3D computational modeling and perceptual analysis of kinetic depth effects

Meng-Yao Cui[1], **Shao-Ping Lu**[1] (✉)**, Miao Wang**[2], **Yong-Liang Yang**[3], **Yu-Kun Lai**[4], **and Paul L. Rosin**[4]

Abstract Humans have the ability to perceive *kinetic depth effects*, i.e., to perceived 3D shapes from 2D projections of rotating 3D objects. This process is based on a variety of visual cues such as lighting and shading effects. However, when such cues are weak or missing, perception can become faulty, as demonstrated by the famous silhouette illusion example of the *spinning dancer*. Inspired by this, we establish objective and subjective evaluation models of rotated 3D objects by taking their projected 2D images as input. We investigate five different cues: ambient luminance, shading, rotation speed, perspective, and color difference between the objects and background. In the objective evaluation model, we first apply 3D reconstruction algorithms to obtain an objective reconstruction quality metric, and then use quadratic stepwise regression analysis to determine weights of depth cues to represent the reconstruction quality. In the subjective evaluation model, we use a comprehensive user study to reveal correlations with reaction time and accuracy, rotation speed, and perspective. The two evaluation models are generally consistent, and potentially of benefit to inter-disciplinary research into visual perception and 3D reconstruction.

Keywords kinetic depth effects; 3D reconstruction; perceptual factor analysis

1 TKLNDST, CS, Nankai University, Tianjin, China. E-mail: M.-Y. Cui, im-clizmy@qq.com; S.-P. Lu, slu@nankai.edu.cn (✉).

2 State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China.

3 Department of Computer Science, University of Bath, UK.

4 School of Computer Science and Informatics, Cardiff University, Wales, UK.

1 Introduction

Mechanisms of human perception of the 3D world have long been studied. In the early 17th century, artists developed a whole system of stimuli of monocular depth perception based especially on shading and transparency [1]. Loss of stimuli related to depth perception leads to a variety of visual illusions, such as the Pulfrich effect [2]. Here, with a dark filter on the right eye, dots moving to the right seem to be closer to participants than dots moving to the left, even though all the dots are actually at the same distance. This is caused by slower human perception of darker objects.

When a 3D object is rotating around a fixed axis, humans are capable of perceiving the shape of the object from its 2D projections. This is called the *kinetic depth effect* [4]. However, when there is no light above the object, humans can only perceive partial 3D information from the varying silhouette of the kinetic object over time, which easily leads to ambiguous understanding of the 3D object. One typical example of this phenomenon is the *spinning dancer* [3, 5] (see Fig. 1 for sample frames). The
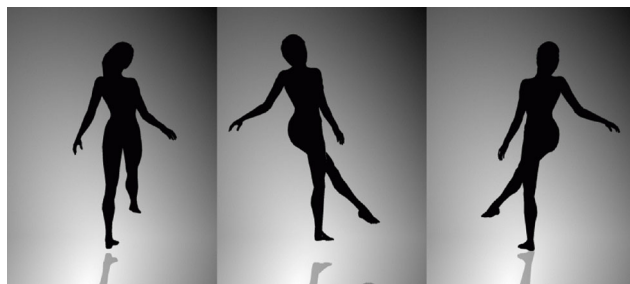
Fig. 1 3 of 34 frames from the original animation of the spinning dancer [3], courtesy of Nobuyuki Kayahara. Due to a lack of visual cues, humans are confused as to whether the dancer is rotating clockwise or counterclockwise.

dancer is observed to be spinning clockwise by some viewers and counterclockwise by others. Such ambiguity implies that more cues are needed for humans to make accurate depth judgements for 3D objects. Visual cues such as occlusion [6], frame timing [7], and speed and axis of rotation [8] have been widely studied by researchers. In addition, perspective effects also affect the accuracy of direction judgements [9].

In this paper, we make in-depth investigations into how visual cues influence the perception of kinetic depth effects, using both objective computational modeling and subjective perceptual analysis. We formulate and quantify visual cues from both 3D objects and their surrounding environment. On one hand, we make a comprehensive subjective evaluation to correlate subjective depth judgement for a 3D object with its visual conditions. On the other hand, as depth perception largely depends on the quality of mental shape reconstruction, we also propose an objective evaluation method based on 3D computational modeling [10]. This allows us to quantify the impacts of the involved visual cues. The impact factors are determined by solving a multivariate quadratic regression problem. Finally, we analyze the interrelations between the proposed subjective and objective evaluation models, and consider the consistency of impacts of visual cues on these models.

In summary, our work makes the following major contributions:

- a novel objective evaluation of kinetic depth effects based on multi-view stereo reconstruction,
- a novel subjective evaluation of kinetic depth effects based on a carefully designed user study, and
- a detailed analysis of how visual cues affect depth perception based on these subjective and objective evaluations.

2   Related work

Our work focuses on objective computational modeling and subjective analysis of 3D perception of kinetic depth effects under different visual conditions. We first discuss related work on visual perception using psychological and computational approaches, and then briefly describe the relevant reconstruction techniques employed in this work.

2.1   Psychological research on shape perception

For monocular vision, shading effects contain rich information [1]. Compared with diffuse shading, specular shading helps more to reduce underestimates of cylinder depth by subjects [11]. However, the shading effect can be ambiguous in some cases. For example, when the illumination direction is unknown, it is hard to disambiguate shape convexities and concavities; humans tend to assume that illumination comes from above [12]. Besides, when the level of overall illumination is low, effect of shadows is generally assumed to be determined by the overall illumination [13].

Motion information also benefits shape perception. The inherent ambiguity of depth order in projected images of 3D objects can be resolved by dynamic occlusion [14]. Perspective also gives rich information about 3D objects during this process [15]. The human visual system can infer 3D shapes from 2D projections of rotated objects [4], interpolating the intervening smooth motion from pairs of images of rotated objects [16].

Color information is very important not only in immersive scene representation [17–20] but also in psychological depth perception. Isono and Yasuda [21] found that chromatic channels can contribute to depth perception using a prototype flicker-free fieldsequential stereoscopic television system. Guibal and Dresp [22] realized that color effects are largely influenced by luminance contrast and stimulus geometry. When shape stimuli are not strong, color can give an illusion of closeness [23].

2.2   Computational visual perception

Computational visual perception has been extensively studied in the computer graphics community. Here we briefly describe the most relevant work on perception-based 2D image processing and 3D modeling.

In terms of 2D images, Chu et al. [24] presented a computational framework to synthesize camouflage images that can hide one or more temporally unnoticed figures in the primary image. Tong et al. [25] proposed a hidden image framework that can embed secondary objects within a primary image as a form of artistic expression. The edges of the object to be hidden are firstly detected, and then an image blending based optimization is applied to perform image transformation as well as object embedding. The study of kinetic depth effects often

uses subjective response [26], and some researchers also use the judgement of the rotation direction as the response [9].

