Naturalness and Image Quality: Saturation and Lightness Variation in Color Images of Natural Scenes*

Huib de Ridder

Institute for Perception Research/IPO, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

The relation between perceived image quality and naturalness was investigated by varying the colorfulness of natural images at various lightness levels. At each lightness level, subjects assessed perceived colorfulness, naturalness, and quality as a function of average saturation by means of direct category scaling. Colorfulness was found to increase monotonically with average saturation. The relation between the quality/naturalness judgments and average saturation could always be described by an inverted U-shaped function. A systematic difference was found between quality and naturalness judgments. This difference, reflecting the subjects' preference for more colorful, but, at the same time, somewhat unnatural images, was most noticeable at the original lightness level and diminished with decreasing lightness, in particular being least at the lowest lightness level investigated.

Journal of Imaging Science and Technology 40:487–493 (1996)

Introduction

For most products in the field of color imaging one of the biggest challenges is to generate high-quality images of natural scenes. Under many circumstances, however, natural images turn out to be very critical test material. A possible explanation for this is that human observers "know" what the appearance of these images should be. Stated otherwise, human observers can incorporate the degree of naturalness of images into their quality judgments. This suggests a positive relation between perceived naturalness and image quality. But what exactly is the relation between naturalness and image quality? And, how does this relation depend on the kind of color transformation employed?

It is generally assumed that a natural image of high quality should at least be perceived as "natural." That is, such an image conforms as much as possible to the ideas and expectations the observers have about the original scene at the time the picture was taken. In general, this ideal version of an image has to be derived from the image itself, because the observers do not have the original scene at their disposal. Moreover, pictures show natural scenes that the observers have probably never seen before. Consequently, the observers must rely on earlier experiences with comparable situations, or, more generally, on their internalized knowledge of the world.[1–3] From this it follows that for natural images the impression of naturalness will reflect the degree of correspondence to memorized reality.

If one accepts this line of reasoning, it becomes clear why naturalness is generally considered a decisive constituent of the perceived quality of color images of natural scenes and why, for example, the presence of familiar objects in a scene is important in optimizing color reproduction.[1,4] The assumed relevance of naturalness is supported by experimental data from Laihanen et al.[5] showing that the impression of increased naturalness of reproduced skin color correlates positively with a quality improvement. Teunissen and Westerink[6] found a strong correlation between the perceived naturalness of colors and overall quality impression for 2- to 4-s video fragments displayed on different television sets. Recently, de Ridder et al.[7] investigated the relation between naturalness and image quality by manipulating the colors in four test scenes through a rotation of their hues around the neutral point in the CIELUV color space.[8] They found an almost perfect correlation between quality and naturalness ratings; both decreased as soon as the hues started to deviate from the ones in the original images. In the same experiment, however, a small but systematic difference between quality and naturalness judgments was observed whenever the chroma values of the colors were changed proportionately to those in the originals. This difference reflected the subjects' preference for more colorful, but, at the same time, somewhat unnatural images.

These results indicate the existence of a strong, positive relationship between perceived naturalness and image quality. However, they also suggest that this relation depends on the kind of color transformation employed. The objective of the present study is to continue the research on the naturalness–quality relation and its dependence on color transformations. To this end, an experiment is described in which the relation between naturalness and image quality was investigated by varying the colorfulness of four natural images at various lightness levels. The variation of colorfulness was created by digitizing the images, subsequently calculating their color point distributions in the CIELUV color space,[8] and finally multiplying the chroma value of each pixel by a constant. Overall lightness was manipulated by placing neutral density filters in front of the CRT monitor on which the images were displayed.

Experimental Method

Subjects. Five male and two female subjects, all students or staff members of the Institute, participated in the experiment. Their ages ranged between 21 and 33 years, with an average of 26.3 years. The subjects had normal vision or vision corrected to normal. Their color vision was checked with the H-R-R Pseudoisochromatic Plates.[9] No color deficiencies were observed.

