

Guidance Method to Allow a User Free Exploration with a Photorealistic View in 3D Reconstructed Virtual Environments

Sho Iwasaki^(✉), Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose

Graduate School of Information Science and Technology, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
{iwasaki, narumi, tani, hirose}@cyber.t.u-tokyo.ac.jp

Abstract. Architectural reconstruction based on photographs is important for digitally archiving historic buildings. In addition, freely exploring a reconstructed Virtual Environment (VE) enhances users' understanding of and interest in the background cultural information. However, reconstructed models often have errors as well as terribly distorted views when viewed from a point distant from where the pictures were taken. Because of this problem, it is difficult to allow users to freely explore in a VE while keeping photorealistic views.

In this study, we propose a new method that enables both free exploration and high rendering quality by implicitly guiding users to the well-rendered view-points. First, we evaluate the rendering quality of the reconstructed model with view-dependent texture mapping. Second, by combining the evaluation results, we create a type of potential field that defines which direction to guide the users. Experimental results suggest that our proposed method can decrease the distortion in the views during VE exploration, and can possibly allow free exploration in VEs.

Keywords: User guidance · Image-based rendering · 3D reconstruction · Virtual environment · Digital museum

1 Introduction

Image-based rendering and modeling makes it possible to digitally archive and explore historic places and architectures. Exploring virtual environments (VEs) with free interaction is expected to not only enable users to intuitively understand the size and atmosphere, but also to promote their interest in the historical and cultural background information [1, 2]. Actually, interactive museum exhibits that use image-based rendering and modeling can effectively convey information [2].

A three-dimensional (3D) reconstruction method makes it possible to explore a virtual environment freely without being constrained by the camera positions [3]. Three-dimensional reconstruction estimates the 3D structure of a target object selected from images and generates a 3D model of it. Structure from Motion (SfM) is an essential 3D reconstruction technique. In the SfM workflow, the feature points in the images are detected, and then the positions in 3D space of these points and the cameras are estimated

geometrically. Generally, further reconstruction, e.g., dense-point reconstruction and surface reconstruction, will be made based on the results of SfM.

Many studies have investigated methods for reconstructing 3D models from images and applying 3D reconstruction to Virtual Reality (VR) [3, 4]. However, 3D reconstruction often makes errors, depending on the conditions in which the pictures were taken and the materials or shapes of the target objects. For example, a glass or wall with a plain texture is very difficult to reconstruct from only images because their feature points cannot be detected. Then, these errors cause a view distortion when exploring the VE and decrease the sense of immersion. Moreover, the decrease in the sense of immersion decreases the users' interest and their understanding of the target.

Incidentally, many studies have been conducted on user guidance in VE, and several studies have proposed methods with which users are guided unconsciously [5]. These techniques enable users to keep the feeling of free exploration while leading them to pre-defined viewpoints in the VE. We think that, with these guidance methods, we can help users avoid locations with distorted views and guide them to photorealistic views as they explore.

Therefore, in this paper, we propose a new guidance method that leads users implicitly and keeps their views photorealistic while allowing free exploration. In this method, we evaluate the rendering quality at each viewpoint in the VE, and generate a type of potential field, which we refer to as a guidance field, whose value is the result of the evaluation. Then, users are guided according to the guidance field. In addition, we conducted preliminary experiments to examine the effectiveness of the guidance field in decreasing the view distortion.

2 Related Work

In this section, we first describe studies on 3D viewers for VR using 3D reconstruction technology. Second, we describe related studies on user-guidance methods in VE.

2.1 3D Viewers for VEs Reconstructed from Multiple Images

“Photosynth” [6] by Microsoft and “Photo tourism” [3] by Snavely et al. are famous 3D viewer applications that enable the interactive exploration of VEs reconstructed using images.

Photosynth is a web service that enables users to easily construct panoramic views from several photos on a web browser. It has two viewer modes. In one mode, users can look around panoramically from a single point; in the other mode, users can move around a single target object and look at it from 360°. However, users can only move to the left or right, and the exploratory degrees of freedom (DOF) are very small.

