Perceptually-Guided Design of Nonperspectives Through Pictorial Depth Cues

Kenichi Yoshida
The Univ. of Tokyo, Japan

Shigeo Takahashi
The Univ. of Tokyo, Japan

Hiroaki Ono
The Univ. of Tokyo, Japan

Issei Fujishiro
Keio Univ., Japan

Masato Okada
The Univ. of Tokyo & RIKEN, Japan

Introducing distortions into perspective views is a popular technique to direct our visual attention to specific objects, as seen in hand-drawn illustrations and cartoon animations. This type of image expression, called nonperspective projection, is feasible in visual communication, because the human visual system can reconstruct the target three-dimensional (3D) scene correctly provided that the corresponding image distortions are within a certain perceptual tolerance. In this paper, we develop a perceptual approach to guiding the design of such nonperspective images by referring to the 3D perception induced by pictorial depth cues. We formulate an acceptable tolerance by investigating how we perceive image distortion according to the change in the configuration of depth cues. The obtained formulation is then incorporated into our new algorithm, with which we can automatically control plausible image deformation by simply modifying the positions and sizes of specific objects in a scene.

***Keywords*-nonperspective projection; pictorial depth cues; psychophysical experiments; perceptually-guided design**

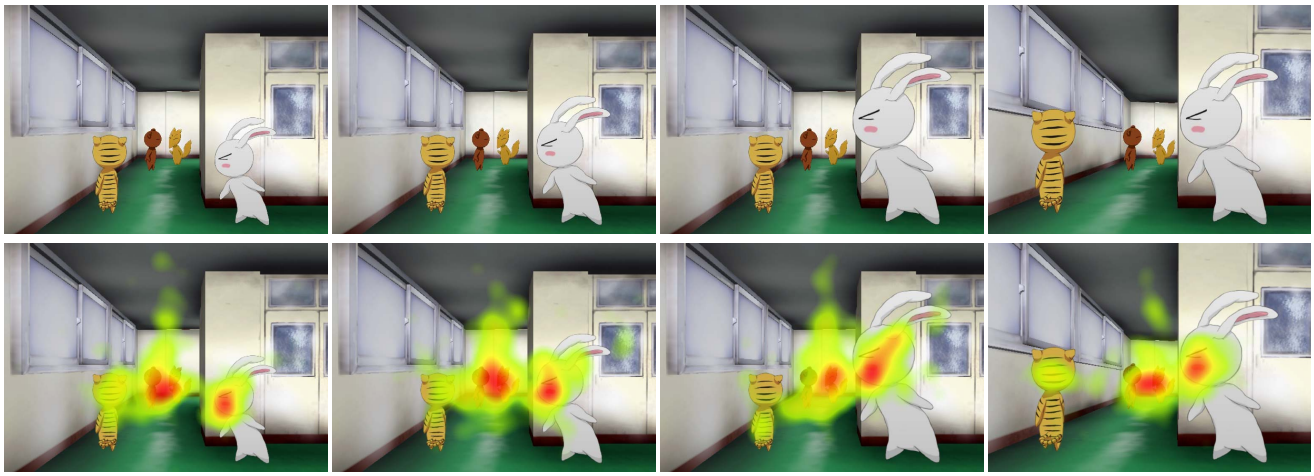
I. Introduction

Perspective projection is the most common tool used to transform a three-dimensional (3D) scene onto a two-dimensional (2D) screen because it is based on the pinhole camera model, and thus represents the depth information of the 3D scene precisely. On the other hand, hand-drawn pictures such as illustrations and cartoons commonly contain plausible distortions in order to emphasize or suppress specific features in the 3D scene. This type of projection, called *nonperspective projection*, has served as important visual media in our daily communication and has been intensively investigated to support its generation process.

Nonperspective projection can be defined as a process of generating a composite image by smoothly synthesizing local perspectives seen from different viewpoints. However, prior research on this projection technique has tended to focus on stitching together or smoothly blending such local perspectives to make the synthesized image visually consistent, but has not provided any control over the acceptable amount of overall deformation of the image. This means that to date the global design of nonperspective deformation has been manually controlled, and thus results in a time-consuming, trial-and-error process even for experienced artists. A desire to avoid the need for such a tedious drawing process leads us to introduce appropriate global constraints for managing plausible deformations of an overall image, which can also be followed by possible local fine adjustments.

