An integration of color and motion information in visual scene analyses

Noguchi, Yasuki
Shimojo, Shinsuke
Kakigi, Ryusuke
Hoshiyama, Minoru

(Citation)
Psychological Science,22(2):153-158

(Issue Date)
2011-02

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90002612

An integration of color and motion information in visual scene analyses

Yasuki Noguchi[1,2,3], Shinsuke Shimojo[1,4], Ryusuke Kakigi[5], and Minoru Hoshiyama[2]

[1]Division of Biology, California Institute of Technology, Pasadena, California, USA, [2]Department of Health Sciences, Faculty of Medicine, Nagoya University, Nagoya, [3]Present address: Department of Psychology, Graduate School of Humanities, Kobe University, Kobe, [4]ERATO Shimojo Implicit Brain Function Project, JST, [5]Department of Integrative Physiology, National Institute for Physiological Sciences, Myodaiji, Okazaki, Japan.

Address correspondence to Yasuki Noguchi, Department of Psychology, Graduate School of Humanities, Kobe University, 1-1 Rokkodai-cho, Nada, Kobe, 657-8501, Japan. E-mail: ynoguchi@lit.kobe-u.ac.jp

This manuscript includes: 12 pages, 2 figures (no table) and Supplementary Material

Abstract

To analyze complex scenes efficiently, our visual system performs perceptual groupings based on various features (e.g. color and motion) of visual elements in the scene. Although previous studies reported those groupings based on a single feature (either color or motion information), here we show that the visual system also performs the scene analyses based on a combination of two features. We presented a mixture of red and green dots moving in various directions. Although the pairings between color and motion information were variable across the dots (e.g. one red dot moved upward while another moved rightward), their subjective color-motion pairings were significantly biased when those dots were flanked by additional dots with coherent color-motion parings. These results indicate that the visual system resolves local ambiguities in the color-motion pairings using unambiguous pairings in surrounds, demonstrating a new type of the scene analysis based on the combination of two featural cues.

Introduction

Mounting evidence has indicated that the visual system of humans performs perceptual inferences about global structures of visual scenes. A typical example would be an analysis of three-dimensional (3D) structures of visual images. Although images projected to the retina are two-dimensional (2D), the neural system in the brain reconstructs them into 3D representations using various depth cues in the images, such as occlusion and convergence. The influences of those perceptual analyses are so strong that they could sometimes change subjective appearances of physically-identical stimuli even in the fovea. For example, a recent study reported that, when a multi-stable object (Necker cube) over a fixation point was flanked by unambiguous 3D cubes in the periphery, the percept of the ambiguous Necker cube was significantly biased so that it was consistent with the unambiguous cubes in the background (Sundareswara & Schrater, 2008). This suggested that the unambiguous objects in the periphery provided the information (cue) about an orientation of the ground plane of the scene, which was used by observers to infer the 3D structure of the ambiguous cube in the center. Other studies also provided evidence that the depth cues in 2D images altered the perceived lightness (Kitazaki, Kobiki, & Maloney, 2008; Anderson & Winawer, 2005) and size (Ono, 1966; Murray, Boyaci, & Kersten, 2006) of stimuli in the fovea.

Those perceptual inferences about the structures of the scenes are induced not only by the depth cues above, but also by simple features of visual elements in the scene, such as motion and color. As shown in **Figure 1a**, when the scene consists of multiple numbers of dots moving in either upward or downward direction, the visual system automatically groups dots moving in the same direction, segregating the whole scene into two surfaces with different directions of motion (the motion-based grouping or transparent motion (Wallach & O'Connell, 1953)). The same type of the perceptual grouping is also seen in color dimension, which decomposes two colors of dots into difference surfaces (**Fig. 1b**).

Although these examples of the visual grouping/segmentation provide the evidence for the analyses of the scene using a single cue (either color or motion information), here we examined whether the visual system could perform the scene analyses based on a combination of multiple cues (i.e. considering both color *and* motion information) embedded in the visual elements.

As shown in **Figure 1c,** our present stimuli consisted of three sets of moving dots located in the left, central, and right visual fields. In the central field, we presented a mixture of red and green dots moving in various directions. On the other hand, the left and right fields in the periphery contained either red or green dots, with all dots in each field moving in the same direction (either upward or downward). The point was that the pairings between color and motion information were unambiguous (coherent across all dots) in the peripheral fields, while they were ambiguous in the central field. Thus, if the visual system performs the analyses of the scene based on the combination of color and motion cues, the clear color-motion pairings in the peripheral fields would affect the perception of the dots in the central field (as in Sundareswara and Schrater (2008) above), changing the ambiguous color-motion pairings of those dots so as to be consistent with the peripheral ones (**Fig. 1d**). We tested below this possibility of the 'cue-combination' effect in the scene analyses by manipulating systematically the coherence of the color-motion paring in the central dots.

