Computational Visual Media
DOI 10.1007/s41095-xxx-xxxx-x
Vol. x, No. x, month year, xx–xx

Research Article

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 3D shapes from 2D projections of rotating 3D objects, which is called Kinetic Depth Effects. This process is based on a variety of visual cues such as lighting and shading effects. However, when such cues are weakened or missing, perception can become faulty, as demonstrated by the famous silhouette illusion example – 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 a quadratic stepwise regression analysis method to determine the weights among depth cues to represent the reconstruction quality. In the subjective evaluation model, we design a comprehensive user study to reveal correlations on the reaction time/accuracy, rotation speed, and the perspective. The two evaluation models are generally consistent, and can largely benefit the inter-disciplinary research of visual perception and 3D reconstruction.

Keywords Kinetic Depth Effects, 3D reconstruction, perceptual factor analysis.

1 TKLNDST, CS, Nankai University, Tianjin, China. Email: imclimmy@qq.com; 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.
Manuscript received: 20xx-xx-xx; accepted: 20xx-xx-xx.

Fig. 1 The Spinning Dancer. Due to the lack of visual cues, it confuses humans as to whether rotation is clockwise or counterclockwise. Here we show 3 of 34 frames from the original animation [**?**]. Image courtesy of Nobuyuki Kayahara.

1 Introduction

The human perception mechanism of the 3D world has long been studied. In the early 17th century, artists developed a whole system of stimuli of monocular depth perception especially on shading and transparency [**?**]. The loss of depth perception related stimuli leads to a variety of visual illusions, such as the Pulfrich effect [**?**]. In this example, 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* [**?**]. However, when the light over the object is disabled, 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 [**?** **?**] (see Fig. **??** for some sample frames). The dancer is observed to be spinning in clockwise or counterclockwise directions by different viewers. Such ambiguity implies that more cues are needed for humans to make accurate depth

judgements for 3D objects. Visual cues such as occlusion [**?**], frame timing [**?**], speed and axis of rotation [**?**] are widely studied by researchers. In addition, the perspective effects also affect the accuracy of direction judgements [**?**].

In this paper, we make in-depth investigations on how visual cues influence the perception of Kinetic Depth Effects from two aspects, including objective computational modeling, and subjective perceptual analysis. We formulate and quantify visual cues from both 3D objects and their surrounding environment. On the one hand, we make a comprehensive subjective evaluation to correlate the subjective depth judgement of a 3D object and its visual conditions. On the other hand, as depth perception largely depends on the quality of shape reconstruction in mind, we also propose an objective evaluation method based on 3D computational modeling [**?**]. This allows us to quantify the impacts of the involved visual cues. The impact factors are achieved by solving a multivariate quadratic regression problem. Finally, we analyze the interrelations between the proposed subjective and objective evaluation models, and reveal the consistent impacts of visual cues on such 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 from a carefully designed user study.
- A detailed analysis of how visual cues affect depth perception based on our 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 through psychological and computational approaches, and then briefly describe the relevant reconstruction techniques employed in this work.

Psychology research on shape perception. For monocular vision, shading effect contains rich information [**?**]. Compared with diffuse shading, specular shading helps to reduce underestimate of cylinder depth by subjects [**?**]. However, the shading effect can be ambiguous in some cases. For example, when the illumination direction is unknown, it is hard to judge shape convexities and concavities, and humans tend to assume that illumination comes from above [**?**]. Besides, when the level of overall illumination is low, the effect of the shadows is tended to be assumed coming from overall illumination [**?**].

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

The color information is very important not only in immersive scene representation [**?** **?** **?** **?**] but also in depth perception of psychology. Isono and Yasuda [**?**] find that chromatic channels can contribute to depth perception using a prototype flickerless field-sequential stereoscopic television system. Guibal and Dresp [**?**] realize that the color effect is largely influenced by luminance contrast and stimulus geometry. When shape stimuli are not strong, color could make an illusion of closeness [**?**].

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

In terms of 2D images, Chu *et al.* [**?**] present a computational framework to synthesize camouflage images that can hide one or more temporally unnoticed figures in the primary image. Tong *et al.* [**?**] propose 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 transform as well as object embedding. The study of Kinetic Depth Effects often uses subjective response [**?**], and some researchers also use the judgement of the rotation direction as the response [**?**].

