Brief communication

Spatio-chromatic properties of natural images and human vision

C.A. Párraga*, T. Troscianko†, D.J. Tolhurst‡

*Department of Experimental Psychology, University of Bristol, 8 Woodland Rd. Bristol, BS8 1TN, UK.
†School of Cognitive and Computing Sciences, University of Sussex, Brighton, BN1 9QH, UK.
‡The Department of Physiology, University of Cambridge, Downing Street, Cambridge, CB2 3EG, UK.

Correspondence: C.A. Párraga

E-Mail: Alej.Parraga@bristol.ac.uk

Running head: Vision is optimised for natural images.

Corresponding author: C.A. Párraga (Alej.Parraga@bris.ac.uk)

Supplementary material in electronic form is available on the following web address:

Abstract

The human visual system shows a relatively greater response to low spatial frequencies of chromatic spatial modulation than of luminance spatial modulation[1]. However, previous work[2,3] has shown that this differential sensitivity to low spatial frequencies is not reflected in any differential luminance and chromatic content of general natural scenes. This is contrary to the prevailing assumption that the spatial properties of human vision ought to reflect the structure of natural scenes[4-6]. Now, colorimetric measures of scenes suggest that red-green (and perhaps blue-yellow) colour discrimination in primates is particularly suited to the encoding of specific scenes: reddish or yellowish objects on a background of leaves[7-10]. We therefore ask whether the *spatial*, as well as chromatic, properties of such scenes are matched to the different spatial encoding properties of colour and luminance modulation in human vision. We show that the spatio-chromatic properties of a wide class of scenes, which contain reddish objects (e.g. fruit), on a background of leaves, correspond well to the properties of the red-green (but not blue-yellow) systems in human vision, at viewing distances commensurate with typical grasping distance. This implies that the red-green system is particularly suited to encoding both the spatial and the chromatic structure of such scenes.

Results and Discussion

Fig.1 shows two very different examples of photographs of natural scenes: (a) a landscape view of a British garden and (b) a close-up of a single ripe tomato seen against foliage. We calculated how the human L, M and S (for long-, medium- and short-wavelength sensitive) cones would have responded at each point in each of the scenes, and we then calculated the spatial form of the luminance signal and the red-green and blue-yellow chromatic signals within the human visual system. A simple way of representing the amount of Fourier amplitude at different spatial frequencies in complex images is given by plotting the spectral slope of the signal. If the amplitudes of all the Fourier components of natural images are plotted on logarithmic axes against their spatial frequencies, one obtains an approximately straight-line plot, with a slope (α) around -1.2[11,12]. This slope is said to have evolutionary importance, since our visual system is optimised for processing the spatial information in natural scenes at this particular value[13] and there are theoretical reasons why there should be a strong relationship between α and cortical receptive field profiles[6]. Essentially, a match between the scene content and the visual system's response maximises the signal-to-

noise ratio of the neural representation. Fig.2a shows just such plots for the luminance and for the two chromatic signals for the garden scene of Fig.1a. It is clear that the slopes (α) are very similar for the three graphs, though it should be noted that the magnitude of the amplitude in the red-green signals was considerably less than in the blue-yellow signal (the luminance has to be plotted on a different scale). Since human colour vision is biased more towards low spatial frequencies than is luminance vision, we might have expected the spectral slope of the chromatic spatial signals to be steeper than that of the luminance signal, given that a steeper spectral slope gives a greater weight to low spatial frequencies. The failure to find such an effect for the scene in Fig.1a confirms the earlier study of Párraga *et al*[2] using hyperspectral images.

Insert Figure1.tif about here

Insert Figure2.tif about here

In that earlier study, the spatio-chromatic properties of 29 natural scenes were analysed, but they mostly consisted of views of foliage and gardens such as that in Fig.1a. However, there is a body of literature [7-10] which suggests that primate (including human) colour channels may optimally encode the chromatic difference between red or yellow fruit and green foliage (or, indeed, of any red or yellow objects on green foliage), so it is important to examine the spectral slopes of similar natural images; i.e. of fruit in foliage, similar to those used by these researchers. Fig.2b shows plots of the Fourier amplitude spectra of the luminance and the two chromatic spatial signals of the close-up image of the ripe tomato of Fig.1b. It is, indeed, clear that the spectral slope (α) of the red-green spatial signal is now steeper than those of the luminance or blue-yellow spatial signals; the total amplitude for the red-green signal is, not surprisingly, much greater than for the general garden scene.

