The role of depth information in object recognition

Ayako SANEYOSHI*, Chikashi MICIMATA*, and Keiji SANEYOSHI**

Sophia University* and Tokyo Institute of Technology**

"Geons" (Biederman, 1987) with different rotation angles were sequentially presented through a haploscope and the participants performed a sequential matching task. The RT increased as the rotation angle increased. The 3D condition, in which the stimulus was stereoscopically presented, was performed faster than the 2D condition. The effect of the rotation angle was equivalent for the 2D and 3D conditions. Furthermore, there was a possibility that depth information is helpful in extracting the qualitative information of an object. These results suggested that the qualitative information extracted from view-dependent 2D representation plays a crucial role in constructing the representation of objects.

Key words: object representation, view-dependency, stereoscopic vision

Tarr, Williams, Hayward, and Gauthier (1998) reported that the recognition of geons was affected by a change of viewpoint. Thus they concluded that the representation of an object would not be a 3D view-independent structural description but 2D view-dependent images. However, the structural description model suggested that the depth information of an object would be very important for reconstructing the representation of a 3D object. Because Tarr et al. (1998) employed 2D images as stimuli, there is a possibility that their result was due to a lack of depth information of the objects. In the present experiment we investigated the effect of depth information in recognizing geons from different viewpoints. We presented the geons stereoscopically using a haploscope.

Methods

Participants Sixteen graduate and undergraduate students served as volunteer participants.

Apparatus The stimuli were presented on a SONY 19-inch CRT monitor connected to an Apple PowerBook G4 computer. A ten-key pad was connected to the computer and served as a two-key response console. Through a haploscope each eye views its respective half side of the monitor (e.g. the left eye sees the left half of monitor). In a 2D condition both eyes receive identical images. In the present investigation when the 3D condition was used two images with binocular disparity were presented.

Stimulus The stimuli were 20 gray-shaded geons (Biederman, 1987) which were rotated around the vertical axis with a step of 15 degrees. All of the stimuli were approximately 6 × 6 degrees of visual angle in size.

Design The experimental design was an orthogonal combination of "View" (2D vs. 3D) and "Rotation Angle" (0°, 15°, 30°, 45°, 60°, 75°, and 90°). Both of the variables were manipulated within the participants.

Task and procedure The participants were asked to decide whether 2 successively presented stimuli were the same geons, ignoring their rotation. In 2 separate sessions the participants observed the geons with a haploscope in both the 2D and 3D condition. In each trial a fixation cross appeared for 500 ms, and this was followed by a 200 ms presentation of the first stimulus. A mask stimulus was then displayed for 500 ms, and this was followed by a 100 ms presentation of the second stimulus.

Results and Discussion

The results are shown in Table 1(a). The reaction time data was subjected to a 2 (View) × 7 (Rotation Angle) analysis of variance with repeated measurements. The main effect of Rotation Angle was sig-
significant \(F(6, 90) = 5.22, p < .01\). As can be seen in Table 1a, RT linearly increased as the Rotation Angle increased, both in the 2D and 3D conditions. The main effect of View was also significant \(F(1, 15) = 4.92, p < .05\), indicating that the 3D condition was performed with shorter RTs than the 2D condition.

The interaction of the Rotation Angle by View was not significant. That is, the Rotation Angle was equally affected in the 2D and 3D conditions. This implies that the view-dependency in recognition of geons, as reported by Tarr et al. (1998), could not be explained by a lack of depth information for the objects. The present results support the theory that the object representation is based on view-dependent 2D images.

On the other hand, it should be noted that the 3D condition was performed with a shorter RT than the 2D condition. It is therefore necessary to consider the role of depth information in constructing 2D-based representation. Hayward and Tarr (1997) suggested that the effect of rotation was larger when the rotation did not preserve qualitative information (e.g. the main axis was curved or straight) than when it did not preserve quantitative information (e.g. the curvature of the main axis was changed, see Figure 1) but did preserve qualitative information. They concluded that qualitative information would be more important than quantitative information for construction of a representation of the object.

We therefore considered the possibility that quantitative information could be extracted much easier when depth information was added and we reanalyzed the data to examine this possibility. The RT data was subjected to a 2 (View: 2D, 3D) \times 2 (Change Type: quantitative, qualitative; see Figure 1) analysis of variance with repeated measurements.

The results are shown in Table 1(b). The main effect of Change Type was not statistically significant \(F(1, 15) = 3.35, p = .09\). The interaction of View by Change Type was also not statistically significant \(F(1, 15) = 3.23, p = .09\), but the quantitative change produced shorter RTs than the qualitative information change in the 2D condition, (Sidak's test: \(p < .05\)). However a significant difference was not revealed in the quantitative change and qualitative change in the 3D condition, (Sidak's test: \(p = .69\)). Consequently, the addition of depth information reduced the cost of the qualitative change. This result implies that depth information is helpful for extracting the qualitative information of an object. These results indicated that the qualitative information extracted from view-dependent 2D images plays a crucial role in constructing the object representation.

### References