Like for image-based content embedding and hiding, 3D objects can be embedded into 2D images [27]; the objects can be easily detected by humans, but not by an automatic method. Researchers have also generated various mosaic effects on both images [28] and 3D surfaces [29]. A computational model for the psychological phenomenon of change blindness is investigated in Ref. [30]. As change blindness is caused by failing to store visual information in short-term memory, the authors model the influence of long-range contextual complexity, allowing them to synthesize images with a given degree of blindness. Illusory motion is also studied as self-animating images in Ref. [31]. In order to computationally model the human motion perception of a static image, repeated asymmetric patterns are optimally generated on streamlines of a specified vector field. Tong et al. [32] created self-moving 3D objects using the hollow-face illusion from input character animation, where the surface's gradient is manipulated to fit the motion illusion. There is also research into rendering, designing, and navigating impossible 3D models [33–35]. In contrast to investigating those seemingly impossible models, our work focuses on evaluating the 3D perception of rotated objects.

2.3 Multi-view stereo reconstruction

Multi-view 3D reconstruction and 3D point cloud registration are fundamental in computer graphics and computer vision. Comprehensive surveys on these topics can be found in Refs. [36, 37]. The well-known structure-from-motion [10] can effectively recover camera poses and further generate a sparse 3D point cloud by making use of multiple images of a scene or objects. Moreover, multi-view stereo algorithms [38] can reconstruct a fully textured surface of the scene. We employ such computational techniques to evaluate 3D reconstruction quality under various environmental conditions.

3 Overview

3.1 Approach

Our goal is to evaluate the influence of various visual conditions on kinetic depth effects, including the ambient luminance, shading, perspective, rotation speed, and color difference between the object and background.

For both the human visual system and image-based 3D reconstruction techniques, the input visual information usually takes the form of projected 2D images. Thus, by using a set of projected 2D images of the 3D objects under the chosen conditions, we investigate the shape perceived by human participants and the shape produced by multi-view stereo 3D reconstruction. As well as measuring the perception of kinetic depth effects using our objective and subjective evaluation models, we also investigate the correlations between these two different methods. Our approach is outlined in Fig. 2.
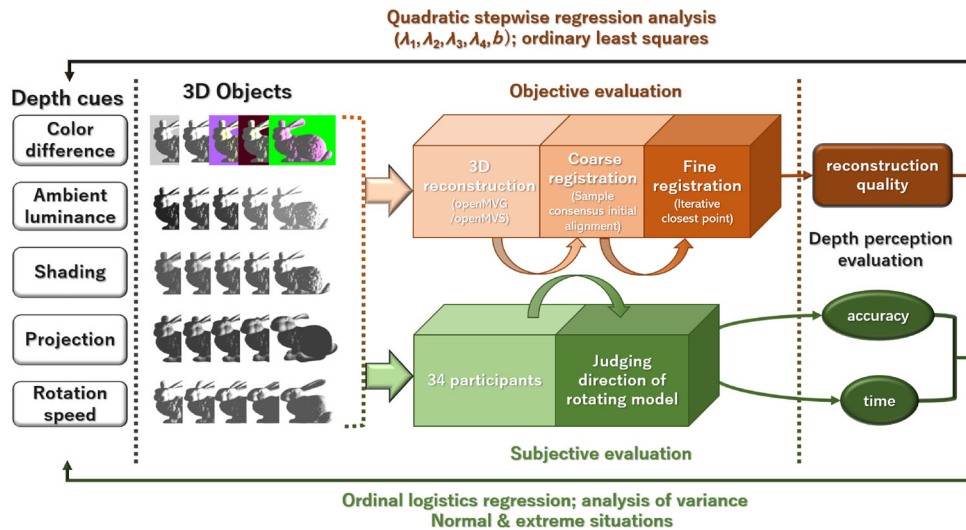
Fig. 2 Our approach. We project the input 3D objects onto 2D image planes under specified conditions (lighting, projection mode, rotation speed, etc.), which are fed into constructed objective and subjective evaluation models. Analysis reveals interesting correlations between depth perception of rotated 3D objects and the visual conditions.

3.2 Dataset

Each 3D object is rotated around a fixed vertical axis passing through the geometric center of the object, and we sample projected 2D images at an angular interval of θ. As the frame rate when displaying projected images is fixed, changing the sampling angle interval causes a changing rotation speed of the object. We can obtain such datasets under different visual conditions as the images are explicitly rendered. Specifically, we manipulate the ambient luminance by adjusting ambient lights, and control shading by changing diffuse lights. We control perspective by selecting either orthogonal or perspective projection mode, which affects perception. We also control the color difference between the object and the background; predefined color pairs are used to generate the colors of the background and the 3D object (see Section 4.1). Table 1 summarizes the parameters used. Using such a dataset generated under controlled conditions, we can then assess the 3D perception of the rotated object using the following two evaluation models.

3.3 Objective evaluation model

This model utilizes the reconstruction quality of the input 3D object as the basis for evaluation. First, using the projected 2D images of the 3D object under specified visual conditions, we reconstruct a point cloud using multi-view stereo reconstruction algorithms. Then, we develop a method to measure the quality of reconstruction of the original 3D object by the point cloud (see Section 4.2). Finally, we analyze the effects of different visual conditions in detail (see Section 4.3).

3.4 Subjective evaluation model

Directly measuring 3D reconstruction in the brains of human subjects is difficult. Based on the observation that if humans successfully mentally reconstruct a rotated object, it is easy for them to tell the direction of rotation; in our study, the time and accuracy of direction judgements are used as proxies

to measure the quality of depth perception. We first display rotating objects with the same set of projected images as used for 3D reconstruction in the objective evaluation, and ask participants to judge the rotation direction of the object. Then we consider extreme situations in which image sequences could not be reconstructed well, including overexposure, low lighting levels, and overly fast rotation. We analyze the results in terms of the accuracy and the reaction time of direction judgements (see Section 5.5).

4 Objective evaluation

Our objective evaluation process includes four steps: generating 2D images of 3D objects under various conditions, reconstructing 3D shapes of objects based on the generated images, quantifying the reconstruction quality of the 3D objects, and fitting weighting factors for depth cues by multivariate quadratic regression optimization.

4.1 Parameter selection and image set generation

In order to generate images of 3D objects under various controlled conditions, we need to select the parameter values to determine the depth cues. Firstly, we normalize the size of each 3D object to have a unit bounding box centered at the origin. Then, we import the object into a virtual scene, display it under orthogonal projection, and set a fixed-point light. The line between the light and the geometric center of the object is perpendicular to the rotation axis, and the distance between the light and the geometric center of the object is ten units. Since in openMVG the focal length is a given parameter, considering perspective projection mode in our objective evaluation is not that meaningful, and in this study we restrict ourselves to use of orthogonal projection.