EXPERIMENT 1

Methods

Six subjects participated in Experiment 1 (5 males and 1 female, age: 21-30). Informed consent was received from each participant after the nature of the study had been explained. All visual stimuli were generated using Matlab Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), and were presented on a CRT monitor (refresh rate: 60 Hz). In each trial, we presented a number of moving dots in a rectangular area on the screen (width: 13.3 deg, length: 6.7 deg) for 500 ms. It consisted of three sets of moving dots (**Fig. 1c**). In the left field (the left surround), all dots (n = 150) were red (10.3 cd/m^2) and moved in the same direction (100 % coherence), either upward or downward (variable across trials). Likewise, the right field (the right surround) had the same number of green dots (10.3 cd/m^2) moving upward or downward (the dots in the left and right surrounds always moved in opposite directions, as described below). On the other hand, the central field over a fixation point was a mixture of red and green dots (n = 150 for each color) moving at the same speed as the dots in the surrounds. As shown in the left panel of **Figure 2a**, a coherence of the central red dots was varied from -100 – 100 % across trials. We named the coherence of those dots as 100 % when all moved upward and as -100 % when all moved downward. In the trials of 0 % coherence, there was no red dots moving upward or downward and all moved in one of six directions as follows; upward-right (45 deg), rightward (90 deg), downward-right (135 deg), downward-left (225 deg), leftward (270 deg), and upward-left (315 deg). On the other hand, directions of the central green dots were always random, being chosen from the eight directions from 0 to 315 degrees. The task of the subjects was to judge a direction of the central red dots (upward or downward), neglecting the green dots. We also instructed them to ignore all dots in the surrounds, informing them that the directions of those surrounding dots were determined independently from the directions of the central dots.

Based on the up/down responses of each subject, we measured the percentage that he/she answered ‘Up’ (%‘Up’) at each of six coherence levels (-100, -60, -20, 20, 60, and 100 %) of the target (central red) dots. As shown in the right panel of **Figure 2a**, the %‘Up’ generally increased as a function of the coherence. We then fitted this increase in %‘Up’ by a psychometric sigmoid function

$$F(x) = \text{Min} + (\text{Max} - \text{Min})/[1+e^{-a(x-b)}]$$

, where x was the coherence of the central red dots, a and b were free parameters estimated with the Nelder-Mead method. The Max and Min indicated the maximum and minimum %'Up' through the six coherence levels, respectively.

We tested the possibility of the cue-combination effect (see **Introduction**) by comparing the 50 % thresholds of those psychometric curves between two types of trials; Red-Up surround trials and Red-Down surround trials. In the Red-Up surround trials, the red dots in the left surround moved upward, while the green dots in the right surround moved downward. In the Red-Down surround trials, those red and green dots moved downward and upward, respectively. Since the central field was identical between those two types of trials, any differences in the subjects' performances could be attributed to a difference in the surrounds. Specifically, our prediction was that, if the color-motion pairings in the surrounds could affect the central ones, the Red-Up surrounds should bias the perceptual direction of the central red dots into upward, resulting in higher %'Up' at all coherence levels (Wu, Kanai, & Shimojo, 2004). The 50 % threshold of the psychometric curve, therefore, should decrease (shift leftward) in the Red-Up surround trials, while it should increase (shift rightward) in the Red-Down surround trials.

One experimental session consisted of 72 trials, including 24 Red-Up surround trials, 24 Red-Down surround trials, and 24 no-surround trials as a control. In the no-surround trial, the central field was solely presented and the subjects simply answered the direction of the red dots. Those three types of trials were randomly intermixed in one session, and the subjects underwent 2-5 sessions in one experiment. To control the difficulty of the task across the subjects, the speed of all dots in the screen was adjusted for each subject (range: 4.4 – 6.1 deg/s), based on a preliminary session (72 trials) before the experiment. We ensured that the subjects maintained the fixation throughout the trial by monitoring their eye positions at 250 Hz using the EyeLink II system (SR Research).

Results

As shown in **Figure 2b**, we found a significant decrease in the 50 % threshold when the surrounding red dots moved upward (Red-Up surround trials, red line) compared to when those moved downward (Red-Down surround trials, green line) (t = 2.96, p = 0.032, paired t-test). Those results indicated that the perceived direction of the central red dots was biased into the same direction as the red dots in the surround, thus supporting the possibility of the cue-combination effect in the visual scene analyses.