Snavely et al. proposed a 3D viewer named “Photo tourism,” which enables the interactive exploration of VEs reconstructed using numerous images on the web. They employed several interfaces to view the VEs. Basically, users smoothly transition between photos and only view them from the positions from which they were taken. Although the authors also implemented the viewer mode using the standard 6-DOF, the

views could not be kept photorealistic and the quality of the views could be low if the users' viewpoints were away from the photos [7].

These days, other 3D reconstruction methods using prior knowledge, or methods using images and a 3D sensor in combination, have been proposed. However, the views are still distorted, depending on the users' viewpoint and view direction, even when using the latest 3D reconstruction methods or 3D sensors [8, 9].

2.2 Guidance Methods for Exploration in VEs

Galyean proposed a guidance method called “the river analogy,” which contains the users' exploration around an “anchor” moving along a predefined path in the VE [10]. In this method, the users are tied to the anchor by a virtual spring; if the user goes too far from the anchor, he is pulled back near it. This method makes it possible to keep users near the predefined path and prevent them from leaving the desired positions. However, with Galyean's original method, since the anchor moves from moment to moment, there is a high possibility that the users will be forced to move and their free exploration may be hindered.

Tanaka et al. proposed another guidance method that slightly alters the user inputs and implicitly leads them to the predefined positions [5]. A type of potential field is generated depending on the distance and direction from the target position, and it defines the alterations against the users' input. If the alteration is very small, the users may not feel a sense of incompatibility or sometimes do not even notice the guidance; thus, this guidance method can keep the feeling of free exploration. We extended this method and propose a new method that generates a potential field and slightly changes the users' viewpoints to less distorted view positions.

3 Design of Proposed Method

3.1 Overview

In this section, we explain the algorithm for our proposed guidance method, based on evaluating the rendering quality, and the manner of generating the guidance field, which defines where to guide the users. Figure 1 shows our method's workflow. Generating the guidance field mainly consists of three processes:

1. 3D reconstruction of VE,
2. Evaluation of rendering quality, and
3. Generation of a guidance field, based on the evaluation results.

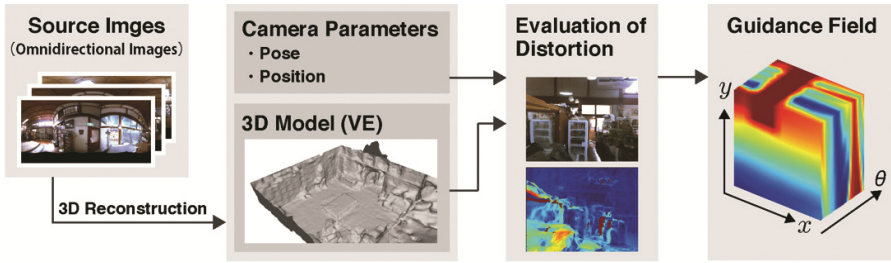


Fig. 1. Our method's workflow.

3.2 3D Reconstruction of VE

Our 3D reconstruction workflow and processes are similar to previous studies [4]. The workflow is as follows:

- (1) Feature points are detected in the images and the feature correspondences are acquired.
- (2) The cameras' poses and positions are estimated from the feature-correspondence results. Simultaneously, the feature points' positions in 3D space are estimated; these are called sparse space points.
- (3) Based on the camera parameters, dense space points are reconstructed using a multi-view stereo method.
- (4) Finally, the surface of the target objects is reconstructed and a 3D model is generated.

We used the Agisoft [11] for processes (1) to (3) and a Poisson Surface Reconstruction [12] implemented on MeshLab [13] for process (4).

3.3 Evaluation of Rendering Quality

To realize photorealistic views, we evaluate the rendering quality. In this paper, we use rendering quality and view distortion to mean opposite things. Before we define our evaluation of the rendering quality, we explain the texture-mapping method used in our 3D viewer. We employ view-dependent texture-mapping [14].

View-dependent texture-mapping is a texture-mapping method which selects a texture image projected onto the model, depending on the current viewpoint and view angle. Specifically, an image that is taken from an angle or position close to the current view angle or position is selected as the texture. Therefore, the texture mapped onto the VE changes dynamically as users walk or look around in the VE. The camera positions used to calculate the distance from the current viewpoint were already recovered in the 3D reconstruction process.

