

# The role of first- and second-order stimulus features for human overt attention

HANS-PETER FREY AND PETER KÖNIG

*University of Osnabrück, Osnabrück, Germany  
and University and Swiss Federal Institute of Technology (ETH), Zürich, Switzerland*

AND

WOLFGANG EINHÄUSER

*University and Swiss Federal Institute of Technology (ETH), Zürich, Switzerland  
and California Institute of Technology, Pasadena, California*

When processing complex visual input, human observers sequentially allocate their attention to different subsets of the stimulus. What are the mechanisms and strategies that guide this selection process? We investigated the influence of various stimulus features on human overt attention—that is, attention related to shifts of gaze with natural color images and modified versions thereof. Our experimental modifications, systematic changes of hue across the entire image, influenced only the global appearance of the stimuli, leaving the local features under investigation unaffected. We demonstrated that these modifications consistently reduce the subjective interpretation of a stimulus as “natural” across observers. By analyzing fixations, we found that first-order features, such as luminance contrast, saturation, and color contrast along either of the cardinal axes, correlated to overt attention in the modified images. In contrast, no such correlation was found in unmodified outdoor images. Second-order luminance contrast (“texture contrast”) correlated to overt attention in all conditions. However, although none of the second-order color contrasts were correlated to overt attention in unmodified images, one of the second-order color contrasts did exhibit a significant correlation in the modified images. These findings imply, on the one hand, that higher-order bottom-up effects—namely, those of second-order luminance contrast—may partially account for human overt attention. On the other hand, these results also demonstrate that global image properties, which correlate to the subjective impression of a scene being “natural,” affect the guidance of human overt attention.

---

The complexity of natural input usually exceeds the parallel processing capacity of the human visual system. Consequently, the visual system sequentially attends to subsets of the input. Under natural viewing conditions, these shifts of attention are usually associated with changes in fixation. This so-called overt attention therefore provides an objective measure of attention shifts.

In the control of attention, two types of signals are to be distinguished, those related to the stimulus (“bottom-up”) and those related to subjects’ expectations and experiences as well as to the task (“top-down”). Interest in the top-down features guiding human fixations dates at least back to Buswell’s (1935) study. Comparing students of art with average observers, he found fixation durations to be slightly shorter for the “expert” group when looking at pictures; during reading (but not with pictures), a similar difference held true between adults and children. Both results highlight the roles of subjective experience and special training in human attention. In addition, Buswell provided some qualitative account of the effect of the task, which three decades later Yarbus’s (1967) seminal study

addressed in detail. Yarbus, in particular, demonstrated the dramatic influence of the task on overt attention in scenes that contain objects of high behavioral relevance to human observers, such as people or faces. On the basis of such results, there seems to be general consensus on the relative importance of experience and task to human attention. Nevertheless, bottom-up features are also assumed to play an important role for the allocation of spatial attention, especially in the absence of an explicit task. Consequently, the investigation of bottom-up models and their interaction with top-down signals provides fruitful insight into the mechanisms underlying human attention.

One of the most influential models to describe bottom-up control of human attention is that of the so-called saliency map (Koch & Ullman, 1985). Though implementations of this model have undergone several modifications since its original formulation, the basic scheme remains unchanged (see Itti & Koch, 2001, for a review). The stimulus is analyzed in various feature channels, such as luminance, color, or orientation. Local differences in each feature are computed, combined across different spatial

scales, and normalized nonlinearly. The resulting maps are summed, yielding the saliency map. Locations of high activity in this saliency map are assumed to be salient—that is, likely to be attended. Recently, Peters, Iyer, Itti, and Koch (2005) performed a thorough comparison between the predictions of saliency map models and human fixations for a large variety of stimuli. Those authors found that the models, especially versions that include additional interactions within the orientation channel, predict human fixation patterns reasonably well. Nevertheless, the performance of the models still remains clearly below the optimum that any bottom-up model may achieve, the human interobserver prediction. The present study does not aim at testing the predictions of saliency map models directly. Instead, we chose two particular local features that are also employed in saliency map models, luminance contrast and color. We analyzed how the effects of these particular features depend on changes in the global appearance of a stimulus when the local features themselves remain unchanged.

In recent years, several studies have examined the influence of individual stimulus features on human overt attention. Mannan, Ruddock, and Wooding (1996) found that luminance contrast, arguably the best-investigated feature, is elevated at fixation locations if the analysis is corrected for general biases in stimuli and fixation patterns. Using a different definition of the feature, Reinagel and Zador (1999) also reported this correlative effect between luminance contrast and human overt attention. Their result was later confirmed in several other studies (e.g., Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Parkhurst & Niebur, 2003), although under some presentation conditions a dependence on spatial frequency may obstruct the effect for unfiltered stimuli (Mannan, Ruddock, & Wooding, 1997; Tatler, Baddeley, & Gilchrist, 2005). Recently, a study in our laboratory also replicated the correlation of luminance contrast to human fixation, but we demonstrated that this effect of luminance contrast is not causal (Einhäuser & König, 2003). Hence, first-order luminance contrast—despite being elevated at fixations—is not causally related to the bottom-up control of human overt attention.

These results for first-order luminance contrast raise the question of the extent that higher-order stimulus features can account for overt attention under natural conditions. On the modeling side, Parkhurst and Niebur (2004) were able to explain Einhäuser and König's (2003) data by including second-order luminance contrast, which they dubbed "texture contrast," in a modified saliency map model. Besides this "texture contrast," which only operates on the luminance channel, second-order effects might also be effective for other features: For example, Walther, Edgington, and Koch (2004) demonstrated that a second-order orientation contrast enhances performance in a machine vision application of the saliency map model. In summary, several lines of evidence—experimental, modeling, and application performance—point toward the relative importance of higher-order stimulus features for the bottom-up control of attention in complex visual stimuli.