Similar to image-based content embedding and hiding, 3D objects can be embedded into 2D images [**?**], where the objects can be easily detected by humans, but not by an automatic method. Researchers also generate various mosaic effects on both images [**?**] and 3D surfaces [**?**]. A computational model for the psychological phenomenon of change blindness is investigated in [**?**]. As change blindness is caused by failing to store visual information in short-term memory, the authors model the influence of long-range context complexity, and synthesize images with a given degree of blindness. Illusory motion is also studied as self-animating images in [**?**]. 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.* [**?**] create 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 are also some research works on rendering, designing and navigating

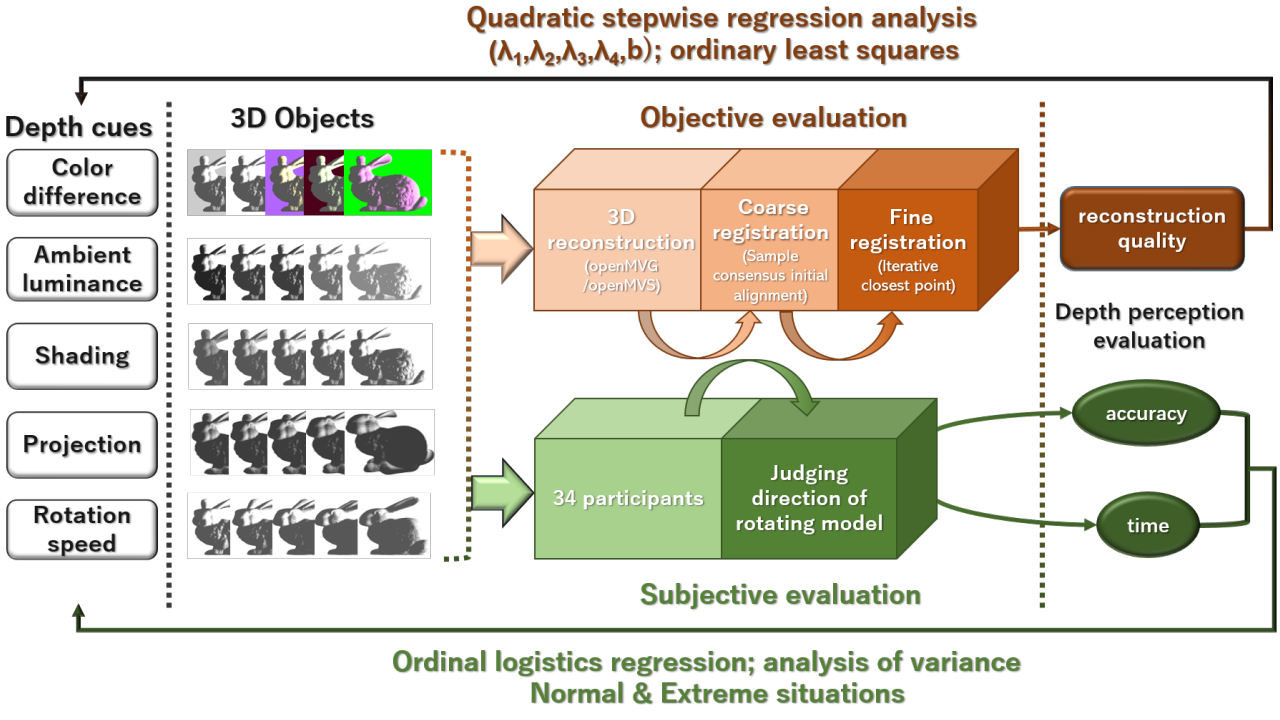

Fig. 2 Overview of our work. We project the input 3D objects onto 2D image planes with some specified conditions (*e.g.,* lighting, projection mode, rotation speed, etc.), based on which we construct objective and subjective evaluation models, respectively. Finally we reveal some interesting correlations between the depth perception of rotated 3D objects and the visual conditions.

impossible 3D models [**?** **?** **?**]. In contrast to investigating those seemingly impossible models, our work focuses on evaluating the 3D perception of rotated objects.

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 [**?** **?**]. Among different techniques, the well-known structure-from-motion [**?**] can effectively recover the camera poses and further generate a sparse 3D point cloud by making use of multiple images of the scene or objects. Moreover, multi-view stereo algorithms [**?**] can reconstruct the fully textured surface of the scene. We employ such computational techniques to evaluate the 3D reconstruction quality under various environmental conditions.

3 Overview

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

For both the human visual system and image-based 3D reconstruction techniques, the input visual information is usually in the form of projected 2D images. Therefore, by using a set of projected 2D images of the 3D objects under the aforementioned conditions, we investigate the perceived shape from human participants and the multi-view stereo reconstruction of 3D objects. Besides measuring the perception of Kinetic Depth Effects using our objective and subjective evaluation models, we further investigate the co-relations between these two different methods. The overview of our work is shown in Fig. **??**.

