OPEN

Correspondence and
requests for materials
should be addressed to
Y.T. (sutokuin8lab@
gmail.com)

Higher resolution stimulus facilitates depth perception: MT+ plays a significant role in *monocular* depth perception

Yoshiaki Tsushima[1,2], Kazuteru Komine[1], Yasuhito Sawahata[1] & Nobuyuki Hiruma[3]

[1]Three-Dimensional Image Research Division, NHK Science and Technology Labs, Tokyo, Japan, [2]Sutokuin Lab., Osaka, Japan, [3]Human Interface Research Division, NHK Science and Technology Labs, Tokyo, Japan.

Today, we human beings are facing with high-quality virtual world of a completely new nature. For example, we have a digital display consisting of a high enough resolution that we cannot distinguish from the real world. However, little is known how such high-quality representation contributes to the sense of realness, especially to depth perception. What is the neural mechanism of processing such fine but virtual representation? Here, we psychophysically and physiologically examined the relationship between stimulus resolution and depth perception, with using luminance-contrast (shading) as a monocular depth cue. As a result, we found that a higher resolution stimulus facilitates depth perception even when the stimulus resolution difference is undetectable. This finding is against the traditional cognitive hierarchy of visual information processing that visual input is processed continuously in a bottom-up cascade of cortical regions that analyze increasingly complex information such as depth information. In addition, functional magnetic resonance imaging (fMRI) results reveal that the human middle temporal (MT+) plays a significant role in *monocular* depth perception. These results might provide us with not only the new insight of our neural mechanism of depth perception but also the future progress of our neural system accompanied by state-of- the-art technologies.

In present day, we humans are living in numerous technologies generating high-quality virtual world we have never seen before. For instance, we have a digital display consisting of a high enough resolution that we are not able to distinguish form the real world[1]. Some people who have watched the images on a higher resolution display reported that they felt more real than the ones on a lower resolution display. However, little is known how such high-quality representation contributes to the sense of realness. What is the neural mechanism of processing such fine but virtual representation?

Depth information of a two dimensional image can greatly help us to feel the realness of the image[1]. Perceiving depth, in other words, creating the third dimension of a view from two dimensional images on our retina is one of the most mysterious functions in our visual system. A great number of studies have suggested a variety of ways to perceive depth, such as binocular disparity[2,3], motion parallax[4,5], shape from shading[6–9] and so on, or combination of these factors[10]. However, it is still unclear how stimulus resolution correlates with depth perception (Figure 1), as well as the neural correlates.

Here, we hypothesize that a visual image on a higher resolution display facilitates depth perception, and investigate the correlation between stimulus resolution and depth perception. To examine the relationship between stimulus resolution and depth perception, here we used luminance-contrast difference/change (shading) as a depth cue[7,11], because the aspect of luminance-contrast difference/change is crucially influenced by the display resolution, on top of that, it allows us to investigate one of simple mechanisms of depth perception excluding complex higher-levels mechanisms[7–9]. We conducted, a series of psychophysical and fMRI experiments. In psychophysical experiments, we ran the depth task in which the participants were asked to report which stimulus they perceived more depth with monocular viewing (Figure 2, also see *Methods*). Also, to investigate whether or not the participants notice the stimulus resolution difference, we conducted the resolution task in which the participants were asked to report which stimulus had higher resolution (see *Methods*). In order to explore the neural mechanism of the monocular depth perception, we measured fMRI activities while the participants were doing the depth task. In addition, to test a possibility that the participants who were asked
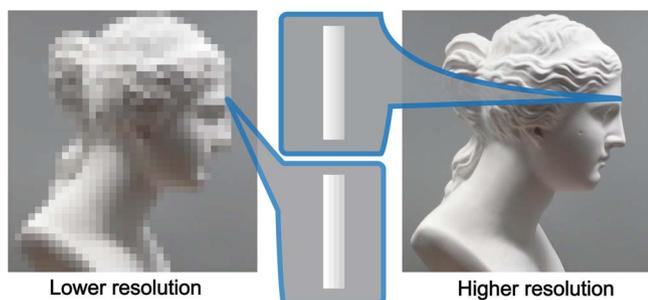
Figure 1 | The same images shown in different spatial quality, lower and higher resolution. Higher resolution image (right) has finer luminance-contrast representation than lower resolution. Which one do we perceive more depth? (The drawings were created by Y.S and Y.T., and modified with using Adobe Illustrator CC, RRID:nlx_157287).

to do the depth task did a different task with focusing on other stimulus features, such as the total or partial amount of luminance-energy differences between two stimuli, we conducted the luminance task, and compared fMRI activities between the depth task and luminance task conditions (see *Methods*).

