



Combination of texture and color cues in visual segmentation

Toni P. Saarela^{*}, Michael S. Landy

Department of Psychology and Center for Neural Science, New York University, New York, NY, USA

ARTICLE INFO

Article history:

Received 8 December 2011
Received in revised form 16 January 2012
Available online 24 February 2012

Keywords:

Visual segmentation
Texture
Color
Cue combination
Signal detection theory

ABSTRACT

The visual system can use various cues to segment the visual scene into figure and background. We studied how human observers combine two of these cues, texture and color, in visual segmentation. In our task, the observers identified the orientation of an edge that was defined by a texture difference, a color difference, or both (cue combination). In a fourth condition, both texture and color information were available, but the texture and color edges were not spatially aligned (cue conflict). Performance markedly improved when the edges were defined by two cues, compared to the single-cue conditions. Observers only benefited from the two cues, however, when they were spatially aligned. A simple signal-detection model that incorporates interactions between texture and color processing accounts for the performance in all conditions. In a second experiment, we studied whether the observers are able to ignore a task-irrelevant cue in the segmentation task or whether it interferes with performance. Observers identified the orientation of an edge defined by one cue and were instructed to ignore the other cue. Three types of trial were intermixed: neutral trials, in which the second cue was absent; congruent trials, in which the second cue signaled the same edge as the target cue; and conflict trials, in which the second cue signaled an edge orthogonal to the target cue. Performance improved when the second cue was congruent with the target cue. Performance was impaired when the second cue was in conflict with the target cue, indicating that observers could not discount the second cue. We conclude that texture and color are not processed independently in visual segmentation.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Humans often combine multiple sensory cues to improve perceptual performance. In many cases, human cue integration is optimal in the sense of achieving maximal reliability (e.g., Ernst & Banks, 2002; Landy et al., 1995). Near-optimal cue integration has been demonstrated for several tasks including estimation of size, slant, shape and location, for multiple visual cues (Hillis et al., 2004; Knill & Saunders, 2003; Landy & Kojima, 2001) and for combinations of cues from more than one sensory modality (Alais & Burr, 2004; Ernst & Banks, 2002; Gepstein & Banks, 2003; Hillis et al., 2002).

In this paper, we examine the task of visual segmentation, i.e., detection and identification of edges. Many cues may help the viewer to segment the visual scene into figure and ground, including differences in luminance, contrast, texture, color, motion, and depth (e.g., Braddick, 1993; Landy & Graham, 2004; Li & Lennie, 2001; Nakayama, Shimojo, & Silverman, 1989). The larger the difference between two stimulus regions along any of these dimen-

sions, the easier they are to segment. For example, when two regions differ in dominant pattern orientation, segmentation is easier when the difference in orientation is increased (Landy & Bergen, 1991; Nothdurft, 1985; Wolfson & Landy, 1995). But how are regions segmented when multiple cues are available to signal the region boundary, e.g., texture and color? A simple, but inefficient, solution is to use only one of the cues and ignore the other, perhaps concentrating on the more reliable cue (choosing one of the “streams” in Fig. 1A). Alternatively, one could process the cues independently, and signal the presence of an edge if any cue signaled that an edge was detected or base the decision on a combination of the outputs (Fig. 1B). Or, the effects of the different cues could summate, so that a texture cue would somehow *add* to a color cue in signaling the boundary. Such a model would require a segmentation mechanism capable of using information from multiple cues (Fig. 1C), as has been demonstrated for combinations of visual (printed words) and auditory (spoken words) for text recognition (Dubois, Poeppel, & Pelli, in preparation). For a 2-alternative multi-cue detection or discrimination task (as in the experiment we discuss below), if the two alternatives share a common Gaussian noise covariance, cue summation is the most effective mechanism for combining multiple cues (Duda, Hart, & Stork, 2001).

^{*} Corresponding author. Present Address: Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA.

E-mail address: saarela@cns.nyu.edu (T.P. Saarela).

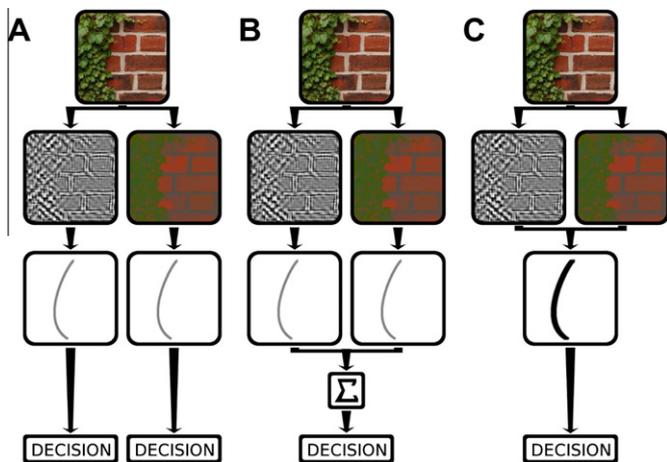


Fig. 1. Three schematic models of the processing of texture and color in visual segmentation. (A) An image (top) contains an edge signaled by changes in both texture and color (second row). Texture and color are processed by separate and independent segmentation mechanisms that extract the edge (third row), with independent decision stages (bottom row). (B) Texture and color are processed by separate and independent segmentation mechanisms, but the outputs of the two mechanisms are subsequently combined. (C) Texture and color are processed by a single segmentation mechanism that sums responses to color and texture differences.

We concentrate here on the combination of color and texture cues. Both color and texture can provide information for visual segmentation. Both are also robust segmentation cues when luminance information is varying; Segmentation based on color is little affected by variations in luminance (Hansen & Gegenfurtner, 2006; Li & Lennie, 2001), and the texture information in natural images is only partially correlated with pure luminance information, so that it likely provides additional segmentation information (Schofield, 2000). We test the independence of texture and color processing in segmentation using a simple task where observers identify the orientation of an edge. In our stimuli, both the texture- and color-defined edges are second-order: The average luminance and chromaticity on either side of each edge is identical, and only the spatial pattern of luminances or chromaticities varies across these second-order boundaries. We first measure performance in the segmentation task with either cue alone. We then measure performance when both cues are present, with the texture and color edges aligned (cue combination) or orthogonal (cue conflict). This design with cue-combination and cue-conflict conditions is similar to those used in testing the independence of spatial frequency and orientation channels in human vision (e.g., Olzak & Thomas, 1991). Independent processing of texture and color would predict equal performance in the cue-combination and cue-conflict conditions. We find substantial improvement in the cue-combination condition relative to the single-cue conditions but no improvement in the cue-conflict condition. This argues against independent processing. Further, observers perform better in the cue-combination condition than the “optimal” independent-processing prediction based on single-cue performance. A signal-detection model incorporating interactions in the processing of texture and color edges can account for all observations. In the second experiment we report, observers perform the segmentation task using a single cue (texture or color). On a given trial, the second cue can be absent, aligned with, or orthogonal to the target cue. We find that observers perform better on trials where the second cue is aligned and worse on trials when the sec-

ond cue is orthogonal to the target cue. The observers thus cannot ignore the other cue even when instructed to do so.