Dataset. For each 3D object, when it rotates around a fixed vertical axis passing through the geometric center of the object, we sample the projected 2D images with an interval of rotation angle θ. As the frame rate when displaying projected images is fixed, changing the sampling angle interval also means changing the rotation speed of the object. Also, we can obtain the dataset of projected 2D images in 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 the perception of perspective. We also control the color difference between the object and the background. In order to define expected color difference, predefined color pairs are used to generate the colors of the background

θ	Angular interval of 2D projection
α	Lightness in HSL color space (0, 0, α) which is used as the intensity of the diffuse light
β	Lightness in HSL color space (0, 0, β) which is used as the intensity of the global ambient light
D	Color difference between objects and background

Tab. 1 Definition of parameters in the controlled generation of 2D projected images.

and the 3D object (Sec. **??**). A summary of parameters is presented in Tab. **??**. Based on the generated dataset under the controlled conditions, we can then measure the 3D perception of rotated objects using the following two evaluation models.

Objective evaluation model. This model utilizes the reconstruction quality of the input 3D objects as the basis for evaluation. First, according to the projected 2D images of the 3D object under specified visual conditions, we reconstruct the point cloud using multi-view stereo reconstruction algorithms. Then, we develop a method to measure the reconstruction quality between the point cloud and the original 3D object (Sec. **??**). Finally, we analyze the effects of different visual conditions in detail (Sec. **??**).

Subjective evaluation model. Directly measuring 3D reconstruction in the brain of human subjects is difficult. Based on the observation that if humans successfully reconstruct a rotated object in their mind, it is easy for them to tell the direction of rotation, the time and the accuracy of direction judgments can be used as proxies to measure the quality of depth perception, as done in our study. 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 with the accuracy and the reaction time of direction judgements (Sec. **??**).

4 Objective Evaluation

Our objective evaluation 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; obtaining the fitted weighting factors of depth cues by solving a multivariate quadratic regression optimization.

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

Tab. 2 Values of α and β that control diffuse and ambient lighting used to generate 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

Tab. 3 Object and background color and corresponding color difference used to generate image sets.

4.1 Parameter selection and image set generation

In order to generate images of the 3D objects with various expected conditions, we need to select some parameter options for the depth-aware cues. Firstly, we normalize the size of all 3D objects with a unit bounding box centered at the origin. Then, we import objects in a virtual scene, display them under orthogonal projection, and set the 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 times of the bounding box. Since in openMVG the focal length is one given parameter, considering perspective projection mode in our objective evaluation is not that meaningful. So we turn to displaying 3D objects under orthogonal projection.

We control the brightness of diffuse and ambient lights. We set the HSL value of the diffuse light as (0, 0, α), and choose seven options for α, corresponding to different luminance levels. We set the HSL value of ambient light as (0, 0, β) with six options for β (see Tab. **??**).

As mentioned before, we sample the projected 2D images when rotating the 3D objects. Here we set the four possible sampling intervals θ as 0.209, 0.157, 0.126, and 0.105. To simplify the test, we also choose five optional pairs of RGB values for the 3D object and the background (see Tab. **??**).

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

$$D(C_B, C_O) = \sqrt{w_r(r_B - r_O)^2 + w_g(g_B - g_O)^2 + w_b(b_B - b_O)^2}. \quad (1)$$

In this equation, C_O is the color of the object, with RGB values of (r_O, g_O, b_O), C_B is the color of the background, with RGB values of (r_B, g_B, b_B), w_r, w_g, w_b are weighting factors, which are empirically set as $(3, 4, 2)$. In order to choose 3D objects, we generate image sets for 15 objects