This paper presents a novel approach to designing nonperspective projections with such global constraints, by formulating a perceptual tolerance of the associated image deformations. To our knowledge, this is the first attempt to introduce such a perceptual measure to formulate the plausibility of deformation inherent in nonperspective projections. Our basic idea is to relate such a perceptual measure with the configuration of depth cues in a 3D scene. This is because the human visual system extracts the depth cues together with their induced depth information from a 2D projection, and combines them to correctly reconstruct the target 3D scene provided that the corresponding image distortions fit within a given perceptual tolerance. In our approach, we first investigate the acceptable amount of deformation in terms of the configuration of depth cues through psychophysical experiments. We then incorporate the obtained perceptual measure into our prototype system to automatically optimize the nonperspective distortion in the target 3D scene.

In this paper, we will focus our attention on two major pictorial cues, that is, *linear perspective* and *relative size*, because they play a primary role in the perception of the object sizes in our target 2D projection. Linear perspective is a form of perspective in drawing in which the relative size of an object is determined by its distance from the viewer. A typical example of this is a scene that contains a set of parallel lines that converge to a vanishing point according to depth. Figure 1 shows 2D projections with different image distortions (top) and their corresponding distribution of visual attention obtained through eye-tracking measurements. In an ordinary perspective image (Figure 1(a)), the distribution of our visual attention is well-balanced. Slightly enlarging the rabbit will not significantly disturb our visual perception, even with the same characters and background scenes in size (Figure 1(b)). However, excessive magnification of the rabbit incurs a biased balance of our gaze fixations since we recognize that the rabbit is unnaturally bigger as compared with the other characters and background scene (Figure 1(c)). Our new approach automatically alleviates this problem by adjusting the relative sizes and positions of the other objects together with the pictorial layout of parallel lines, and thus allows us



(a) (b) (c) (d)

Figure 1. 2D projections with different image distortions (top) and their corresponding distribution of gaze fixation period (bottom): (a) Original perspective image. (b) The rabbit is slightly enlarged. (c) Excessive distortion is applied to the rabbit. (d) Image distortion perceptually optimized using our approach. The red color corresponds to the long gaze fixation time.

to recover the well-balance distribution of visual attention over the image (Figure 1(d)).

The rest of this paper is organized as follows: Section II provides a survey on related work. Section III formulates a perceptual tolerance of image deformation through psychophysical experiments. Section IV presents an algorithm for calculating such plausible deformations in terms of the configuration of depth cues. Section V gives several results, which will be followed by the conclusion in Section VI.

II. Related Work

A. Pictorial Depth Cues

Depth perception through pictorial cues has been intensively investigated as an important research theme in the field of visual psychology. One of the major propositions is size constancy, which claims that the size of a known object tends to remain relatively constant even when its retinal size changes considerably with distance [1]. Nonetheless, our visual system often tends to overestimate the sizes of objects slightly when viewed up close [2]. Further research has been performed to explore how the perception of relative size is affected by the existence of other pictorial cues such as linear or texture perspectives [3]. However, little work has been done to analyze how the configuration of such depth cues affects our understanding of 2D projections. In this paper, we design several psychophysical experiments to evaluate the influence of the image distortion induced by the pictorial perspective cues on our visual perception (Section III).

B. Nonperspective Projection

Nonperspective projection consists of techniques for composing a single image from multiple views, and has become a hot topic in the computer graphics and visualization community. Panoramic image generation is the most widely used application example of nonperspective projection [4], [5], [6]. Bending sight rays with different types of lenses enables magnification effects in 2D projection [7], [8]. Deforming the target 3D objects allows us to simulate the effect of composing nonperspective images as seen from multiple viewpoints [9], [10], [11]. Projection with multiple cameras is an effective approach for generating nonperspective images [12], [13], [14]. which was followed by an interactive interface that directly manipulates 2D image deformations [14]. Our approach related to the last category, while the associated image deformation is rather automated because we take full advantage of the perceptual tolerance of nonperspective deformation in terms of the configuration of pictorial perspective cues.

