A Monocular Depth Estimator to Perceive Crater Illusions in Several Characteristics

Ryota Mima\(^1\), Yuki Kubota\(^1\) and Masahiko Inami\(^1\)

\(^1\)Information Somatics Lab, RCAST, the University of Tokyo, Japan
mamii07180@g.ecc.u-tokyo.ac.jp

Abstract

How machines perceive visual illusions is essential to clarify the differences between artificial and human perception. Crater illusions cause depth perception in the texture by the intensity gradient of the disks. In this study, we investigated whether a monocular depth estimator perceives the crater illusions. We adopted MonoDepth as the estimator and conducted two experiments using the illusions with planar images and attached to a cube. As a result, the estimator perceived the convexity on the top-lit disk, while a little concavity on the bottom-lit disk. We also found the vertical heterogeneity for the depth perception of it. Additionally, the machine perceived the illusion attached to the cube, although the indeterminacy for object scale was revealed. Our findings would bridge the research between computer vision and cognitive science in terms of depth perception.

Introduction

Visual illusions are known as the manifested discrepancies between physical and perceptual images, and how machines perceive these illusions is essential to investigate the differences between artificial and human perception (Hecht, 2014; Watanabe et al., 2018). The characteristics of depth cues for reconstructing depth information from two-dimensional retinal images (Landy et al., 1995) often produce the illusions in relation to depth perception (Gregory, 1970; Fieandt, 1938). In particular, the crater illusion (Figure 1(a)) makes depth perception by intensity gradient of the disks owing to a shading cue; a top-lit disk tends to be perceived as convex, and a bottom-lit disk as concave.

Monocular depth estimators with Convolutional Neural Networks (CNNs) have been widely studied and applied in augmented reality and automated driving technology. However, due to the absence of training two-dimensional images, including typical illusion figures, how well the monocular depth estimator with CNNs can perceive the depth illusions have not been investigated.

In this study, we verified whether a monocular depth estimator could perceive the crater illusions. Two experiments were conducted using the illusions with planar images and attached to a cube, and the output of the estimator was compared with human perception.

Experiment 1: Crater illusions with planar images

In Experiment 1, we investigated how crater illusions with planar images were perceived by MonoDepth (Godard et al., 2017), one of the unsupervised monocular depth estimator with CNNs, used as a depth estimator. MonoDepth has a simple network structure with guaranteed accuracy on natural images, and does not explicitly learn shading cues. Besides, MonoDepth learns about disparity consistency using stereo images rather than depth images, and infers from monocular images.

Methods

A pre-trained model with the KITTI dataset (Geiger et al., 2012) and city landscape images was used. The input image was rotated by 30\(^\circ\), and the output depth image from the model was qualitatively evaluated. Since monocular depth estimation is scale-invariant, the minimum and maximum depth values are normalized on the \([0, 1]\) interval.

Results and Discussion

![Figure 1: Example of the depth estimation result with a crater illusion. (a) Input figure; (b) Output depth map.](image)

When Figure 1(a) was input to MonoDepth, the output depth map was obtained as shown in Figure 1(b). The shift between the background and disks in this depth map indicates that top-lit disks were perceived as convex and bottom-lit disks as concave. Besides, it was observed that the perceived amount was greater in top-lit disks than in bottom-lit disks.
disks, and in the upper region than in the lower region. Additionally, we observed that a uniform gray background in the input image was transformed into the non-uniform depth map shown in Figure 1(b).

The results of the crater illusion after rotation are shown in Figure 2. The images show that the larger rotation from top-lit and bottom-lit disks caused the smaller perceived amount of depth. The non-uniform depth map in the background was also observed in these images, although the distribution of these abnormalities of depth was different from Figure 1(b). Moreover, the difference in perception by the direction of rotation was not significant, as can be seen in the comparison between (a) and (b) in Figure 2.

These results suggest that the crater illusion was perceived by MonoDepth to some extent. The asymmetry between the top-lit and bottom-lit disks is consistent with the previous studies on visual search tasks (Kleffner and Ramachandran, 1992; Kawabe and Miura, 2004). The rotational angular dependence is also compatible with the previous study (Sun and Perona, 1998).

The vertical region asymmetry of the perception may be caused by the training images’ bias that the lower regions are closer than the upper ones. Additionally, the background surface’s non-uniform depth map was observed due to the training data only with natural images, which makes it difficult to perceive two-dimensional patterns.

**Experiment 2: Crater illusions attached to cubes**

The results of Experiment 1 reveal the problem of a non-uniform depth map in the background. In Experiment 2, we propose a method of attaching to a cube to make more natural images than two-dimensional patterns.

**Methods**

We attached the crater illusion on the cube faces and moved the camera diagonally above the cube to capture the images as inputs. The depth estimator and output analysis are the same as Experiment 1.

**Results and Discussion**

An example of the results are shown in Figure 3. The results indicate that the convex and concave were observed in the top-lit and bottom-lit disks attached to the cube, respectively. As in the two-dimensional image in Experiment 1, the top-lit disks had a larger depth amount than the bottom-lit disks. The amount of the crater illusion was different between the ceiling and side faces.

Note that the scale normalization constants differ depending on the cube’s orientation, making it difficult to compare different inputs. The depth of the cube itself may affect the estimated depth according to the rotation angle of the cube; the output depth map just represents the vertical distance from the camera.

**Conclusions**

We verified whether MonoDepth, one of the monocular depth estimators, perceives crater illusions with two experiments. From these results, the estimator would perceive the crater illusion in several characteristics both planar and cubic images including the asymmetry between the top-lit and bottom-lit images. However, we observed the background’s non-uniformity and the difficulty of making comparisons among the outputs due to scale indeterminacy of input monocular images. These problems are essential for more quantitative evaluation of perceptual quantities.

The finding of this study that a depth estimator trained with binocular depth cues implicitly learned monocular depth cues will further the research on the relationship between each cue in artificial depth perception. We hope that this research will bridge the gap between computer vision and cognitive science in terms of depth estimation.

**Acknowledgements**

This work is financially supported by JST ERATO Grant Number JPMJER1701, Japan. The authors would like to thank Sohei Wakisaka for useful discussion.
References