We employ view-dependent texture-mapping because the method is compatible with the models generated by 3D reconstruction. It is almost impossible to reconstruct object structures in detail; therefore, reconstructed 3D models sometimes have rough meshes and errors. View-dependent texture-mapping can visually remove the roughness and the

errors. In other words, by switching textures based on the view positions and angles, the views look natural even if the 3D model contains errors. Specifically, if the texture is projected onto the mesh from almost the same angle as the current view angle, the generated view will be very natural, regardless of the roughness of the 3D model. Thus, view-dependent texture-mapping is very effective for reconstructed models.

Let us now return to the evaluation of rendering quality. The rendering quality of the mesh whose texture was selected by the view-dependent texture-mapping depends on the following three elements: the current viewpoint, the target mesh position, and the point where the texture image was taken. Thus, the evaluation of rendering quality, which is defined for each mesh, can be defined using the geometric information. Buehler et al. [15] proposed using the angular error, expressed as follows (Fig. 2(a)):

$$\varphi_k^{ULG} = \cos^{-1}(\hat{\mathbf{v}}_c \cdot \hat{\mathbf{v}}_v), \quad (1)$$

where $\hat{\mathbf{v}}_v$ denotes the direction vectors between the current view points and the k -th mesh centers, and $\hat{\mathbf{v}}_c$ denotes the direction vectors between the camera centers and the k -th mesh centers. When the texture is projected obliquely against the mesh face, the texture image will be stretched and the value calculated by Eq. 1 will sometimes not correspond to the visually observed rendering quality. Thus, we propose a new evaluation, which is a simple change from Eq. 1. We consider the normal vector of the mesh, i.e. (Fig. 2(b)):

$$\varphi_k^{proposed} = \cos^{-1}(\hat{\mathbf{v}}_c \cdot \hat{\mathbf{v}}_v) \cos^{-1}(-\mathbf{n} \cdot \hat{\mathbf{v}}_c), \quad (2)$$

where \mathbf{n} denotes the normal unit vector of the k -th mesh. If the texture is obliquely projected onto the mesh, the angle between $-\mathbf{n}$ and $\hat{\mathbf{v}}_c$ increases, as does the value of $\varphi_k^{proposed}$. Moreover, the value of $\varphi_k^{proposed}$ decreases if the rendering quality is high.

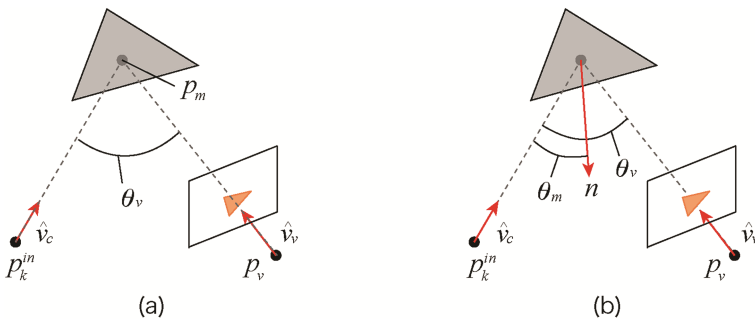


Fig. 2. Evaluation of rendering quality. p_v, p_k^n and p_c are respectively the positions of the user and the camera with which the texture image was taken, and p_k is the position of the k -th mesh. (a) The evaluation used in a previous study [15] employs θ_v , the angle between the vectors $\hat{\mathbf{v}}_c$ and $\hat{\mathbf{v}}_v$, which are direction vectors from p_v and p_c to p_k , respectively. (b) Our proposed method considers the normal unit vector \mathbf{n} of the mesh and uses the angle θ_m between $\hat{\mathbf{v}}_c$ and \mathbf{n} .

In this paper, because we generate the texture from a single omnidirectional image, only the meshes visible from the camera position are rendered with the texture.

Thus, the rendering quality can't be defined for the meshes invisible from the cameras, and we defined their rendering quality as the constant value $C_{invisible}$. Figure 3 shows the result of our proposed evaluation.