III. Psychophysical Experiments

In our approach, we formulate a visually plausible deformation of a projected image by taking into account the configuration of pictorial perspective cues. As described earlier, we focus on 3D perception induced by the pictorial arrangement of parallel lines and objects in the 2D projection. We accomplish this goal by quantitatively evaluating acceptable size and position of an object with respect to the 2D arrangement of parallel lines. Note that we conducted the experiments under several specific viewing conditions because a difference in viewing condition is considered to have no significant effect [15].

A. Experimental Setup

Figure 2 exemplifies how the pictorial arrangement of parallel lines restricts the acceptable sizes and positions of

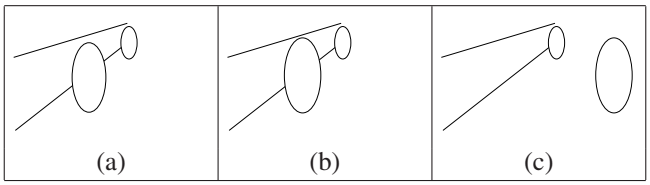


Figure 2. Visual illusion depending on the change in the pictorial arrangement of parallel lines and objects.

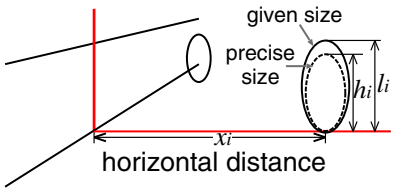


Figure 3. Experimental setup for one-point perspectives, where l_i, h_i, and x_i are the height of the given front ellipse at the i-th trial, its corresponding precise height obtained from the exact perspective projection, and horizontal distance from the parallel lines. Here, two ellipses are assumed to be the same in size.

two objects. Suppose that we have front and rear ellipses of the same size on the level ground, and place two parallel lines to their left in 3D space (Figure 2(a)). If we enlarge the size of the front ellipse while its position remains fixed (Figure 2(b)), we perceive the front ellipse to be larger than the rear one. However, when we translate the front ellipse to the right (Figure 2(c)), we are likely to miss the difference in size between the two ellipses. This visual illusion suggests that the positional tolerance of the parallel lines tends to increase as the corresponding object moves horizontally away from them in 2D projected image. We validate this hypothesis by conducting the psychophysical experiment described below.

In our experimental setup, we generated a set of 3D scenes, each of which was composed of front and rear ellipses on the ground and two parallel lines. Figure 3 shows this experimental setup, where we employed a line drawing in order to exclude other pictorial cues such as occlusion, shade and shadow, and texture gradient. Here, the sizes of the two ellipses and the orientation of the upper line were randomly specified by the system, while the lower line was always fixed on the ground in the 3D scene to make the heights of the parallel lines and ellipses visually measurable from the projected image. Furthermore, the front ellipse was also randomly placed along the horizontal line (in red) whereas the rear ellipse remained to be fixed. Participants were asked to return a yes/no answer according to whether the given 3D scene was perceptually acceptable under the condition that the two ellipses are identical in size, after the scene was displayed for 2 seconds.

Here, each participant had 400 different patterns of two ellipses and a pair of lines, while the lines were placed equally on the left and right in the scene to remove any bias

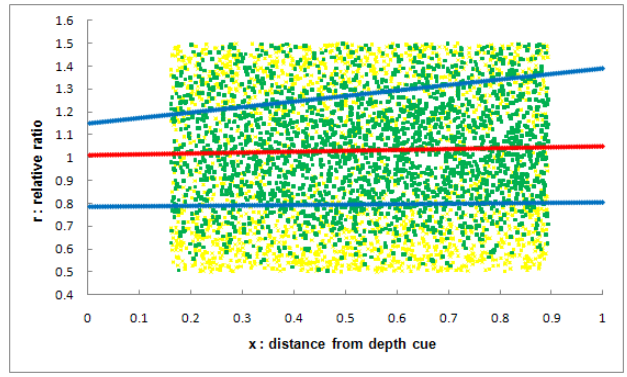


Figure 4. Distribution of samples in the domain of 2D horizontal distance x and relative size ratio r. Green plots represent acceptable samples while yellow plots unacceptable ones. The average ratio is almost constant with respect to x (red) and the acceptable range of relative ratios increases (blue).

due to asymmetric distribution of the depth cues. For each pattern, the system recorded the relative ratio of the two heights $r_i = l_i/h_i\,(i = 1,2,\ldots)$ together with the horizontal distance x_i and its corresponding yes/no answer, where l_i was the height of the given front ellipse at the i-th trial, and h_i was the corresponding precise height obtained from the exact perspective projection if two ellipses were the same in size (Figure 3). Note that both heights were evaluated at the depth of the front ellipses. We asked participants to first perform the overall experimental task once for rehearsal, and then to tackle the task again as their final trial. We recruited 14 participants (12 males and 2 females) ranging in age from 20 to 40 where all the participants worked on image synthesis problems, and thus we accumulated 5,600 samples in total.