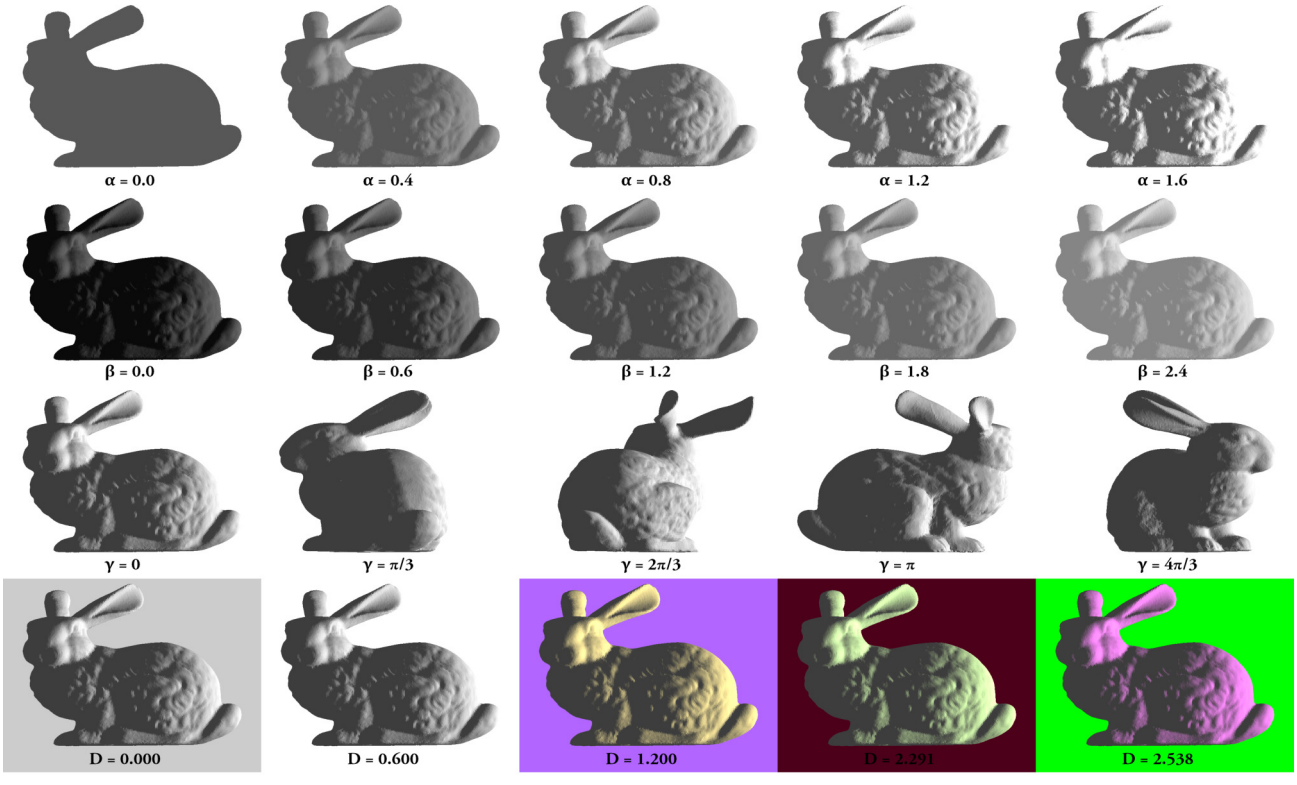

Fig. 3 Some examples of projected 2D images. The first row shows the changing of diffuse light. The second row shows the changing of ambient light. The third row shows five continuous images taken at the angular interval of $\theta = \pi/3$, where γ shows the angle between the initial orientation and current orientation. The last row shows the changing of color difference.

with different conditions. Then we choose three of them which have high reconstruction success rate (30%). Finally, for each object to be tested, we generate image sets for 7 (Shading) × 6 (Ambient Luminance) × 4 (Rotation Speed) × 5 (Color Difference) conditions, each of which has a separate image set. The size of every image is set to 800×600 pixels. Some examples are shown in Fig. **??**.

4.2 3D reconstruction and quality assessment

We employ openMVG [**?**] and openMVS [**?**] to process image sequences, and take the reconstructed point clouds as input. We normalize the size of all point clouds with the same bounding box as used in normalizing 3D objects. Then we match the reconstructed point clouds and the original objects. More specifically, we use the Sample Consensus Initial Alignment (SAC-IA) method [**?**] for initial alignment, and Iterative Closest Point (ICP) for refined alignment [**?**]. Finally, we compute the Euclidean fitness score μ between the reconstructed point cloud and the original object.

4.3 Objective evaluation results

We perform 2520 3D reconstruction cases using different image sets, and 929 of them generate 929 point clouds, while 1591 of them fail. We use the following equation to measure the reconstruction quality s between a pair of reconstructed point cloud and original point cloud, based on the point cloud distance μ:

$$s = -\lg(\mu). \tag{2}$$

We use the logarithmic processing to make the residuals of our model normally distributed. The reconstruction quality values are linearly normalized to range $[0,1]$. Given a set of reconstruction quality samples $S = \{s_1, s_2, \ldots, s_n\}$, we formulate the factor analysis model with the following quadratic stepwise regression:

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

where $\lambda = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ are weighting coefficients to balance the corresponding impacts, and b is a constant value. We fit the coefficients in the model using the standard least squares method. Results are shown in Tab. **??**.

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 individual visual cues is analyzed as follows:

