Influence of local scene colour on target detection tested by global rearrangement of natural scenes

Kinjiro Amano, David H. Foster, Matthew S. Mould, and John P. Oakley; School of Electrical and Electronic Engineering, University of Manchester; Manchester, UK

Abstract

Local scene colour can influence the visual detectability of an object or target, but so can the familiarity, meaning, and global organisation of the scene. The aim of this study was to test whether the effects of local scene colour on target detectability are secondary to global effects. A target-detection task was undertaken by human observers with coloured images of natural scenes that were cut into quarters, randomly rearranged, and then reassembled. The target was a small, shaded, neutral grey sphere located randomly within the scene and matched in mean luminance to its local surround. It was found that observers' target-detection performance with the rearranged images was about as good as with the original images. The combination of local colour properties, namely, lightness and the red-green and yellow-blue components of chroma, accounted, respectively, for 55% and 50% of observers' detection performance with the original and rearranged images. Despite the disruption of global organisation, local scene colour continued to influence target detection.

Introduction

Although an everyday task, visually detecting an object or target in a scene presents a challenge for analysis. This is because performance depends on the structure of the scene being searched as well as on its contents [1-3], both of which can vary markedly in the natural world.

It is often assumed that spatial achromatic image features of scenes, such as edges, texture, orientation, and contrast [4-6] are the main determinants of target detectability, but local scene colour can also be influential, accounting for about 60% of the variance in observers' performance with natural scenes [7]. Yet it is unclear whether the effects of local colour properties in natural scenes are truly local, in the sense that detection still takes place within a coherently structured environment, with all the advantages of familiarity, meaning, and global scene organisation that are immediately apparent to an observer [8-10]. It is possible, therefore, that the effects of local colour properties on target detection are secondary to the effects of the more global properties of natural scenes. The aim of the present study was to test this alternative interpretation.

A target-detection task was undertaken by human observers with coloured images of natural scenes that were cut into quarters, randomly rearranged, and then reassembled [8-10]. Figure 1 shows the original and rearranged images of two example scenes. The target was a small, shaded, neutral grey sphere located randomly within the scene and matched in mean luminance to its local surround. Target-detection performance was quantified by the discrimination index d' from signal-detection theory [11], which is intended to reduce the effects of observer bias. The spatial variation of d' over each scene was compared with the spatial variation of local colour properties quantified in the colour space CIECAM02 [12-14].

It was found that observers could detect the target in the rearranged images almost as well as in the original images. When combined, the local colour properties of lightness and the red-green and yellow-blue components of chroma accounted on average for 50% of the adjusted variation in observers' detection performance when images were rearranged and for 55% when they were not. These levels are similar to levels reported elsewhere for the predictability of target detection and of eye fixations [7], with the remaining variation usually attributed to cognitive or neural factors.

Methods

Apparatus

As in [7], images of natural scenes were simulated on a 20-inch cathode-ray tube display (GDM-F520, Sony Corp., Tokyo, Japan) controlled by a graphics system (Fuel, Silicon Graphics Inc., CA, USA) with spatial resolution 1600 × 1200 pixels, refresh rate approx. 60 Hz, and intensity resolution 10 bits on each of the red, green, and blue channels. The display system was calibrated with a telespectroradiometer (SpectraColorimeter PR-650, Photo Research Inc., Chatsworth, CA, USA) and photometer (LMT, L1003, Lichtmesstechnik GmbH, Berlin, Germany), whose accuracy was verified against other calibrated instruments. The colorimetric accuracy of the system was tested in each experimental session and the system regularly recalibrated. Errors in a coloured test patch were <0.005 in CIE (Commission Internationale de l'Eclairage) 1931 (x, y) chromaticity coordinates and <5% in luminance.

Stimuli

Again as in [7], images of twenty natural scenes were obtained from a set of hyperspectral data [15], which allowed accurate calculation of colour properties under constant illumination, namely, a daylight of correlated colour temperature 6500 K, corresponding to an average daylight [16]. The images subtended approximately 17 × 13 degree of visual angle at the viewing distance of 1 m. The mean luminance of the images on the screen was 3.6 cd m^{-2} (range 0–61.4 cd m^{-2}).