In the statistical analysis phase, we classified all the samples into an acceptable and unacceptable groups $\{r_j^{\text{yes}}\}\,(j = 1,2,\ldots)$ and $\{r_k^{\text{no}}\}\,(k = 1,2,\ldots)$, respectively, according to the corresponding yes/no answer, and applied a linear regression to the acceptable samples $\{r_i^{\text{yes}}\}$ to discover how the average acceptable ratio μ changes according to the horizontal distance x, as follows:

$$\mu(x) = 0.0396x + 1.01. \tag{1}$$

Figure 4 shows such a result where the average ratio is represented in red while the acceptable and unacceptable samples are depicted in green and yellow. We also formulated the probability of the tolerable image deformation p, by first classifying each sample set (r_i, x_i) into two groups according to whether $r_i > \mu(x_i)$, and then applying a logistic regression separately to the two groups, as follows:

$$p(r,x) = \begin{cases} 1/(1+\exp(-5.58r+1.11x+6.54)) & r \geq \mu(x) \\ 1/(1+\exp(-7.76r-0.158x+6.27)) & r < \mu(x). \end{cases}$$

This lets us bound the acceptable range of the relative size ratio by setting $p(r,x) = 0.5$, as

$$\alpha(x) = 0.199x + 1.17 \text{ and } \beta(x) = -0.0203x + 0.808. \tag{2}$$

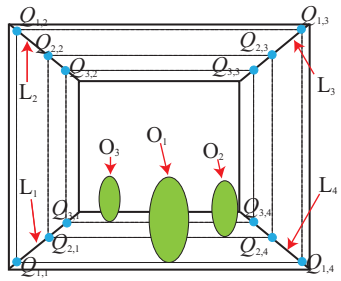


Figure 5. Original perspective scene consisting of a box-like container background where small foreground objects and a set of linear perspective cues are involved.

Note that the acceptable range $[\beta(x), \alpha(x)]$ becomes large as the distance x increases as shown in Figure 4, where its boundaries are outlined in blue. This fact supports our hypothesis that the difference in object size gradually becomes less perceivable as the front object moves away from the linear perspective cues.

IV. Generating Nonperspective Deformations

The psychophysical formulation we have just obtained allows us to understand how the acceptable amount of image deformation depends on the configuration of depth cues. In this section, we construct an algorithm for generating nonperspective projection using this formulation.

A. Pictorial Arrangement of Parallel Lines

Our design system is assumed to take as input the 3D models of background scene and foreground objects, and provide an ordinary perspective as a starting point for further design of nonperspective images. After having obtained the 2D positions and sizes of specific objects on the 2D projection from the user, the system calculates the global constraints imposed on the parallel lines in the scene. The final image deformation is then produced automatically by optimizing the error between the actual positions of the parallel lines and their constraints, while referring to the perceptual tolerance derived from the psychophysical experiments.

We now explain how to calculate the pictorial arrangement of parallel lines from the user-specified positions and sizes of objects. Figure 5 shows an ordinary perspective projection of a 3D scene consisting of a box-like container background, which contains small foreground objects $\mathrm{O}_i (i = 1, \ldots, n)$ such as characters and parallel lines $\mathrm{L}_j (j = 1, \ldots, m)$ such as intersections between walls, floors, and ceilings. Note that n and m denote the numbers of (constrained) foreground objects and parallel lines, respectively. For each foreground object O_i, we define the point on the parallel lines L_j at the same depth of O_i as $Q_{i,j}$. Then suppose that the user changes the positions and sizes of the foreground objects O_i by the 2D displacement vectors D_i and the scale factor s_i on the screen, while the 2D screen coordinates are normalized to $[-1,1] \times [-1,1]$.