Shading. Shading and reconstruction quality follow the

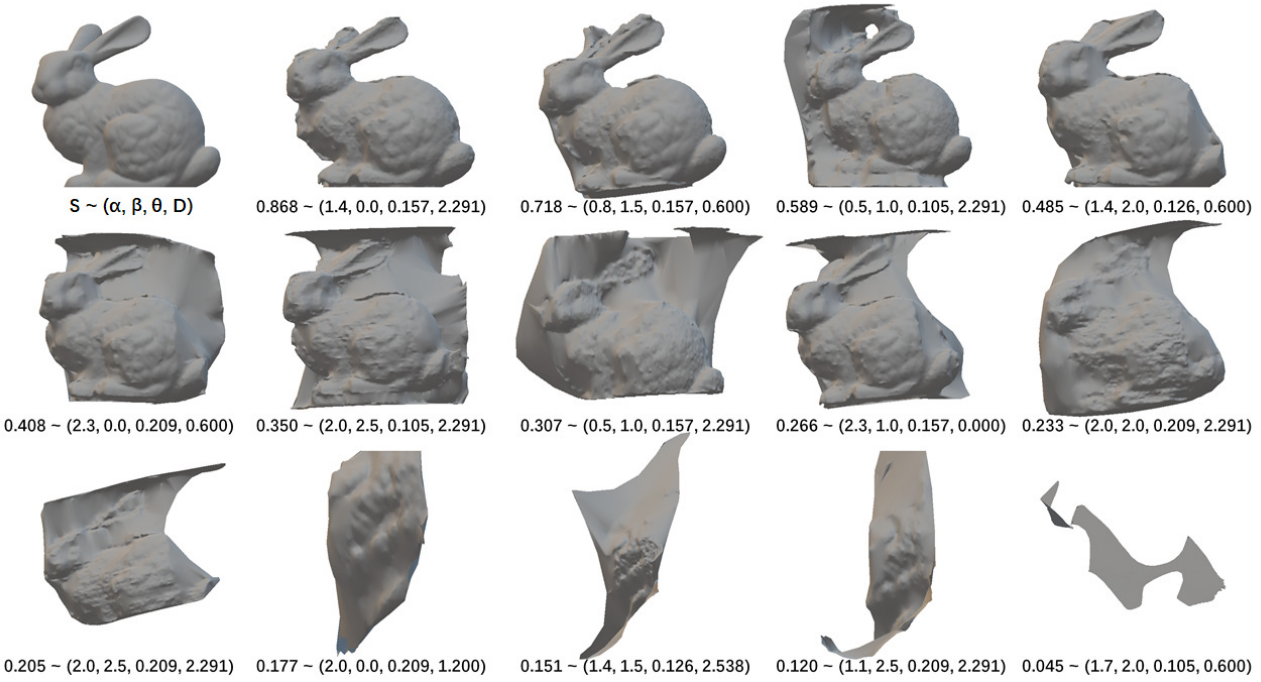

Fig. 4 Given a 3D object (top-left) and specified visual conditions, we generate the corresponding projected 2D images, and reconstruct the 3D shapes (others) using existing multi-view stereo algorithms. With the reconstructed and original objects, we then quantitatively measure the reconstruction quality for shape perception analysis. For each reconstruction, we report the reconstruction quality measure and the corresponding rendering setting.

Coefficients	Values	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-squared	0.103	

* p<0.01

Tab. 4 Objective evaluating model results. p represents the confidence probability of the parameter based on the standard student's t-test

.

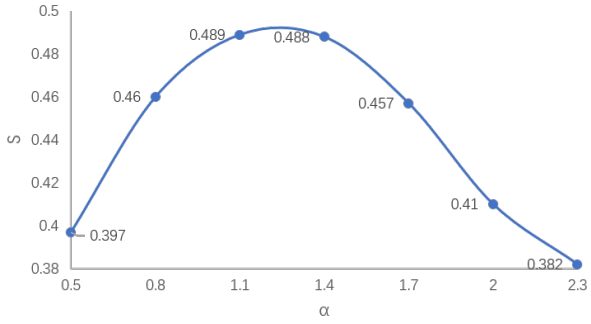

Fig. 5 Correlation between the 3D reconstruction quality and the shading.

quadratic function relationship (Fig. **??**), with $\lambda_1 = 0.3593$ (p<0.01), $\lambda_3 = -0.1234$ (p<0.01).

Ambient luminance. The *Ambient Luminance* × *Shading* interaction is significant with $\lambda_4 = -0.0278$ (p<0.01). High *Ambient Luminance* × *Shading* levels contribute to low reconstruction qualities. As shown in Fig. **??**, the fitted lines in different ambient light levels are not parallel.

Rotation speed. High rotation speeds significantly contribute to low reconstruction qualities ($\lambda_2 = -0.6361$, 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.

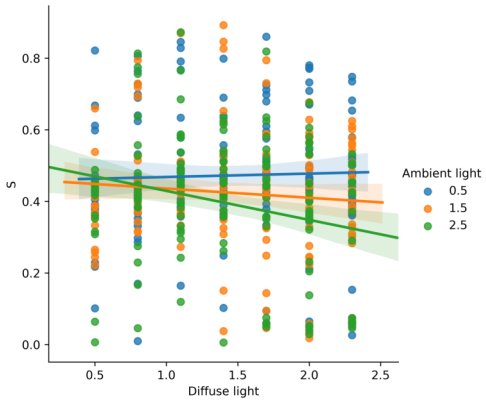

Fig. 6 Significant interaction of *Ambient Luminance* × *Shading*.

Color difference. The color difference does not significantly affect the reconstruction quality.