Fig.3 shows the generality of our finding that the amplitude spectrum of the red-green spatial signal is steep for close-up images of reddish objects viewed against foliage. Fig.3a plots the ratio between the spectral slopes obtained for the luminance spatial signal and the red-green spatial signal for each scene as a function of how many "red" pixels (in %) there were in the scene (see Methods for definition of a "red" pixel). It can be seen that the results for all scenes follow a similar trend, and are differentiated on the basis of red pixel content (note that scenes consisting of a background of foliage and a reddish object such as fruit appear on the right side of the graph, while scenes of landscapes with no red objects appear on the left). The important predictor of slope ratio appears to be the proportion of red pixels on the leafy background. In general, the

spectral slope of the red-green signal is greater than that of the luminance signal whenever the scene contains many red pixels. When those red pixels are grouped together to represent a single object (like Fig.1b), the slopes of the spectra are especially steep and the slope ratio deviates most from unity. When the pixels are clustered into several groups, representing several distinct fruits, the slopes of the amplitude spectra are not as steep, but the slope for the red-green signal is always steeper than for the other two signals.

Insert Figure3.tif about here

In contrast, Fig.3b shows that the spectral slope of the blue-yellow spatial signal does not differ systematically from that of the luminance signal. There is no evidence in this set of images to suggest that there is any advantage in the bias of the human blue-yellow system to low spatial frequencies. There may be other constraints on the development of the spatial characteristics of blue-yellow colour vision in primates. For example, it is possible that low sensitivity to high blue-yellow spatial frequencies is a consequence of chromatic aberration in the eye and/or the sparse array of short-wave sensitive cones (see Conclusions).

If the human contrast sensitivity function is optimised to match the spatial frequency content of natural images, then we should find that chromatic images contain more low spatial frequency amplitude than luminance images, or in other words, that the ratio ($\alpha_{lum}/\alpha_{chrom}$) should be less than unity. We calculated the optimal value for this ratio, assuming optimal capture by the contrast sensitivity functions[1]. This optimal value is around 0.76 (see Methods). Fig.4 plots the ratio of the spectral slopes, (for the luminance image and the red-green chromatic image), against the distance of the main objects in the scene from the camera. It is clear that the distant scenes like Fig.1a have slope ratios around unity, but close views of red objects against foliage have lower slope ratios. The ratio matches the calculated optimal value for the human csf at a viewing or grasping distance of about 0.4 m. This estimate does require some assumptions and would be influenced, for instance, if we had taken images with a different focal length of camera lens. However, our estimate of 0.4 m will be close to the distance at which the human red-green spatio-chromatic system will be optimised for identifying red or yellow fruits against foliage.

Insert Figure4.tif about here

Although our stimuli were pictures of fruit and leaves obtained in England, and not in a forest in which primate colour vision and relevant vegetation are thought to have co-evolved, a wide variety of types of scene can match the human contrast sensitivity data well. For this reason, we believe that this pattern of results would extend to any similar imaging situation, be it located in another garden or in a relevant forest. This also implies that we cannot argue for the optimality of encoding of any particular type of reddish or yellowish object on a leafy background – any such object would do just as well, whether it be a red fruit[7-9], a reddish human face, or a leaf which is more yellow than the rest[10]. Thus, these results do not specify exactly which kinds of images and tasks were associated with the evolution of colour vision in primates. We simply find that human (and presumably other primate) colour vision is efficient at encoding images of reddish or yellowish fruit, viewed up close, against a leafy background. It may be that (as with the sampling of blue light) the low-frequency bias of the red-green system has been determined by factors such as chromatic aberration, but we find that such a bias would not disadvantage a foraging primate and it might even aid the task of picking ripe fruit or leaves.