Figure 6(a) illustrates how the user-provided changes affect the layout of parallel lines L_j. Here, the 2D positional constraint of $Q_{i,j}$ is defined to be:

$$Q'_{i,j} = P_i + D_i + \widetilde{s_i}(Q_{i,j} - P_i) \tag{3}$$
$$\widetilde{s_i} = \begin{cases} s_i - \alpha(x_{i,j}) + \mu(x_{i,j}) & s_i > \alpha(x_{i,j}) \\ \mu(x_{i,j}) & \alpha(x_{i,j}) \geq s_i \geq \beta(x_{i,j}) \\ s_i - \beta(x_{i,j}) + \mu(x_{i,j}) & \beta(x_{i,j}) > s_i \end{cases}$$

where $\mu(x_{i,j}), \alpha(x_{i,j})$ and $\beta(x_{i,j})$ represent the average, upper limit, and lower limit of the relative size ratio as in Eq. (1) and Eq. (2), where $x_{i,j}$ represents the horizontal distance of the object O_i from the parallel line L_j (Figure 6(a)). This definition implies that the new scale factor $\widetilde{s_i}$ remains to be the average $\mu(x_{i,j})$ as long as the difference of s_i from $\mu(x_{i,j})$ stays within the acceptable range $[\beta(x), \alpha(x)]$. However, once it exceeds the limit, the new scale factor $\widetilde{s_i}$ changes proportionally to reduce unnatural distortion of the image caused by the constrained positions and sizes of the foreground objects O_i's (Figure 6(b)).

In practice, the excessive number of constraints make the associated linear equations overconstrained. In this case, we minimize the total sum of squared distances between the constrained position $Q'_{i,j}$ and its corresponding position $Q_{i,j}$ on L_j, in a weighted-least square sense (Figure 6(c)). In this minimization, we use the width of the acceptable range $\sigma_{i,j} = |\alpha(x_{i,j}) - \beta(x_{i,j})|$ as the weight value for the constrained position $Q_{i,j}$, where $x_{i,j}$ again represents the horizontal distance of O_i from L_j.

Suppose that the parallel line L_j is represented by a function of the parameter t as $Q_j(t)$ and passes through the two endpoints U_j and V_j at $t = 0$ and $t = 1$, respectively. When $Q_{i,j} = Q_j(t_{i,j}) = (1 - t_{i,j})U_j + t_{i,j}V_j$, the layout of L_j is obtained by minimizing

$$\sum_{i=1}^{n} (Q_j(t_{i,j}) - Q'_{i,j})^2 / {\sigma_{i,j}}^2. \tag{4}$$

This amounts to finding the positions of U_j and V_j as:

$$U_j = \frac{CD - BE}{AC - B^2} \text{ and } V_j = \frac{AE - BD}{AC - B^2}, \text{ where}$$
$$A = \sum_{i=1}^{n} (1 - t_{i,j})^2 / {\sigma_{i,j}}^2,\ B = \sum_{i=1}^{n} (1 - t_{i,j}) t_{i,j} / {\sigma_{i,j}}^2,$$
$$C = \sum_{i=1}^{n} {t_{i,j}}^2 / {\sigma_{i,j}}^2, D = \sum_{i=1}^{n} (1 - t_{i,j}) Q'_{i,j} / \sigma_{i,j},$$
$$\text{and } E = \sum_{i=1}^{n} t_{i,j} Q'_{i,j} / \sigma_{i,j}. \tag{5}$$

B. Camera Parameters on Parallel Lines

For ordinary perspective images, the camera parameters are identical for all the sample points on the target 3D scene. On the other hand, as described in Section I, the

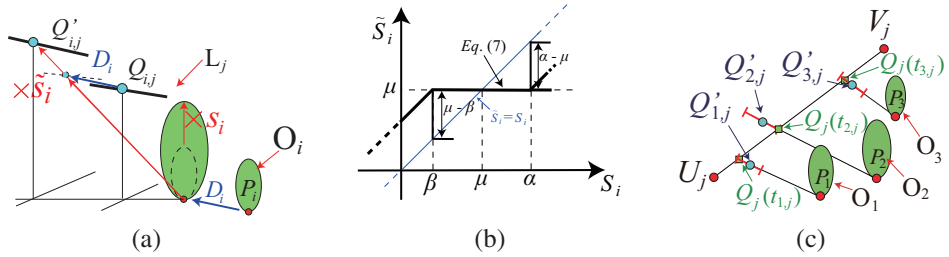


Figure 6. Optimization process of the pictorial arrangement of parallel lines from the given object sizes and positions: (a) Positions of constraints on linear perspective cues are calculated by the 2D displacement vectors and the scale ratios of objects. (b) The new scale factor $\tilde{s}_i$ changes proportionally when s_i exceeds the limit of the acceptable tolerance of scale ratios, (c) The new pictorial arrangement of parallel line L_j is calculated by minimizing the associated error from the constraints in a weighted-least square sense.