5 Subjective Evaluation

As mentioned before, humans can recover 3D rotated objects from their 2D projections. The rotation direction of 3D objects can be an important clue to judge the quality of the shape reconstruction in their mind. Based on this, the following multi-factor experiment is designed.

5.1 Participants

We recruited 35 participants and achieved results from 34 participants (19 males and 15 females) who successfully finished the test.

5.2 Procedure and materials

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

Study A. Here we explore the depth cue effect in general situations. The range of cues are the same as the objective evaluation model, but we choose fewer values for each cue (see Tab. **??**) to ensure participants can concentrate during the study. The projection can either be 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 condition, we display three different objects.

Study B. Here we consider more extreme situations, including low lighting levels, overexposure and high speed rotation, where we vary each condition while keeping other cues fixed (see Tab. **??**). The variables used to represent each situation are shown in Tab. **??**. After that we generate new test image sets (see Fig. **??** and Fig. **??**). To simplify the problem, we only consider the orthogonal projection situations.

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

Tab. 5 Values of each cue used in Study A of the subjective evaluation model.

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)

Tab. 6 Values of each cue used in Study B of the subjective evaluation model for extreme situations. From top to bottom: conditions of low lighting levels (with minimum value of β); overexposure conditions (with a relatively high value of α); high speed rotation conditions (with normal values for α and β).

Every participant was asked to judge rotation direction of all image sets generated in Studies A and B, and each image set was judged only once. The display order of each image set was random, so was the rotation direction of 3D objects. To exclude "viewing-from-above" bias [**?**], we define rotation direction as *Left* and *Right*. From the perspective of the participants, the rotation direction is right if the close part of 3D objects is moving to the right, otherwise the direction is left. The images were displayed at 24 FPS. The maximum display time for one image set did not exceed 5 seconds. Participants were required to judge whether it is rotating on the left or right direction. The participants were given time to practise before the formal experiment. The entire experiment took about $15-20$ minutes.

5.3 Subjective evaluation results

We record the judgments and reaction time of all participants. We rank all cases of the reaction time in ascending order and calculate the standard scores (here

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

Tab. 7 Values of varying cues used in Study B of the subjective evaluation model for extreme situations. From top to bottom: conditions of low lighting levels, overexposure, and high speed rotation.

we use τ to denote it), which correspond to the estimated cumulative proportion of the reaction time. We use the repeated measures Analysis of Variance (ANOVA) method to determine the effect of cues on τ under different conditions. We calculate each participant's judgement accuracy under each condition. Since three objects are tested in each condition, the participant's judgement accuracy take four values, and do not follow a normal distribution. Therefore, we use the ordinal logistic regression models to test the effect of cues on the participant's judgement accuracy. In particular, we choose the complementary log-log link function, since most of the participant's judgement accuracy lie in $0.67-1.00$ [**?**]:

$$\Phi(x) = log(-log(1-x)). \tag{4}$$

We establish the ordinal logistic regression models for all situations, while we only show those models with significant results and pass other tests of parallel lines.

5.3.1 Analysis of study A

A five-way ANOVA method reveals the main effect of *Rotation Speed* ($F(1,5028) = 38.11$, $p<0.01$) on τ. Participants react faster at the high rotation speed condition ($M = -0.08$, $SD = 0.97$) than at the low rotation speed condition ($M = 0.05$, $SD = 1.00$). We also find a significant *Perspective* $\times$ *Rotation Speed* interaction ($F(1,5028) = 6.19$, $p<0.05$) on τ. At the high rotation speed condition, τ is significantly lower under perspective projection mode ($M = -0.109$, $SD = 0.95$) than under orthogonal projection mode ($M = -0.06$, $SD = 1.00$). Because *Rotation Speed* and *Perspective* only have two levels, there is no need for Mauchly's test of sphericity. Apart from the above phenomena, for the investigated five cues in our experiments, we find no significant effect on the participant's judgement accuracy.

5.3.2 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 establish an ordinal logistic regression model as follows:

$$\begin{cases} \Phi(p_1) = \varepsilon_1 - \eta * \alpha, \\ \Phi(p_1+p_2) = \varepsilon_2 - \eta * \alpha, \\ \Phi(p_1+p_2+p_3) = \varepsilon_3 - \eta * \alpha, \\ p_1+p_2+p_3+p_4 = 1, \end{cases} \tag{5}$$

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 $\varepsilon = \{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$ are constant values. For each case, the value with highest probability is the predicted value of the participant's judgement accuracy.

Coefficients	Values	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-squared	0.093	

* $p<0.01$

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

.

We fit the coefficients in the model and the results are shown in Tab. **??**. High α significantly contributes to high judgement accuracy ($\eta = 4.229$, $p<0.01$). Hence using strong shading in low lighting levels conditions strengthens the accuracy and accelerates the reaction (see Fig. **??**).