Original mansucript received March 4, 1996.

* Presented in part at the 3rd IS&T/SID Color Imaging Conference, November 7–10, 1995, Scottsdale, Arizona.

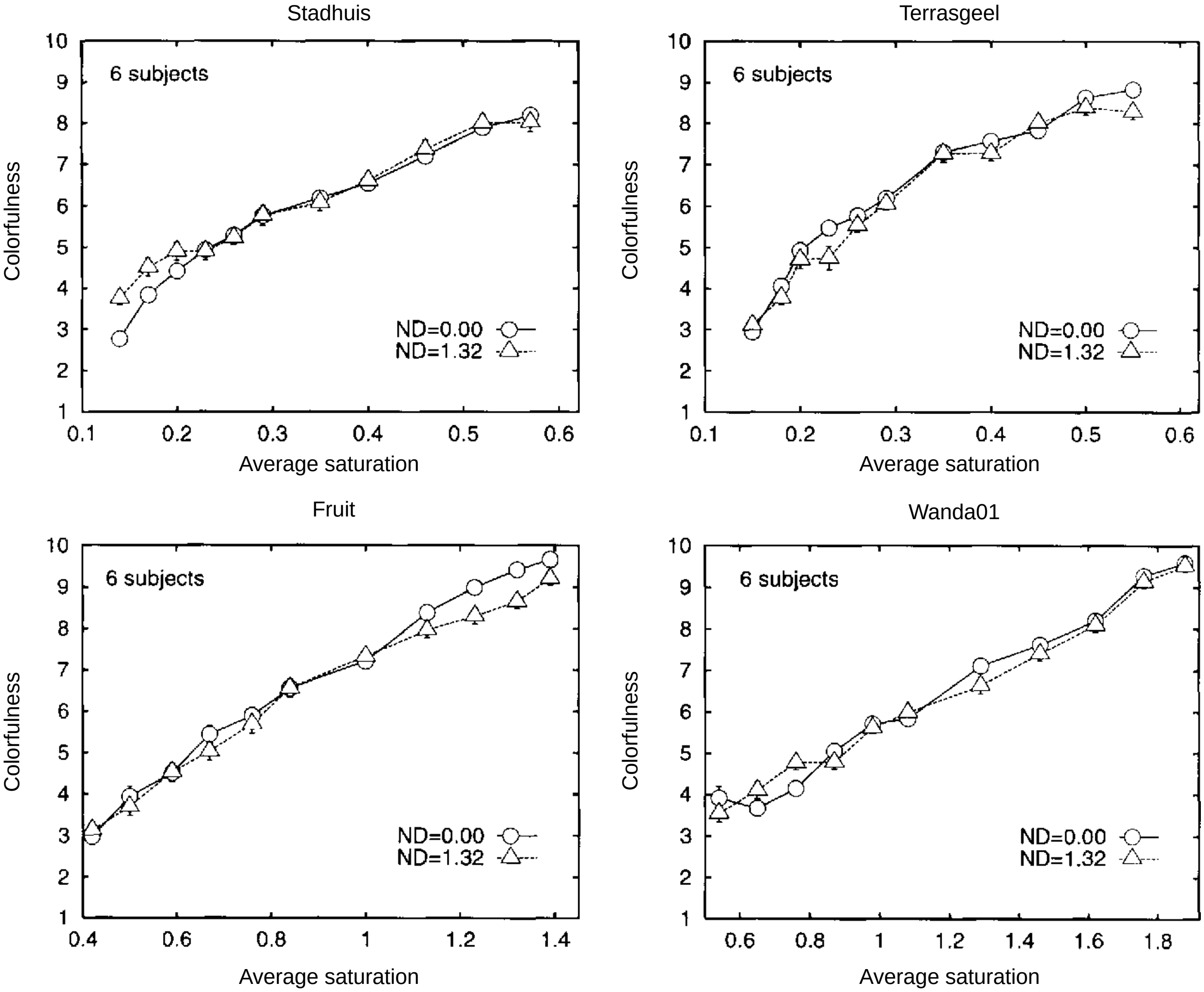

Figure 1. Colorfulness judgments versus average saturation at two lightness levels. *Circles*: original lightness level (ND = 0.00). *Triangles*: lowest lightness level investigated (ND = 1.32). Each panel presents the data for one scene. In this and the following figures, vertical bars denote twice the standard error of the mean.