EXPERIMENT 2

The scheme in **Fig. 1d** assumed that the bias in the perceived direction of the central dots was induced by color-based linkages (grouping) between central and peripheral fields. A specific prediction from this scheme is that the direction bias in the central field would be diminished if the subjective linkages of those fields are disconnected. In Experiment 2, we examined this prediction by manipulating perceptual continuities between the central and peripheral fields.

Methods

Seven subjects participated in Experiment 2 (2 males and 5 females, age: 20-30). In 50 % of trials, vertical white lines (78 cd/m^2, 0.03 × 6.7 deg) were placed at the boundaries among the three fields, so that perceptual continuities among those fields were reduced (divided condition). No line was placed in the remaining half trials (uniform condition, same as Experiment 1). The task of the subjects was identical to Experiment 1. In each of the divided and uniform condition, we set the Red-Up and Red-Down surround trials. The resultant four types of trials (uniform/divided × Red-Up/Red-Down surrounds) were randomly intermixed in one session of 72 trials, and each subject performed up to 5 sessions. Other details were identical to Experiment 1.

Results

Results were shown in **Figure 2c**. For each subject, we quantified the magnitude of the direction bias by measuring the difference in the 50 % threshold (Δthreshold) between the Red-Up and Red-Down surround trials. In the uniform condition, those Δthresholds (Red-Down surround – Red-Up surround) were significantly larger than 0 (t = 3.3, p = 0.02, one-group t-test), replicating the results in Experiment 1. In contrast, the Δthresholds in the divided condition did not reach significance (t = 2.6, p > 0.05). A direct comparison between the uniform and divided conditions revealed a significant difference (t = 3.17, p = 0.025, paired t-test), indicating that the bias in the perceived direction of the central dots was substantially diminished in the divided condition These results showed that the perceptual continuities among the central field and surrounds played a key role in inducing the cue-combination effect observed in Experiment 1.

Discussion

In the present study, we observed that the perceived direction of the central red dots was biased into the same direction as the red dots in the surround. This indicated that the visual system resolved the ambiguity of the color-motion pairings in the central field using those pairings in the unambiguous peripheral fields. Compared to previous studies on the perceptual groupings and segmentations based on either color (Baylis & Driver, 1992; Harms & Bundesen, 1983; Kim & Cave, 2001) or motion information (Driver & Baylis, 1989; Wallach & O'Connell, D. N., 1953), our present results showed a new type of the visual scene analysis based on the combination of those two features.

An important point for an interpretation of the present results is that coherent (unambiguous) motion signals at one spot in the visual field generally bias an incoherent (ambiguous) motion at another spot into an *opposite* direction to the coherent motion. We confirmed this point in our supplementary experiments. In Additional Experiment 1b, we removed all green dots in the right surround, presenting only the left (red) surround with the ambiguous central field (**Fig. S1b** in Supplementary Material). The data showed that the perceived direction of the central red dots was strongly biased into an opposite direction to the surround (i.e. the upward red motion in the surround biased the perceptual direction of the central red dots into downward, not upward). Those results in the supplementary experiment would reflect the mechanism of induced motion (Duncker, 1929) in which coherent motion signals in a background induce illusory motion of static stimuli in an opposite direction to the background. In the case of the present study, strong (unambiguous) motion signals in the red surround induced the illusory motion of the central red dots with a weak (ambiguous) motion signals, producing the results as **Fig. S1**.

This effect of the induced motion, however, was eliminated in Experiments 1 and 2 (**Fig. 2**), because opposite directions of motions in the left and right surrounds in those experiments cancelled the background motion. Consequently, another mechanism based on the perceptual groupings biased the percepts of the central red dots into the *same* direction as the surround. As illustrated in **Fig. 1d**, presenting two colors of dots initially provoked a segregation of two surfaces for each color. Subsequent signals of motion information, therefore, could work as an additional cue for this surface segregation. If clear (100 % coherent) motion signals with different directions were presented in the two surrounds, the segregation of red and green surfaces should be strengthened. Thus, the unambiguous motion signals in the surrounds spread into the ambiguous central field through the surface-based (color-based) linkages, biasing the motion of the central dots into the same, not opposite, direction as the dots in the surround.

Another important point in the present study is that the direction bias in the central red dots was substantially diminished when the perceptual continuities between the central and peripheral fields were disconnected (Experiment 2). This provided further evidence that the present cue-combination effect was mediated by the mechanism of the perceptual groupings. Although a classical framework of psychology argues that an integration of color and motion information requires attention (Treisman & Gelade, 1980), our present results suggest a possibility that the information about color and motion of the visual elements could be (partly) combined without attention, which helps the visual system analyze the structures of the scenes. In fact, we conducted several additional experiments to investigate an involvement of attention in the present effect (see Supplementary Material and **Fig. S3-S4**). The results of those supplementary experiments altogether indicated a limited role of attention in the present phenomenon, supporting a possibility of a pre-attentive combination of the color and motion cues in the scene analyses.