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 overexposure conditions delays reaction (see Fig. **??**). We find 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 speed rotation conditions, lower rotation speed leads to faster reaction. We establish an ordinal logistic regression model again as follows:

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

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 the 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 fit the coefficients in the model and the results are shown in Tab. **??**. High θ significantly contributes to low judgement accuracy ($\kappa = -1.353$, $p<0.01$). In high speed rotation conditions, increasing the rotation speed reduces the judgement accuracy and delays the reaction (see Fig. **??**).

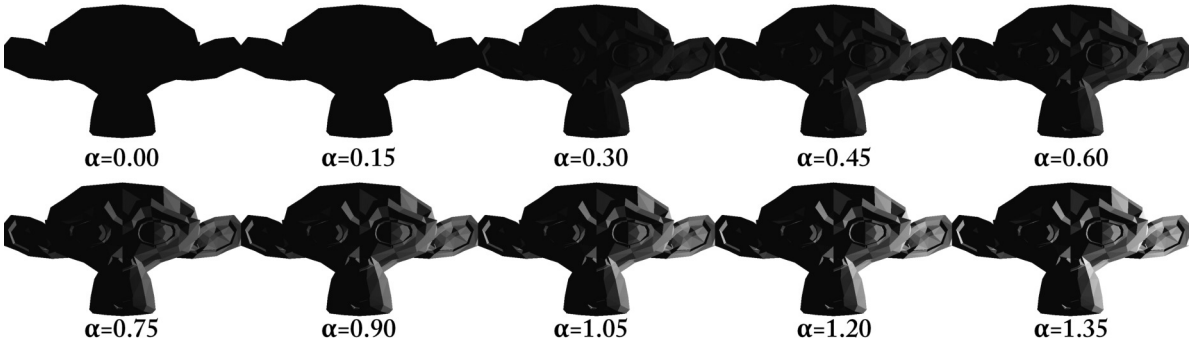

Fig. 7 Examples of 2D images under each individual low lighting levels condition.

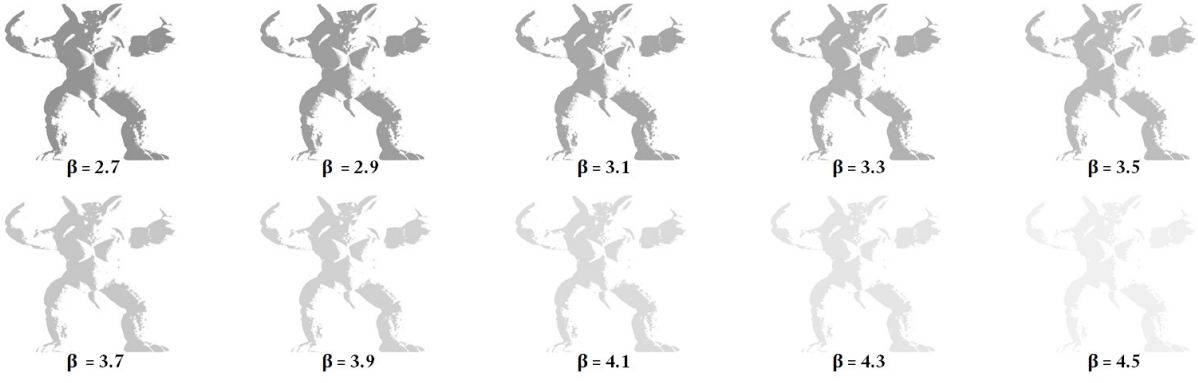

Fig. 8 Examples of 2D images under each individual low lighting levels condition.

Coefficients	Values	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-squared	0.030	

* p<0.01

Tab. 9 Results of ordinal logistic regression under the high speed situation. p represents the confidence probability of the parameter based on the Wald Test
.

6 Joint Objective and Subjective Analysis

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

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

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* $\times$ *Ambient Luminance* levels contribute to low reconstruction qualities. In the subjective evaluation, high ambient luminances in overexposure cases can increase observers' reaction time.

Rotation speed. The rotation speed plays an important

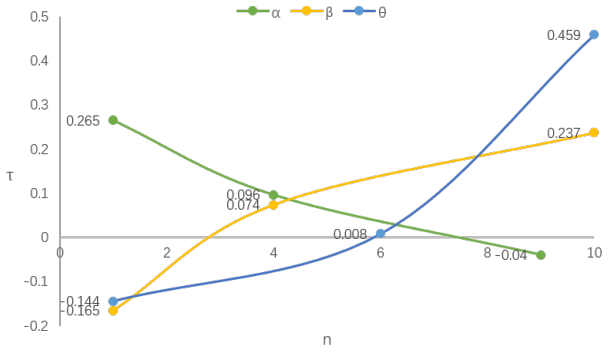

Fig. 9 Mean values of τ under nth level of shading, ambient luminance and speed. The marked points present significant differences under pairwise comparison.