Several electrophysiological studies have demonstrated the representation of saliency in various brain regions, such as the pulvinar (Posner & Petersen, 1990; Robinson & Petersen, 1992), the superior colliculus (Horowitz & Newsome, 1999; Kustov & Robinson, 1996; McPeck & Keller, 2002; Posner & Petersen, 1990), the frontal eye field (Thompson, Bichot, & Schall, 1997), and the lateral intraparietal area (Gottlieb, Kusunoki, & Goldberg, 1998). In the visual cortex, regions as early as primary visual cortex have been suggested as locations for the computation of saliency (Li, 2002). A recent study by Mazer and Gallant (2003) of macaque monkeys viewing grayscale images of natural scenes suggested that higher ventral areas, such as V4 and IT, are strongly involved in the computation of saliency. Since macaque V4 (Zeki, 1983) and IT (Gegenfurtner & Kiper, 2003; Komatsu, Ideura, Kaji, & Yamane, 1992) are also associated with the processing of color, the relation of color to attention is of specific interest.

In the present study, we used color images and modifications thereof. By consistently changing hue across the entire stimulus, we manipulated the global stimulus appearance without affecting the statistics of the local features under investigation. First, we verified whether these modifications affect the subjective perception of an image as "natural." In the main part of the study, we investigated the influence of different first- and second-order stimulus features, such as luminance contrast, saturation, color contrast, second-order luminance contrast, and second-order color contrasts, on human overt attention. Although we did not address saliency map models directly, our choice of features was obviously inspired by such models. In the case of color, saliency map models typically employ center-surround mechanisms akin to color-opponent cells in the retina or lateral geniculate nucleus (LGN), and this connection serves as a rationale for performing all analysis in a color space that is based on the excitations of these cells. Using our particular setting, we directly assessed the contribution of some bottom-up features that are typically used in models of human attention. Furthermore, we investigated how a global modification, which also relates to the subjective impression of a stimulus being "natural," modulates the effect of these local features. Since the local features were identical in our unmodified and modified stimuli, any observed difference would point to mechanisms of overt attention beyond local bottom-up features.

## METHOD

### Color Representation

When dealing with color stimuli, the choice of an appropriate color space is critical. Here, our choice was guided by (1) the fact that most models of bottom-up attention use color-opponent channels akin to those in the retina and LGN, and (2) the need to define color features independently of assumptions about the observer or the spatial scale of the stimulus. Consequently, we decided to use a physiological color space defined on the basis of relative cone excitations, rather than a psychophysically defined color space. The so-called Derrington-Krauskopf-Lennie (DKL) color space (Derrington, Krauskopf, & Lennie, 1984) is based on the relative excitations of the three cone types (L, M, and S) in the primate retina. Two

orthogonal color axes, “constant blue” (cb, defined by the difference between L and M cone excitations) and “tritanopic confusion” (tc, defined as  $L + M - S$  cone excitations), and a “luminance” axis span the DKL space (Figure 1A). Isoluminant planes are spanned by the orthogonal color axes. The azimuth in this polar space represents a color’s hue ( $0^\circ$  at  $L + M = \max$ ,  $L + M - S = 0$ ). A color’s saturation is defined by the length of the projection onto the isoluminant plane through the origin. The luminance is given by the value along the luminance axis. For clarity of presentation, the maximum luminance of the setup is normalized to 1, the minimum to  $-1$ , so that all represented colors fall within a unit sphere. Note that the manifold of colors that can actually be displayed—the gamut—depends on the presentation setup and is smaller than (i.e., lies fully inside) this unit sphere.

### Stimuli

Stimuli were based on images of the local outdoor environment, which were taken with a digital camera (3.3-megapixel color mosaic CCD, Nikon Coolpix 995, Tokyo) and down-sampled to a resolution of  $1,024 \times 768$  using bicubic interpolation. To reduce the semantic content of the stimuli as much as possible, 12 images (out of about 100 in the database) were selected that contain neither manmade objects nor “blurry” regions of extremely low luminance contrast (Figure 1B). They were presented in four different conditions: unmodified; converted to grayscale (Figure 1C); and with the azimuth (hue) of each pixel in DKL space rotated by  $90^\circ$  either counterclockwise or clockwise (Figures 1D and 1E, respectively). For analysis, the latter two conditions are not distinguished and will be referred to as *color rotated* throughout. In order to ensure that each pixel had the same saturation in all three colored conditions, individual pixels that exceeded the gamut of the screen in at least one condition were reduced in saturation in all conditions accordingly. The luminance of each pixel was constant across conditions. Consequently, the conditions differed from each other only in azimuth (hue).