Acknowledgments - We thank Mr. Y. Takeshima (National Institute for Physiological Sciences, Japan) and Dr. Yohsuke Ohtsubo (Kobe University, Japan) for their technical supports. This work was supported by grants from the Japan Society for the Promotion of Science for Young

Scientists to Y.N and ERATO project, JST to S.S.

References

Anderson, B.L., & Winawer, J. (2005). Image segmentation and lightness perception. *Nature, 434*, 79-83.

Baylis, G.C., & Driver, J. (1992). Visual parsing and response competition: the effect of grouping factors. *Perception & Psychophysics, 51*, 145-162.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10,* 433-436.

Driver, J., & Baylis, G.C. (1989). Movement and visual attention: the spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception & Performance, 15,* 448-456.

Duncker, K. (1929). Uber induzierte Bewegung. *Psychologische Forschung, 12*, 180-259.

Harms, L., & Bundesen, C. (1983). Color segregation and selective attention in a nonsearch task. *Perception & Psychophysics, 33*, 11-19.

Kim, M.S., & Cave, K.R. (2001). Perceptual grouping via spatial selection in a focused-attention task. *Vision Research*, *41*, 611-624.

Kitazaki, M., Kobiki, H., & Maloney, L.T. (2008). Effect of pictorial depth cues, binocular disparity cues and motion parallax depth cues on lightness perception in three-dimensional virtual scenes. *PLoS.One., 3,* e3177.

Moutoussis, K., & Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision.

Proceedings of the Royal Society B: Biological Sciences, 264, 393-399.

Murray, S.O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience, 9*, 429-434.

Ono, H. (1966). Distal and proximal size under reduced and non-reduced viewing conditions. *American Journal of Psychology, 79,* 234-241.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Sundareswara, R., & Schrater, P.R. (2008). Perceptual multistability predicted by search model for Bayesian decisions. *Journal of Vision, 8*, 12-19.

Treisman, A.M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Wallach, H., & O'Connell, D.N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45*, 205-217.

Wu, D.A., Kanai, R., & Shimojo, S. (2004). Vision: steady-state misbinding of colour and motion. *Nature, 429*, 262.