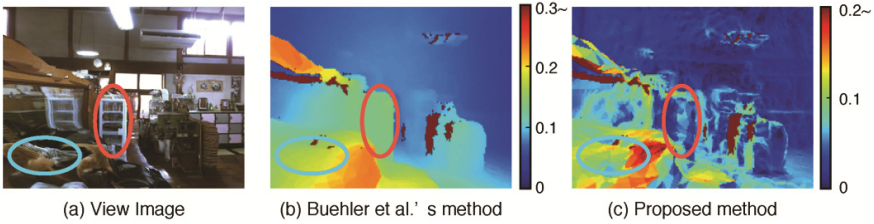


Fig. 3. Result of evaluating the rendering quality. (a) View image. The area circled in red is natural and the area circled in blue is distorted. (b) The result of applying Buehler et al.'s method [15] to the view. The rendering quality of the two circled areas is almost the same. (c) Result of our proposed method. The values correlate well with the visual distortion. (Color figure online)

3.4 Guidance Field Based on the Rendering-Quality Evaluation

We next describe how to generate a guidance field based on the rendering-quality evaluation and how to lead users according to the guidance field. We defined the view distortion at each viewpoint as the average evaluated rendering quality in the view there. Then, we used it as the value of the viewpoint guidance field. Consequently, the guidance field is defined as follows:

$$\Phi = \frac{1}{m} \sum_{p \in D} \varphi_k^p, \quad (3)$$

where p denotes a pixel in the view screen, and D and m denote a set of pixels in the view screen and the total number of pixels, respectively. φ_k^p is the value of the rendering quality of the k -th mesh, which is projected onto pixel p . That is, the value of the guidance field is the sum of the rendering quality on the screen at that viewpoint. Finally, the user's locomotion and rotation are altered by the following equations, respectively:

$$\mathbf{v}_{guide} = -\nabla\Phi, \quad (4)$$

$$\mathbf{v}_{output} = \mathbf{v}_{input} + g(\Phi)\mathbf{G}\mathbf{v}_{guide}, \quad (5)$$

where \mathbf{v}_{guide} denotes the vector indicating which direction to guide the user, \mathbf{v}_{input} denotes the user's input vector, and \mathbf{v}_{output} denotes the vector indicating the total changes of the user's position and rotation. The function $g(\Phi)$ defines the strength of the guidance and the matrix \mathbf{G} defines the strength balance between locomotion and rotation. With guidance according to Eq. 4, the user will be guided to less distorted view points.

4 Experiment

To evaluate our proposed guidance-field method, we conducted preliminary experiments by applying our method to real data. First, we describe the VE used in the experiment. Second, we describe the experimental procedures and results.

4.1 VE Used in the Experiment

3D Reconstruction and Model. We reconstructed “Kikusuiyu,” one of the most famous Japanese public bathhouses. The bathhouse was closed in September 2015. We photographed the bathhouse with an omnidirectional camera, the Ladybug3, with a resolution of 4096×2048 pixels (Fig. 4(a)). We took pictures along the five straight paths (Fig. 4(b)). For 3D reconstruction, we used 44 omnidirectional images, which are each about 30 cm away from the next picture.

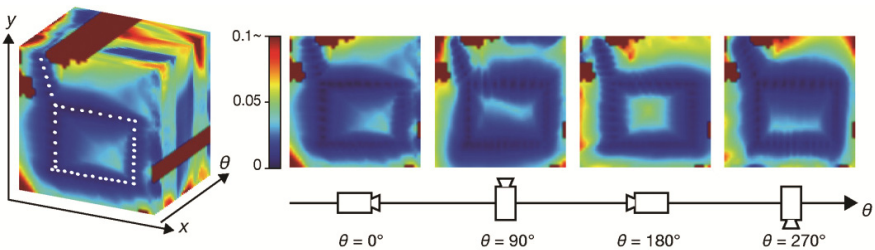


Fig. 4. Guidance field. (*left*) Overview of the guidance field $\Phi(x, y, \theta)$. The white dots indicate the positions where the texture images were taken. (*right*) Cross sections of the field at four typical directions.