We control the brightness of diffuse and ambient lights as follows. We set the HSL value of the diffuse light to be $(0, 0, \alpha)$, with seven set values for α, corresponding to different luminance levels. We set the HSL value of the ambient light to $(0, 0, \beta)$ with six set values for β (see Table 2).

Table 1 Control parameters for generating 2D projected images

θ	Angular interval of 2D projection
α	Lightness in HSL color space $(0, 0, \alpha)$, used as diffuse light intensity
β	Lightness in HSL color space $(0, 0, \beta)$, used as global ambient light intensity
D	Color difference between object and background

Table 2 Values of α and β used to control diffuse and ambient lighting when generating image sets

α	0.5	0.8	1.1	1.4	1.7	2.0	2.3
β	0.0	0.5	1.0	1.5	2.0	2.5	

We sample projected 2D images as the 3D object is rotated. We use four different sampling intervals $\theta = 0.209, 0.157, 0.126,$ and 0.105. For simplicity, we use five fixed pairs of RGB values for the 3D object and the background (see Table 3).

We calculate the difference of the chosen color pairs using the following equation:

$$D(C_{\mathrm{B}}, C_{\mathrm{O}}) = \sqrt{w_{\mathrm{r}}(r_{\mathrm{B}} - r_{\mathrm{O}})^2 + w_{\mathrm{g}}(g_{\mathrm{B}} - g_{\mathrm{O}})^2 + w_{\mathrm{b}}(b_{\mathrm{B}} - b_{\mathrm{O}})^2} \tag{1}$$

Here, $C_{\mathrm{O}} = (r_{\mathrm{O}}, g_{\mathrm{O}}, b_{\mathrm{O}})$ is the RGB color of the object, $C_{\mathrm{B}} = (r_{\mathrm{B}}, g_{\mathrm{B}}, b_{\mathrm{B}})$ is the color of the background, and $w_{\mathrm{r}}, w_{\mathrm{g}}, w_{\mathrm{b}}$ are weighting factors, empirically set to $(3, 4, 2)$. We generated image sets for 15 different 3D objects under various conditions,

Table 3 Object and background colors and corresponding color differences used when generating image sets

| Object color	Background color	Color difference
(0.8, 0.8, 0.8)	(0.8, 0.8, 0.8)	0.000
(0.8, 0.8, 0.8)	(1.0, 1.0, 1.0)	0.600
(1.0, 0.9, 0.5)	(0.7, 0.4, 1.0)	1.200
(0.8, 1.0, 0.6)	(0.3, 0.0, 0.1)	2.291
(1.0, 0.4, 1.0)	(0.0, 1.0, 0.0)	2.538

and selected three objects with a high reconstruction success rate (30%). Finally, for each test object, we generated a separate image set for each of 7 (shading) × 6 (ambient luminance) × 4 (rotation speed) × 5 (color difference) conditions, at an image resolution of 800×600 pixels. Some examples are shown in Fig. 3.

4.2 3D reconstruction and quality assessment

We employ openMVG [39] and openMVS [38] to process image sequences, and take the reconstructed point clouds as input. We normalize the size of each point cloud to the same bounding box as for normalizing 3D objects. Then we match each reconstructed point cloud to the original object. Specifically, we use the sample consensus initial alignment (SAC-IA) method [40] to provide an initial alignment, and the iterative closest point (ICP) to refine the alignment [41]. Finally, we compute the Euclidean fitness score μ between the reconstructed point cloud and the original object.