### Stimulus Presentation

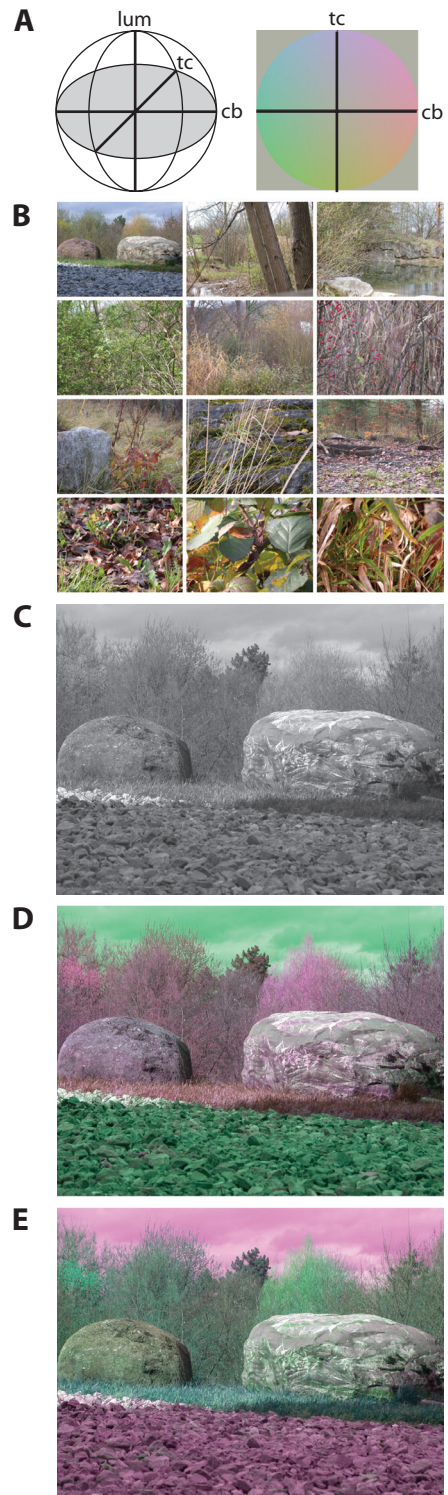
Stimuli were generated using a Macintosh G4/800 computer (Apple, Inc., Cupertino, CA) running MATLAB (The MathWorks, Natick, MA), including its Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). Stimuli were presented on a 19-in. computer screen (Hitachi CM772E, Tokyo, Japan; CIE coordinates of the phosphors: red, .625/.34; green, .285/.605; blue, .15/.065) located 57 cm from the subject. The monitor’s resolution was set to the stimulus size of  $1,024 \times 768$  pixels with a color depth of 24 bits at a frame rate of 120 Hz. Maximum luminance of the screen (white) was set to  $85.9 \text{ cd/m}^2$ , the minimum (black) below  $0.1 \text{ cd/m}^2$ . The gamma of the screen was corrected in order to achieve a linear mapping of the DKL luminance axis to actual luminance.

### Subjects

Six volunteers with uncorrected normal vision participated in the main eyetracking experiment. Five additional volunteers with uncorrected or corrected-to-normal vision participated in the behavioral control experiment. All subjects had normal color vision as assessed by Ishihara plates. Two of the subjects of the main experiment were authors (H.-P.F. and P.K.), and all others were naive to the purpose of the experiment. The experiments conformed to the National and Institutional Guidelines for the use of human subjects and to the Declaration of Helsinki. All subjects gave written informed consent to participate in the study.

### Eyetracking

We recorded eye position using a noninvasive infrared oculometer (Dr. Bouis, Karlsruhe, Germany; Bach, Bouis, & Fischer, 1983). The setup and calibration protocol were as described in our previous study (Einhäuser & König, 2003). In brief: Subjects’ heads were stabilized with a chinrest and a bitebar made of thermoplastic impression material. Before each experimental block, a calibration trial was



**Figure 1. Stimuli.** (A) Left: The DKL color space, comprising the two chromatic axes cb and tc and the achromatic luminance (lum) axis (see the Method section for details). The gray-filled area represents the isoluminant plane through the origin of the color space. Right: The colors mapped to the isoluminant plane. (B) All 12 outdoor images used as stimuli in the unmodified color condition. (C) A stimulus in the grayscale condition. (D) The stimulus rotated counterclockwise ( $+90^\circ$ ) in DKL space. (E) The stimulus rotated clockwise ( $-90^\circ$ ) in DKL space.

conducted in which subjects had to fixate points presented on the screen. From this calibration, a bilinear coordinate transform was computed offline to map oculometer output voltages to eye position for the subsequent block. Before the onset of each experimental trial, subjects had to fixate a single central fixation point, which was used to verify the calibration. If the distance between this fixation point and the computed eye position (*fixation error*) exceeded a threshold, the trial was excluded from analysis. If the mean distance of all fixation points of a calibration trial (*calibration error*) exceeded a threshold, the subsequent block was excluded from analysis.

## Experimental Design

**Main experiment.** In the main experiment, subjects viewed stimuli while their eye position was continuously recorded. To minimize any instruction-related biases, the subjects were merely instructed to “study the images carefully.” In total, each subject viewed each of the 12 images in each of the four conditions 10 times, for 6 sec each. This yielded a total number of 480 experimental trials for each subject. We restricted our analysis to well-calibrated trials, in which the fixation and calibration errors (as defined above) were smaller than 40 pixels (1.5°). This resulted in 200 valid experimental trials out of 480 (41.7%) for subject A.S., 269/480 trials (56.0%) for H.-P.F., 159/480 trials (33.1%) for M.K., 325/480 trials (67.7%) for P.K., 50/480 trials (10.4%) for P.R.K., and 49/480 trials (10.2%) for S.O. Although our quality criterion was rather conservative, because 10 repetitions were performed for each stimulus, our design ensured that data were sufficient to perform a robust analysis for each individual subject. For the remainder of the article, the term *trial* will refer to the aggregate data for each stimulus from the up to 10 repetitions. Changing the quality criterion to (for instance) 80 pixels at either threshold did not qualitatively affect the results.

Since periods of fixation account for 71.8% of the valid data, we restricted all analysis to the points of fixation. In order to reduce potential effects of the screen borders and to operate the oculometer within its linear range, analysis was also restricted to the central  $600 \times 600$  pixel region ( $23^\circ \times 23^\circ$ ) of each stimulus, containing 77.3% of all fixations. Using for analysis the whole image,  $1,024 \times 768$  pixels wide ( $38^\circ \times 29^\circ$ ), did not yield any qualitative difference in results.

**Behavioral control: Subjective perception of a stimulus as “natural.”** In addition to the main (eyetracking) experiment, we aimed to assess the extent to which our global modifications influenced the subjective perception of a scene being “natural.” We presented the same images as in the main experiment and asked 5 new subjects “how natural the images appeared” to them. The subjects had to respond on a scale from 1 (*unnatural*) to 9 (*natural*). To sample the range within which a scene is perceived as “natural” more densely, in this experiment we included—in addition to the conditions from the main experiment—intermediate hue rotation angles ( $10^\circ$ ,  $30^\circ$ ,  $60^\circ$ ,  $120^\circ$ , and  $150^\circ$  clockwise and counterclockwise), as well as  $180^\circ$ . Analogous to the presentation of the stimuli in the main experiment, each stimulus was preceded by a fixation cross presented for 0.5 sec. Stimuli were presented for 1 sec, and the next trial was started only after the subject had provided a response.