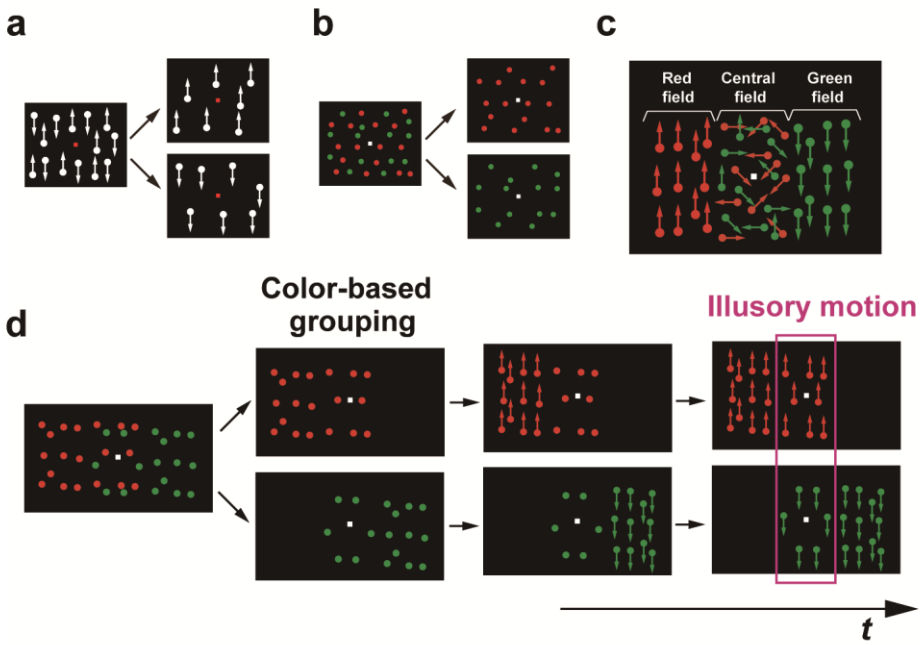

Figure 1. Visual groupings and segmentations based on motion and color cues. (**a**) The motion-based grouping (transparent motion). (**b**) The color-based grouping of multiple dots. (**c**) An example display of the present stimuli to examine the possibility of the cue-combination effect (see **Introduction**). A motion field consisting of red and green dots over the fixation was flanked by the surrounds where a set of red or green dots moved in 100 % coherence (upward or downward). (**d**) A prediction from the cue-combination effect in visual scene analyses. We assume here that the neural processing of color information of the stimuli precedes that of motion information, in conformity to a previous study (Moutoussis & Zeki, 1997). When the stimuli as **Figure 1c** are given, the visual system initially groups dots with the same color, segregating a whole scene into two surfaces of different colors. When motion processing is conducted later in each surface, the clear motion signals in the surround could spread into the ambiguous central field *through a color-based linkage or grouping*, producing illusory motions in the central vision. On the other hand, if there is no interaction between color- and motion-based processing, such spreading of motion signals should not be expected, because no color-based linkage between the central field and surrounds is available for the motion processing.

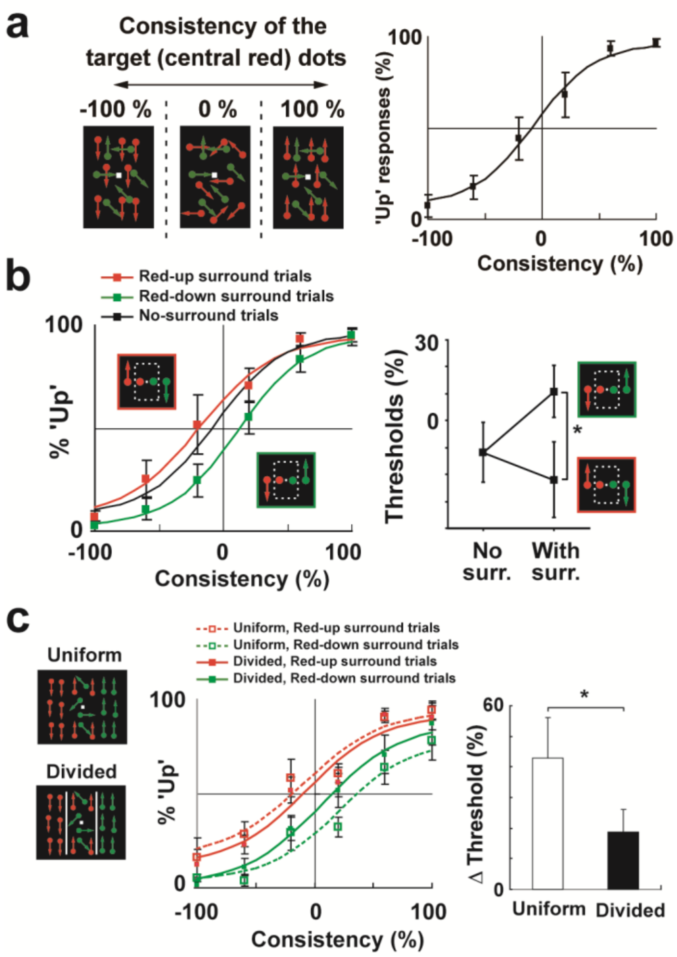

Figure 2. Experiment 1 and 2. (**a**) Structures of the central field. The coherence of the target (central red) dots was manipulated from -100 % (all downward) to 100 % (all upward) across trials, while the non-target (central green) dots always moved in random directions (0 % coherence). The percentage that the subjects answered 'Up' (%'Up') increased as a function of the coherence of the target dots (right panel). (**b**) Experiment 1. Compared to no-surround trials (black line, the same data as **Fig. 2a**), the surrounds with upward red (and downward green) dots generally increased %'Up' (red line), whereas the surrounds with downward red (and upward green) dots decreased %'Up' (green line). Those results were consistent with the prediction from the cue-combination effect (**Fig. 1d**). The right panel showed the 50 % thresholds of the psychometric curves in those three conditions. (**c**) Experiment 2. While the stimuli in the uniform condition were identical to Experiment 1, two white lines were placed at the boundaries among the three fields in the divided condition,

so that perceptual continuities across fields were diminished. Middle panel: The psychometric curves in four types of trials (uniform/divided × Red-Up/Red-Down surrounds). Right panel: Differences in the 50 % threshold (Δthreshold) between Red-Up and Red-down surround trials, as indices for magnitudes of the direction bias in the central red dots. The bias was significantly diminished in the divided condition, showing that visual grouping (based on the perceptual continuities between the central field and surrounds) played a critical role in inducing the cue-combination effect. All error bars denote SE across the subjects. * $p < 0.05$