4.3 Objective evaluation results

We generated 2520 image sets for 3D reconstruction, 929 of which provided point clouds, while for 1591, reconstruction failed. We use the following measure
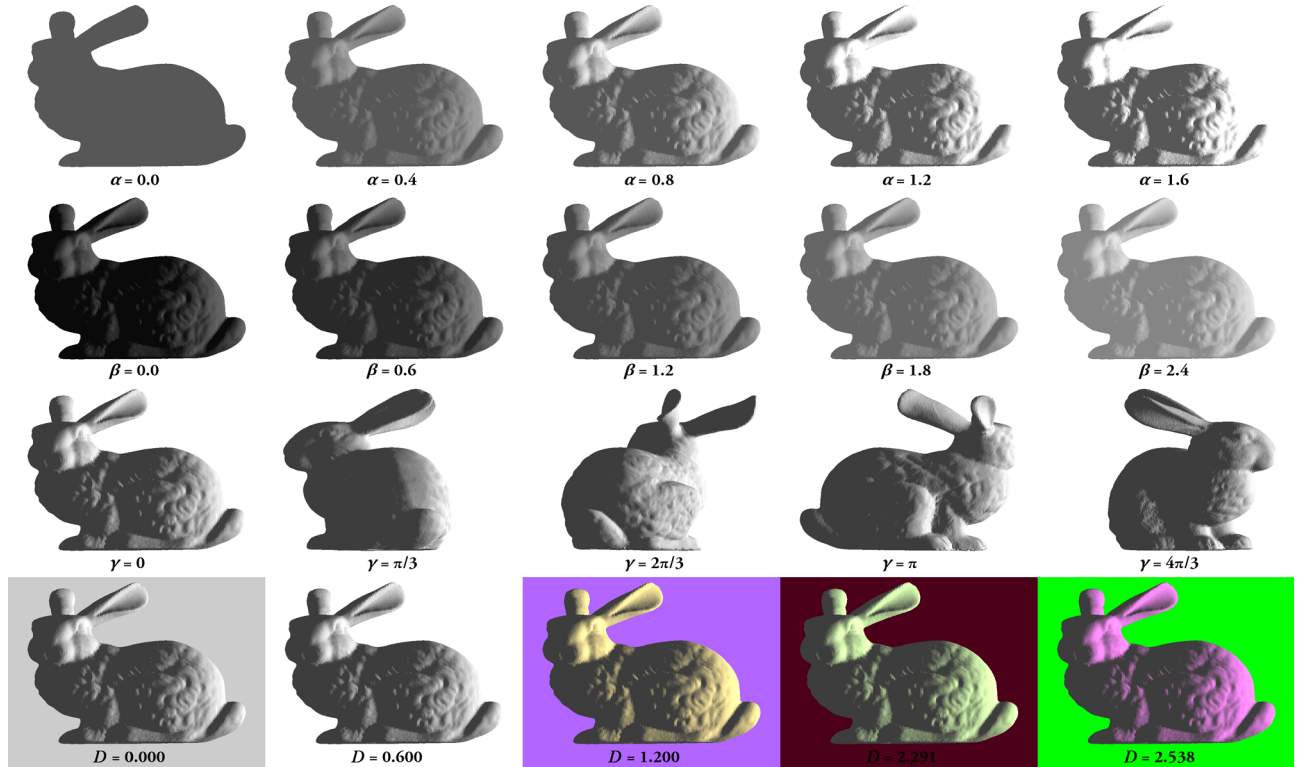
Fig. 3 Example projected 2D images. First row: variations in diffuse light. Second row: variations in ambient light. Third row: images at an angular interval of $\theta = \pi/3$; γ is rotation from the initial orientation. Last row: various color differences.
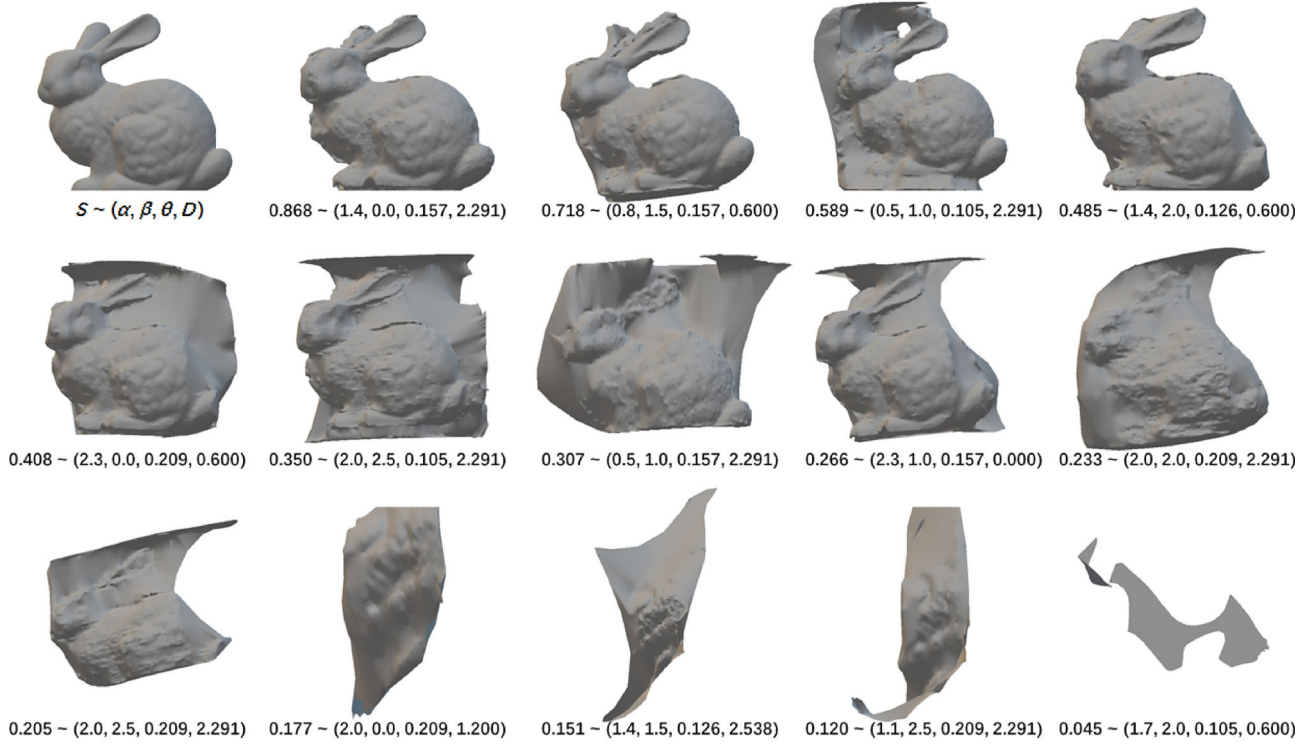
Fig. 4 Given a 3D object (top-left) and specified visual conditions, we generate corresponding projected 2D images, and reconstruct 3D shapes (other views) using multi-view stereo algorithms. We quantitatively measure the reconstruction quality for shape perception analysis. Each view above gives the reconstruction quality and the corresponding rendering parameters.

s of reconstruction quality, based on the point cloud distance μ between a reconstructed point cloud and the corresponding original point cloud:

$$s = -\lg(\mu) \qquad (2)$$

Logarithmic processing is used to make the residuals of our model normally distributed. Reconstruction quality values are linearly normalized to the range $[0, 1]$. Given a set of reconstruction quality samples $S = \{s_1, \ldots, s_n\}$, we formulate a factor analysis model with the following quadratic stepwise regression:

$$\lambda^* = \underset{\lambda}{\operatorname{argmin}}(S - (\lambda_1 \alpha + \lambda_2 \theta + \lambda_3 \alpha^2 + \lambda_4 \alpha\beta + b)) \quad (3)$$

where $\lambda = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ are weighting coefficients to balance the impact of the various control parameters, and b is a constant. We determined the coefficients in the model using the standard least squares method. Results are given in Table 4.

It can be seen that the model accounts for 10.3% of the variation in reconstruction quality. Since the reconstruction algorithm used here is not always stable, the explanatory power of the model is limited. The impact of each individual visual cues is now analyzed in turn.

Table 4 Objective evaluation model results. p represents the confidence probability of the parameter based on the standard Student's t-test

Coefficient	Value	Std. Err.
b	0.3678*	0.041
λ_1	0.3593*	0.056
λ_2	−0.6361*	0.147
λ_3	−0.1234*	0.019
λ_4	−0.0278*	0.004
Observations	929	
R^2	0.103	

* $p < 0.01$

4.4 Shading

Shading and reconstruction quality follow a quadratic function relationship (see Fig. 5), with $\lambda_1 = 0.3593$ ($p<0.01$), $\lambda_3 = -0.1234$ ($p<0.01$).

4.5 Ambient luminance

The (ambient luminance × shading) interaction is significant with $\lambda_4 = -0.0278$ ($p<0.01$). High (ambient luminance × shading) levels contribute to poor quality reconstruction. As shown in Fig. 6, the fitted lines at different ambient light levels are not parallel.
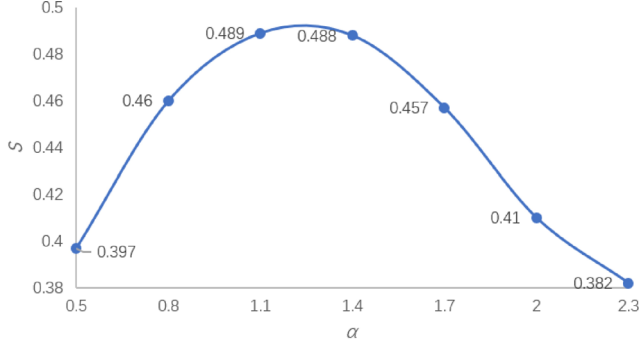
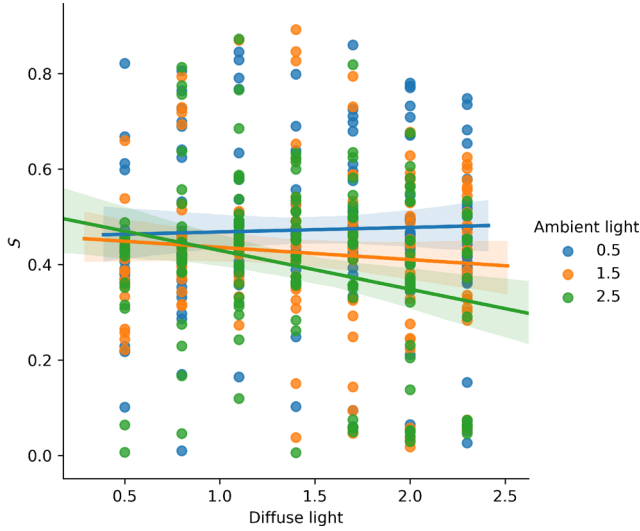**Fig. 5**   Correlation between 3D reconstruction quality and shading.