Conclusions

Our results therefore suggest that it is the red-green system whose spatio-chromatic properties are optimised for detecting reddish or yellowish objects against a background of foliage; while the blue-yellow system retains the same spectral slope as the luminance system. Returning to Fig.1b, we can see why this might be. The red-green system eliminates the dappled background of leaves[10], giving a representation of a *uniform* fruit on a *uniform* background. This allows the fruit to "pop out" from the background. Perhaps surprisingly, the blue-yellow image does not provide a uniform background – possibly since shadow regions have indirect illumination from skylight (which is blue), but also because green leaves differ in their blueness[10]. This behaviour of the blue-yellow system implies that it is not optimised for detecting objects in foliage. Many questions remain. For example, the human blue-yellow system also has a low-pass transfer characteristic – and yet this seems to have little connection with the efficient detection of fruit or any other target that we have discovered. It may be that there is a different class of images which will reveal similar optimisation of the blue-yellow system, or there may be other constraints on the spatiochromatic properties of this system, such as the sparse sampling by the mosaic of short-wavelength sensitive cones in the retina, which may be related to the fact that short wavelengths are out of focus on the retina due to chromatic aberration. We have not investigated these issues

here. We suggest that a reasonable, although clearly tentative, conclusion from the present work is that the spatio-chromatic properties of the red-green system of human colour vision may be optimised for the encoding of any reddish or yellowish objects on a background of foliage, at relatively small viewing distances.

Methods

Obtaining the images:

We carefully calibrated a digital camera (Nikon Coolpix 950) so that each pixel in the (uncompressed) output image represented the capture of each of the three human cone types. The calibration process utilized a set of narrowband optical interference filters spanning the range from 400 to 700 nm, a constant-current light source, and a Topcon SR-1 telespectroradiometer calibrated against a National Physical Laboratory standard light source. This procedure removed the various nonlinearities introduced by the camera electronics, and it compensated for the spectral responses of the camera's CCD. The output of this process was a trichromatic representation of each scene giving, for each pixel, the *relative* capture ratios of the three human cone types, based on the Smith and Pokorny[14] cone fundamentals. The three planes in these images (labelled L, M and S) were then combined to obtain luminance and chromatic representations as shown in Fig.1. The L and M cone activations by red fruit and leaves obtained with this method were compared to similar output of the SR-1 in controlled conditions for a variety of red fruit and green leaves and the results differed by less than 13 % in all cases.

We obtained 124 images with a spatial resolution of 512x512 pixels by 24 bits. These were taken under different conditions of illumination (sunny and cloudy), at different distances from the objects, and with many different kinds of non-green objects among foliage (large red fruit such as apples and tomatoes; small red berries; flowers). Almost all were taken using the minimum aperture (f/11 – f/11.4), i.e. the maximum depth of field available to avoid differences in focus within the same picture. 66 images contained "red" pixels and 58 did not. We defined a "red" pixel as one where the estimated activation of the L cone was 1.5 times or more the activation of the M cone; the ratio of cone activations followed a clear bimodal distribution, with a ratio of 1.5 in the trough. 27 images were of landscapes where objects were at distances in the range 20 - 500 m. 32 images were close-ups of plants with no red objects. All of the pictures were taken in England, in various gardens. Most of the pictures were taken with the zoom lens set to its telephoto (19 – 20.4 mm focal length) setting. This avoided having

to bring the camera very close to the objects, thus affecting illumination. Linear perspective is somewhat affected by this procedure compared to natural viewing but this was found not to affect the spectral slope values in any systematic way. Square images with angular subtense of 13.6 deg were cropped from the rectangular images, and these were reduced to a size of 512 by 512 pixels. We checked for, and did not find, significant effects of chromatic aberration of the camera lens.