We applied 3D reconstruction, mentioned in Sect. 3.1, to these images, and simplified the 3D model by applying a Quadric Edge Collapse Decimation, implemented in MeshLab. The resulting 3D model contains 30,022 vertices and 59,987 faces (Fig. 4(c)).

Guidance Field. We generated the guidance field for the VE according to Eq. 3. In this experiment, the user can only move horizontally, and only rotate around the vertical axis. Therefore, the guidance field is expressed as $\Phi(x, y, \theta)$.

To generate the guidance field, we calculated the rendering quality for 1600 points (40×40 grid with an interval of almost 15 cm), and 36 directions with a 10-degree interval at each point. $C_{invisible}$ was 0.25. The aspect ratio and field of view are 4×3 , and 70° , respectively, which is the same as that used in our 3D viewer. We excluded 64 points where the user can see beyond the 3D model and outside of the VE. Since the rendering quality differs even at the same position and angle, depending on the texture image projected onto the meshes, we generated guidance fields for each of the 44 texture images. Then, we integrated these guidance fields by employing the smallest distortion value at each position and angle. The final generated guidance field is shown in Fig. 5.

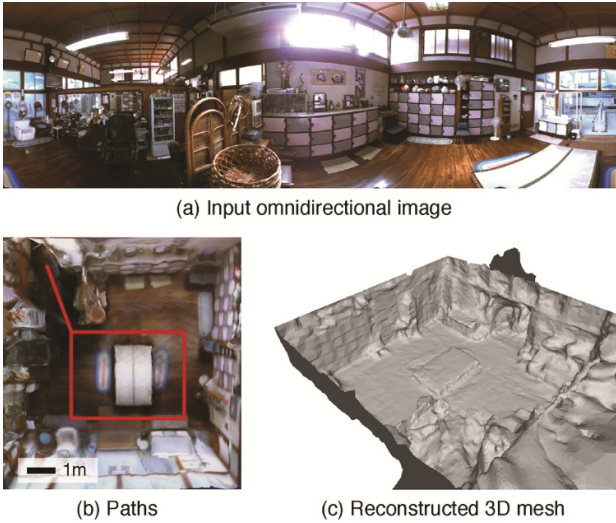


Fig. 5. Input data and result of the 3D reconstruction. (Color figure online)

In Fig. 5, the view distortion is small in the blue areas; conversely, the rendering quality is low in the red areas. The white points denote the positions where the texture images were taken. It can be seen in the figure that, near these points, the guidance field value is small; i.e., the rendering quality is high.

Texture Mapping. We employed view-dependent texture-mapping, but used a different method to select the texture image for the with- and without-guidance conditions. For the with-guidance condition, the texture with the smallest rendering quality at the viewpoint was selected. For the without-guidance condition, the texture nearest the current viewpoint was selected.

4.2 Detailed Procedure

We prepared two typical scenes that assume exploration in the VE, which we call Scene 1 and Scene 2, and compared the views in each scene with and without our proposed guidance method.

In Scene 1, the virtual user moves straight without rotating. This assumes simple locomotion. The start position is $(x, y) = (3.7, 4.2)$ and the start direction is 225° from the x-axis. The moving speed is constant at 0.83 m/s. In Scene 2, the virtual user rotates without locomotion. This assumes that the user is simply looking around. The start position is $(x, y) = (2.9, 0.5)$ and the start direction is the positive y-axis direction, which is 90° from the x-axis. The virtual user rotates clockwise at a constant speed of $40^\circ/\text{s}$.

Through the experiments, the aspect ratio is 4×3 and the field of view is 70° . The function $g(\Phi)$ and the matrix \mathbf{G} in Eq. 5 are defined as follows: $g(\Phi) = \alpha|\Phi|$, where $\alpha = 4.8 \times 10^3$, and $\mathbf{G} = \text{diag}(1, 1, 25)$, respectively. The units are meters and degrees, also respectively.

4.3 Results and Discussion

Figures 6 and 7 show the viewpoint transitions and the transitions of the evaluated distortion values for the views in Scene 1 and Scene 2, respectively.