2. Methods

2.1. Observers

Three observers (ages 21–33 years, 2 female) participated in the experiments. All observers had normal, uncorrected visual acuity and color vision.

2.2. Equipment

The stimuli were presented on a Mitsubishi Diamond Pro 900u CRT monitor that was driven by a 10-bit, NVIDIA GeForce 7300 GT graphics card. The screen had a resolution of 1024×768 pixels, and an 85 Hz refresh rate. From the viewing distance of 57 cm used in the experiment, the screen subtended 34×25.5 deg. The monitor phosphor spectra were measured with a Photo Research PR-650 SpectraScan spectroradiometer. The gamma function for each gun was measured using a Minolta LS-100 photometer. Mean luminance of the screen was about 30 cd/m^2 .

2.3. Stimuli

Stimuli were 12×12 arrays of Gabor patches. The patch in array position (i, j) was defined as

$$G_{ij}(x, y) = A \exp(-(x'^2 + y'^2)/2\sigma^2) \sin(2\pi f x'),$$

$$x' = (x - x_j) \cos(\theta_{ij}) - (y - y_i) \sin(\theta_{ij}),$$

$$y' = (x - x_j) \sin(\theta_{ij}) + (y - y_i) \cos(\theta_{ij}),$$
(1)

where G_{ij} describes the modulation in a given direction in cone-contrast space (see below) of a Gabor that is centered at array element location (x_j, y_i) , and has an orientation of θ_{ij} and modulation contrast controlled by A . The spatial frequency f of the Gabors was 1 cycle/deg, the space constant σ was 0.35° , and the center-to-center spacing of the Gabors in the array was 2° .

In each stimulus, there were two types of Gabors (four in the cue-conflict stimulus, see below), which differed from each other in orientation, in color, or both. These Gabors were arranged in alternating vertical or horizontal stripes of each Gabor type (each stripe was three Gabors wide, giving a spatial frequency of 2 cycle/image, and stripe phase was chosen randomly from among the six possible phases). When there was no orientation difference between the Gabors in adjacent stripes, all Gabor orientations were 45° . When there was an orientation difference, the two Gabor orientations were $45 \pm \Delta\theta$ deg. When there was no color difference between the Gabors in adjacent stripes, all Gabors were achromatic with a carrier luminance contrast of 0.25 (corresponding to a peak luminance contrast of 0.17). The color difference was a red-green modulation added to the luminance modulation. The color modulation was in opposite phase between adjacent stripes, resulting in dark-red/bright-green and bright-red/dark-green Gabors. Thus, the average luminance of each sine-phase Gabor was equal to the background luminance, and the average chromaticity of each Gabor was achromatic. In this sense, the stripes were a second-order stimulus, differing only in the pattern of luminance and chromaticity, but not in their average luminance or chromaticity (see below how the stimulus colors were defined in cone-contrast space).

There were four stimulus conditions. In the single-cue conditions, the stripes were defined by a single cue, texture or color (Fig. 2A and B). In the cue-combination condition (Fig. 2C), the

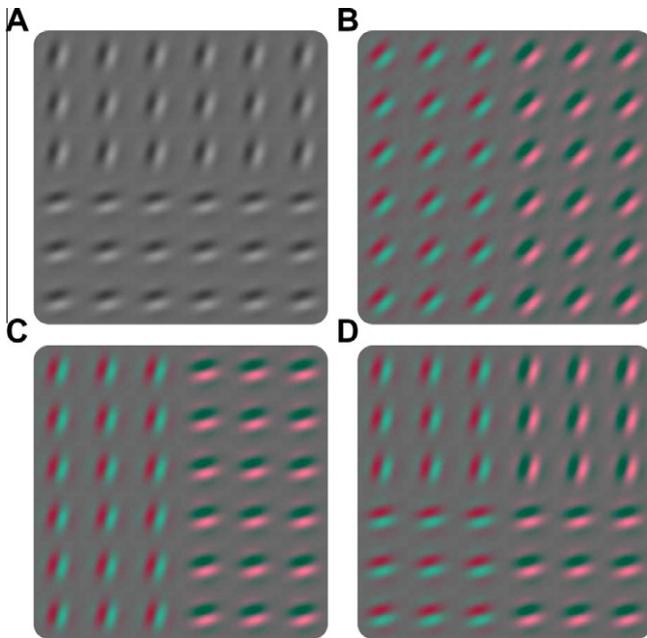


Fig. 2. Example stimuli. (A) Texture-cue only. The edges are second-order edges defined by an orientation difference. (B) Color-cue only. All the Gabor patches have the same orientation, and the edges are defined by color differences (dark-red/bright-green vs. bright-red/dark-green Gabors). (C) Cue combination. The edges are defined by *both* texture and color cues. (D) Cue conflict. Both texture and color cues are present, but the edges they define are orthogonal to each other. Note that the cue-combination and cue-conflict stimuli contain the same texture and color differences, only the alignment of the edges differs. All panels show cropped versions of the actual stimuli used, which were 12×12 arrays of Gabors. The “stripes” were three Gabors wide, so there were four stripes—or two cycles—per stimulus. The phases of the texture/color edges were randomized during the experiments.

stripes were defined by both texture and color cues (i.e., the two cues were spatially aligned). In the cue-conflict condition (Fig. 2D), there were four types of Gabors (both orientations, each with both color phases) arranged so that the orientation cue created one stripe orientation (e.g., horizontal) and the color cue was arranged in orthogonal stripes (in this case, vertical). Note that the cue-combination and cue-conflict stimuli nonetheless have exactly the same texture and color differences, only the alignment of the cues differs. As we have two stimulus dimensions of interest (texture and color), the four stimulus conditions can be represented in a two-dimensional stimulus space (Fig. 3A). In this stimulus space, color contrast of an edge is represented on the x-axis

and orientation contrast of an edge is represented on the y-axis. Positive values correspond to vertical edges and negative values correspond to horizontal edges. The single-cue conditions lie on the two axes (as the value of the other cue is zero), and the cue-combination and cue-conflict conditions lie in the four quadrants.