Measuring α

In order to avoid edge effects, all images had their borders smoothed with a gaussian roll-off (SD = 15 pixels). A two-dimensional fast-Fourier transform was used to derive the amplitude spectrum for a given luminance or chromatic spatial representation. The DC component was removed. Spectral slopes were measured by dividing the Fourier space into 9 circularly-symmetric, logarithmically-spaced, one-octave spatial-frequency bands, and then averaging the Fourier content within each of the bands. The averages were then plotted against the mid-spatial frequency of the band on log-log co-ordinates, and the slope (α) of this line calculated by linear regression. This prevented any bias due to oversampling of the high spatial frequencies. These methods are explained in greater detail elsewhere[2,15].

Calculating the optimal ratio of luminance to red-green colour spectral slopes:

We compared the csf's measured for sinusoidal spatial modulations of luminance and red-green chromatic contrast[1]. The total area under the luminance csf was normalised to unity, as was the area under the equivalent csf for chromatic modulation. We divided the spatial frequency axis into logarithmic bands (coincident with the ones used for measuring α) and calculated the areas under each of the two csf's for every band. We then calculated the area ratios of luminance-csf to colour-csf in each band. We found that a ratio of 0.76 in the spectral slopes ($\alpha_{lum}/\alpha_{chrom}$) gives approximately the same normalised area ratio.

Acknowledgements

Correspondence and request for materials should be addressed to C.A.P. (alej.parraga@bris.ac.uk).

References

1.   Mullen, K. T.: **The contrast sensitivity of human color vision to red-green and blue-yellow chromatic gratings.** *J Physiol Lond* 1985, **359**:381-400.

2.   Párraga, C. A., Brelstaff, G., Troscianko, T. & Moorhead, I. R.: **Color and luminance information in natural scenes.** *J Opt Soc Am A* 1998, **15**:563-569.

3.   Webster, M. A. & Mollon, J. D.: **Adaptation and the color statistics of natural images.** *Vis Res* 1997, **37**:3283-3298.

4.   Barlow, H. B. in *Sensory Communications* (ed. Rosenblith, W. A.) 217-234 (MIT Press, Cambridge, Mass., 1961).

5.   Marr, D. *Vision : a computational investigation into the human representation and processing of visual information* (W.H. Freeman, San Francisco, 1982).

6.   Field, D. J.: **Relations between the statistics of natural scenes and the response properties of cortical-cells.** *J Opt Soc Am A* 1987, **4**:2379-2394.

7.   Osorio, D. & Vorobyev, M.: **Colour vision as an adaptation to frugivory in primates.** *Proc Roy Soc Lond B* 1996, **263**:593-599.

8.   Regan, B. C., Julliot, C., Simmen, B., Vienot, F., Charles-Dominique, P. & Mollon, J. D.: **Frugivory and colour vision in Alouatta Seniculus, a trichromatic platyrrhine monkey.** *Vis Res* 1998, **38**:3321-3328.

9.   Sumner, P. & Mollon, J. D.: **Catarrhine photopigments are optimized for detecting targets against a foliage background.** *J Exp Biol* 2000, **203**:1963-1986.

10.   Dominy, N. J. & Lucas, P. W.: **Ecological importance of trichromatic vision to primates.** *Nature* 2001, **410**:363-365.

11.   Field, D. J.: **What the statistics of natural images tell us about visual coding?**
SPIE - Human vision, visual processing and digital displays 1989, **1077**:269-276.

12.   Tolhurst, D. J., Tadmor, Y. & Chao, T.: **Amplitude spectra of natural images.**
Ophthalm Physiol Opt 1992, **12**:229-232.

13.   Párraga, C. A., Troscianko, T. & Tolhurst, D. J.: **The human visual system is optimised for processing the spatial information in natural visual images.**
Curr Biol 2000, **10**:35-38.

14.   Smith, V. C. & Pokorny, J.: **Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm.** *Vis Res* 1975, **15**:161-171.