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

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

Perspective. In the subjective evaluation, *Perspective* $\times$ *Rotation Speed* interaction is significant. Compared with orthogonal projection, participants react faster under perspective projection conditions.

Color difference. We have not found significant effects caused by color difference between objects and background in either objective evaluation or subjective evaluation model. As future work, we will test more color combinations to further explore possible effects by color differences.

7 Discussion

We analyze the effect of different depth cues on 3D perception of rotated 3D objects, which broadens the scope of previous studies. We also design an objective evaluation and a subjective evaluation to make a thorough analysis.

However, there are also some flaws in our design. In our objective evaluation, when the depth cues in images were extremely weaken, 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-squared = 10.3%). Moreover, the subjective evaluation only uses the judgement of the direction of rotated objects as the response. In the future, we could use more 3D information as response. In our experiments, the reconstruction quality is closely related to the kind 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 to get good reconstruction results for both humans and computers, such as rendering under certain lighting. The objective evaluation also reveals the limitations of existing algorithms. On the other hand, when combining with recent deep learning-based techniques, such as CNN-SLAM [**?**] and deep stereo matching [**?**], our solution could further benefit once there are more accurate depth prediction and 3D reconstruction in various challenging environments.

8 Conclusion and Future Work

We have proposed two approaches to measure the quality of depth perception of Kinetic Depth Effects, where we made a detailed analysis of how visual cues affect depth perception. Firstly, we generated a dataset of images from rotating objects considering five depth cues: ambient luminance, shading, rotation speed, and the color difference between objects and background. In the objective evaluation, we applied 3D reconstruction and measured reconstruction quality between reconstructed and original objects. In the subjective evaluation, we invited participants to judge the rotating direction of 3D objects by showing the projected 2D images. We inferred the perception quality by their reaction time and accuracy. In our study, we found both strong and dim shadings significantly undermine the perception of depth in our experiments. High ambient illumination $\times$ shading level, rotation speed, and orthogonal projection can also reduce the depth perception quality. It is also interesting that the color difference does not have significant effect on the depth perception in our experiments. In the future, we will take more depth cues into consideration and develop a more precise quantitative model for more complex situations. Taking 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-discipline research on robust 3D reconstruction and human visual perception.

Acknowledgements

This work was supported by Tianjin NSF (18JCYBJC41300 and 18ZXZNGX00110), NSFC (61972216), and the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (VRLAB2019B04). Shao-Ping Lu is the corresponding author of the paper.

Meng-Yao Cui is currently a Bachelor student at Nankai University, majoring in Computer Science and minoring in Psychology. Her research interests include visual perception and computing, human-computer interaction and machine learning.

Shao-Ping Lu is an associate professor of Computer Science at Nankai University in Tianjin, China. He had been a senior/postdoc researcher at Vrije Universiteit Brussel (VUB). He received his PhD degree in 2012 at Tsinghua University, China. He also spent two years on high-performance SOC/ASIC design in industry in Shanghai. His research interests lie primarily in the intersection of visual computing, with particular focus on 3D video processing, computational photography, visual scene analysis and machine learning.

Miao Wang is an assistant professor with the State Key Laboratory of Virtual Reality Technology and Systems, Research Institute for Frontier Science, Beihang University, and Peng Cheng Laboratory, China. He received a Ph.D. degree from Tsinghua University in 2016. During 2013-2014, he visited the Visual Computing Group in Cardiff University as a joint PhD student. In 2016-2018, he worked as a postdoc researcher at Tsinghua University. His research interests lie in virtual reality and computer graphics, with particular focus on content creation for virtual reality.

Yong-Liang Yang received his Ph.D. degree in computer science from Tsinghua University in 2009. He worked as post-doctoral fellow and research scientist in King Abdullah University of Science and Technology (KAUST) from 2009 to 2014. He is currently a senior lecturer (associate professor) at the Department of Computer Science, University of Bath. His research interests include geometric modeling, computational design, interactive techniques, and applied machine learning.

Yu-Kun Lai received his bachelor's and Ph.D. degrees in computer science from Tsinghua University, China, in 2003 and 2008, respectively. He is currently a reader at the School of Computer Science & Informatics, Cardiff University. His research interests include Computer Graphics, Computer Vision, Geometry Processing and Image Processing.

Paul L. Rosin is a professor at the School of Computer Science & Informatics, Cardiff University. His research interests include the representation, segmentation, and grouping of curves, knowledge-based vision systems, early image representations, low level image processing, machine vision approaches to remote sensing, methods for evaluation of approximation algorithms, medical and biological image analysis, mesh processing, non-photorealistic rendering and the analysis of shape in art and architecture.