Results

Psychophysical experiments of Depth task and Resolution Task. As a result, the participants reported that they perceived more depth with the higher resolution stimulus in all stimulus set (red lines in Figure 3). In addition, we found that such phenomenon becomes saturated between the higher resolutions (e.g. there are no significant mean Bradley-Terry score (BTS) differences between two highest resolution stimuli in Δ3). According to the results of the resolution task, no participants realized the resolution difference between the stimuli (blue lines in Figure 3). From the results of the visual acuity test (see *Methods*), we found that there are no significant visual acuity differences between Depth Task and Resolution Task groups. Taken together, the participants perceived more depth with the higher resolution stimulus *without* noticing the resolution difference.

However, there is a possibility that the participants who were asked to do the depth task did a different task with focusing on other stimulus features, such as luminance-energy differences between two stimuli (see Figure 2). To check this possibility and examine the neural correlates of the psychophysical results, we asked participants to do the depth and luminance task in fMRI, and compared fMRI activities between the depth task and luminance task conditions (see *Methods*).

fMRI experiments of Depth task and Luminance task. We obtained the same pattern of behavioral results as the previous depth task conducted outside of fMRI scanner (Higher resolution stimulus facilitates depth perception. See *Supplementary Information*). Also, we found the fMRI activities in the visual areas at both the depth and luminance task conditions. On the other hand, the fMRI results show that the amount of activity of the human middle temporal (MT+) was significantly higher at the depth task condition than at the luminance task condition (Figure 4, $n = 10$, $p < .0001$,
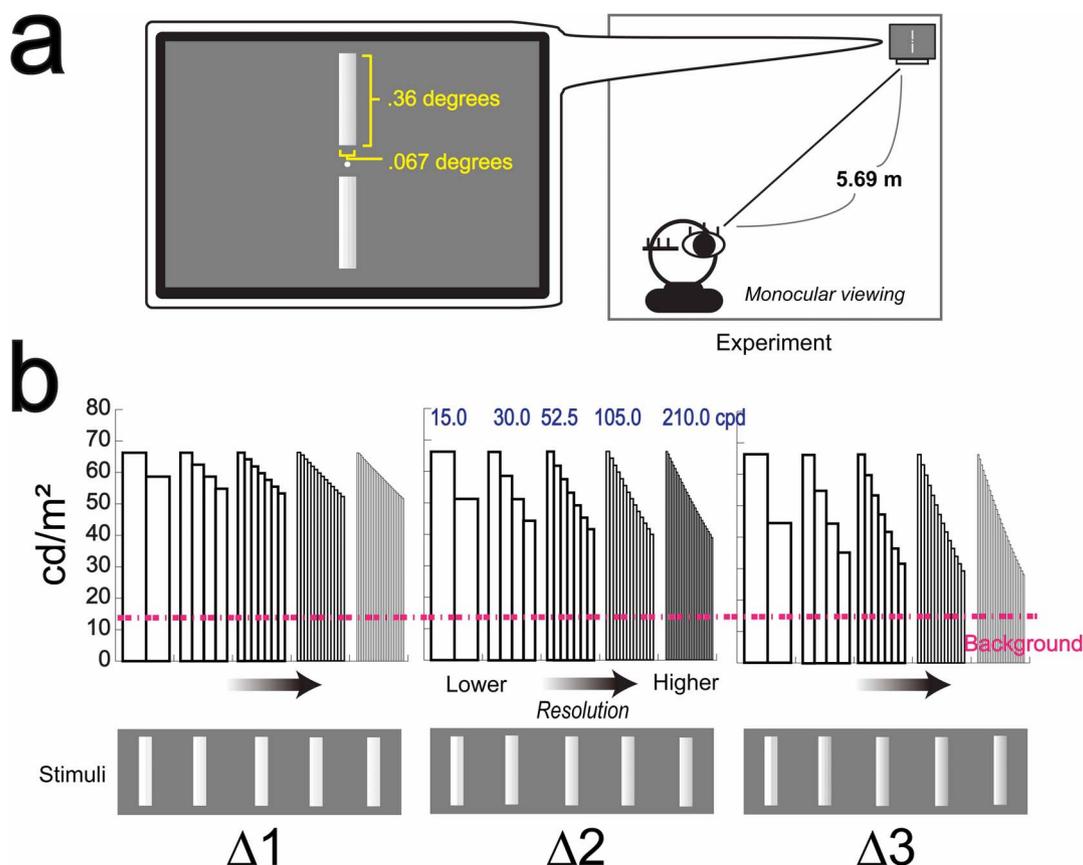