Fig. 6   Significant (ambient luminance × shading) interaction.

4.6   Rotation speed

High rotation speed significantly contributes to poor reconstruction quality ($\lambda_2 = -0.6361$, $p < 0.01$). When $\theta = 0.105$, the mean value of S is 0.489; when $\theta = 0.209$, the mean value of S is 0.400.

4.7   Color difference

Color difference does not significantly affect the reconstruction quality.

5   Subjective evaluation

As noted, humans can recover rotated 3D objects from their 2D projections. The rotation direction of 3D objects can be an important clue to judging the quality of their mental shape reconstruction. Based on this, the following multi-factor experiment was designed.

5.1   Participants

We recruited 35 participants; 34 participants (19 males and 15 females) successfully finished the test.

5.2   Procedure and materials

In the experiment, a set of images was continuously displayed in full screen mode. The experiment was conducted on a laptop with an Intel i5 8250U CPU and 8 GB memory. We designed two types of study as follows.

5.3   Study A

Here we explored the depth cue effect in general situations. The range of cues was the same as for the objective evaluation model, but we chose fewer values for each cue (see Table 5) to ensure participants could concentrate during the study. The projection could be either orthogonal or perspective. Overall we considered 144 conditions consisting of 3 (*shading*) × 3 (*ambient luminance*) × 2 (*rotation speed*) × 4 (*color difference*) × 2 (*projection mode*). For each set of conditions, we displayed three different objects.

5.4   Study B

Here we considered more extreme situations, including low lighting levels, overexposure, and high speed rotation, where we varied each condition while keeping other cues fixed (see Table 6). The variables used to represent each situation are shown in Table 7. Test image sets generated are illustrated in Figs. 7 and 8. To simplify the problem, we only considered orthogonal projection situations.

Every participant was asked to judge rotation direction for each image set generated in Studies A and B; each image set was judged only once. The display order of each image set was random, as was the rotation direction of each 3D object.

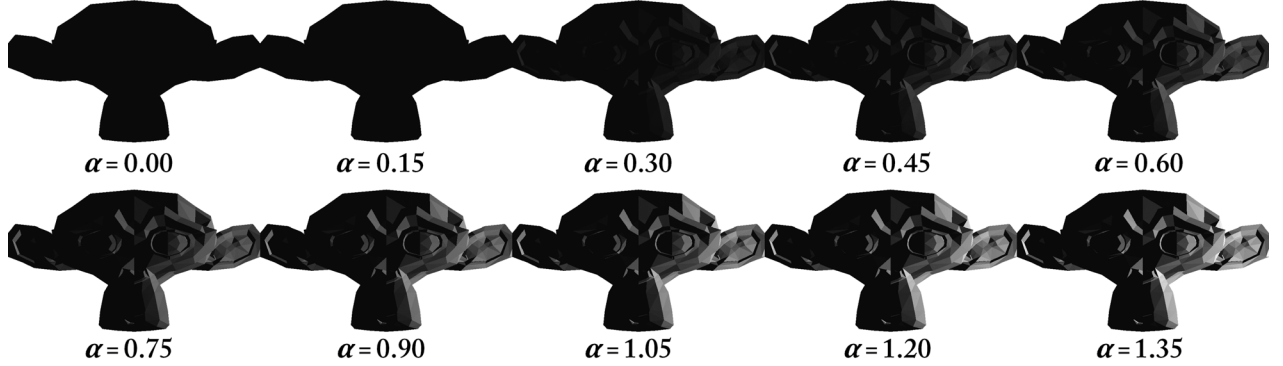**Table 5**   Parameter values used in subjective evaluation model Study A

α	0.5	1.4	2.3	
β	0.0	1.0	2.0	
θ	0.105	0.209		
D	0.000	0.600	1.200	2.291

Table 6   Parameter values used in subjective evaluation model Study B (for extreme conditions). Top to bottom: low lighting (β zero), overexposed (high value of α), fast rotation (normal values of α, β)

α	β	θ	Colors for object & background
varying	0.0	0.157	(0.8, 0.8, 0.8) , (1.0, 1.0, 1.0)
2.3	varying	0.157	(0.8, 0.8, 0.8) , (1.0, 1.0, 1.0)
1.7	1.5	varying	(0.8, 0.8, 0.8) , (1.0, 1.0, 1.0)

Table 7 Parameter values used in the subjective evaluation model Study B (for extreme situations). Top to bottom: low lighting levels, overexposure, and high speed rotation

α	0.00	0.05	0.10	0.15	0.20	0.30	0.35	0.40	0.45	
β	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5
θ	1.047	0.785	0.628	0.524	0.449	0.393	0.349	0.314	0.286	0.262

$\alpha = 0.00$ $\alpha = 0.15$ $\alpha = 0.30$ $\alpha = 0.45$ $\alpha = 0.60$

$\alpha = 0.75$ $\alpha = 0.90$ $\alpha = 1.05$ $\alpha = 1.20$ $\alpha = 1.35$

Fig. 7 2D images under each different low lighting conditions.

$\beta = 2.7$ $\beta = 2.9$ $\beta = 3.1$ $\beta = 3.3$ $\beta = 3.5$

$\beta = 3.7$ $\beta = 3.9$ $\beta = 4.1$ $\beta = 4.3$ $\beta = 4.5$
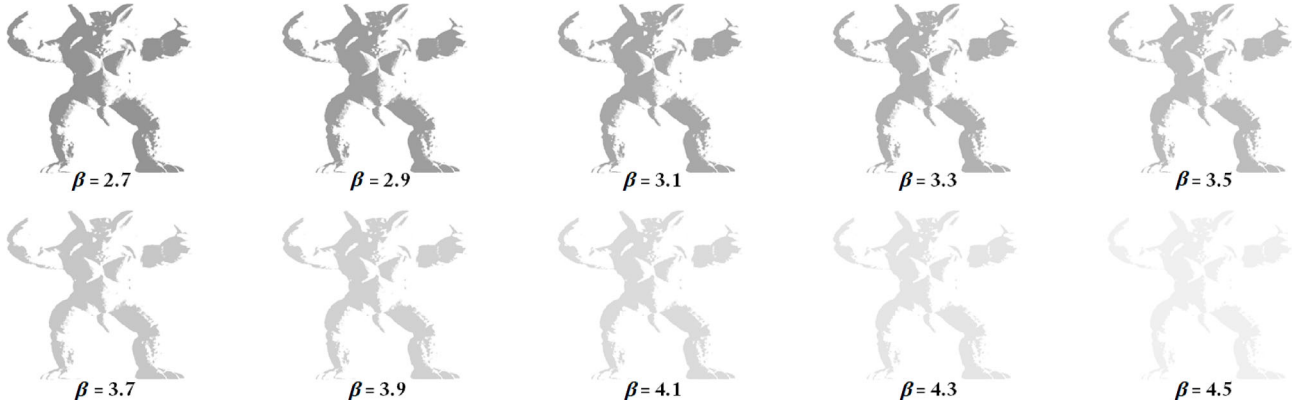
Fig. 8 2D images under each individual overexposed condition.