## Definition of Features

**Luminance contrast.** In line with earlier studies (e.g., Reinagel & Zador, 1999), we define luminance contrast at a point as the standard deviation of luminance in a square patch around this point, normalized by the mean luminance of the image. For analysis, the size of this patch was chosen to be  $80 \times 80$  pixels. Different patch sizes (we also tested  $60 \times 60$  and  $120 \times 120$ ) or normalization by the patch mean instead of image mean did not yield qualitatively different results.

**Saturation and color contrast.** The projection of a pixel in DKL space onto the isoluminant plane through the origin is independent of the pixel’s luminance but fully maintains the chromatic properties (hue and saturation) of the pixel. Hence, we will refer to this projection as the *chromatic content* of a pixel throughout. The

absolute value of the chromatic content defines the saturation of a pixel. We define the saturation of a patch as the mean saturation of all pixels inside the patch. Using the median instead of the mean did not influence the results.

In order to examine the contribution of the two color-opponent processes to overt attention independently, we measured color contrast for both cardinal color axes separately (instead of using a combined measure, such as RMS color contrast). In close analogy to the definition of luminance contrast, we define the cb and tc color contrasts at a point as the standard deviation of the chromatic content in an  $80 \times 80$  pixel patch around this point, projected on the respective cardinal axis.

**Higher-order contrasts.** In order to probe the effect of higher-order structure of the stimuli on overt attention, we canonically extend the definitions of first-order contrasts. We define second-order luminance contrast at one point as the standard deviation of the luminance contrasts in a patch around this point divided by the mean luminance contrast of the whole image. This is analogous to our definition of luminance contrast. In the domain of color, we examined second-order contrasts for both cardinal axes. Second-order color contrast is defined analogously to second-order luminance contrast, as the standard deviation of the color contrasts along each cardinal color axis in a square patch around a point, divided by the mean contrast along the corresponding axis. For second-order contrasts, we chose the same patch size as for first-order contrasts,  $80 \times 80$  pixels.

## Statistical Analysis

In order to assess whether or not any particular feature is related to overt attention, we applied the following procedure, which avoids the potential confound of “central bias” (see Mannan et al., 1996; Tatler et al., 2005), to all features under investigation. In each trial, we defined for each stimulus its “actual” value as the median of the feature values over all fixations on the stimulus. This actual value was then compared with a baseline for each subject and condition separately. This baseline was designed so that potential general biases in the subjects’ eye positions did not confound the analysis. Therefore, we defined control fixations of a subject as all fixations of the same subject obtained on all other stimuli in the same condition. Calculating the median of the feature values at these control fixations on the actual image yielded the control value. This control value served as the unbiased baseline that the corresponding actual value was compared with. The actual value would be different from the control value for a feature if and only if the feature had an effect on overt attention. Since we could not assume that actual and control values would follow a normal distribution, we tested the significance of this difference by a nonparametric statistical test, the Wilcoxon signed-rank test, which tests against the null hypothesis that the median of actual and control values over trials was identical.

In addition to giving the Wilcoxon test value, we also state the percentage of trials in which the actual value exceeded the control value and plot this percentage in all of the overview graphs. Since we presented the stimulus in the main experiment for a prolonged period of time, we also analyzed whether there was a systematic effect of viewing time on these percentages. When we performed the same analysis on single fixations (first fixation, second fixation, etc.), we did not find any systematic dependence on fixation number. Hence, we only report the data based on the median of all fixations in each trial.

## RESULTS

### Subjective Perception of a Stimulus As “Natural”

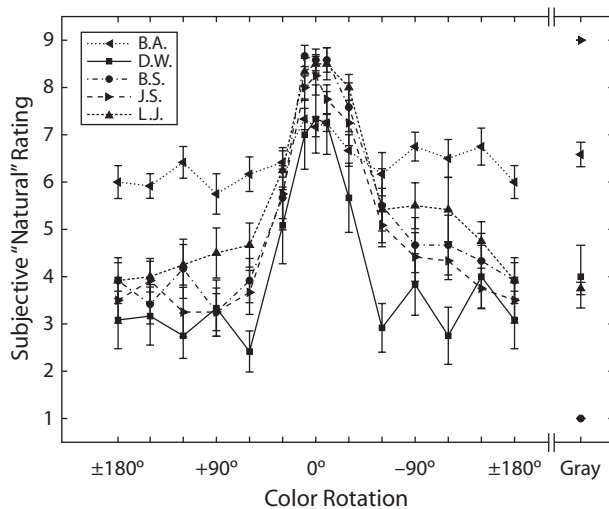
In the rating experiment, we assessed how color rotation affects the subjective perception of a scene as natural. The responses of each individual were significantly modulated by color rotation angle [B.A.,  $F(13) = 2.22$ ,  $p = .011$ ; D.W.,  $F(13) = 7.93$ ,  $p \ll 10^{-10}$ ; B.S.,  $F(13) =$

14.6,  $p \ll 10^{-10}$ ; J.S.,  $F(13) = 18.16$ ,  $p \ll 10^{-10}$ ; L.J.,  $F(13) = 17.24$ ,  $p \ll 10^{-10}$ ; one-way ANOVAs across all 14 rotation angles, excluding grayscale] and peaked at or around the unmodified image (Figure 2). In particular, the mean ratings for the  $\pm 90^\circ$  color-rotated images used in the main experiment were smaller than the mean ratings for unmodified images in all subjects. This result was significant in all 5 subjects (B.A.,  $p = .036$ ; D.W.,  $p = .000041$ ; B.S.,  $p < 10^{-5}$ ; J.S.,  $p < 10^{-5}$ ; L.J.,  $p < 10^{-5}$ ; post hoc  $t$  tests). One subject (J.S.) rated all grayscale images as maximally natural, but 3 out of the remaining 4 subjects rated grayscale images as significantly less natural than the unmodified images (B.A.,  $p = .093$ ; D.W.,  $p = .0026$ ; B.S.,  $p < 10^{-5}$ ; L.J.,  $p < 10^{-5}$ ; post hoc  $t$  tests). Since the distributions were not necessarily normal, in addition to ANOVAs and  $t$  tests we performed the nonparametric Kruskal–Wallis test and the Wilcoxon signed-rank test, which yield the same results. These behavioral data demonstrate that color-rotated images are consistently perceived as less “natural” than unmodified images.