The stimulus colors were defined in cone-contrast space, where each of the three axes corresponds to the relative activation of each of the three cone classes with respect to the background (Cole, Hine, & McIlhagga, 1993; Sankeralli & Mullen, 1996, 1997). Cone excitations were computed using the Stockman and Sharpe (2000) 10-degree cone fundamentals. Cone contrasts C_L , C_M , and C_S (for long-, middle-, and short-wavelength-sensitive cones, respectively) were computed for cone excitations L , M , and S as $C_L = (L - L_0)/L_0$, $C_M = (M - M_0)/M_0$, and $C_S = (S - S_0)/S_0$; where L_0 , M_0 , and S_0 are the cone excitations in response to the gray background. Achromatic Gabors were defined along the direction $(C_L + C_M + C_S)$, which isolates the $L + M$ (“luminance”) mechanism. For the chromatic Gabors, a chromatic modulation was added to the achromatic one. The direction of this chromatic modulation was the $L - M$, or “red–green” isolating direction, which was determined individually for each subject (see below). Choosing the chromatic modulation as the direction that isolates the $L - M$ mechanism assured that the chromatic and achromatic Gabors had the same luminance contrast (the $L - M$ isolating direction is by definition a “null” direction for the luminance mechanism).

2.4. Determining $L - M$ isolating stimuli

The $L - M$ isolating direction in cone-contrast space was determined using the minimum-motion paradigm (Anstis & Cavanagh, 1983). The stimulus was a 12×12 array of Gabor patches with the same spatial parameters as in the main experiment. The Gabor envelopes were static, but their sine wave carriers drifted at 1 Hz. The observer adjusted the relative weights of C_L and C_M to minimize perceived motion. The value of C_S was always chosen so that the modulation was orthogonal to the $S - 0.5 * (L + M)$, or “blue–yellow”, mechanism (that is, the colors were confined to a plane orthogonal to the direction $C_S - 0.5 * (C_L + C_M)$). The pooled cone contrast $\left((C_L^2 + C_M^2 + C_S^2)^{0.5} \right)$ was kept constant during the adjustment, so changing the C_L and C_M weights rotated a vector on that plane. Three cone-contrast levels were used, and the observer made 10 adjustments at each level. The agreement between these 30 adjustments was very good when plotted in cone-contrast space. The $L - M$ isolating direction was determined by fitting a straight line to the data in cone-contrast space, which gave a good fit ($r^2 = 0.79\text{--}0.91$).

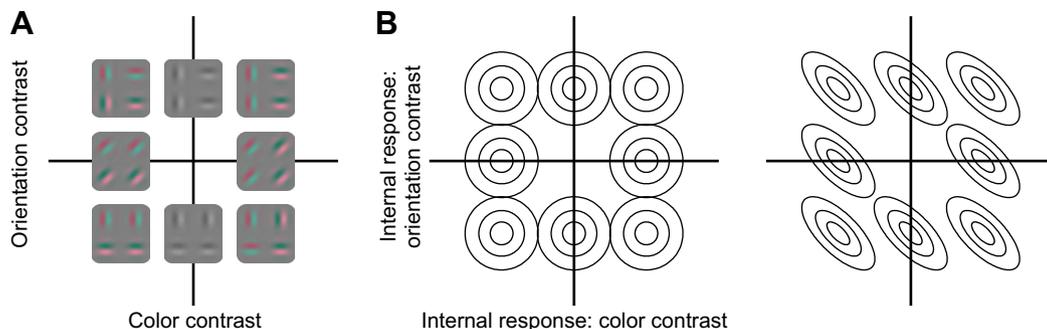


Fig. 3. Signal-detection models for our tasks. (A) Stimulus space for the experiment. The abscissa indicates the color contrast (positive for vertical edges, negative for horizontal edges) and the ordinate indicates the orientation difference. The observer’s task was to discriminate between pairs of stimuli located along the two axes (single-cue conditions) or along the diagonals (two-cue conditions). (B) Two model decision spaces. The horizontal axes show the differential response to the vertical and horizontal color-defined edges, the vertical axis for the texture-defined edges. Left panel: independent processing of the texture and color cues; right panel: non-independence of the two cues. The non-independence is reflected in the non-zero covariance, which makes the iso-probability contours elliptical and “tilted” in this decision space.

2.5. Procedure

2.5.1. General procedure

The observer was seated 57 cm from the monitor in a dark room. The observer's task was to identify the orientation—vertical or horizontal—of the stripes defined by texture, color, or both. Performance was measured using a single-interval design. On each trial, a single stimulus was presented for 247 ms (21 refresh cycles) in the center of the screen. The observer made a binary judgment about which of the two possible stimuli had been presented. The inter-trial interval was 1000 ms, during which a fixation dot was visible in the middle of the screen. Auditory feedback was provided after each trial.

There were equal numbers of vertical and horizontal stimuli in each block of trials, and their order of presentation was randomized. The phase of the stripes was randomized across trials.

2.5.2. Determining orientation and color contrast

In a preliminary experiment, observers practiced extensively on the single-cue conditions. The orientation and color differences were varied using two interleaved staircases in each block (one 1-up-2-down and one 1-up-3-down staircase were used to adjust orientation contrast $\Delta\theta$ or chromatic contrast) during these preliminary sessions. Psychometric functions (Weibull) were fit to the resulting data by maximum likelihood (Wichmann & Hill, 2001) and the fits were used to estimate values for the orientation and chromatic contrast that would lead to a performance level of $d' = 1$ (which corresponds to 69% correct in our single-interval task). These values were used in the main experiments.

2.5.3. Experiment 1

Each block of trials consisted of 10 practice trials, which were not included in the final analysis, followed by 120 experimental trials. The four stimulus conditions (texture-only, color-only, cue-combination, and cue-conflict) were blocked, that is, the cue did not vary from trial to trial within a block, and the observer always knew what the relevant cue was. Each block was repeated 6–10 times, depending on observer's availability.

In the cue-conflict condition it is not possible to give a simple vertical/horizontal response because the stimulus has both a vertical and a horizontal edge, one defined by texture, the other defined by color. However, the conditions were blocked so that the observer always knew what the two possible stimuli were. The task used in all conditions is essentially an identification task (two possible stimuli and two possible response categories). For the cue-conflict condition these categories were color-vertical/texture-horizontal and color-horizontal/texture-vertical. The observers were trained in the identification task before the actual experiments, and the response time was not limited, giving observers adequate time to determine their response. The assignment of response buttons in the cue-conflict blocks was consistent with the texture-only task, which might favor attending only to the texture cue in the conflict task. However, none of the observers reported any confusion about the task or the mapping of the response keys in the cue-conflict blocks.