Figure 2 | Experiment and Stimuli set. (a), Two same-size but different-resolution bars were vertically presented for 1.0 second. The distance between two stimuli was .12 degree. To avoid that the participants easily detect the stimulus resolution difference, they were shown very small, .36 × .067 degrees. After the stimuli disappeared, the participants reported which stimulus they perceived more depth, upper or bottom. The combination of these stimuli resolution was varied from trial to trial. (b), Five kinds of resolution were used, 15.0, 30.0, 52.5, 105.0, 210.0 cpd. Also, there were three different kinds of stimuli set, Δ1, Δ2, and Δ3 (see *Methods*). Y axes represent the actual luminance value of each stimulus. The pink dashed line indicates the luminance value of the gray background (14.4 cd/m²). The presented stimuli are drawn by the same algorithm used in the real experiments.
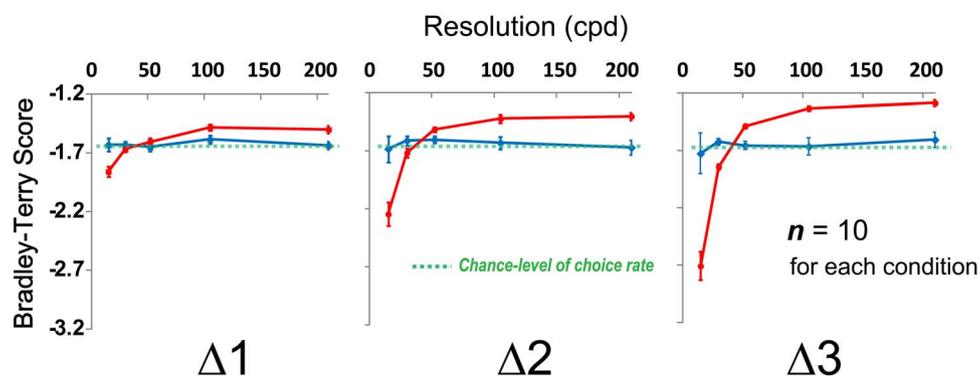
Figure 3 | Results of Depth Task and Resolution Task. Mean Bradley-Terry score (BTS)[12] for each stimulus in the depth task (red lines) as a function of the resolution (cpd) (n = 10). Mean BTS for each stimulus in the resolution task (blue lines) as a function of the resolution (n = 10). Green dashed lines represent the chance-level choice rate (50% choice rate in two alternative forced choice). Vertical error bars, ±1 SEM. Higher resolution significantly facilitated the participants' depth perception in all data set (e.g. mean BTS at the highest resolution (210.0 cpd) was significantly higher than at the lowest resolution (15.0 cpd) in $\Delta 1$, $n = 10$, $p < .001$, t test with Bonferroni correction). There were no significant difference between the performance results of the resolution task (blue lines) and the chance-level of choice rate (green dashed lines), that is, they were not able to detect the resolution difference.

uncorrected, also see *Methods* and *Supplementary Information*), although significant activities difference in other visual cortices (e.g. V1 or V2) and other areas were not found. This result reveals two important things; first, the fMRI activities at the depth task condition were different from the ones at the luminance task condition, that is to say, it is unlikely that the participants who were directed to do the depth task did the luminance task. Second, it is indicated that MT+ played a significant role on engaging in the depth task.