### First-Order Features and Overt Attention

In the main experiment, we recorded eye position in 6 human subjects and analyzed the effect of various local stimulus features on their fixation behavior in modified and unmodified stimuli. Of all local stimulus features, the effect of luminance contrast on overt attention has been the one investigated best experimentally. However, in most of these studies, grayscale images were used. Consequently, we first analyzed whether the previously described correlative effect of luminance contrast is dependent on that choice of stimuli.

For each subject, we compared the actual fixations in one stimulus with all fixations from other stimuli (the “control”). Figure 3A shows the actual (green) and control

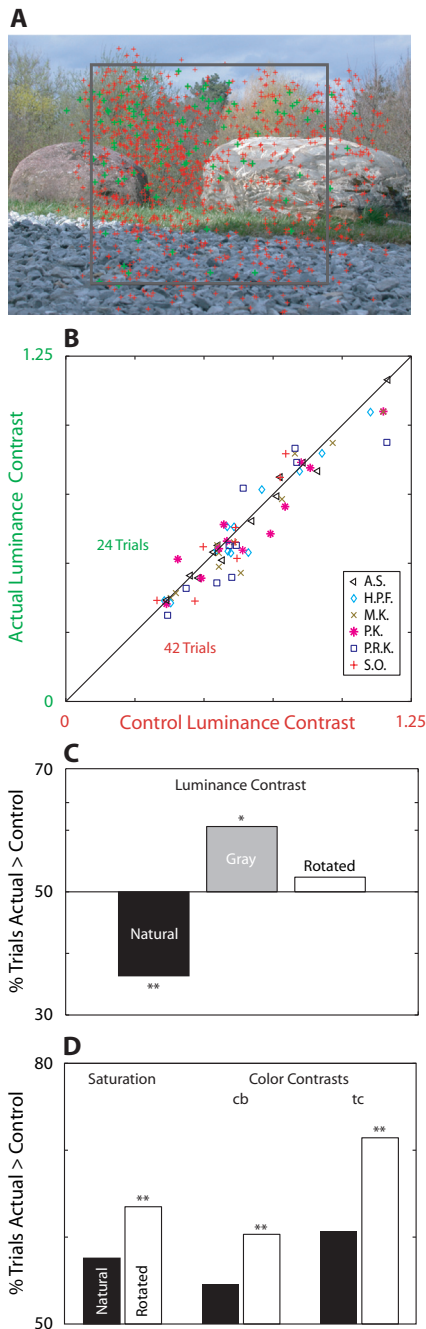


**Figure 2.** Subjective perception of a stimulus as “natural” on a scale from 1 (unnatural) to 9 (natural). The ordinate depicts mean  $\pm$  SEM over images for each individual. The abscissa depicts different color rotation angles; note that the scale is periodic ( $+180^\circ = -180^\circ$ ). The rightmost data points are for the grayscale condition.

(red) fixations for a single subject on one outdoor scene. For this type of stimulus, actual luminance contrast was lower than control in 42 trials, and the opposite was the case in 24 trials (Figure 3B). Indeed, the median actual luminance contrast over all subjects and trials was significantly smaller than control ( $p = .01$ , Wilcoxon signed-rank test). In contrast, for grayscale images we found actual luminance contrast to be larger than control in 60.6% of all stimuli ( $p = .04$ ; Figure 3C), which is in agreement with previous reports. In color-rotated images, we did not find any significant difference between actual and control luminance contrast ( $p = .21$ ; Figure 3C). These results show that the correlative effect of luminance contrast vanishes or reverses as soon as color is present. However, the difference between naturally colored stimuli and color-rotated stimuli suggests that this effect cannot be fully explained by the color feature as such, but is also influenced by the global stimulus appearance.

The luminance of each pixel of an image was identical across all conditions. Hence, one can directly compare actual luminance contrast between corresponding stimuli of different conditions. Comparing unmodified color stimuli to grayscale stimuli yielded an actual luminance contrast that was larger in grayscale images for 66.7% of the trials across all subjects. A similar fraction (69.8%) was found for the comparison of unmodified stimuli with color-rotated stimuli. In both cases, median actual luminance contrast across trials was significantly smaller for unmodified stimuli than for those in the other conditions ( $p = .003$  for grayscale,  $p = .0006$  for the color-rotated stimuli; Wilcoxon signed-rank test). In contrast, there were about as many stimuli for which luminance contrast was larger in the grayscale than in the color-rotated conditions (average across stimuli and subjects, 52.4%) as there were for the reverse conditions ( $p = 1$ ). Irrespective of the choice of a specific baseline, these data demonstrate a distinction in the relation of luminance contrast to human fixation between unmodified images, on the one hand, and color-rotated and grayscale images, on the other.

