree in Computer Science from Tianjin University in 2016.

He is currently an Assistant Professor with the School of Computer Science and Technology, Tianjin University. His current research interests include machine learning, data mining, and computer vision.



Xuguang Mei received the B.S. degree in electronic information engineering from Tianjin Polytechnic University, Tianjin, China, in 2014.

He is currently pursuing the M.S. degree with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. His research interests include 3-D image processing and 3-D display.



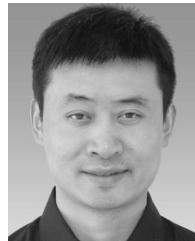
Xiaochun Cao (SM'14) received the B.E. and M.E. degrees in computer science from Beihang University, Beijing, China, and the Ph.D. degree in computer science from the University of Central Florida, Orlando, FL, USA.

He has been a Professor with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China, since 2012. He spent about three years with Object Video Inc., as a Research Scientist. From 2008 to 2012, he was a Professor with Tianjin University, Tianjin, China. He has authored and co-authored over 120 journal and conference papers. He is a fellow of the IET. He is on the Editorial Board of the IEEE TRANSACTIONS ON IMAGE PROCESSING. His dissertation was nominated for the University of Central Florida's university level Outstanding Dissertation Award. In 2004 and 2010, he was a recipient of the Piero Camperoni Best Student Paper Award at the International Conference on Pattern Recognition.



Xiaoting Fan received the M.S. degree in electronic engineering from Xi'an Polytechnic University, Xi'an, China, in 2017.

She is currently pursuing the Ph.D. degree with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. Her research interests include 3-D image processing and 3-D display.



Xuelong Li (M'02–SM'07–F'12) is currently a Full Professor with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, and also with The University of Chinese Academy of Sciences, Beijing, China.