Discussion

The results of the present study demonstrate two important points. First, higher resolution stimulus facilitates depth perception even when the stimulus resolution difference is undetectable. This finding might contradict the traditional view that visual input is processed continuously in a bottom-up cascade of cortical regions that analyze gradually complex information, because the participants perceived more depth with higher resolution stimulus, *without* noticing the resolution difference. More specifically, the participants did not
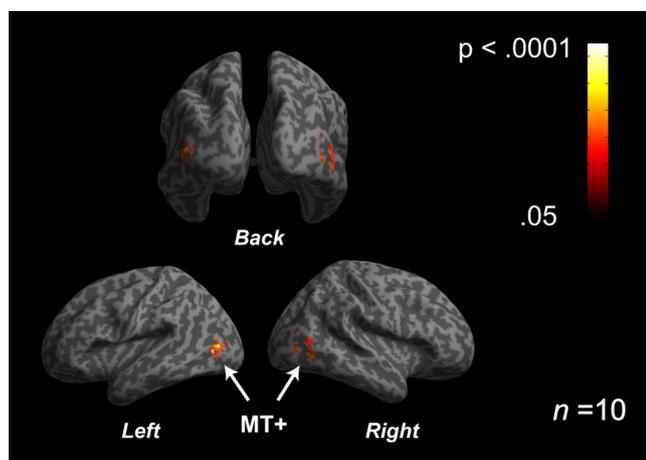

Figure 4 | Results of fMRI data from Depth Task and Luminance Task. With the same stimulus set, the participants asked to do Depth Task or Luminance Task in the separate sessions (All participant used their right eye.). Mean fMRI activities specifically for the depth task, which were calculated by subtracting the luminance task condition (baseline) from the depth task condition ($n = 10$, $p < .0001$, uncorrected, also See *Supplementary Information*), was shown. The amount of MT+ activity at the depth task was significantly higher than the one at the luminance task.

realize the primitive visual information (i.e. the resolution difference) but recognized the higher cognitive information such as depth information. On top of that, this phenomenon suggests the existence of two types of visual information in our visual system, consciously available and unavailable information. To uncover the aspect of visual awareness, further behavioral and physiological investigation must be needed[13].

Second, the fMRI results reveal an important role of MT+ activities. It is well known that MT+ and the monkey MT are largely specialized for motion processing[14–16], and they have been suggested to be crucially involved in extracting depth information by *binocular* disparity[17,18]. Here, we found that the amount of MT+ activity at the depth task condition was significantly higher than the one at the luminance task. It might indicate the new view on MT+ role in which MT+ plays an important role in not only *binocular* but also *monocular* depth perception with luminance-contrast difference/change. Furthermore, MT+ activities are thought to be involved in more of cognitive processing than perceptual processing[18–21], because the participants who engaged in the depth task made a decision based on not just perceptual information such as resolution or luminance-contrast difference but the task requirement or/and prior knowledge in which luminance-contrast difference/change was useful as a depth cue[22]. Actually, this is also in accordance with our findings, in which there are no significant fMRI activities difference between the depth and luminance task conditions at lower-level visual areas processing more perceptual information. However, one of alternative possible explanations of the results found in our experiment is that MT+ activities at the depth task condition could be more effectively regulated by top-down modulations such as attention[20,21,23]. For example, the amount of MT+ activities at the depth task was higher if the participants paid more attention to the stimuli or task than the luminance task (because the depth task required higher cognitive processing than the luminance task). Therefore, additional studies must be needed to clarify this hypothetical view.

Our results suggest that higher resolution stimulus facilitates depth perception. However, since the stimuli used in our experiment were quite simple for investigating the relation between stimulus resolution and depth perception, in the future study, it should be tested by the stimulus on basis of the feature of visual information processing (e.g. Gabor patch) if possible. Also, further physiological researches should be conducted to understand more detailed mechanism of this phenomenon, especially on the neural interactions between MT+ and other cortical areas, because MT neurons are functionally and anatomically linked to other areas[18,19,24,25]. Although significant fMRI activities in the other cortical areas were

not found in the present study, it is highly possible that higher-level cortical area such as orbitofrontal cortex is involved in this kind of cognitive faculty[26,27]. If activities in the orbitofrontal cortex are significantly related to our depth task, the spatial frequency of a visual stimulus manipulated by the display resolution would be a very important factor that causes our behavioral data[26,27]. Finally, we believe that investigation of neural system accompanied by new technologies will show us the future progress of our neural system as well as the new insight of our neural mechanism.

Methods

Psychophysical experiments of Depth task and Resolution Task. *Participants.* 20 participants, aged from 20 to 39 (13 females and 7 males), with normal or corrected vision, participated in a series of experiments. All participants gave written informed consent and the study was approved by the Ethics Committee of the NHK Science and Technology Research Laboratory, and in compliance with Declaration of Helsinki. 10 participants were assigned for Depth Task (6 females and 4 males), the other ten participants for Resolution Task (7 females and 3 males).