2.5.4. Experiment 2

The procedure in Experiment 2 was similar to that in Experiment 1 with the following exceptions. There were two kinds of trial block. In half of the blocks observers performed the task based on the texture cue alone. In the other half, they performed the task using only the color cue. Within a block, there were three types of trial. As an example, consider the blocks where the target cue was texture. First, there were “neutral” trials, on which the color cue was absent (single-cue, texture-only stimuli). Second, there were “congruent” trials, where the color cue was present and

defined the same edges as the texture cue (cue-combination stimuli). Finally, there were “conflict” trials, where the color cue was present and defined edges orthogonal to the texture edges (cue-conflict stimuli). In the blocks where color was the target cue, the three types of trial were analogous to the ones described above for the texture blocks. There were an equal number of each trial type within a block. The observers knew this and they were told to respond only to the target cue and ignore the other cue. The three types of trial were randomly intermixed within a block.

2.6. Data analysis

2.6.1. Experiment 1

The data from each condition were pooled across blocks. We computed d' for each condition in the standard way:

$$d' = \Phi^{-1}(P_{\cdot V|V}) - \Phi^{-1}(P_{\cdot H|H}), \quad (2)$$

where “V”|V indicates responding “vertical” given a vertical stimulus (a “hit”), “V”|H indicates responding “vertical” given a horizontal stimulus (a “false alarm”), and Φ^{-1} is the inverse cumulative normal distribution. Note that with our identification task, the assignment of hits and false alarms is arbitrary (and one would get the same result using $P_{\cdot H|H}$ and $P_{\cdot V|V}$). We bootstrapped 95% confidence intervals for each d' by resampling the data with replacement 10,000 times and computing a distribution of d' -values from the resampled hit and false alarm rates.

We compared the performance in the two-cue conditions against two predictions computed from the single-cue conditions. First, to test whether the data are consistent with independent processing of texture and color, we computed a prediction for the two-cue conditions based on single-cue performance:

$d'_{\text{ind}} = \sqrt{d'_{\text{texture}}^2 + d'_{\text{color}}^2}$ (Green & Swets, 1988). This prediction is based on an assumption of independent processing and optimal integration of the two cues. It gives an upper bound for performance if the processing of the cues is independent. Better performance indicates non-independence of the two cues. The second prediction assumes perfect summation of the cues, that is,

$$d'_{\text{sum}} = d'_{\text{texture}} + d'_{\text{color}}.$$

We also fit a two-dimensional signal-detection model to the observed response rates from all conditions (Fig. 3B). The two dimensions of the model correspond to the internal responses to color and texture edges. We assume that for both color and texture, there are detectors responding to vertical and horizontal edges. Sensitivity on each dimension is determined by the difference in the responses of these two detectors, so that responses to vertical edges in the model are positive and responses to horizontal edges are negative. Sensory responses on a single trial, in which a single stimulus is presented, correspond to a point in this space, and the observer's task is to decide which of the distributions gave rise to that particular response. The internal response distributions associated with stimulus presentations were modeled as bivariate normal distributions. For example, for the single-cue color-only condition there is one distribution corresponding to each possible color-cue-only stimulus. The means of these distributions are $(\mu_{\text{color}}, 0)$ for the vertical and $(-\mu_{\text{color}}, 0)$ for the horizontal stimulus. Similarly, the means of the distributions corresponding to the presentation of the texture-cue-only stimuli are $(0, \mu_{\text{texture}})$ and $(0, -\mu_{\text{texture}})$. The means of the cue-combination distributions are then $(\mu_{\text{color}}, \mu_{\text{texture}})$ and $(-\mu_{\text{color}}, -\mu_{\text{texture}})$, and those of the cue-conflict distributions $(\mu_{\text{color}}, -\mu_{\text{texture}})$ and $(-\mu_{\text{color}}, \mu_{\text{texture}})$. In this space, the distribution corresponding to an achromatic stimulus with no texture edge would be centered at the origin. The marginal variances were fixed at unity. We compared two models: (1) a “separable-dimensions” model, with zero covariance between the two dimensions, that is, independent processing of the two cues

(Fig. 3B, left panel), and (2) an “integral-dimensions” model, in which the covariance was allowed to take on nonzero values and thus reflect non-independent processing (Fig. 3B, right panel).

In each experimental condition, the observer indicated which of two possible stimuli had been presented. We modeled the decision as based on the likelihood ratio of the two possible stimuli, with no response bias. Thus, there were two free parameters in the separable model (the means, μ_{color} and μ_{texture}) and three free parameters in the integral model (including the covariance, ρ). The model was fit to the data by maximum likelihood. The two models are nested (the separable model is a constrained version of the integral model), and we compared their ability to account for the data with a likelihood ratio test (Mood & Graybill, 1963).

Although the model axes give the distances between the responses to the vertical and horizontal stimuli, the use of this decision space does not require the assumption of an explicit stage where a difference between vertical and horizontal detector responses is computed. A model with four dimensions—corresponding to responses of mechanisms tuned to vertical color edges, horizontal color edges, vertical texture edges, and horizontal texture edges—would still predict, for example, the square-root improvement in the two-cue conditions for d'_{ind} . The space used here is effectively a projection of that four-dimensional space onto two dimensions that show the distances relevant to the task.

2.6.2. Experiment 2

The data from each condition were pooled across blocks. We computed d' and confidence intervals as in Experiment 1.

3. Results

3.1. Experiment 1

Performance in the two single-cue conditions was roughly equal for each subject and near $d' = 1$, as intended (Fig. 4). Performance in the cue-combination condition was better than in the single-cue conditions. Performance in the cue-conflict condition was not, however, improved compared to the single-cue conditions. Based on the performance in the single-cue conditions, we computed two predictions for the two-cue conditions. The predictions are indicated by horizontal lines in Fig. 4. The dashed line shows d'_{ind} , which assumes independent processing and optimal integration of the two cues. The solid line shows d'_{sum} , which assumes perfect summation of the two cues (Green & Swets, 1988). The bootstrapped 95% confidence intervals are shown by the error bars (for data points) and shaded areas (for predictions).

We tested for significant differences from the predictions with a Monte Carlo permutation test. We resampled the observed hit and false-alarm data 10,000 times. On each iteration, we computed d'

predictions for the cue-combination and cue-conflict conditions based on the single-cue data, as well as new d' values from the resampled cue-combination and cue-conflict data. We then constructed a distribution for the differences in d' between predictions and data. From this distribution, we computed 95% confidence intervals for the differences and tested whether this interval contained zero. All three observers performed significantly better than the two-cue prediction d'_{ind} (two-tailed $p < 0.05$) in the cue-combination condition, and significantly worse than the prediction (two-tailed $p < 0.05$) in the cue-conflict condition, although these two stimuli (combination and conflict) had exactly the same texture and color differences. On the other hand, none of the observers reached the perfect-summation prediction d'_{sum} in the cue-combination condition (the difference was significant for one of the three observers, two-tailed $p < 0.05$); observed performance was always between d'_{ind} and d'_{sum} .