Stimuli. The colorfulness of four different kinds of natural images was manipulated, using a Gould deAnza Image Processing System IPS8400. Pictures of the following four scenes were used: a portrait of a female model (WANDA01), a terrace scene with a yellow parasol in the foreground and people in the background (TERRASGEEL), fruit displayed in front of a greengrocer's shop (FRUIT), and an abstract sculpture with bushes on either side (STADHUIS). Each of the RGB signals obtained by scanning slides of these scenes was digitized with 8 bits/pixel on a grid of 512 × 512 pixels. To avoid the inclusion of the black area surrounding the original images when describing these images by their color point distributions in the CIELUV color space, only a central region of 456 × 450 pixels was used. Each color point corresponded to one pixel. Reference white was D_{65}.[8] New images were generated by multiplying the chroma value of each pixel by a constant. During the chroma transformation the lightness and hue angle of each pixel were kept constant. For each scene this resulted in five images in which chroma *decreased* (multiplication factors ranging from 0.5 to 0.9) and five images in which chroma *increased* (multiplication factors ranging from 1.2 to 2.0). If, during the processing of the images, calculated values were outside the color gamut of the monitor, the nearest possible value of chroma was used (clipping). During the experiment, the original images were also included, creating an experimental set of 44 images.

Images were displayed on a 70-Hz Barco CCID7351B monitor placed in a dark room. The monitor was corrected such that the screen luminance was linearly related to the optical density of the original slides. The maximum attainable luminance for white was 60 cd/m^2. Using this luminance as a reference,[10,11] the average lightnesses of the scenes were calculated to be 37.89 (FRUIT), 42.18 (STADHUIS), 43.54 (WANDA01), and 46.04 (TERRASGEEL). Overall lightness was decreased by placing neutral density (ND) filters (Cinemoid No. 60 Pale Gray) in front of the monitor. This way of changing lightness was chosen because varying lightness by means of image processing led to clearly visible artifacts due to quantization errors at the lower lightness levels. The transmission of the ND filters was measured to be the same for the three phosphors of the monitor. By placing one ND filter in front of the monitor, luminance decreased to 22% of its original value (ND = 0.66), resulting in a lightness level that was 46% of the original lightness level. Placing two of these filters in front of the monitor reduced the luminance to 5% of its original value (ND = 1.32), resulting in a lightness level of 17% of the original.

Procedure. The experiment consisted of eight sessions. No ND filters were placed in front of the monitor during the first three sessions, two filters were used during the following three sessions, and one filter during the final two sessions. In the course of a session, all 44 images were displayed four times in a random sequence, except that the same scene never appeared on two consecutive trials. For each trial, a single image was presented for 5 s, after which a 9 cd/m^2 adaptation field appeared on the screen. The luminance of this adaptation field was also attenuated by the ND filters. The subjects viewed the monitor at a distance of about 1.7 m. The pixel size is about 1 min of arc at this distance. In each session, the subjects rated either the perceptual quality, the naturalness, or the colorfulness of all images on a 10-point numerical category scale ranging from 1 (lowest quality/naturalness/colorfulness) to 10 (highest quality/naturalness/colorfulness). The instructions given to the subjects defined perceptual image quality as "degree of excellence of the image," naturalness as "degree of correspondence between the reproduced image and reality (that is, the original scene as it is according to the viewer)," and colorfulness as "presence and vividness of colors in the whole picture." Four subjects carried out the whole experiment. In addition, two subjects took part in the first six sessions and one subject in the last two sessions. During the first three sessions, i.e., the condition with no ND filters, three subjects first rated quality, then colorfulness, and finally naturalness. The order of rating was reversed for the other three subjects. In the next three sessions, i.e., the condition with two ND filters, three subjects first rated quality, then naturalness, and finally colorfulness. The other three subjects assessed quality and naturalness in the reverse order. In the final two sessions, i.e., the condition with one ND filter, quality was assessed before naturalness by three out of the five subjects. Before starting a session, subjects always judged a training series of 12 images.