Apparatus. A 27″ IPS-TFT color LCD Monitor (ColorEdge CG275W,EIZO Nanao Corp.) was used to present stimuli. The display had an area of 2560×1440 pixels with the pixel size of 0.2331×0.2331 mm and the contrast ratio of 850 : 1. Color calibration was performed before experiments to correct color balance and display gamma. We used 256 gray levels (8-bit color depth) to present the stimuli. Visual stimuli were presented by using Psychtoolbox 3 (Psychophysics Toolbox, RRID:rid_000041) on Windows 7.

Stimuli and procedure

Visual Acuity Test. To confirm that participants have normal vision, we examined each participants' visual acuity with using Snellen eye chart. We used only "E" shape presented on the same setting (equipments and environment) as the main experiment.

Depth Task. In order to make the same images shown in different resolutions (Figure 1), we presented two same-size but different-resolution visual stimuli on the computer display (Figure 2a), which had the gradual luminance-contrast change from one side to the other for containing depth information (from right to left, or left to right)[7,11]. Although Masaoka et al. suggest less effects of "realness" with more than 60.0 cycle/degree (cpd) resolution stimulus[1], according to the result of our preliminary experiment, we presented five kinds of resolution stimuli, 15.0, 30.0, 52.5, 105.0, and 210.0 cpd, and three different contrast-step of stimuli sets, $\Delta 1$, $\Delta 2$, and $\Delta 3$(Figure 2b). We set the highest resolution stimulus as the base stimulus (the 210.0 cpd stimulus consisting of 28 sub-bars, each sub-bar was the same in size; right stimuli in each stimulus set at Figure 2b), and then it was downconverted using linear interpolation of the contrast, with fixing the highest luminance-contrast (left edge of each stimulus in Figure 2b). Here, we denoted the number of sub-bar by n, the contrast-step by a (from 1 to 3, $\Delta 1$ to $\Delta 3$ respectively), and the contrast value of i-th sub-bar by $Ci = 256 - (i - 1)(28a/n)$. For example, the highest resolution stimulus in $\Delta 1$ was (right in $\Delta 1$): the contrast change was from 256 to 229 gray levels (from 66.3 to 52.0 cd/m²). The lowest resolution stimulus in $\Delta 1$ was (left in $\Delta 1$): the contrast change was from 256 to 242 gray levels (from 66.3 to 58.6 cd/m²) (Figure 2b).

In each trial, two kinds of resolution stimuli were randomly chosen within each stimuli set, $\Delta 1$, $\Delta 2$, or $\Delta 3$. Before starting the experiment, the experimenter told the participants the idea of 'shading' as a depth cue, and we confirmed that they understood it. The participants were instructed to fixate the center and report which stimulus they perceived more depth with monocular viewing (Figure 2a). In a complete experiment, each stimulus set was repeated 32 times, so that a total experiment consisted of $_5C_2 \times$ three stimulus set ($\Delta 1$, $\Delta 2$, and $\Delta 3$) $\times$ 32 repetitions = 960 trials. The order of presentation of these conditions was randomly determined for each participant. In order to avoid that participants noticed the resolution difference between two stimuli, we did not tell them that there was resolution difference. No feedback was given to the participants.

Resolution Task. It was identical to Depth Task except that the participants were asked to report which stimulus had higher resolution. Before starting the experiment, the experimenter told what "resolution" is, and we confirmed that they understood it.

Measurement of the actual luminance. We used a photometer, Luminance Colorimeter (BM-7, Topcon Technohouse, Tokyo Japan) and measured the luminance of each component of the stimulus used in the experiment for five times, then calculated the mean value for each point of observation. Those values are shown in Figure 2b.

Psychophysical data analysis. To examine the relationship between depth perception and stimulus resolution, we used the method of paired comparison in which each stimulus is matched one-on-one with each of the other stimulus in our experiment. Thurstone-Mosteller (TM) model[28,29] (Case V) and Bradley-Terry (BT) model[29] are well-known paired comparison model that can convert the paired comparison data to psychophysical scale rating. Since BT model is more mathematically developed[30] and produces more robust estimates of confidence intervals than Thustone's Case V model[31], we used BT model to analyze the psychophysical data in this study.