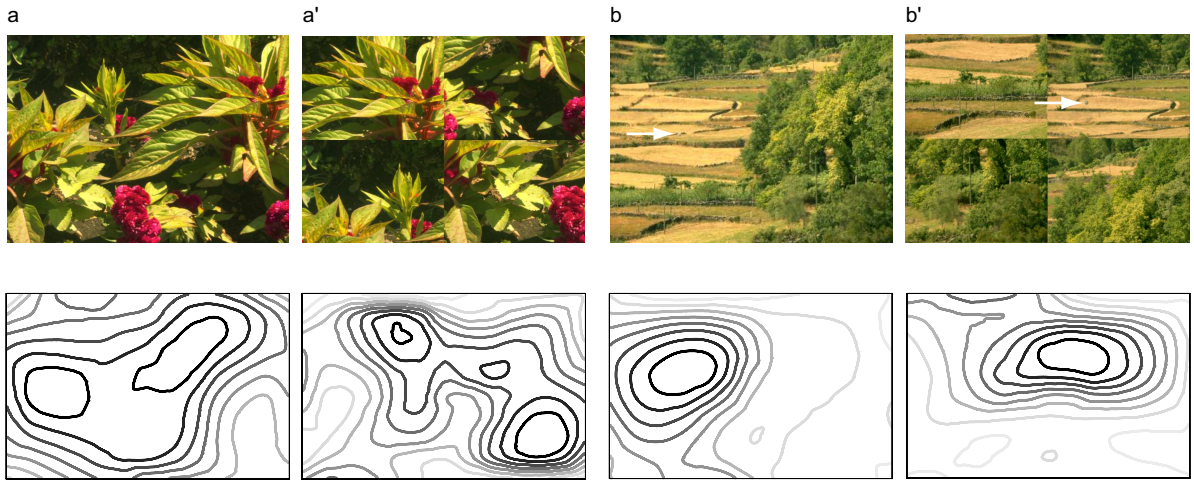

Figure 1. *Two example scenes and observers' detection performance. In the top row, images of the scenes are shown without spatial rearrangement (a and b), and with spatial rearrangement (a' and b'). The targets are indicated by the arrows in b and b'. The bottom row shows the spatial distributions of observers' smoothed detection performance d' from signal-detection theory [11] for the corresponding images in the top row. Higher values of d' are indicated by darker contours. The loess smoothing bandwidth was 0.2.*

Images of the scenes were cut into quarters, randomly rearranged by translations only, and then reassembled [8]. Division into quarters rather than into e.g. eighths or smaller portions was determined by the size of the estimated functional visual field in object-search tasks and in contrast-sensitivity measurements [17-19], and by the known fixation patterns of observers [7].

The target was a shaded sphere, indicated by the arrows in Fig. 1 b and b', top row. Its surface was spectrally neutral (Munsell N7) and it subtended approximately 0.25 degree of visual angle in the images of each of the twenty scenes. It appeared randomly in each image at one of 130 possible locations, defined by an imaginary 13 × 10 grid. Since the sphere was physically placed in the scene at the time of hyperspectral imaging, the illumination and shading on the sphere were consistent with the illumination on the scene. Trivial brightness cues to detection were eliminated by matching the average luminance of the target at each location to the average luminance of its local surround (< 1.0 degree extent), but because of its spherical shape and variation in shading, the target was not isoluminant with its local surround. The relatively small angular subtense of the target was chosen, as a result of preliminary measurements, to encourage observers to inspect the entire image [20] and avoid the so-called central bias in gaze behaviour [21, 22].

Procedure

In each trial, observers were presented with the image of a particular scene for 1 s. They had to indicate whether they saw the target or not by pressing buttons on a computer mouse. Observers were allowed to move their eyes during the trial and had unlimited time to respond.

The two kinds of images of each scene, i.e. the original and rearranged images, appeared repeatedly in separate experimental blocks to reveal scene-specific effects [7, 23]. Each block consisted of two sub-blocks, each of 130 trials. Half of the trials contained the target and the other half did not. In total, each observer performed no less than 10,400 trials (2 image arrangements × 20 scenes × 260 trials).

Observers

Six observers (one female and five male, aged 22–26 yr) took part in the experiment. All had normal or corrected-to-normal visual acuity and normal colour vision verified with a series of colour vision tests (Farnsworth-Munsell 100-Hue test, Ishihara pseudoisochromatic plates, Rayleigh and Moreland anomaloscopy with luminance test). All except one (coauthor MSM) were unaware of the purpose of the experiment. The experimental procedure was approved by the University of Manchester Committee on the Ethics of Research on Human Beings, which operated in accord with the principles of the Declaration of Helsinki.