The fact that performance in the cue-conflict condition was worse than in the cue-combination condition indicates that observers could only integrate texture and color information when the two were spatially aligned. The fact that performance in the cue-combination condition was better than predicted assuming independent processing suggests that texture and color are not processed independently of each other. To further investigate this possibility, we fit two nested versions of a two-dimensional signal-detection model to the data by maximum likelihood (Fig. 3B). Fig. 5 shows how well each model accounted for the sensitivity data. The fit of the integral model is extremely good for each observer: all open symbols lie on or near the diagonal. The model accurately accounts for the greater sensitivity in the cue-combination condition and the lack of improvement in the cue-conflict condition. The separable model, on the other hand, consistently predicts too low a sensitivity in the cue-combination condition and too high a sensitivity in the cue-conflict condition. Fig. 6 shows the observed and modeled probabilities for hits and false alarms. Here, the agreement between the model and data is better for both models, and all the points lie near the diagonal. The reason why these small deviations from the diagonal lead to large differences in sensitivity with the separable model is that it tends to under-estimate $p(\text{hit})$ and simultaneously over-estimate $p(\text{FA})$ for cue-combination, and vice versa for cue-conflict. The integral-model fits are significantly better than the separable-model fits as compared using a likelihood ratio test ($p < 0.01$ for each observer). We also fit both models with an additional bias parameter to account for non-optimal criterion placement. Adding this parameter did not significantly improve the overall fits and did not change the main finding: the integral model fits the data better than the separable model.

Fig. 7 shows the best-fitting decision space for each observer. The correlation between the two dimensions is also shown for each observer (the correlation is the same as the covariance because the variance along each axis was fixed at 1). The correlation is non-zero

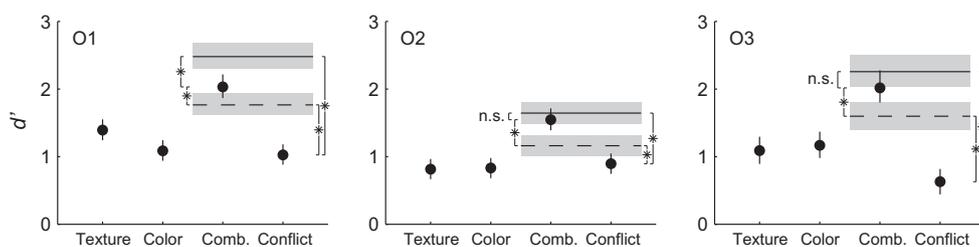


Fig. 4. Sensitivity (d') in the segmentation task for the three observers in the four stimulus conditions. The dashed horizontal line shows the predicted sensitivity for the two-cue conditions assuming independent processing and optimal integration of texture and color. The solid horizontal line shows the prediction based on linear summation of sensitivities to the single cues. The predictions are the same for the cue-combination and cue-conflict stimuli. In all cases, sensitivity in the cue-combination condition was significantly higher, and sensitivity in the cue-conflict condition was significantly lower than predicted based on independent processing of the two cues. Error bars and shaded areas show 95% confidence intervals. Asterisks indicate significant differences (two-tailed $p < 0.05$).

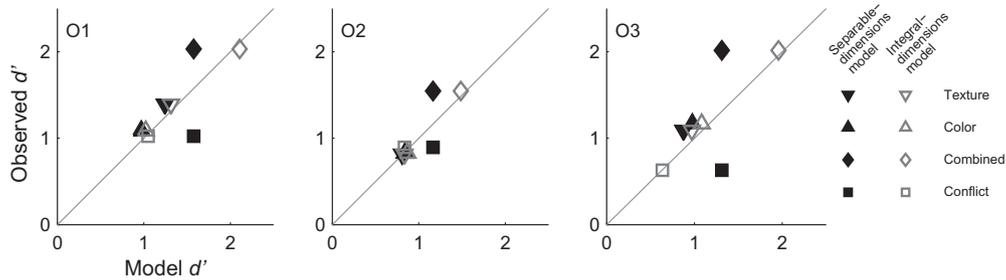


Fig. 5. Measured vs. model sensitivity. The separable model, which assumes independent processing of texture and color, consistently predicts too-low sensitivity in the cue-combination condition (filled diamonds) and too-high sensitivity in the cue-conflict condition (filled squares). The integral model, which allows for interdependence of color and texture processing, fits the data well, with all points lying on or near the diagonal.

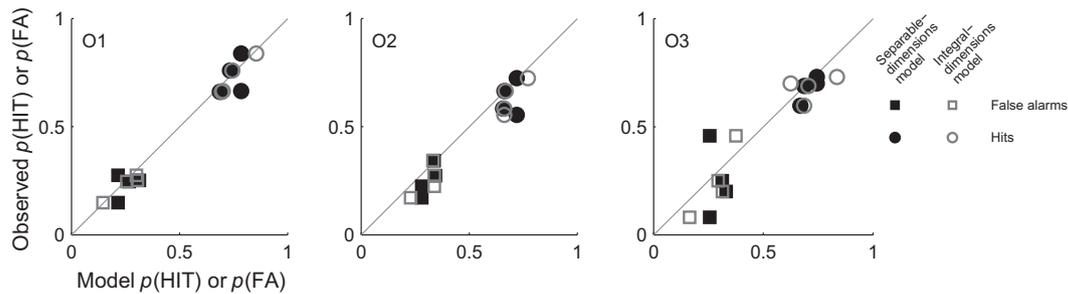


Fig. 6. Measured vs. model hits and false alarms. Similar to Fig. 5, but probabilities of hits and false alarms are plotted instead of sensitivity.

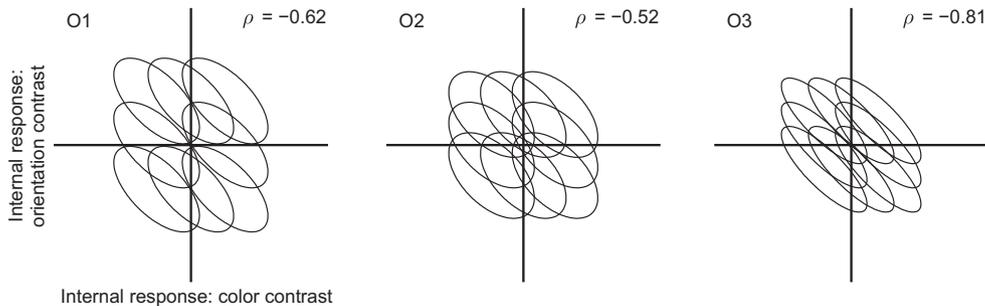


Fig. 7. The best-fitting decision spaces for the three observers for the integral model. The distributions are shown as contour plots. The pairs of distributions lying on the axes correspond to the single-cue conditions. The pair of distributions near the positive diagonal corresponds to the cue-combination condition, and the pair of distributions near the negative diagonal corresponds to the cue-conflict condition. The covariance in each of the best-fitting models was non-zero and negative, resulting in an elliptical, “tilted” distribution. The non-zero, negative covariance effectively reduces the variance along the direction on which the cue-combination distributions lie, reflecting the greater sensitivity in this condition. The same does not happen along the direction on which the cue-conflict distributions lie.

and negative in each case, making the equal-probability contours elliptical and “tilted” in the decision space. The negative correlation between the dimensions effectively improves the signal-to-noise ratio in the cue-combination condition: the distance of the two cue-combination distributions (upper right and lower left) relative to the amount of noise along the relevant direction is increased relative to the single-cue conditions. The amount of noise relative to the distance between the two cue-conflict distributions (upper left and lower right), however, is not changed substantially relative to single-cue conditions because the noise is increased but the distance between the two stimuli (i.e., the signal) is increased by a similar amount, reflecting the lack of improvement in this condition.