MRI experiments of Depth task and Luminance task. *Participants.* 10 participants, aged from 20 to 39 (6 females and 4 males), with normal or corrected vision, participated in a series of experiments. All participants gave written informed consent and the study was approved by the Ethics Committee of the NHK Science and Technology Research Laboratory, and in compliance with Declaration of Helsinki. All participants did both depth and luminance tasks.

Apparatus. The experimental equipments were the same as the previous psychophysical experiments. However, the monitor was placed outside of the fMRI scanner, and the subjects viewed the stimuli by the mirror.

Stimuli and procedure

Depth Task. It was identical to the Depth Task used out of fMRI scanner, except that only 15.0, 30.0, 210.0 cpd in $\Delta 1$ and $\Delta 3$ were used. In a complete experiment, each stimulus set was repeated 60 times, so that a total experiment consisted of $_3C_2 \times$ two stimulus set ($\Delta 1$ and $\Delta 3$) $\times$ 60 repetitions = 360 trials.

Luminance Task. It was identical to the Depth Task in fMRI scanner except that the participants were asked to report which stimulus was darker.

fMRI Data acquisition. MRI data were obtained using a 3T MRI scanner (MAGNETOM Trio A Tim; Siemens, Erlangen, Germany) using a 12 channels head coil at the ATR Brain Activity Imaging Center (Kyoto, Japan). An interleaved T2*-weighted gradient-echo planar imaging (EPI) scan was performed to acquire functional images to cover the entire brain (TR, 2000 ms; TE, 30 ms; flip angle, 80°; FOV, 192×192 mm; voxel size, $3.5 \times 3.5 \times 4$; slice gap, 1 mm; number of slices, 30). T2-weighted turbo spin echo images were scanned to acquire high-resolution anatomical images of the same slices used for the EPI (TR, 6000 ms; TE, 57 ms; flip angle, 90°;FOV, 256×256 mm; voxel size, $0.88 \times 0.88 \times 4.0$ mm). T1-weighted magnetization-prepared rapid-acquisition gradient echo (MP-RAGE) fine-structural images of the whole head were also acquired (TR, 2250 ms; TE, 3.06 ms; TI, 900 ms; flip angle, 9°; FOV, 256×256 mm; voxel size, $1.0 \times 1.0 \times 1.0$ mm).

fMRI Data Analyses. Image preprocessing and statistical analyses were run by SPM8 (The Wellcome Trust Centre for Neuroimaging, UCL, SPM, RRID:nif-0000-00343). The acquired fMRI data underwent slice-timing correction and motion correction by SPM8. The data were then coregistered to the within-session high-resolution anatomical image of the same slices used for EPI and subsequently to the whole-head high-resolution anatomical image. Then, the images were spatially normalized to MNI template and their resampled voxel size as $3 \times 3 \times 3$ mm voxels. Finally, they were smoothed (The size of FWHM was 8 mm). Anatomical labels for region of interest analysis were defined by SPM anatomy toolbox (SPM Anatomy Toolbox, RRID:nif-0000-10477). Although the spatial location of the human middle temporal (MT+) was functionally defined in some studies, we anatomically determined it based on the previous research[32].

To compare fMRI activities at the depth task with ones at the luminance task, for each participant, fMRI activities at the luminance task were subtracted from ones at the depth task. They were analyzed only when the stimuli were presented on a display. Those individual data were used for group analyses.

1. Masaoka, K., Nishida, Y., Sugawara, M., Nakasu, E. & Nojiri, Y. Sensation of Realness from High-Resolution Images of Real Objects. *Trans.Broadcast., IEEE* **59**, 72–83 (2013).
2. Barlow, H. B., Blakemore, C. & Pettigrew, J. D. The neural mechanism of binocular depth discrimination. *J. Physiol. (Lond.)* **193**, 327–342 (1967).
3. Pettigrew, J. D., Nikara, T. & Bishop, P. O. Binocular interaction on single units in cat striate cortex simultaneous stimulation by single moving slit with receptive fields in correspondence. *Exp. Brain Res.* **6**, 391–410 (1968).
4. Ferris, S. H. Motion parallax and absolute distance. *J. Exp. Psychol.* **95**, 258–63 (1972).
5. Reichardt, W. & Poggio, T. Figure-ground discrimination by relative movement in the visual system of the fly. *Biol. Cybern.* **35**, 91–100 (1979).
6. Horn, B. K. P. Obtaining shape from shading information. *In*: The Psychology of Computer Vision, ed. P.H. Winston. McGraw-Hill (1975).
7. Ramachandran VS. Perception of shape from shading. *Nature* **331**, 163–6 (1988).
8. Hou *et al.*"Neural correlates of shape-from-shading." *Vision Res.* **46**, 1080–90 (2006).
9. Sawada, T. & Kaneko H. "Smooth-shape assumption for perceiving shapes from shading." *Perception* **36**, 403–15 (2007).
10. Welchman, A. E. *et al.* 3D shape perception from combined depth cues in human visual cortex. *Nat. Neurosci.* **8**, 820–27 (2005).
11. O'Shea, R., Blackburn, S. & Ono, H. Contrast as a depth cue. *Vision Res.* **34**, 1595–1604 (1994).
12. Bradley, R. A. Paired comparisons: Some basic procedures and examples, in *Handbook of Statistics*, vol. 4, Krishnaiah, P. R. & Sen, P. K. Eds. Amsterdam, The Netherlands: Elsevier, pp. 299–326 (1984).
13. Tong, F. Primary visual cortex and visual awareness, *Nat. Rev Neurosci.* **4**, 219–229 (2003).
14. Newsome, W. T. & Pare, E. B. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* **8**, 2201–211 (1988).