To exclude *viewing-from-above* bias [1], we defined rotation direction as *Left* and *Right*. From the participants point of view, the rotation direction is right if the closer part of a 3D object is moving to the right, and vice versa. The images were displayed at 24 FPS. The maximum display time for each image set did not exceed 5 s. The participants were given time to practise before the formal experiment. The entire experiment took about 15–20 min.

5.5 Subjective evaluation results

We recorded the judgements and reaction time of all participants. We rank all reaction time in ascending order and calculated the standard scores (denoted τ), which correspond to the estimated cumulative proportion of the reaction time. We use the repeated measure analysis of variance (ANOVA) method to determine the effect of cues on τ under different conditions. We calculated each participant's accuracy of judgement under each condition. Since three objects were tested in each condition, the participant's judgement accuracy has four values, and does not follow a normal distribution. Therefore, we used ordinal logistic regression models to test the effect of cues on the participant's judgement accuracy. In particular, we chose the complementary log–log link function, since the participant's judgement accuracy mostly lay in 0.67–1.00 [42]:

$$\Phi(x) = \log(-\log(1-x)) \qquad (4)$$

We establish ordinal logistic regression models for all situations, but only show those models with significant results and skip the remainder.

5.5.1 Analysis of Study A

A five-way ANOVA method revealed the main effect of rotation speed ($F(1, 5028) = 38.11$, $p<0.01$) on τ.

Participants react faster under high rotation speed conditions ($M = -0.08$, $SD = 0.97$) than at low rotation speed ($M = 0.05$, $SD = 1.00$). We also found a significant (perspective $\times$ rotation speed) interaction ($F(1, 5028) = 6.19$, $p<0.05$) on τ. At high rotation speed, τ is significantly lower with perspective projection ($M = -0.109$, $SD = 0.95$) than with orthogonal projection ($M = -0.06$, $SD = 1.00$). Because rotation speed and perspective only have two levels, there is no need for Mauchly's test of sphericity. Other than the above phenomena, we found no significant effect on the participant's judgement accuracy for the five cues considered in our experiments.

5.5.2 Analysis of Study B

Low lighting levels. A one-way ANOVA reveals the main effect of shading on τ ($F(9, 340) = 2.668$, $p<0.01$), and Mauchly's test of sphericity is not significant ($p = 0.579$). Here higher α leads to lower reaction time. We established an ordinal logistic regression model as follows:

$$\begin{cases} \Phi(p_1) = \epsilon_1 - \eta * \alpha \\ \Phi(p_1 + p_2) = \epsilon_2 - \eta * \alpha \\ \Phi(p_1 + p_2 + p_3) = \epsilon_3 - \eta * \alpha \\ p_1 + p_2 + p_3 + p_4 = 1 \end{cases} \quad (5)$$

where $p = \{p_1, p_2, p_3, p_4\}$ are the probabilities of each value of the participant's judgement accuracy (from low to high), η is the weighting coefficient, and $\epsilon = \{\epsilon_1, \epsilon_2, \epsilon_3\}$ are constant values. For each case, the value with highest probability is the predicted value of the participant's judgement accuracy.

We fitted the coefficients in the model with results shown in Table 8. High α significantly contributes to high judgement accuracy ($\eta = 4.229$, $p<0.01$). Hence using strong shading in low lighting levels conditions improves accuracy and accelerates reactions (see Fig. 9).

Table 8 Results of ordinal logistic regression under low lighting levels. p represents the confidence probability of the parameter based on the Wald test

Coefficient	Value	Std. Err.
ϵ_1	-3.682^*	0.515
ϵ_2	-2.039^*	0.257
ϵ_3	-0.578^*	0.177
η	4.229^*	0.872
Observations	348	
Nagelkerke's R^2	0.093	

* $p < 0.01$
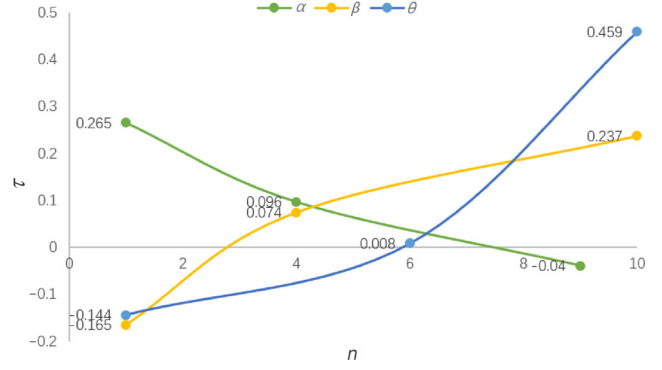
Fig. 9 Variation of mean values of τ with shading, ambient luminance, and speed. Marked points represent significant differences under pairwise comparison.

Overexposure. A one-way ANOVA reveals the main effect of ambient luminance) ($F(9, 340) = 2.661$, $p<0.01$) on τ, and Mauchly's test of sphericity is not significant ($p = 0.350$). This means that the covariance matrix assumption is met, and the result of repeated measures ANOVA is robust. Participants react faster when $\beta = 2.7$ ($M = -0.17$, $SD = 1.07$) than when $\beta = 4.5$ ($M = 0.24$, $SD = 0.94$), which implies that the higher ambient luminance in overexposed conditions delays reactions (see Fig. 9). We found no significant effect of ambient luminance on judgement accuracy.

High speed rotation. Rotation speed has a significant effect on τ ($F(9, 340) = 7.627$, $p<0.01$), and Mauchly's test of sphericity is not significant ($p = 0.162$). In high rotation speed conditions, lower rotation speed leads to faster reactions. We established an ordinal logistic regression model again as follows:

$$\begin{cases} \Phi(p_1) = \varepsilon_1 - \kappa * \theta \\ \Phi(p_1 + p_2) = \varepsilon_2 - \kappa * \theta \\ \Phi(p_1 + p_2 + p_3) = \varepsilon_3 - \kappa * \theta \\ p_1 + p_2 + p_3 + p_4 = 1 \end{cases} \quad (6)$$

where $p = \{p_1, p_2, p_3, p_4\}$ are the probabilities of each value of the participant's judgement accuracy (from low to high), κ is a weighting coefficient, and $\varepsilon = \{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$ are constant values. For each case, the value with highest probability is the predicted value of judgement accuracy.