3.2. Experiment 2

Experiment 2 was designed to look for interactions in the processing of texture and color when observers were instructed to use only one of the cues to do the task. This experiment, in other

words, tested whether color information interferes with texture processing and vice versa.

Fig. 8 shows the performance in Experiment 2. Different panels correspond to different observers. The d' -values for the texture blocks are on the x-axis, and the d' -values for the color blocks are on the y-axis. There are two shaded areas in each plot. First, the area labeled “mutual facilitation” is the region where performance is above the 95% confidence interval (of neutral-trial performance) in both the texture task and the color task. If a data point falls in this region, then in that condition the presence of the color cue tends to improve performance with the texture cue and texture also improves color performance. The second shaded area is labeled “mutual masking”. It shows the region where performance is below the 95% confidence interval (of neutral-trial performance) in both the texture task and the color task. A data point falling in this area indicates that in that condition, color tends to impair performance when observers are using the texture cue and vice versa.

In the “congruent” condition, the second cue was present and in agreement with the target cue. Each of the data points from the

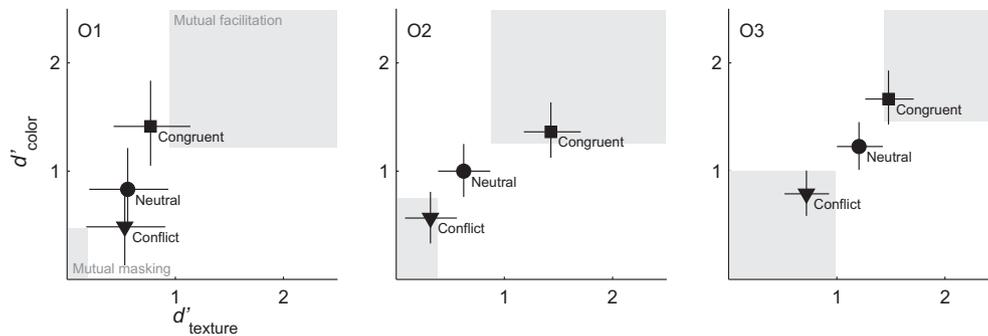


Fig. 8. Masking and facilitation between texture and color cues in the segmentation task. The x -axis shows the sensitivity (d') when the observer used only the texture cue, and the y -axis shows the sensitivity when the observer used the only color cue to do the task. Neutral trials: the second cue was absent. Congruent trials: the second cue defined an edge in the same orientation as the target cue. Conflict trials: the second cue defined an orthogonal edge. The area labeled “mutual facilitation” indicates a region where performance in both tasks (color and texture) tends to be better than in the neutral condition. A point falling inside this region indicates that color improved performance in the texture blocks and texture improved color performance. The area labeled “mutual masking” indicates a region where performance in both tasks (color and texture) tends to be weaker than in the neutral condition. A point falling inside this region indicates that color impaired performance in the texture blocks and vice versa. Error bars show 95% confidence intervals and shaded areas are based on the error bars for the neutral condition (see text for details).

“congruent” condition (squares) falls in or near the “mutual facilitation” region, indicating that observers benefited from having the second cue present (for observer O1, performance in the color task was improved, but performance in the texture task was not). In the “conflict” condition, the second cue was also present but it defined an edge orthogonal to the target cue. In this case, the second cue generally impaired performance and the data points (triangles) fall in or near the “mutual masking” region (again, with the exception of O1 in the texture task).

We tested for significant differences between the congruent and conflict conditions with a permutation test. We resampled the hits and false alarms 10,000 times and computed d' for the congruent and conflict trials (separately for color and texture), and took their difference. From the resulting distribution of differences in d' values, we computed the 95% confidence interval and tested whether this interval contained zero. We thus had six tests for the difference in d' (two tasks, color and texture, and three observers). The difference was not significant for observer O1 in the texture task (the x -axis locations of the square and triangle did not differ significantly in the first panel, Fig. 8; two-tailed $p = 0.36$). Due to the observer’s time restrictions, O1 was able to complete fewer sessions than the other two observers (with a total of 200 trials per d' estimate, compared to 440 and 500 trials for O2 and O3, respectively). In the five other cases—color task for O1 and both tasks for O2 and O3—the difference between congruent and conflict conditions was significant (two-tailed $p < 0.05$). Thus, the presence of a congruent cue, as compared to a conflicting cue, improved performance, even though the additional cue was not being judged.

4. Discussion

Human observers can integrate two different cues, texture and color, to improve visual segmentation. Visual segmentation is significantly better when both texture and color cues are available, compared to conditions where only one cue is available. Performance only improves, however, when the cues are spatially aligned; when both texture and color edges are present but orthogonal to each other, observers perform similarly to the single-cue conditions (Experiment 1). Further, observers cannot completely discount the other cue when segmenting the stimulus based on only one cue: they perform better when the second cue is congruent with the target cue compared to when the second cue is in conflict with the target cue (Experiment 2).

The results presented above argue against complete independence of texture and color processing in visual segmentation.

Consider, first, the cue-combination and cue-conflict conditions of Experiment 1. In both conditions, the stimuli to be discriminated contained texture- and color-defined edges. The only difference was the alignment: In the cue-combination condition, the texture and color edges coincided, whereas in the cue-conflict condition, the texture and color edges were orthogonal to each other. If texture and color were processed independently of each other, the alignment should not matter, and performance should be equal in these two conditions. The observers, however, did much better in the cue-combination condition (aligned cues) than in the cue-conflict condition (orthogonal cues). In fact, none of the observers showed any improvement in the cue-conflict condition over the single-cue conditions. Second, we compared the performance in the cue-combination condition to predictions based on independent processing of texture and color. The predictions were calculated from the performance in the single-cue conditions, assuming independent processing and optimal integration of the cues. All observers performed significantly better than predicted with the cue-combination stimulus (and significantly worse than predicted with the cue-conflict stimulus).