camera parameters for nonperspective images are assumed to change smoothly in the 3D scene. This lets us to introduce locally different camera parameters to each of the parallel lines, by referring to their pictorial arrangement optimized through our perception-based approach. In our approach, this has been accomplished by preparing a 3D field of camera parameters that fully contains the target 3D scene [16]. In our framework, the camera parameters are adaptively sampled in the 3D field while they are identical with those for generating ordinary perspective projection in the initial stage. The difference from these initial values along the parallel lines is then incorporated so that they satisfy the global constraints specified by the user. Such displacements are smoothly propagated over the 3D field by introducing the natural neighbor interpolation [17], in order to generate visually-plausible nonperspective images without any unexpected local creases and elongations.

V. Design Examples and Discussions

Our prototype system is implemented on a laptop PC with an Intel Core2Duo T9500 CPU (2.6GHz) and 4GB RAM. We adjust the 2D positions and sizes of foreground objects as intended, and the system automatically generates perceptually plausible deformation of the corresponding projected image at almost interactive rate (8-10 fps).

Our primary challenge is to introduce perceptual criteria for deforming projected images. Figure 7(a) shows a traditional Japanese woodblock print called "ukiyo-e," as an example of a distorted perspective image, where the pictorial arrangement of parallel lines and objects is intentionally specified. We simulated the nonperspective distortion effects with our design system. First we prepared an ordinary perspective image from which we started our design process as shown in Figure 7(b). We then dragged the positions of the performer on the corridor to the bottom left corner of the image, and the wooden bucket and box on the stage to the appropriate positions, so as to synthesize the nonperspective image as shown in Figure 7(c). Here, the stage in the rear inclines to our side and thus we can clearly see the performance on the stage, while we can still consistently perceive the overall 3D structure of the architecture together with the arrangement of audience.

Our approach enables intuitive design of nonperspective animation once we have specified the positions and sizes of the associated objects at specific times as keyframes, without changing any camera parameters explicitly. Figure 8 shows such an example where Figures 8(a), (c), and (e) represent the given keyframes, and Figures 8(b) and (d) correspond to the interpolated frames. In this animation sequence, we look at the rabbit first (Figure 8(a)), and next the bear and fox talking to each other from rather far away (Figure 8(c)). We then step forward to have a close look at the rabbit (Figure 8(e)).

We also conducted an eye-tracking experiment in order to evaluate how the temporal distribution of gaze fixation changes in accordance with image deformation. For this experiment, we asked 10 participants (9 males and 1 female) to view each image in the top row of Figure 1 for 2 seconds, and tracked their eye-gaze movements using a Tobii X120 eye-tracker system. The images at the bottom of Figure 1 shows temporal distributions of gaze fixation, each of which corresponds to that at the top. As described in Section I, we can easily confirm that our optimization scheme successfully controls the distribution of gaze fixation period in a plausible manner without losing the good balance of its distribution.

VI. Conclusion

This paper has presented a new perceptual approach to guiding the design of nonperspective images by taking advantage of the configuration of pictorial depth cues. To our knowledge, this is the first attempt to synthesize nonperspective images by taking into account the acceptable tolerance of associated image deformation. The plausible amount of such image deformation has been investigated through psychological experiments, in order to formulate how our perception of 3D scenes changes according to the pictorial arrangement of parallel lines and objects. This formulation has enabled us to design nonperspective images by simply editing the positions and sizes of objects involved in the target scenes. Various nonperspective images together

(a) (b) (c)

Figure 7. (a) "Shibai Ukie" by Masanobu Okumura (downloaded from Wikimedia Commons). (b) An ordinary perspective of the scene we reconstructed in our design system. (c) The generated nonperspective scene to simulate the deformation effects.