Chroma and Saturation. Fedorovskaya et al.[12] have shown that the average of the chroma distribution is a convenient correlate of colorfulness within CIELUV for the kind of chroma transformation employed in the present study. Chroma, however, is proportional to lightness and thus depends on the choice of the luminance of the reference white. For self-luminous displays it is customary to use the maximum attainable luminance as a reference.[10,11] It can be argued that in the present study this will lead to a separate reference for each experimental condition, namely, references with luminances of 60, 13, and 3 cd/m^2 for the conditions with zero, one and two ND filters, respectively. To enable a direct comparison of these experimental conditions within one color space, we decided to use one reference with a luminance of 60 cd/m^2 and to replace chroma by saturation. The latter makes sense because *CIE 1976 u,v saturation,* being chroma divided by lightness, is invariant with changes in luminance[11] and thus, unlike chroma, does not scale with luminance.

Results and Discussion

For each image attribute (colorfulness, quality, naturalness) and lightness level, the presence of possible individual differences was checked by correlating the average judgments of each subject with those of all other subjects. Cluster analyses of the resulting intersubject correlation matrices never yielded clearly separated groups of subjects. This finding suggests that the order of judging quality and naturalness did not affect the judgments. In general, no systematic differences were observed between the subjects. Accordingly, the naturalness–quality relation was investigated on the basis of the judgments averaged across the subjects.

Colorfulness. Figure 1 shows the colorfulness judgments, averaged across the subjects, as a function of average saturation for the two extreme lightness levels. We can see that colorfulness always increased monotonically with average saturation. This result implies that the subjects could perceive colors in the whole lightness range investigated. The results at the original lightness level (ND = 0.00) agree with those of an earlier study carried out by Fedorovskaya et al.[12] For one scene (WANDA01) the relationship between the colorfulness judgments and saturation is linear, but for the other scenes this relationship is slightly compressive. Per scene, the colorfulness–saturation relationship appears to be independent of lightness level. Apparently, the subjects compensated for the perceived differences in overall brightness. This phenomenon supports the above-mentioned choice of replacing chroma by saturation.

Image Quality. Fedorovskaya et al.[12] and de Ridder et al.[7,13] have already shown that at the original lightness level the perceived quality of the transformed images is nonmonotonically related to the average chroma and that this relation can be described by an inverted U-shaped function. Figure 2 reveals that their results are confirmed in the present study and generalized to the other two lightness levels. Figure 2 also shows that, per scene, the functions at the different lightness levels are almost identical when the quality judgments are plotted versus saturation. To facilitate this comparison, the judgments at the two lowest lightness levels were linearly transformed such that they had the same range and overall mean as those at the original lightness level. This procedure had no substantial effect on the quality judgments except in the case of the STADHUIS scene, because for that scene the range systematically decreased with decreasing lightness level. The overall mean always varied around the middle of the category scale, i.e., between 5 and 6.