Analysis

The discrimination index d' was calculated as follows [11]. Suppose that HR is the hit rate, i.e. the frequency with which the target was reported as being present when it was present, and suppose that FAR is the false-alarm rate, i.e. the frequency with which the target was reported as being present when it was not. Then $d' = \Phi^{-1}(\mathrm{HR}) - \Phi^{-1}(\mathrm{FAR})$, where Φ is the normal cumulative distribution function. The HR was recorded at each of the 130 locations within the image. Since the position of the target was not defined when the target was absent, the FAR was assumed to be constant over each scene, although different from scene to scene. To ensure that there were enough trials in each scene at each location, observers' detection responses were pooled over observers.

To further improve the accuracy of the d' estimates at each of the 130 locations, the distribution of raw d' values over each scene was smoothed by a locally weighted quadratic polynomial regression, loess [24, 25]. Its bandwidth, which defines the size of the local neighbourhood, that is, the proportion of data included in each local fit, was set to 0.15, 0.2, and 0.3; in general, the larger the bandwidth, the less detailed the smoothed spatial profile. In fact, there was little difference in the resulting proportions of variance accounted for by colour properties.

The choice of the CIE colour appearance model CIECAM02 [12-14] for quantifying the local colour properties

of each scene was for computational convenience, deriving from the approximate perceptual uniformity of CIECAM02, rather than for any particular appearance attributes. This approach was simpler than the alternative of using a separate non-uniform colour space, e.g. CIELAB [16], and an associated non-Euclidean metric based on e.g. the colour-difference formula CIEDE2000 [16, 26].

Among other attributes, CIECAM02 provides lightness (J), chroma (C), and hue (h) in a polar coordinate system, or, equivalently, lightness (J) and the red-green chroma component (a_C) and yellow-blue chroma component (b_C) in a Cartesian coordinate system. The parameters [14] of the CIECAM02 model were set so that the white point was D65, the luminance of the background 20% of the white level, and the surround "average".

Results

Mean target detection

Figure 1, bottom row, shows the smoothed spatial distribution of d' values for the corresponding images in the top row. Higher values of d' are indicated by the darker contours. It can be seen that there are systematic effects of scene structure on detection performance (examined in detail in the next subsection), but there are also variations in mean d' values from scene to scene.

Figure 2 shows values of mean d' for the rearranged images plotted against the corresponding mean values for the original images. Values for the particular scenes in Fig 1, top row, are arrowed. Although mean detection performance varied markedly over scenes, crucially it was similar whether images were rearranged or not. A formal test of the effects of rearrangement showed it not to be significant ($t(37) = 1.01$, $p > 0.3$), and the correlation between the two sets of d' values was high, with Pearson's $r = 0.85$.

Two scenes produced extremely low mean d' values, which (by chance) were actually negative (bottom left of Fig. 2). One of these scenes was a medium-distance view of buildings with arrays of windows, which may have masked the target. The other scene was a distant view of uniform green terrain, in which the target appeared to merge with the background.

Regression on colour properties

The colour properties at each point of each image of each scene were, as noted earlier, quantified in the approximately uniform colour space CIECAM02. The spatial distributions of lightness J and the red-green and yellow-blue components of chroma a_C, and b_C, respectively, were smoothed by the same locally weighted quadratic regression as for d' (Fig. 1, bottom row). Figure 3 shows the smoothed spatial distribution of J, a_C, and b_C in the top, middle, and bottom rows, respectively, for the scenes a and b in Fig. 1, top row. Higher values of the variable are again indicated by darker contours.

The capacity of local colour properties to explain observers' target-detection performance was quantified by regressing the smoothed distribution of d' values (as in Fig. 1, bottom row) on the smoothed distribution of J, a_C, and b_C values (as in Fig. 3, top, middle, and bottom rows, respectively) for each scene. Thus, if $\hat{d}'(x,y)$, $\hat{J}(x,y)$, $\hat{a}_C(x,y)$, and $\hat{b}_C(x,y)$ represent the smoothed values of d', J, a_C, and b_C, respectively, at location (x, y), then the full regression equation had the following form:

$$\mathrm{E}\left[\hat{d}'(x,y)\right] = \beta_1\hat{J}(x,y) + \beta_2\hat{a}_C(x,y) + \beta_3\hat{b}_C(x,y) + \alpha, \qquad (1)$$

where E is the expectation, and β_1, β_2, β_3, and α are scalars.