Retinal ganglion and LGN cells are modulated by saturation and by color contrast between the center and surround of a visual stimulus. Therefore, we examined the influence of these features on overt visual attention. For unmodified images, actual saturation was larger than control in 57.6% of the trials. The medians across trials were not significantly different between actual and control ( $p = .28$  in a Wilcoxon signed-rank test; bars on the left of Figure 3D). In contrast, actual saturation was highly significantly different from control in the color-rotated images (63.5% of trials,  $p = .005$ ; Figure 3D, left). In summary, although the color-rotated images demonstrate that saturation is a potentially salient feature, there is no relation between saturation and overt attention under natural conditions. Along the cb-axis, actual color contrast was larger than control in 54.5% of the trials with unmodified stimuli, and the medians were not significantly different ( $p = .13$  in a Wilcoxon signed-rank test; bars in the middle of Figure 3D). In contrast, 60.3% of trials exhibited larger actual cb contrast than control for color-rotated images, and the medians were significantly dif-



**Figure 3. First-order features.** (A) Actual (green) and control (red) fixations for a single subject (H.-P.F.) on a naturally colored stimulus. The black rectangle indicates the  $600 \times 600$  region used for analysis. (B) Actual luminance contrast (i.e., median luminance contrast over actual fixations) plotted versus control luminance contrast (i.e., median luminance contrast over control fixations) for naturally colored images. Each data point corresponds to one “trial” (i.e., the aggregate fixation data for up to 10 repetitions of the same stimulus; see the Method section). (C) Percentages of trials for which actual was larger than control luminance contrast for the three different stimulus conditions. (D) Percentages of trials for which actual saturation, constant blue contrast (cb), and tritanopic confusion contrast (tc) were larger than their respective control values, for naturally colored stimuli (black) and color-rotated stimuli (white). All significance values are from Wilcoxon signed-rank tests.  $*p < .05$ .  $**p < .01$ .

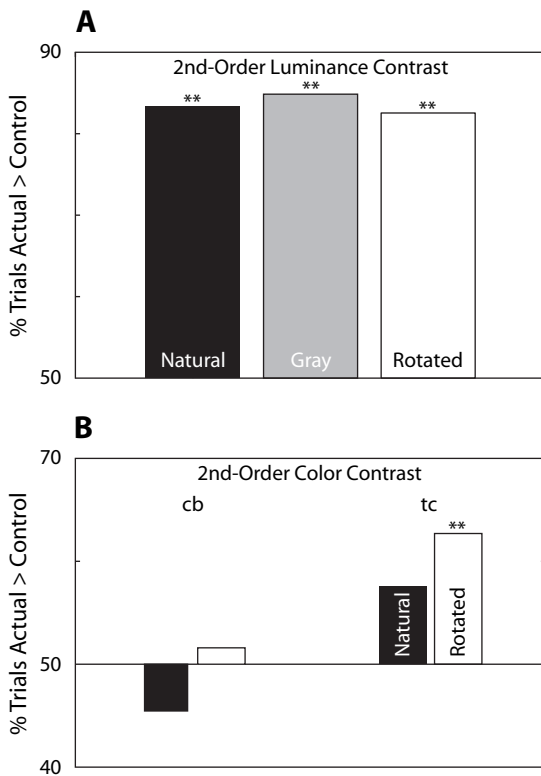
ferent ( $p = .001$ ; Figure 3D, middle). Along the tc-axis, actual color contrast was larger than control for 60.6% of trials with unmodified images ( $p = .053$ ; bars on the right of Figure 3D), but for color-rotated images this fraction amounted to 71.4% ( $p = 2 * 10^{-6}$ ; Figure 3D, right). Both color contrasts exhibited a significant effect on overt attention in color-rotated images but showed no such effect in unmodified images.

All first-order features—luminance contrast, saturation, and both color contrasts—are potentially salient, since they exhibited a significant effect on overt attention in at least one modified (grayscale or color-rotated) condition. However, none of these features showed a significant relation to overt attention in unmodified images. This finding provides evidence that no first-order feature in either the color or the luminance channel influences human overt attention in outdoor scenes.

### Second-Order Features and Overt Attention

In order to assess the influence of higher-order features on overt attention, we canonically extended the definitions of luminance contrast and color contrast. We defined second-order luminance contrast on the map of luminance contrast analogously to luminance contrast on the map of luminance. For unmodified images, we found actual second-order luminance contrast to be larger than control in 83.3% of all trials ( $p \ll 10^{-10}$  in a Wilcoxon signed-rank test; Figure 4A). The same relation also held in 84.9% of trials with grayscale images ( $p \ll 10^{-10}$ ) and 82.5% of trials with color-rotated images ( $p < 10^{-5}$ ). The direct comparison between conditions for actual second-order luminance contrast did not yield significant differences ( $p > .84$  for all pairwise comparisons in a Wilcoxon signed-rank test). These results show that second-order luminance contrast is correlated to overt attention, irrespective of whether or not the stimulus contains color.

In the domain of color, we examined second-order contrasts for both cardinal color axes. Analogously to the definition of second-order luminance contrast, we defined second-order color contrast on the two color-contrast maps. Second-order color contrast on the cb-axis was larger than control in 45.5% of the trials for unmodified images ( $p = .77$  in a Wilcoxon signed-rank test; bars on the left of Figure 4B) and in 51.6% of trials for color-rotated images ( $p = .51$ ; Figure 4B, left). On the tc-axis, we also found no difference between actual and control for unmodified images (57.6% of trials,  $p = .40$ ; bars on the right of Figure 4B). However, actual second-order color contrast on the tc-axis was significantly larger than control in 62.7% of trials for color-rotated images ( $p = .0008$ ; Figure 4B, right). Summarizing the results from first- and second-order color contrasts, they demonstrate that color contrasts are potentially salient features, since three out of the four contrasts showed a significant effect on overt attention in color-rotated images. The fact that none of these contrasts showed any significant relation to fixation in unmodified images, however, implies that color contrasts do not contribute to the guidance of human overt attention in outdoor scenes.