15.   Brelstaff, G. J., Párraga, C. A., Troscianko, T. & Carr, D.: **Hyperspectral camera system: acquisition and analysis.** *Geographic information systems, photogrammetry, and geological/geophysical remote sensing. SPIE - Human vision, visual processing and digital displays.* 1995, **2587**:150-159.
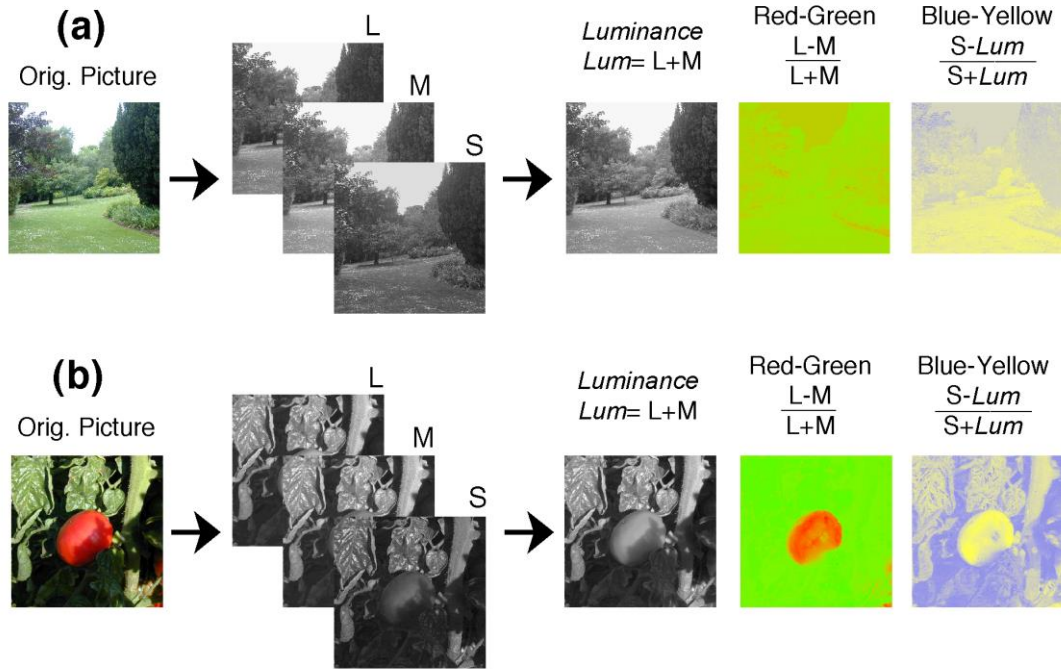
Figure 1: Decomposition of an original image (left) into L, M, S cone sampled images, and the derivation from these of the three final representations: luminance, red-green chromatic, and blue-yellow chromatic. Panel (a) shows the analysis for a picture similar to those used in previous work. Panel (b) shows the same analysis for a picture of a red tomato on a background of green leaves. Note that the luminance and chromatic representations have been scaled in order to maximise the available dynamic range. In fact, the red-green representation in case (a) actually has much less amplitude than shown. In case (b) the variability of colour and brightness within the background of leaves is removed, making the uniform red tomato pop out. Note that the divisive term in the calculation of the red-green and blue-yellow representations means that the pixel values range between –1 and 1, whilst the scaling factor for the luminance representation is essentially arbitrary.
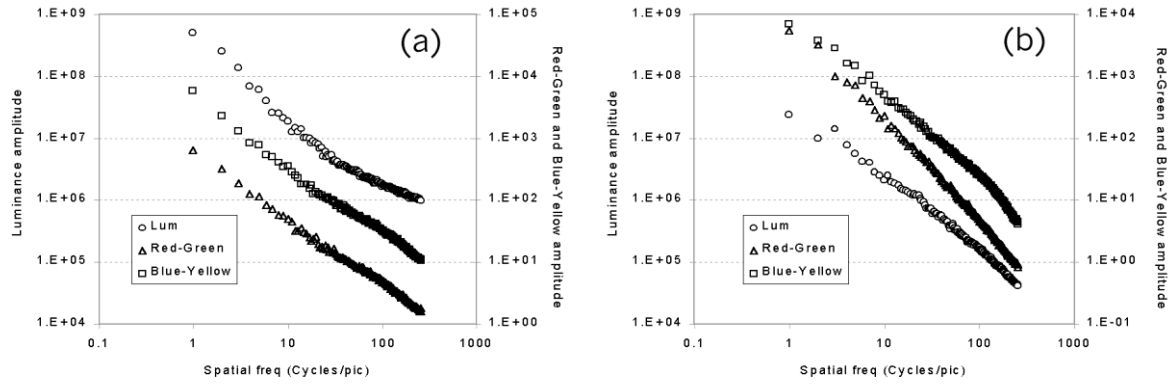