In each scene the optimum quality seems to occur at the same saturation level, independent of lightness. An exception is TERRASGEEL, because for this scene the quality function has the tendency to shift to higher saturation values when lightness decreases. To quantify these trends, second- to fourth-order polynomials were fitted to the quality ratings ($r^2 > 0.95$). The saturation values at which these polynomials reached their maxima were (in order of decreasing lightness): 0.43, 0.41, 0.44 (STADHUIS); 0.30, 0.32, 0.36 (TERRASGEEL); 0.96, 0.97, 0.97 (FRUIT); 1.25, 1,25, 123 (WANDA01). These values are systematically higher than the saturation levels of the original images (STADHUIS: 0.29; TERRASGEEL: 0.29; FRUIT: 0.84; WANDA01: 1.08). The subjects apparently preferred more colorful images to the original ones.

Naturalness. Figure 3 demonstrates that, in essence, the same results were obtained for the naturalness judgments as for the quality judgments. The above-mentioned linear transformation was also applied to the naturalness judgments at the two lowest lightness levels. Again, no substantial changes were observed, with the exception that for STADHUIS the range of naturalness judgments had to be changed because it systematically decreased with decreasing lightness. Polynomials were fitted to the naturalness judgments in order to determine the saturation levels at which the subjects judged the images to be most natural. The resulting values were (in order of decreasing lightness): 0.37, 0.36, 0.41 (STADHUIS); 0.28, 0.31, 0.34 (TER.-RASGEEL); 0.85, 0.87, 0.92 (FRUIT); 1.17, 1.18, 1.23