We fitted the coefficients in the model with results shown in Table 9. High θ significantly contributes to low judgement accuracy ($\kappa = -1.353$, $p<0.01$). In high speed rotation conditions, increasing the rotation speed reduces judgement accuracy and delays reactions (see Fig. 9).

Table 9   Results of ordinal logistic regression at high speed. p represents the confidence probability of the parameter based on the Wald test

Coefficient	Value	Std. Err
ε_1	-5.191^*	0.581
ε_2	-4.266^*	0.432
ε_3	-2.480^*	0.320
κ	-1.353^*	0.499
Observations	346	
Nagelkerke's R^2	0.030	

* $p < 0.01$

6   Joint objective and subjective analysis

Based on objective computational modeling and subjective perceptual evaluations, we next performed a joint analysis on the 3D perception of rotated objects.

6.1   Shading

For both objective and subjective evaluations, shading has a significant effect on depth perception. In the objective evaluation, shading and reconstruction quality are correlated by a quadratic function. As shading increases, the reconstruction quality first improves and then declines. This coincides with subjective evaluation as under low lighting levels, greater shading improves judgement accuracy and accelerates the observer's reactions.

6.2   Ambient luminance

The depth cue from ambient luminance is also effective in both objective and subjective evaluations. In objective evaluation, the interaction of shading and ambient luminance is significant. High (shading × ambient luminance) levels contribute to poor reconstruction. In the subjective evaluation, high ambient luminance in overexposed cases can increase observers' reaction time.

6.3   Rotation speed

Rotation speed plays an important role in both objective and subjective evaluations. In the objective evaluation, increasing rotation speed decreases reconstruction quality, which coincides with the result of subjective evaluation that, in high speed conditions, higher rotation speed decreases judgement accuracy.

However, in the subjective evaluation, increasing the rotation speed accelerates users' reaction time. A possible reason is that, with higher rotation speeds, participants receive more information within the same time interval, stimulating the participants to make a decision faster. In our experiments under general situations, this acceleration is stronger than the delay caused by uncertainty.

6.4   Perspective

In the subjective evaluation, (perspective × rotation speed) interaction is significant. Participants react faster under perspective projection conditions than orthogonal.

6.5   Color difference

We found no significant effects caused by color difference between objects and background in either objective evaluation or subjective evaluation models. As future work, we will test more color combinations to further explore possible effects of color differences.

7   Discussion

We have analyzed the effects of different depth cues on 3D perception of rotated 3D objects, broadening the scope of previous studies. We also designed an objective evaluation and a subjective evaluation to make a thorough analysis.

However, there are also some shortcomings in our design. In our objective evaluation, when the depth cues in images were extremely weak, 3D reconstruction based on structure-from-motion would be unstable caused by unexpected feature matching. This common challenge limits the space of our analysis model ($R^2 = 10.3\%$). Moreover, the subjective evaluation only uses judgement of the direction of rotated objects as the response. In future, we could use more 3D information. In our experiments, reconstruction quality is closely related to the kinds of 3D objects. This specific type of influence on shape perception could also be further analysed.

The analysis of the effect of depth cues guides us how good reconstruction results can be achieved both for humans and computers, such as rendering under certain lighting conditions. The objective evaluation also reveals the limitations of existing algorithms.

Our approach could benefit from more accurate depth prediction and 3D reconstruction in various challenging environments, which could potentially be provided by recent deep learning-based techniques, such as CNN-SLAM [43] and deep stereo matching [44].

8 Conclusions and future work

We have proposed two approaches to measuring the quality of depth perception of kinetic depth effects, with a detailed analysis of how visual cues affect depth perception. Firstly, we generated a dataset of images of rotating objects considering five depth cues: ambient luminance, shading, rotation speed, and color difference between objects and background. In the objective evaluation, we applied 3D reconstruction and measured reconstruction quality via distances between reconstructed and original objects. In the subjective evaluation, we invited participants to judge the rotation direction of 3D objects by showing them projected 2D images, and inferred perception quality by their reaction time and accuracy. In our experiments, we found both strong and dim lighting significantly undermined the perception of depth. High ambient illumination × shading level, rotation speed, and orthogonal projection can also reduce depth perception quality. Yet it is also interesting that color difference does not have a significant effect on depth perception in our experiments. In future, we will take more depth cues into consideration and develop a more precise quantitative model for more complex situations. Using our new observations to guide other 3D computational modeling would also be an interesting avenue of future work. We hope our study will inspire more inter-disciplinary research into robust 3D reconstruction and human visual perception.

References

[1] Pashler, H.; Yantis S. *Stevens' Handbook of Experimental Psychology, Volume 1: Sensation and Perception*, 3rd edn. John Wiley & Sons, Inc., 2002.

[2] Bach, M. The Pulfrich effect. 1922. Available at `https://pulfrich.siu.edu/Pulfrich{\_}Pages/lit{\_}pulf/1922{\_}Pulfrich.htm`.

[3] Kayahara, N. Spinning dancer. Available at `https://en.wikipedia.org/wiki/Spinning{\_}Dancer`.

[4] Wallach, H.; O'Connell, D. N. The kinetic depth effect. *Journal of Experimental Psychology* Vol. 45, No. 4, 205–217, 1953.

[5] Troje, N. F.; Mcadam, M. The viewing-from-above bias and the silhouette illusion. *i-Perception* Vol. 1, No. 3, 143–148, 2010.

[6] Andersen, G. J.; Braunstein, M. L. Dynamic occlusion in the perception of rotation in depth. *Perception & Psychophysics* Vol. 34, No. 4, 356–362, 1983.

[7] Timothy Petersik, J. The effects of spatial and temporal factors on the perception of stroboscopic rotation simulations. *Perception* Vol. 9, No. 3, 271–283, 1980.

[8] Green, B. F. Figure coherence in the kinetic depth effect. *Journal of Experimental Psychology* Vol. 62, No. 3, 272–282, 1961.

[9] Braunstein, M. L. Perceived direction of rotation of simulated three-dimensional patterns. *Perception & Psychophysics* Vol. 21, No. 6, 553–557, 1977.

[10] Moulon, P.; Monasse, P.; Marlet, R. Global fusion of relative motions for robust, accurate and scalable structure from motion. In: Proceedings of the IEEE International Conference on Computer Vision, 3248–3255, 2013.

[11] Todd, J. T.; Mingolla, E. Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology Human Perception & Performance* Vol. 9, No. 4, 583–595, 1983.

[12] Howard, I. P.; Bergström, S. S.; Ohmi, M. Shape from shading in different frames of reference. *Perception* Vol. 19, No. 4, 523–530, 1990.