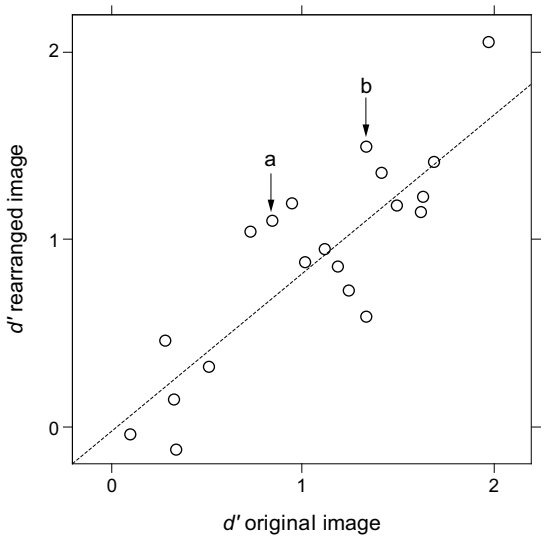

Figure 2. *Observers' mean target-detection performance d' for the rearranged images plotted against the corresponding mean values for the original images. The dashed line is a linear regression fit with Pearson's r = 0.85. Values for the particular scenes in Fig 1, top row, are indicated.*

Goodness of fit was summarized by R^2, which represents the proportion of variance accounted for. Values of R^2 were adjusted for the loss in degrees of freedom (d.f.) in smoothing [25] and in fitting the regressor variables [27]. The adjusted value of R^2 was defined as $1 - (1 - R^2)(n - 1)/(n - k)$, where n is the d.f. of the smooth [25] and k is the number of regressor variables considered (e.g. $k = 2$ for the regression on J and $k = 4$ for the regression on J, a_C, and b_C).

Table 1 lists the adjusted R^2 values, averaged over the twenty scenes, for each of the local colour properties J, a_C, and b_C, and their combination for the original and rearranged images. Results for all three loess smoothing bandwidths 0.15, 0.2, and 0.3 are shown.

The adjusted R^2 values for each of the colour properties J, a_C, and b_C ranged over 27–33% with the original images and 27–32% with the rearranged images, depending on bandwidth. But when all three properties were combined, the adjusted R^2 values reached 55–56% with the original images (slightly lower than in an earlier study with different observers [7]) and 49–52% with the rearranged images, again depending on bandwidth. A formal test of the effect of rearrangement showed it not to be significant, either for individual values of J, a_C, and b_C ($t(37) \le 0.7$, $p \ge 0.5$) or for their combination ($t(37) \le 0.75$, $p \ge 0.5$). Notice that the sum of the R^2 values from each of the colour properties was less than when they were combined, showing that in spite of their orthogonality in CIECAM02, there were redundancies in their contribution to target detection.