**Figure 4. Second-order features. (A) Percentages of trials for which actual second-order luminance contrast (“texture contrast”) was larger than control for the three different stimulus conditions. (B) Percentages of trials for which actual second-order constant blue contrast (cb) and tritanopic confusion contrast (tc) were larger than their respective control values, for naturally colored stimuli (black) and color-rotated stimuli (white). Significance values are from Wilcoxon signed-rank tests.  $*p < .05$ .  $**p < .01$ .**

The qualitative difference between unmodified and color-rotated images implies that human overt attention depends not only on local features, but also to a large extent on the global appearance of the stimulus.

## DISCUSSION

In the present study, we demonstrated that first-order features are unrelated to human overt attention in stimuli depicting outdoor scenes. Of the second-order features investigated, only second-order luminance contrast contributed to overt attention. Because the absence of an effect of first-order features is specific to unmodified stimuli, it cannot be explained by first-order features lacking saliency in general. Rather, our results suggest that global stimulus appearance modulates the effect of local features. Since this global stimulus context in our case also correlates with the subjective perception of a scene being “natural,” it is tempting to speculate that this interpretation itself might modulate the effect of local features on human attention.

The choice of features we investigated was guided by the features most commonly used in experimental and model-

ing studies of human attention. Attention models for static scenes, such as the saliency map and related schemes, employ the features of luminance, color, and/or orientation (Itti & Koch, 2001). Among these, luminance contrast is the feature that has been studied most thoroughly with respect to natural scenes. In grayscale images, luminance contrast correlates with human overt attention (Krieger et al., 2000; Mannan et al., 1996, 1997; Parkhurst & Niebur, 2003; Reinagel & Zador, 1999; Tatler et al., 2005), although this relation is not causal (Einhäuser & König, 2003). In this study, we demonstrated that this correlative effect is valid only for grayscale stimuli, and is absent or reversed for stimuli that contain color. This result is especially remarkable since luminance contrast as such did not differ between these conditions. It is quite conceivable, however, that *perceived* luminance contrast depended on color, in spite of the physical identity of luminosity across conditions. Irrespective of the extent to which color rotation affects perceived luminance, the present results provide further evidence that any effect of luminance contrast on overt attention is strongly modulated by contextual factors. Although those factors may be based on global appearance and/or other (local) features that affect perceived luminance, our findings in either case highlight the importance of context in modulating the effects of local features on attention.

In order to account for the absence of a causal effect of first-order luminance contrast on human overt attention, Parkhurst and Niebur (2004) extended the saliency map model of Itti, Koch, and Niebur (1998) by incorporating a second layer in the luminance channel. They found that second-order luminance contrast (“texture contrast,” in their terms) is about 10 times more relevant than first-order luminance contrast. For our study, we used a definition of second-order luminance contrast that is insensitive to specific model assumptions and generalizes the definition of luminance contrast canonically. We demonstrated directly that second-order luminance contrast is indeed correlated with human overt attention. Unlike the first-order effect, the effect of second-order luminance contrast does not depend on whether or not color is present. Although this result does not necessarily imply that second-order luminance contrast attracts attention causally, it does nevertheless underline the relative importance of higher-order features to overt attention.

In the color channel, saliency map models typically employ color-opponent mechanisms (Itti & Koch, 2001) reminiscent of those observed in retinal ganglion cells and cells of the lateral geniculate nucleus. The activity in these color channels depends on saturation and/or on color contrasts. The use of these features in models of attention inspired our own investigation of them and provided the rationale for the choice of a physiological color space. For unmodified outdoor scenes, we did not observe an effect for either first- or second-order color features. In contrast, we did find saturation, in terms of both the first-order color contrasts and one second-order contrast, to be related to human overt attention in color-rotated stimuli. This is striking, since the configuration of the color features investigated was identical in both conditions. In line

with this result, luminance contrast affects overt attention differently in unmodified images than in images from either the grayscale or color-rotated conditions. Whether this difference is a direct consequence, or just a correlate, of the difference in the subjective perception of a stimulus being “natural” will be an interesting issue for further research. Nonetheless, it is evident from the present results that the local features we investigated are not alone sufficient to account for the effects we observed on overt attention; global stimulus context must also be taken into account.

The stimuli in the present study were deliberately chosen to represent outdoor scenes with few or no human-made objects. Consequently, they only represent a small subset of the stimuli a contemporary human typically encounters. Stimulus category, however, is known to affect the prediction performance of different bottom-up models (Privitera & Stark, 2000) as well as the extent to which low-level features relate to attention (Parkhurst, Law, & Niebur, 2002). By using this particular subset of stimuli, we aimed at reducing potential implicit tasks and the effect of individual experience with particular images and objects as much as possible. The fact that we found, even in such stimuli, a predominant effect of higher-order features and global stimulus appearance therefore highlights the relative importance of such high-level cues.

In the present study, we investigated the effects of bottom-up features and the extent to which these effects are modulated by the global appearance of the stimulus. Although our behavioral data suggest that the latter influence might be related to cognitive interpretation of the stimulus, and therefore to top-down signals, we did not aim at investigating explicit top-down cues, such as subjective experience or the task. Expert experience with a particular set of stimuli has a profound influence on fixation durations (Buswell, 1935). To our knowledge, however, there have been no systematic comparisons between experts and novices regarding the spatial distribution of fixations. This will be an interesting issue for future research, but it is unlikely that experience or expert knowledge plays a role for our set of outdoor stimuli. More important here (and a factor that has been better investigated) is the role of the task on human fixations (Henderson, 2003; Land, Mennie, & Rusted, 1999; Yarbus, 1967). Although we instructed our subjects merely to “study the images carefully” and thus minimized explicit tasks, any complete description of human attention must take task-related information into account. For a visual search task with noise stimuli whose power spectrum matches those of our natural scenes, Najemnik and Geisler (2005) demonstrated that the assumption of an ideal Bayesian observer predicts human fixation patterns well. Several recent models of overt attention in natural scenes (Navalpakkam & Itti, 2005; Oliva, Torralba, Castelano, & Henderson, 2003; Torralba, 2003) have also included prior knowledge of the task or of stimulus category, and adding the component of prior knowledge does indeed increase the prediction performance of their models. However, under free viewing conditions and in the absence of an explicit (search) task, the choice of the priors for such models is far from obvious. Further research along

the lines of the present study on the interaction between global stimulus appearance and local features will thus be needed to constrain such models. Our present findings, which demonstrate the importance of global stimulus context, even in the absence of an explicit task and for stimuli with little semantic content, underline the importance of this investigation in the search for a complete description of human visual attention.