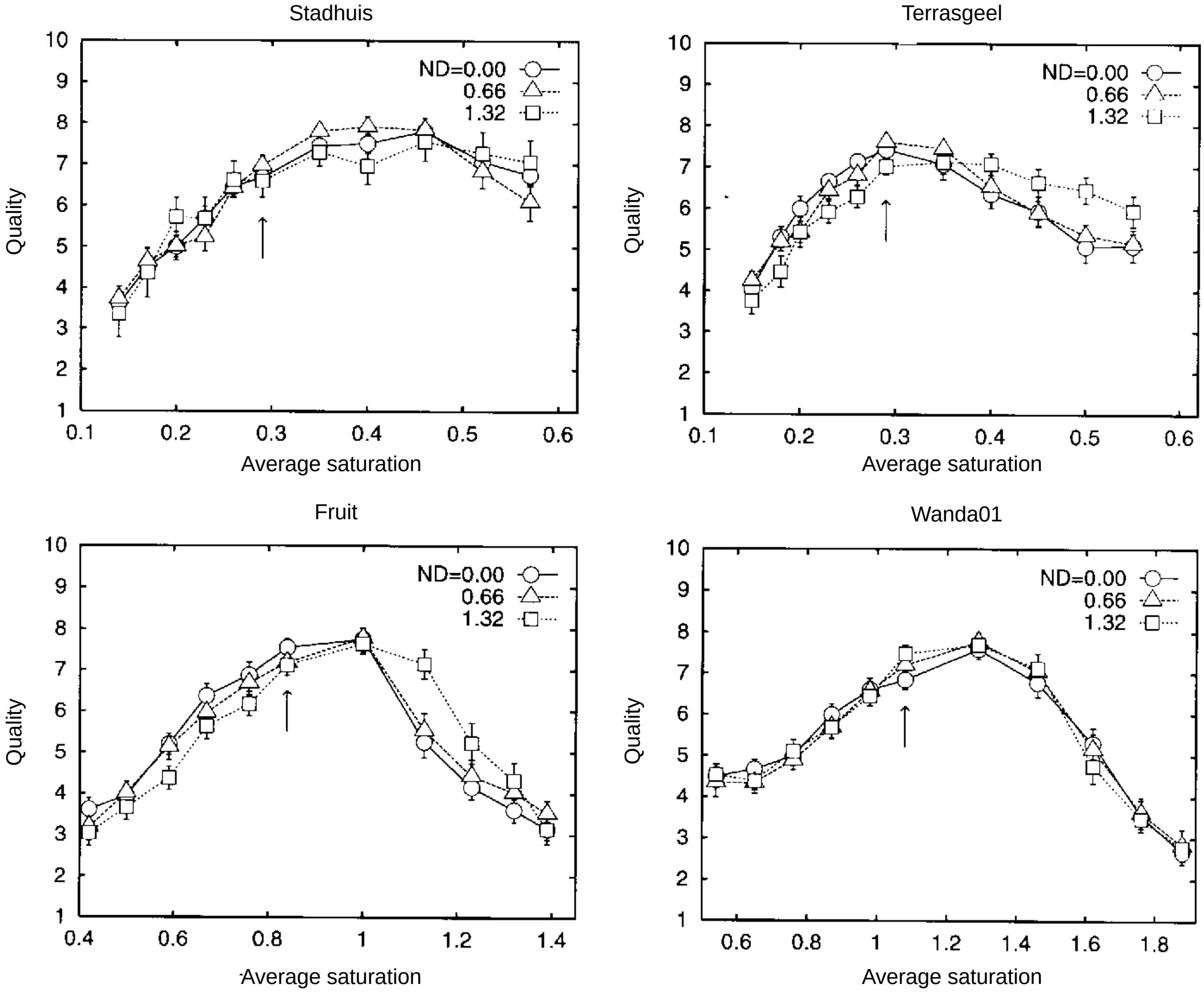

Figure 2. Quality judgments versus average saturation at three lightness levels (ND = 0.00, 0.66, 1.32). Each panel presents the data for one scene. In this and the following figures, arrows point to the judgments of the images at the original saturation level.

(WANDA01). In contrast with what was observed for the quality judgments, the saturation levels at which the images were judged to look most natural systematically increased with decreasing lightness, although the main increase occurred between the two lowest lightness levels.

Naturalness and Image Quality. Figure 4 presents the naturalness ratings as a function of scaled image quality for each lightness level separately. A small but systematic deviation from linearity can be seen for all scenes and lightness levels but is, in general, most noticeable at the original lightness level and least noticeable at the lowest lightness level. Comparable results were obtained in previous studies employing chroma variation at the original lightness level only.[7,12,13] In one of these studies,[12,13] the observed deviation was interpreted as follows. With increasing saturation, both naturalness and quality increase linearly as a function of colorfulness. Above a certain saturation value, however, naturalness starts to decrease while quality remains relatively high. An alternative interpretation is that the inverted U-shaped function for the quality judgments has been shifted to higher average saturation values relative to the function for the naturalness judgments. This interpretation is consistent with the finding by de Ridder et al.[7] that quality and naturalness judgments can be fitted by the same second-order polynomial after the saturation values belonging to the naturalness function have been multiplied by a constant varying between 1.06 and 1.15.

A consequence of the observed nonlinearity is that the saturation values at which naturalness and quality are estimated to be optimal differ. A comparison of these values shows that, in general, the qualitatively optimal images are more colorful than the images considered to be the most natural ones (Fig. 5). Apparently, the subjects used different criteria to assess naturalness and image quality; the subjective preference in quality was biased toward more colorful images, although the subjects realized that these images looked somewhat unnatural. This difference between quality and naturalness judgments was most noticeable at the original lightness level and diminished with decreasing lightness, especially at the lowest lightness level. This difference is mainly caused by the naturalness function shifting toward higher saturation levels with decreasing lightness, whereas the quality function is hardly affected by such lightness changes.