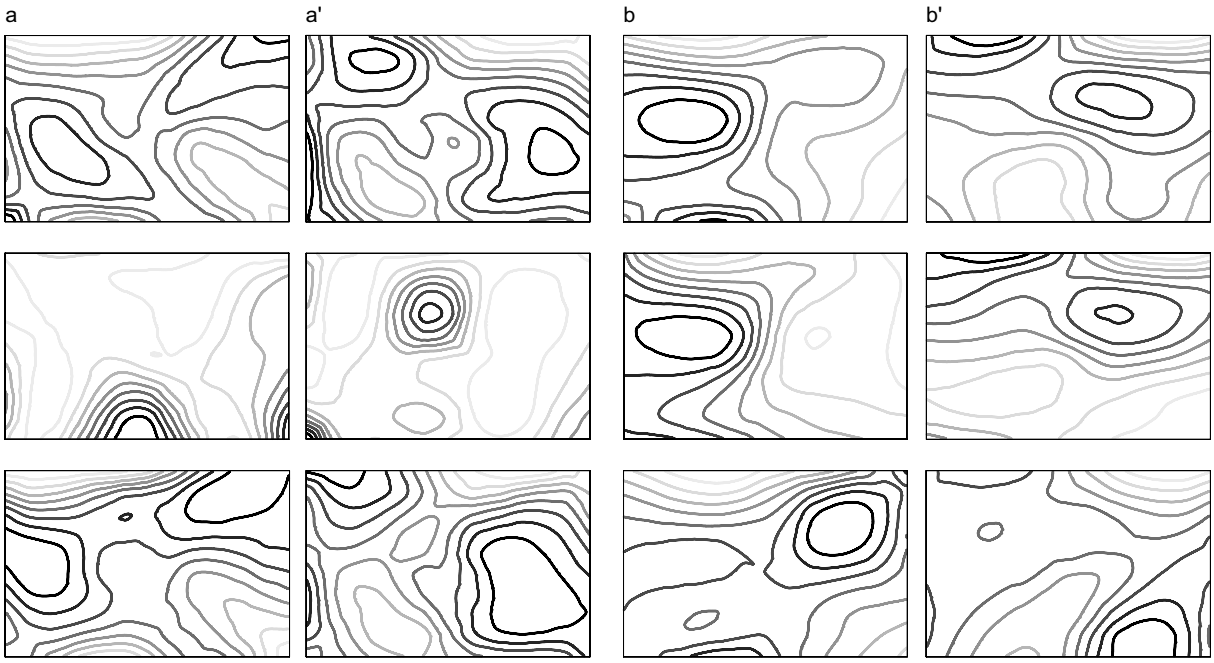

Figure 3. *Spatial distributions of colour properties for each of the corresponding scenes in Fig. 1. The top row shows smoothed lightness J; the middle row, the smoothed red-green chroma component a_C; and the bottom row, the smoothed yellow-blue chroma component b_C. Higher values of the variable are indicated by darker contours. The loess smoothing bandwidth was 0.2.*

Table1. Regression of observers' target-detection performance on local colour properties with original and rearranged images. Values of R^2 are shown for each of the local colour properties and their combinations, with the values of R^2 adjusted for loss in degrees of freedom in smoothing [25] and in fitting [27]. Values of the loess smoothing bandwidth were 0.15, 0.2, and 0.3, as indicated. Means and standard errors (SEs) were obtained over 20 natural scenes.

		Original image		Rearranged image	
Bandwidth	Colour property	Mean %	SE %	Mean %	SE %
0.15	Lightness *J*	32.9	5.9	30.4	5.5
	Chroma a_C	27.3	6.1	26.9	5.3
	Chroma b_C	32.4	5.1	28.2	4.7
	Combined $J+a_C+b_C$	54.9	5.6	48.9	4.9
0.2	Lightness *J*	33.3	6.1	31.0	5.5
	Chroma a_C	27.7	6.0	27.8	5.2
	Chroma b_C	32.3	5.0	27.9	4.6
	Combined $J+a_C+b_C$	55.6	5.5	50.2	4.7
0.3	Lightness *J*	33.4	6.3	31.8	5.4
	Chroma a_C	27.9	5.8	27.8	5.1
	Chroma b_C	31.4	4.9	26.8	4.4
	Combined $J+a_C+b_C$	55.7	5.3	51.9	4.3

Conclusion

Observers' overall ability to detect a target in natural scenes was little affected by the global rearrangement of the scenes. Local colour properties of the scenes accounted, on average, for about 55% of the variation in performance with the original images and about 50% with the rearranged images. These levels are similar to the maximum level reported for the effects of spatial scene structure on eye movements in free viewing of black-and-white video images [28].

Although in all these kinds of measurements, there is a large proportion of variation unexplained by image features, it is interesting that local scene colour can be as influential as spatial achromatic features in determining observers' target-detection performance, and that this influence persists despite the disruption of global organisation produced by image rearrangement.

Acknowledegments

We thank Gaoyang Feng for critically reading the manuscript. This work was supported by the Engineering and

Physical Sciences Research Council, UK (grant EP/F023669/1). A partial report of the results was presented at the European Conference on Visual Perception (ECVP), Regensburg, Germany, August, 2009.