#### AUTHOR NOTE

We thank D. C. Kiper for his invaluable support and fruitful discussions. The work was financially supported by Honda RI Europe, EU/BW (“AMOUSE Project”), and the Swiss National Science Foundation (Grant PBEZ2-107367 to W.E.). Correspondence relating to this article may be sent to H.-P. Frey, Institute of Cognitive Science, Department of Neurobiopsychology, University of Osnabrück, Albrechtstrasse 28, 49076 Osnabrück, Germany (e-mail: hfrey@uos.de).

#### REFERENCES

- BACH, M., BOUIS, D., & FISCHER, B. (1983). An accurate and linear infrared oculometer. *Journal of Neuroscience Methods*, *9*, 9-14.
- BRAINARD, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433-436.
- BUSWELL, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago: University of Chicago Press.
- DERRINGTON, A. M., KRAUSKOPF, J., & LENNIE, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *Journal of Physiology*, *357*, 241-265.
- EINHÄUSER, W., & KÖNIG, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience*, *17*, 1089-1097.
- GEGENFURTNER, K. R., & KIPER, D. C. (2003). Color vision. *Annual Review of Neuroscience*, *26*, 181-206.
- GOTTLIEB, J. P., KUSUNOKI, M., & GOLDBERG, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*, 481-484.
- HENDERSON, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498-504.
- HORWITZ, G. D., & NEWSOME, W. T. (1999). Separate signals for target selection and movement specification in the superior colliculus. *Science*, *284*, 1158-1161.
- ITTI, L., & KOCH, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194-203.
- ITTI, L., KOCH, C., & NIEBUR, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *20*, 1254-1259.
- KOCH, C., & ULLMAN, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219-227.
- KOMATSU, H., IDEURA, Y., KAJI, S., & YAMANE, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *Journal of Neuroscience*, *12*, 408-424.
- KRIEGER, G., RENTSCHLER, I., HAUSKE, G., SCHILL, K., & ZETZSCHE, C. (2000). Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. *Spatial Vision*, *13*, 201-214.
- KUSTOV, A. A., & ROBINSON, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, *384*, 74-77.
- LAND, M., MENNIE, N., & RUSTED, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311-1328.
- LI, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*, 9-16.
- MANNAN, S. K., RUDDOCK, K. H., & WOODING, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*, 165-188.
- MANNAN, S. K., RUDDOCK, K. H., & WOODING, D. S. (1997). Fixation patterns made during brief examination of two-dimensional images. *Perception*, *26*, 1059-1072.



- MAZER, J. A., & GALLANT, J. L. (2003). Goal-related activity in V4 during free viewing visual search: Evidence for a ventral stream visual salience map. *Neuron*, **40**, 1241-1250.
- MCPEEK, R. M., & KELLER, E. L. (2002). Superior colliculus activity related to concurrent processing of saccade goals in a visual search task. *Journal of Neurophysiology*, **87**, 1805-1815.
- NAJEMNIK, J., & GEISLER, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, **434**, 387-391.
- NAVALPAKKAM, V., & ITTI, L. (2005). Modeling the influence of task on attention. *Vision Research*, **45**, 205-231.
- OLIVA, A., TORRALBA, A., CASTELHANO, M. S., & HENDERSON, J. M. (2003). Top-down control of visual attention in object detection. In *Proceedings of the 2003 International Conference on Image Processing (ICIP 2003)* (Vol. 1, pp. 253-256). Washington, DC: IEEE Press.
- PARKHURST, D. [J.], LAW, K., & NIEBUR, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, **42**, 107-123.
- PARKHURST, D. J., & NIEBUR, E. (2003). Scene content selected by active vision. *Spatial Vision*, **16**, 125-154.
- PARKHURST, D. J., & NIEBUR, E. (2004). Texture contrast attracts overt visual attention in natural scenes. *European Journal of Neuroscience*, **19**, 783-789.
- PELLI, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.
- PETERS, R. J., IYER, A., ITTI, L., & KOCH, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, **45**, 2397-2416.
- POSNER, M. I., & PETERSEN, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, **13**, 25-42.
- PRIVITERA, C. M., & STARK, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **22**, 970-982.
- REINAGEL, P., & ZADOR, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, **10**, 341-350.
- ROBINSON, D. L., & PETERSEN, S. E. (1992). The pulvinar and visual salience. *Trends in Neurosciences*, **15**, 127-132.
- TATLER, B. W., BADDELEY, R. J., & GILCHRIST, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, **45**, 643-659.
- THOMPSON, K. G., BICHOT, N. P., & SCHALL, J. D. (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *Journal of Neurophysiology*, **77**, 1046-1050.
- TORRALBA, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America A*, **20**, 1407-1418.
- WALTHER, D., EDGINGTON, D. R., & KOCH, C. (2004). Detection and tracking of objects in underwater video. In *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)* (Vol. 1, pp. 544-549). Washington, DC: IEEE Press.
- YARBUS, A. L. (1967). *Eye movements and vision* (B. Haigh, Trans.). New York: Plenum.
- ZEKI, S. (1983). The distribution of wavelength and orientation selective cells in different areas of monkey visual cortex. *Proceedings of the Royal Society of London: Series B*, **217**, 449-470.

(Manuscript received April 8, 2005;  
revision accepted for publication April 11, 2006.)