15. Movshon, J. A. & Newsome, W. T. Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J. Neurosci.* **16**, 7733–41 (1996).

16. Rees, G., Friston, K. & Koch, C. A direct quantitative relationship between the functional properties of human and macaque V5. *Nat. Neurosci.* **3**, 716–23 (2000).

17. DeAngelis, G. C., Cumming, B. G. & Newsome, W. T. Cortical area MT andthe perception of stereoscopic depth. *Science* **394**, 677–80 (1998).

18. Uka, T. & DeAngelis, G. C. Contribution of Area MT to stereoscopic Depth Perception: Choice-Related Response Modulation Reflect Task Strategy. *Neuron* Vol.42, 297–310 (2004).

19. Uka, T. & DeAngelis, G. C. Linking neural representation to function in stereoscopic depth perception: roles of the middle temporal area in coarse versus fine disparity discrimination. *J. Neurosci.* **26**, 6791–802 (2006).

20. Treue, S. & Maunsell, J. H. R. Attentional modulation of visual motion processing in cortical area MT and MST. *Nature* **382**, 539–54 (1996).

21. Seidemann, E. & Newsome, W. T. Effect of spatial attention on the responses of are MT neurons. *J. Neurophysiol.* **81**, 1783–1794 (1999).

22. Gibson, J. J. *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin (1979).

23. Eger, E., Henson, R. N., Driver, J. & Dolan, R. J. Mechanisms of top-down facilitation in perception of visual objects studied by fMRI. *Cereb Cortex.* **17**, 2123–133 (2007).

24. Parker, A. J. & Newsome, W. T. Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* **21**, 227–77 (1998).

25. Tsushima, Y., Sasaki, Y. & Watanabe, T. Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science* **314**, 1786–788 (2006).

26. Bar, M. A cortical Mechanism for Triggering Top-Down facilitation in Visual Object Recognition. *J. Cognit. Neurosci.* **15**, 600–609 (2003).

27. Bar *et al.* Top-down facilitation of visual recognition. *Proc. Natl Acad. Sci. U.S.A.* **103**, 449–54 (2006).

28. Thurstone, L. L. Psychophysical analysis. *American J.Psychol.* **38**, 368–89 (1927).

29. Thurstone, L. L. A low of comparative judgment. *Psychol. Rev.* **101**, 266–70 (1994).

30. Handley, J. C. Comparative analysis of Bradley-Terry and Thurstone-Mosteller paired comparison models for image quality assessment. *Proc. IS&T PICS Conf.* Montreal, Qubec, Canada, 108–12 (2001).

31. Morovic, J. [Psychovisual methods] *Color Gamut Mapping* [58–70] (Chichester, England: Wiley, 2008).

32. Malkovic, A. *et al.* Cytoarchitechtonic analysis of the human extrastriate cortex in the region of V5/MT+: a probabilistic, stereotaxic map of area hOc5. *Cereb Cortex* **17**, 562–74 (2007).

Acknowledgments

Author contributions

N.H. was involved in designing a research. Y.S. analyzed fMRI data. K.K. designed the study, collected and analyzed behavioral data. Y.T. designed the study, collected, analyzed behavioral and fMRI data, and wrote the paper. All authors discussed the results and commented on the manuscript.

Additional information