Another extreme possibility, opposed to complete independence, is complete cue-invariance with respect to texture and color in visual segmentation. Complete cue-invariance predicts that sensitivity (d') in the cue-combination condition is a sum of the single-cue sensitivities. We tested for cue-invariance by comparing the cue-combination performance to this prediction. The observed d' values were lower than predicted for all three observers (significantly so for one), showing that processing is not completely cue-invariant either.

Comparison of the two signal-detection models also speaks for interdependence of texture and color. The “integral” model, which allows for interactions between texture and color, fit the data significantly better than the “separable” model, which assumes strict perceptual independence. The separable model always predicts equal performance in the cue-combination and cue-conflict conditions. The integral model, on the other hand, accurately captures the difference between these conditions with one additional parameter, the covariance.

Two of the observers, O1 and O2, performed in the cue-conflict condition in a roughly similar way to the single-cue conditions. Thus, these observers might have used only one of the cues when the cues were conflicting (that is, they might have attended to one of the two cues alone). The third observer, on the other hand, performed worse with the conflict stimulus than with either of the single cues alone, as if the conflict induced cross-orientation

masking between the two second-order gratings. With all observers, the signal-detection model with correlated cues accounts for the performance in all four conditions simultaneously, with no need to postulate attentional restrictions or a change in strategy—cue exclusion in this case—in the cue-conflict condition.

In fact, the results from Experiment 2 make cue exclusion an unlikely possibility. In this experiment, we directly tested whether the observers are able to ignore a task-irrelevant cue when performing the segmentation task. Observers knew that on one-third of the trials the second cue would not be there (neutral trials), on one-third it would be in conflict with the target cue (conflict trials), and only on one-third of the trials it would be informative (congruent trials). Nonetheless, their sensitivity tended to be lower on the conflict trials and higher on congruent trials compared to the neutral trials, as if the conflicting cue masked and the congruent cue facilitated the detection of the edge. The observers seem to be unable to filter out the other cue even when they know it will not help them do the task.

The non-zero covariance in the best-fitting signal-detection model, together with the pattern of facilitation and masking in Experiment 2, is consistent with non-independent processing of texture and color. There are several possible underlying causes for this interdependence. The non-zero covariance could reflect correlated noise in two separate mechanisms, one responsive to texture and the other for color. Second, these two separate mechanisms, one for texture and the other for color, could interact when both are activated at the same time. Third, some mechanisms could be tuned for both texture and color signals. Double-opponent neurons that are jointly tuned for local orientation and chromaticity (Johnson, Hawken, & Shapley, 2008) are one possible underlying neural mechanism for our observations. This would also be reasonable in accord with the study by Pearson and Kingdom (2002), who found subthreshold summation between color and luminance contrast in an orientation-modulated texture discrimination task.

Earlier studies on visual segmentation have investigated the processing of texture and first-order color cues. The findings are compatible with our observations. Reaction times are faster and performance improves when both texture and color cues are available (Callaghan, Lasaga, & Garner, 1986; Gorea & Papathomas, 1991; Zhaoping & May, 2007). It is, however, difficult to judge whether the improvement is great enough to suggest non-independent processing of the cues. Reaction times increase and performance worsens in a texture-segmentation task in the presence of task-irrelevant color variability (Callaghan, Lasaga, & Garner, 1986; Gorea & Papathomas, 1991; Morgan, Adam, & Mollon, 1992; Snowden, 1998; Zhaoping & Snowden, 2006; Zhaoping & May, 2007, but see Gorea & Papathomas, 1993). This interference seems to be asymmetrical: color segmentation is not similarly affected by texture (orientation) noise (Snowden, 1998; Zhaoping & May, 2007; Zhaoping & Snowden, 2006), although the results are mixed (Callaghan, Lasaga, & Garner, 1986). Similarly, in visual search, having both orientation and first-order color cues available speeds up responses (Koene & Zhaoping, 2007; Poom, 2009) and improves performance (Monnier, 2006). As to the independence of orientation and color cues in visual search, the results are mixed: although some studies suggest summation of orientation and color (Koene & Zhaoping, 2007), others have found no evidence for it (Monnier, 2006; Poom, 2009).

Rivest and Cavanagh (1996) studied the localization of edges defined either by single cues (luminance, color, or texture) or their combination. Their results were consistent with independent processing followed by cue integration with equal weights for each cue, whereas our results suggest non-independent processing of texture and color. A somewhat similar discrepancy seems to hold for the combination of two texture cues, orientation and spatial frequency: Edge-localization performance is consistent with

independent processing and subsequent (optimal) integration (Landy & Kojima, 2001), whereas performance in a coarser task of texture detection indicates non-independent processing (Meinhardt et al., 2004). The “feature synergy”—the extent to which the cues summate—also depends on the strength of the cues. The weaker the cues are, the higher the advantage of having several cues (Meinhardt et al., 2004; Persike & Meinhardt, 2006). The cues in our tasks had a low contrast, which would make them more likely to reveal cue-interaction effects. Also, in edge-localization experiments the different cues are often mis-aligned, whereas our results suggest that cue-integration is best with aligned cues.

Recent neuro-imaging evidence suggests that color and texture are processed either partially (Cant, Arnott, & Goodale, 2009; Cant & Goodale, 2007) or completely (Cavina-Pratesi et al., 2010) separately of each other when discriminating or identifying surface properties of objects. This is backed by behavioral data by Cant et al. (2008): Texture does not interfere with surface color identification, and color does not interfere with texture identification. This difference—Cant et al. (2008) found independence whereas we found interdependence of texture and color—can probably be explained by the tasks used: the studies mentioned above were interested in the identification of the surface properties themselves, not in how those properties are used to detect edges or shapes. For example, our cue-combination condition is very different from the experiment by Cant et al. (2008): In the cue-combination condition the observers can and should make use of both color and texture information to do the task, whereas in their experiment the task was to identify the color or texture and ignore the other. Our observations are also consistent with psychophysical and imaging studies demonstrating joint selectivity in the processing of stimulus properties when they are used for segmentation or shape recognition. Møller and Hurlbert (1997) demonstrated interactions between color and motion signals in visual segmentation. Self and Zeki (2005) found cue-invariance for motion and color, and Grill-Spector et al. (1998) found cue-invariance for luminance, texture, and motion in shape processing in area LOC.

In summary, we find that texture-segmentation performance improves for edges signaled by two cues as compared to single-cue boundaries, but only if the two cues signal identical boundaries. When the texture and color edges are not spatially aligned, performance does not improve. This, together with the amount of improvement with aligned cues suggests that texture and color are not processed independently of each other in visual segmentation.