References

[1] A. Treisman and S. Sato, Conjunction Search Revisited, J. Exp. Psychol.-Hum. Percept. Perform., 16, 459 (1990).
[2] S. D. Buluswar and B. A. Draper, Color Machine Vision for Autonomous Vehicles, Eng. Appl. Artif. Intell., 11, 245 (1998).
[3] J. M. Wolfe, M. L.-H. Võ, K. K. Evans, and M. R. Greene, Visual Search in Scenes Involves Selective and Nonselective Pathways, Trends Cogn. Sci., 15, 77 (2011).
[4] D. H. Foster and P. A. Ward, Asymmetries in Oriented-Line Detection Indicate Two Orthogonal Filters in Early Vision, Proc. R. Soc. Lond. B, 243, 75 (1991).
[5] J. M. Wolfe, Visual Search in Continuous, Naturalistic Stimuli, Vision Res., 34, 1187 (1994).
[6] L. Itti, C. Koch, and E. Niebur, A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, IEEE Trans. Pattern Anal. Mach. Intell., 20, 1254 (1998).
[7] K. Amano, D. H. Foster, M. S. Mould, and J. P. Oakley, Visual Search in Natural Scenes Explained by Local Color Properties, J. Opt. Soc. Am. A-Opt. Image Sci. Vis., 29, A194 (2012).
[8] I. Biederman, Perceiving Real-World Scenes, Science, 177, 77 (1972).
[9] T. Foulsham, R. Alan, and A. Kingstone, Scrambled Eyes? Disrupting Scene Structure Impedes Focal Processing and Increases Bottom-up Guidance, Atten. Percept. Psychophys., 73, 2008 (2011).
[10] M. R. Greene and J. M. Wolfe, Global Image Properties Do Not Guide Visual Search, J. Vision, 11(6):18, 1 (2011).
[11] D. M. Green and J. A. Swets, Signal Detection Theory and Psychophysics (Wiley, New York, 1966).
[12] M. R. Luo, G. Cui, and C. Li, Uniform Colour Spaces Based on CIECAM02 Colour Appearance Model, Color Res. Appl., 31, 320 (2006).
[13] M. Melgosa, R. Huertas, and R. S. Berns, Performance of Recent Advanced Color-Difference Formulas Using the Standardized Residual Sum of Squares Index, J. Opt. Soc. Am. A-Opt. Image Sci. Vis., 25, 1828 (2008).
[14] CIE, Technical Committee 8-01, A Colour Appearance Model for Colour Management Systems: CIECAM02, CIE 159:2004 (Commission Internationale de l'Éclairage, Vienna, Austria, 2004).
[15] D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, Frequency of Metamerism in Natural Scenes, J. Opt. Soc. Am. A-Opt. Image Sci. Vis., 23, 2359 (2006).
[16] CIE, Technical Committee 1-48, Colorimetry, 3rd Edition, CIE 015:2004 (Commission Internationale de l'Éclairage, Vienna, Austria, 2004).
[17] D. Parkhurst, E. Culurciello, and E. Niebur, Evaluating Variable Resolution Displays with Visual Search: Task Performance and Eye Movements, Proc. ETRA 2000, pg. 105 (2000).
[18] P. Reinagel and A. M. Zador, Natural Scene Statistics at the Centre of Gaze, Netw.-Comput. Neural Syst., 10, 341 (1999).
[19] J. Rovamo and V. Virsu, An Estimation and Application of the Human Cortical Magnification Factor, Exp. Brain Res., 37, 495 (1979).
[20] H. Corbin, J. Carter, E. P. Reese, and J. Volkmann, Experiments on Visual Search 1956-1957 (Psychological Research Unit, Mount Holyoke College, 1958).
[21] S. K. Mannan, K. H. Ruddock, and D. S. Wooding, The Relationship between the Locations of Spatial Features and Those of Fixations Made During Visual Examination of Briefly Presented Images, Spatial Vis., 10, 165 (1996).
[22] B. W. Tatler, The Central Fixation Bias in Scene Viewing: Selecting an Optimal Viewing Position Independently of Motor Biases and Image Feature Distributions, J. Vision, 7(14):4, 1 (2007).
[23] K. Kaspar and P. König, Viewing Behavior and the Impact of Low-Level Image Properties across Repeated Presentations of Complex Scenes, J. Vision, 11(13):26, 1 (2011).
[24] W. S. Cleveland, Robust Locally Weighted Regression and Smoothing Scatterplots, J. Am. Stat. Assoc., 74, 829 (1979).
[25] T. J. Hastie and R. J. Tibshirani, Generalized Additive Models (Chapman and Hall/CRC, London, 1990).
[26] M. R. Luo, G. Cui, and B. Rigg, The Development of the CIE 2000 Colour-Difference Formula: CIEDE2000, Color Res. Appl., 26, 340 (2001).
[27] D. B. Montgomery and D. G. Morrison, A Note on Adjusting R^2, J. Financ., 28, 1009 (1973).
[28] W. Kienzle, B. Schölkopf, F. A. Wichmann, and M. O. Franz, How to Find Interesting Locations in Video: A Spatiotemporal Interest Point Detector Learned from Human Eye Movements, Pattern Recognition: Lecture Notes in Computer Science, vol. 4713 (Springer Berlin / Heidelberg, Berlin, Germany, 2007), pg.405.

Author Biography

Kinjiro Amano (k.amano@manchester.ac.uk) received his Ph.D. in 1998 from the Tokyo Institute of Technology, Japan. He has worked at Aston University, Birmingham, and UMIST, Manchester, and is currently at the University of Manchester. His research work has concentrated on the psychophysics of human colour perception.