[13] Wisessing, P.; Zibrek, K.; Cunningham, D. W.; McDonnell, R. A psychophysical model to control the brightness and key-to-fill ratio in CG cartoon character lighting. In: Proceedings of the ACM Symposium on Applied Perception, Article No. 2, 2019.

[14] Braunstein, M. L.; Andersen, G. J.; Riefer, D. M. The use of occlusion to resolve ambiguity in parallel projections. *Perception & Psychophysics* Vol. 31, No. 3, 261–267, 1982.

[15] Gibsen, J. J. *The Ecological Approach to Visual Perception.* New York: Psychology Press, 1986.

[16] Young, L. R.; Oman, C. M. Model for vestibular adaptation to horizontal rotation. *Aerospace Medicine* Vol. 40, No. 10, 1076–1080, 1969.

[17] Lu, S. P.; Ceulemans, B.; Munteanu, A.; Schelkens, P. Spatio-temporally consistent color and structure optimization for multiview video color correction. *IEEE Transactions on Multimedia* Vol. 17, No. 5, 577–590, 2015.

[18] Lu, S. P.; Dauphin, G.; Lafruit, G.; Munteanu, A. Color retargeting: Interactive time-varying color image composition from time-lapse sequences. *Computational Visual Media* Vol. 1, No. 4, 321–330, 2015.

[19] Ceulemans, B.; Lu, S.-P.; Lafruit, G.; Schelkens, P.; Munteanu, A. Efficient MRF-based disocclusion inpainting in multiview video. In: Proceedings of the IEEE International Conference on Multimedia and Expo, 1–6, 2016.

[20] Ceulemans, B.; Lu, S. P.; Lafruit, G.; Munteanu, A. Robust multiview synthesis for wide-baseline camera arrays. *IEEE Transactions on Multimedia* Vol. 20, No. 9, 2235–2248, 2018.

[21] Isono, H.; Yasuda, M. Stereoscopic depth perception of isoluminant color random–dot stereograms. *Systems and Computers in Japan* Vol. 19, No. 9, 32–40, 1988.

[22] Guibal, C. R. C.; Dresp, B. Interaction of color and geometric cues in depth perception: When does "red" mean "near"? *Psychological Research Psychologische Forschung* Vol. 69, Nos. 1–2, 30–40, 2004.

[23] Pisanpeeti, A. P.; Dinet, E. Transparent objects: Influence of shape and color on depth perception. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, 1867–1871, 2017.

[24] Chu, H.-K.; Hsu, W.-H.; Mitra, N. J.; Cohen-Or, D.; Wong, T.-T.; Lee, T.-Y. Camouflage images. *ACM Transactions on Graphics* Vol. 29, No. 4, Article No. 51, 2010.

[25] Tong, Q.; Zhang, S.-H.; Hu, S.-M.; Martin, R. R. Hidden images. In: Proceedings of the NPAR, 27–34, 2011.

[26] Dosher, B. A.; Landy, M. S.; Sperling, G. Ratings of kinetic depth in multidot displays. *Journal of Experimental Psychology: Human Perception and Performance* Vol. 15, No. 4, 816–825, 1989.

[27] Mitra, N. J.; Chu, H.-K.; Lee, T.-Y.; Wolf, L.; Yeshurun, H.; Cohen-Or, D. Emerging images. *ACM Transactions on Graphics* Vol. 28, No. 5, Article No. 163, 2009.

[28] Xu, P. F.; Ding, J. Q.; Zhang, H.; Huang, H. Discernible image mosaic with edge-aware adaptive tiles. *Computational Visual Media* Vol. 5, No. 1, 45–58, 2019.

[29] Lai, Y. K.; Hu, S. M.; Martin, R. R. Surface mosaics. *The Visual Computer* Vol. 22, Nos. 9–11, 604–611, 2006.

[30] Ma, L.-Q.; Xu, K.; Wong, T.-T.; Jiang, B.-Y.; Hu, S.-M. Change blindness images. *IEEE Transactions on Visualization and Computer Graphics* Vol. 19, No. 11, 1808–1819, 2013.

[31] Chi, M. T.; Lee, T. Y., Qu, Y. G.; Wong, T. T. Self-animating images: Illusory motion using repeated asymmetric patterns. *ACM Transactions on Graphics* Vol. 27, No. 3, Article No. 62, 2008.

[32] Tong, J.; Liu, L.; Zhou, J.; Pan, Z. Mona Lisa Alive: Create self-moving objects using hollow-face illusion. *The Visual Computer* Vol. 29, Nos. 6–8, 535–544, 2013.

[33] Elber, G. Modeling (seemingly) impossible models. *Computers & Graphics* Vol. 35, No. 3, 632–638, 2011.

[34] Lai, C.-F. W.; Yeung, S.-K.; Yan, X.; Fu, C.-W.; Tang, C.-K. 3D navigation on impossible figures via dynamically reconfigurable maze. *IEEE Transactions on Visualization and Computer Graphics* Vol. 22, No. 10, 2275–2288, 2016.

[35] Wu, T.-P.; Fu, C.-W.; Yeung, S.-K.; Jia, J.; Tang, C.-K. Modeling and rendering of impossible figures. *ACM Transactions on Graphics* Vol. 29, No. 2, Article No. 13, 2010.

[36] Tam, G. K. L.; Cheng, Z.-Q.; Lai, Y.-K.; Langbein, F. C.; Liu, Y.; Marshall, A. D.; Martin, R. R.; Sun, X.; Rosin, P. L. Registration of 3D point clouds and meshes: A survey from rigid to nonrigid. *IEEE Transactions on Visualization and Computer Graphics* Vol. 19, No. 7, 1199–1217, 2013.

[37] Chen, K.; Lai, Y. K.; Hu, S. M. 3D indoor scene modeling from RGB-D data: A survey. *Computational Visual Media* Vol. 1, No. 4, 267–278, 2015.

[38] `https://gitee.com/xiaoyangyang2013/openMVS`.

[39] Moulon, P.; Monasse, P.; Marlet, R. OpenMVG: An open multiple view geometry library. Available at `https://github.com/openMVG/openMVG`.

[40] Rusu, R. B.; Blodow, N.; Beetz, M. Fast point feature histograms (FPFH) for 3D registration. In: Proceedings of the IEEE International Conference on Robotics and Automation 3212–3217, 2009.

[41] Besl, P. J.; McKay, N. D. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 14, No. 2. 239–256, 1992.

[42] Agresti, A. Logistic regression models using cumulative logits. In: *Analysis of Ordinal Categorical Data.* Agresti, A. Ed. John Wiley & Sons, Inc., 44–87, 2012.

[43] Tateno, K.; Tombari, F.; Laina, I.; Navab, N. CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 6565–6574, 2017.

[44] Kar, A.; Hne, C.; Malik, J. Learning a multi-view stereo machine. In: Proceedings of the 31st Conference on Neural Information Processing Systems, 2017.