(a) (b) (c) (d) (e)

Figure 8. Nonperspective keyframe animation. Given keyframes ((a), (c), and (e)) and interpolated frames ((b) and (d)).

with keyframe animations were generated to demonstrate the feasibility of our approach.

Acknowledgment

This work has been partially supported by Japan Society of the Promotion of Science under Grants-in-Aid for Scientific Research (A) No. 20240020, Scientific Research (B) No. 20300033 and No. 21300033, and Young Scientists (B) No. 17700092, and the Casio Science Promotion Foundation.

References

[1] W. Epstein, J. Park, and A. Casey, "The current status of the size-distance hypothesis," *Psychological Bulletin*, vol. 58, no. 6, pp. 491–514, 1961.

[2] V. R. Carlson, "Size-constancy judgement and perceptual compromise," *Journal of Experimental Psychology*, vol. 63, no. 1, pp. 68–73, 1962.

[3] İ. Oruç, L. T. Maloney, and M. S. Landy, "Weighted linear cue combination with possibly correlated error," *Vision Research*, vol. 43, pp. 2451–2468, 2003.

[4] D. N. Wood, A. Finkelstein, J. F. Hughes, S. E. Thayer, and D. H. Salesin, "Multiperspective panoramas for cel animation," in *Proc. ACM SIGGRAPH 97*, 1997, pp. 243–250.

[5] J. Yu and L. McMillan, "A framework for multiperspective rendering," in *Proc. Eurographics Symp. Rendering 2004*, 2004, pp. 61–68.

[6] A. Agarwala, M. Agrawala, M. F. Cohen, D. Salesin, and R. Szeliski, "Photographing long scenes with multi-viewpoint panoramas," *ACM Trans. Graphics*, vol. 25, no. 3, pp. 853–861, 2006.

[7] Y. Kurzion and R. Yagel, "Interactive space deformation with hardware-assisted rendering," *IEEE Computer Graphics & Applications*, vol. 17, no. 5, pp. 66–77, 1997.

[8] J. Brosz, F. F. Samavati, M. T. C. Sheelagh, and M. C. Sousa, "Single camera flexible projection," in *Proc. 5th Int'l Symp. Non-photorealistic Animation and Rendering (NPAR2007)*, 2007, pp. 33–42.

[9] D. Martín, S. García, and J. C. Torres, "Observer dependent deformations in illustration," in *Proc. 1st Int'l Symp. Non-Photorealistic Animation and Rendering (NPAR2000)*, 2000, pp. 75–82.

[10] S. Takahashi, N. Ohta, H. Nakamura, Y. Takeshima, and I. Fujishiro, "Modeling surperspective projection of landscapes for geographical guide-map generation," *Computer Graphics Forum*, vol. 21, no. 3, pp. 259–268, 2002.

[11] S. Takahashi, K. Yoshida, K. Shimada, and T. Nishita, "Occlusion-free animation of driving routes for car navigation systems," *IEEE Trans. Visualization and Computer Graphics*, vol. 12, no. 5, pp. 1141–1148, 2006.

[12] M. Agrawala, D. Zorin, and T. Munzner, "Artistic multi-projection rendering," in *Eurographics Rendering Workshop 2000*, 2000, pp. 125–136.

[13] K. Singh, "A fresh perspective," in *Proc. Graphics Interface 2002*, 2002, pp. 17–24.

[14] P. Coleman, K. Singh, L. Barrett, N. Sudarsanam, and C. Grimm, "3D scene-space widgets for non-linear projection," in *Proc. GRAPHITE 2005*, 2005, pp. 221–228.

[15] M. A. Hagen and R. Glick, "Pictorial perspective: Perception of size, linear, and texture perspective in children and adults," *Perception*, vol. 6, pp. 675–684, 1977.

[16] K. Mashio, K. Yoshida, S. Takahashi, and M. Okada, "Automatic blending of multiple perspective views for aesthetic composition," in *Proc. Smart Graphics 2010*, ser. LNCS, vol. 6133, no. 220–231. Springer, 2010.

[17] H. Ledoux and C. Gold, "An efficient natural neighbour interpolation algorithm for geoscientific modelling," in *Proc. 11th International Symposium on Spatial Data Handling*, 2004, pp. 97–108.