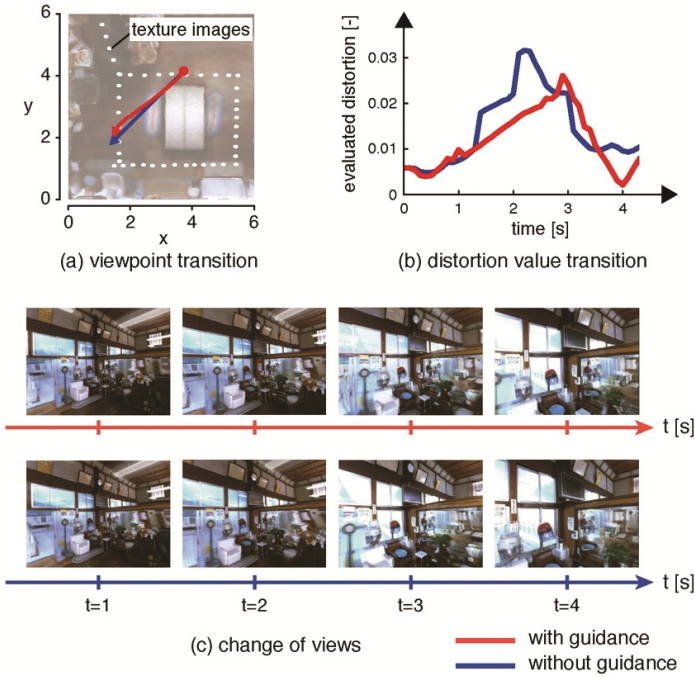


Fig. 6. Views with and without guidance in Scene 1. The white dots in (a) depict the positions where the texture images were taken.

Figure 6 shows that, in Scene 1, our proposed guidance method slightly altered the virtual user’s moving path, and decreased the distortion in the views; the average evaluated distortion value four seconds from the start was 0.015 without guidance and 0.012 with guidance. Figure 6(a) shows that the viewpoints are led near the positions where the texture images were taken. Consequently, the evaluated distortion was kept lower with the guidance than without it, from $t = 1.5$ to $t = 2.5$. However, the value was reversed at $t = 3$. We think it is mainly because the proposed method only leads users to current lower potential viewpoints according to Eq. 4, and does not consider time-series information or path planning. Figure 6(c) shows that the guidance altered not only the moving path but also the view directions.

Figure 7 shows that, in Scene 2 as well, the guidance method slightly altered the virtual user’s viewpoints and kept the evaluated distortion in the views. In Scene 2, although the virtual user’s input was only rotation without locomotion, the virtual user was led by the guidance near the positions where the textures were taken (Fig. 7(a)). The average evaluated distortion value for four seconds from the start was 0.021 without

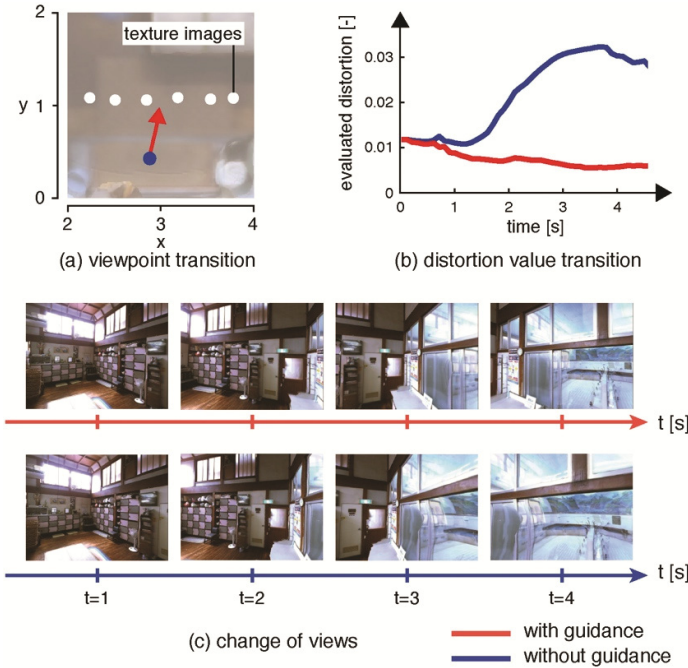


Fig. 7. Views with and without guidance in Scene 2. The white dots in (a) depict the positions where the texture images were taken.

guidance and 0.0078 with guidance. In Fig. 7(c), the view at $t = 3$ without guidance is terribly distorted, but the view with guidance is not.