Acknowledgments

This work was supported in part by NIH Grant EY16165. TS was supported by the Swiss National Science Foundation fellowship PBELP1-125415. We would like to acknowledge the help of Angel Patel and the helpful comments of John Ackermann and Zack Westrick on earlier drafts of this manuscript.

References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262.
- Anstis, S. M., & Cavanagh, P. (1983). A minimum motion technique for judging equiluminance. In J. D. Mollon & L. T. Sharpe (Eds.), *Colour vision: Psychophysics and physiology* (pp. 155–166). New York, NY: Academic Press.
- Braddick, O. (1993). Segmentation versus integration in visual motion processing. *Trends in Neurosciences*, *16*, 263–268.
- Callaghan, T. C., Lasaga, M. I., & Garner, W. R. (1986). Visual texture segregation based on orientation and hue. *Perception and Psychophysics*, *39*, 32–38.
- Cant, J. S., Arnott, S. R., & Goodale, M. A. (2009). fMRI-adaptation reveals separate processing regions for the perception of form and texture in the human ventral stream. *Experimental Brain Research*, *192*, 391–405.
- Cant, J. S., & Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral Cortex*, *17*, 713–731.

- Cant, J. S., Large, M.-E., McCall, L., & Goodale, M. A. (2008). Independent processing of form, colour, and texture in object perception. *Perception*, 37, 57–78.
- Cavina-Pratesi, C., Kentridge, R. W., Heywood, C. A., & Milner, A. D. (2010). Separate channels for processing form, texture, and color: Evidence from fMRI adaptation and visual object agnosia. *Cerebral Cortex*, 20, 2319–2332.
- Cole, G. R., Hine, T., & McIlhagga, W. (1993). Detection mechanisms in l-, m-, and s-cone contrast space. *Journal of the Optical Society of America A*, 10, 38–51.
- Dubois, M., Poeppel, D., & Pelli, D. G. (in preparation). The cost of combining features to see and hear a word.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York, NY: Wiley.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433.
- Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, 13, 483–488.
- Gorea, A., & Papathomas, T. V. (1991). Texture segregation by chromatic and achromatic visual pathways: An analogy with motion processing. *Journal of the Optical Society of America A*, 8, 386–393.
- Gorea, A., & Papathomas, T. V. (1993). Double opponency as a generalized concept in texture segregation illustrated with stimuli defined by color, luminance, and orientation. *Journal of the Optical Society of America A*, 10, 1450–1462.
- Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics*. Los Altos Hills, CA: Peninsula Publishing.
- Grill-Spector, K., Kushnir, T., Edelman, S., Itzhak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*, 21, 191–202.
- Hansen, T., & Gegenfurtner, K. R. (2006). Higher level chromatic mechanisms for image segmentation. *Journal of Vision*, 6, 239–259.
- Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, 298, 1627–1630.
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4, 967–992.
- Johnson, E. N., Hawken, M. J., & Shapley, R. (2008). The orientation selectivity of color-responsive neurons in macaque V1. *Journal of Neuroscience*, 28, 8096–8106.
- Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43, 2539–2558.
- Koene, A. R., & Zhaoping, L. (2007). Feature-specific interactions in saliency from combined feature contrasts: Evidence for a bottom-up saliency map in V1. *Journal of Vision*, 7(7), 1–14. 6.
- Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, 31, 679–691.
- Landy, M. S., & Graham, N. (2004). Visual perception of texture. In L. M. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 1106–1118). Cambridge, MA: MIT press.
- Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A*, 18, 2307–2320.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35, 389–412.
- Li, A., & Lennie, P. (2001). Importance of color in the segmentation of variegated surfaces. *Journal of the Optical Society of America A*, 18, 1240–1251.
- Meinhardt, G., Schmidt, M., Persike, M., & Rösers, B. (2004). Feature synergy depends on feature contrast and objecthood. *Vision Research*, 44, 1843–1850.
- Møller, P., & Hurlbert, A. (1997). Interactions between colour and motion in image segmentation. *Current Biology*, 7, 105–111.
- Monnier, P. (2006). Detection of multidimensional targets in visual search. *Vision Research*, 46, 4083–4090.
- Mood, A. M., & Graybill, F. A. (1963). *Introduction to the theory of statistics* (2nd ed.). New York: McGraw-Hill Book Company.
- Morgan, M. J., Adam, A., & Mollon, J. D. (1992). Dichromats detect colour-camouflaged objects that are not detected by trichromats. *Proceedings of the Royal Society of London, Series B*, 248, 291–295.
- Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, 18, 55–68.
- Nothdurft, H. C. (1985). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research*, 25, 1957–1968.
- Olzak, L. A., & Thomas, J. P. (1991). When orthogonal orientations are not processed independently. *Vision Research*, 31, 51–57.
- Pearson, P. M., & Kingdom, F. A. A. (2002). Texture-orientation mechanisms pool colour and luminance contrast. *Vision Research*, 42, 1547–1558.
- Persike, M., & Meinhardt, G. (2006). Synergy of features enables detection of texture defined figures. *Spatial Vision*, 19, 77–102.
- Poom, L. (2009). Integration of colour, motion, orientation, and spatial frequency in visual search. *Perception*, 38, 708–718.
- Rivest, J., & Cavanagh, P. (1996). Localizing contours defined by more than one attribute. *Vision Research*, 36, 53–66.
- Sankeralli, M. J., & Mullen, K. T. (1996). Estimation of the L-, M-, and S-cone weights of the postreceptoral detection mechanisms. *Journal of the Optical Society of America A*, 13, 906–915.
- Sankeralli, M. J., & Mullen, K. T. (1997). Postreceptoral chromatic detection mechanisms revealed by noise masking in three-dimensional cone contrast space. *Journal of the Optical Society of America A*, 14, 2633–2646.
- Schofield, A. J. (2000). What does second-order vision see in an image? *Perception*, 29, 1071–1086.
- Self, M. W., & Zeki, S. (2005). The integration of colour and motion by the human visual brain. *Cerebral Cortex*, 15, 1270–1279.
- Snowden, R. J. (1998). Texture segregation and visual search: A comparison of the effects of random variations along irrelevant dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1354–1367.
- Stockman, A., & Sharpe, L. T. (2000). The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research*, 40, 1711–1737.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics*, 63, 1293–1313.
- Wolfson, S. S., & Landy, M. S. (1995). Discrimination of orientation-defined texture edges. *Vision Research*, 35, 2863–2877.
- Zhaoping, L., & May, K. A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Computational Biology*, 3, e62.
- Zhaoping, L., & Snowden, R. J. (2006). A theory of a saliency map in primary visual cortex (V1) tested by psychophysics of colour-orientation interference in texture segmentation. *Visual Cognition*, 14, 911–933.