Figure 2: The graphs show examples of the luminance and chromatic spectral slopes of our natural scenes. Circles correspond to the luminance, triangles to the red-green and squares to the blue-yellow signals. The Fourier amplitude (averaged across orientation) is shown as a function of spatial frequency for the luminance and chromatic representations of the outdoor image with various kinds of foliage (Fig.1a), and the tomato image (Fig.1b). When logarithmic coordinates are used on both axes, the data fall on straight lines whose slopes are called the spectral slope (α), which usually has a value in the range -0.7 to -1.6. Different y-scales are used for the luminance data compared to the chromatic data, because of the different scaling in the pixel values of the two kinds of representation. The (a) plot shows no difference between the luminance and chromatic slopes. The (b) plot shows a steeper slope for the red-green image than the blue-yellow and luminance images.
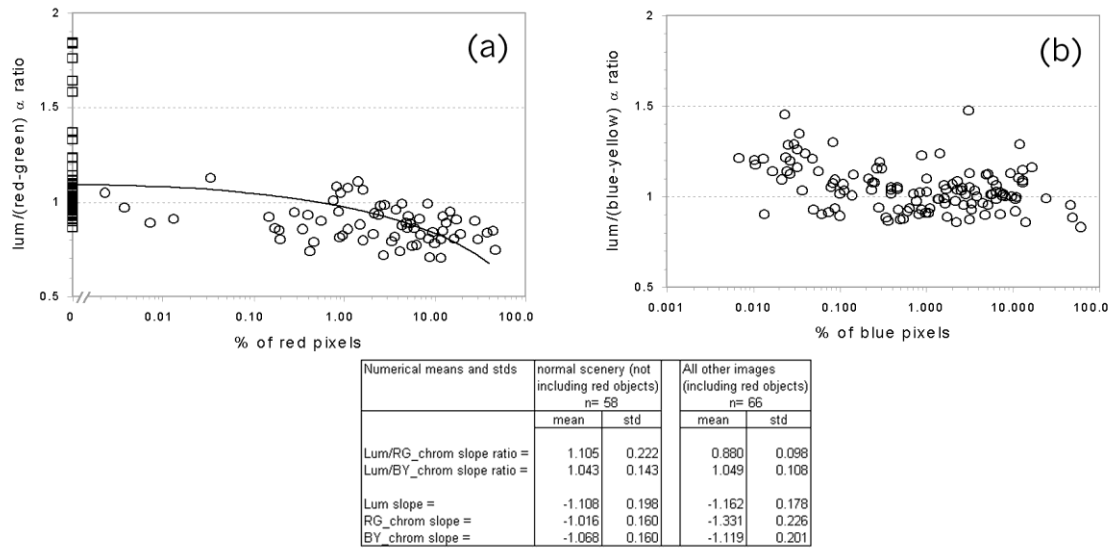
Numerical means and stds	normal scenery (not including red objects) n= 58		All other images (including red objects) n= 66	
	mean	std	mean	std
Lum/RG_chrom slope ratio =	1.105	0.222	0.880	0.098
Lum/BY_chrom slope ratio =	1.043	0.143	1.049	0.108
Lum slope =	-1.108	0.198	-1.162	0.178
RG_chrom slope =	-1.016	0.160	-1.331	0.226
BY_chrom slope =	-1.068	0.160	-1.119	0.201