These results suggest that our proposed guidance method efficiently decreases the distortion and keeps the rendering quality in the views. We also think that the alterations of the users' positions and directions each time are small enough to not hinder free exploration.

5 Conclusion

Both free exploration and high visual quality are important for exploration in VEs. In this paper, we proposed a new method that maintains high rendering quality while allowing free exploration by slightly altering the users' viewpoints and view direction according to the guidance field generated by evaluating the rendering quality.

Experimental results showed that our proposed guidance method could maintain the view quality and suggested that it could provide both free exploration and high view quality.

In future work, we plan to examine whether the guidance-based alteration disturbs the users' exploration in real scenes. It is also important to investigate what types of alteration the users notice easily. We think that the users will more easily notice the direction change while moving straight, as in Scene 1, than the viewpoint change while

rotating, as in Scene 2, because the former alteration sometimes results in moving the users away from their destination in VE. For real use, we must consider these things when designing the guidance method.

Acknowledgements. This research and development work was supported by the MIC/SCOPE#162303009.

References

1. Cairncross, S., Mannion, M.: Interactive multimedia and learning: realizing the benefits. *Innovations Educ. Teach. Int.* **38**(2), 156–164 (2001)
2. Narumi, T., Hayashi, O., Kasada, K., Yamazaki, M., Tanikawa, T., Hirose, M.: Digital diorama: AR exhibition system to convey background information for museums. In: Shumaker, R. (ed.) *VMR 2011. LNCS*, vol. 6773, pp. 76–86. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-22021-0_10](https://doi.org/10.1007/978-3-642-22021-0_10)
3. Snively, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3D. In: *ACM Transactions on Graphics (TOG)*, vol. 25, pp. 835–846. ACM (2006)
4. Musialski, P., Wonka, P., Aliaga, D.G., Wimmer, M., Gool, L.V., Purgathofer, W.: A survey of urban reconstruction. In: *Computer Graphics Forum*, vol. 32, no. 6, pp. 146–177. Wiley Online Library (2013)
5. Tanaka, R., Narumi, T., Tanikawa, T., Hirose, M.: Guidance field: potential field to guide users to target locations in virtual environments. In: *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 39–48. IEEE (2016)
6. Microsoft Corporation. Photosynth. <https://photosynth.net/>
7. Snively, N., Garg, R., Seitz, S.M., Szeliski, R.: Finding paths through the world’s photos. In: *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, p. 15 (2008)
8. Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R.: Reconstructing building interiors from images. In: *2009 IEEE 12th International Conference on Computer Vision*, pp. 80–87. IEEE (2009)
9. Xiao, J., Furukawa, Y.: Reconstructing the world’s museums. *Int. J. Comput. Vis.* **110**(3), 243–258 (2014)
10. Galyean, T.A.: Guided navigation of virtual environments. In: *Proceedings of the 1995 Symposium on Interactive 3D Graphics*, p. 103-ff. ACM (1995)
11. Agisoft PhotoScan User Manual, Professional Edition, Version 1.2. http://www.agisoft.com/pdf/photoscan-pro_1_2_en.pdf
12. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, vol. 7, pp. 61–70 (2006)
13. Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G.: MeshLab: an open-source mesh processing tool. In: Scarano, V., De Chiara, R., Erra, U. (eds.) *Eurographics Italian Chapter Conference. The Eurographics Association* (2008)
14. Debevec, P.E., Taylor, C.J., Malik, J.: Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, pp. 11–20. ACM. (1996)
15. Buehler, C., Bosse, M., McMillan, L., Gortler, S., Cohen, M.: Unstructured lumigraph rendering. In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 425–432. ACM (2001)