Figure 3: The ratio of the spectral slopes of the luminance and chromatic representations, for different image types, as a function of the percentage of pixels signalling the presence of a "red" object. The background always consisted of green foliage. The data plotted on the left correspond to pictures of scenery without a strong presence of red objects (normal scenery—distant landscapes or close ups of foliage) and their $\alpha_{lum}/\alpha_{chrom}$ slope ratio near unity is consistent with previous findings[2]. The squares correspond to scenes with no presence of "red" pixels (i.e. they are plotted on the y-axis). The line in (a) represents the approximate data trend and reflects the fact that the average $\alpha_{lum}/\alpha_{chrom}$ slope ratio for normal scenery is about 1.1 (see Table). The Table shows the numerical means and standard deviations of the slope ratios. Note that there is a tendency for the slope ratio to fall with increasing proportion of red content for the red-green (a), but not for the blue-yellow images against the proportion of blue content (b). A high red content indicated the presence of fruit or similar object.
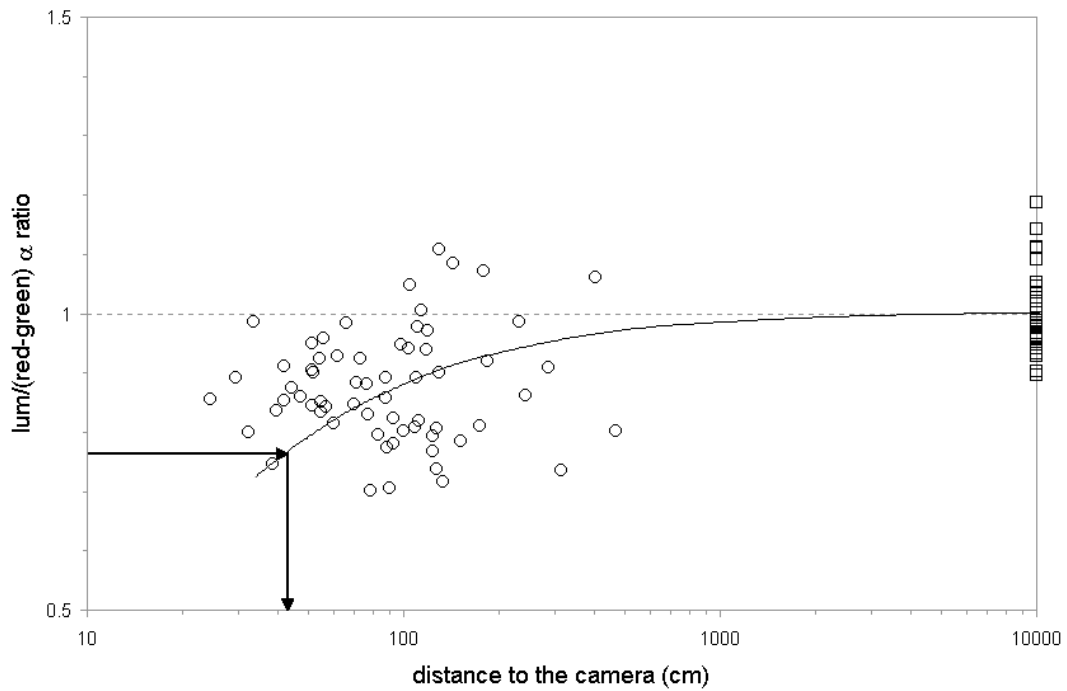
Figure 4: The relationship between the ratio of the spectral slopes (of the luminance and red-green chromatic representations) and distance from the camera to the main object in that scene, assuming a 13.6° total subtense. The squares represent "distant landscape" scenes, i.e. scenes where objects (trees, rocks, bushes, etc.) were numerous and their location was further away than 50 m. The line represents the data trend and reflects the fact that average values of $\alpha_{lum}/\alpha_{chrom}$ slope ratio are equal to 1.0 for these landscapes. The log x-axis was chosen to best show the geometrical relationship between the variables. Distances were estimated from the real size of the object (in cm), the size of the object on the final picture (in pixels) and the subtended angle of our pictures. The optimal viewing distance (where the slope ratio is around 0.78) is therefore approximately 40 cm. It could be conservatively argued that this calculation suggests that the optimal viewing distance is of the order of magnitude of typical primate grasping distance. Since we are only making order-of-magnitude claims here, the trend-line was drawn by eye.