The difference between naturalness and quality appears to be rather robust; it was found for all scenes and was

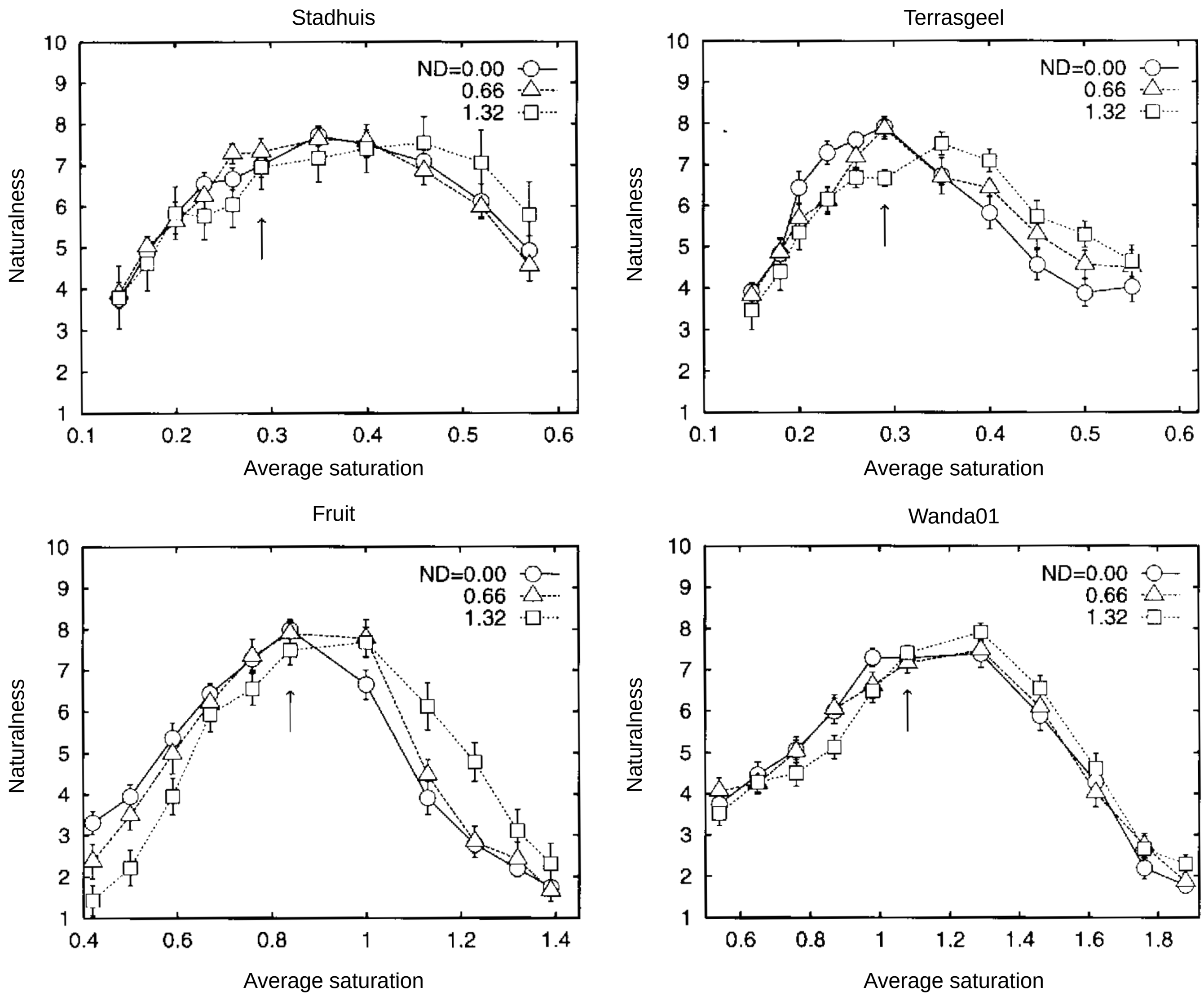

Figure 3. Naturalness judgments versus average saturation at three lightness levels (ND = 0.00, 0.66, 1.32). Each panel presents the data for one scene.

only slightly influenced by the level of overall lightness (Figs. 4 and 5). Furthermore, it is insensitive to whether the images are viewed in a completely darkened room, as in the present study, or on a monitor placed in front of an illuminated white background.[7] Finally, it seems independent of the absolute value of the preferred saturation level. As an illustration, Fig. 6 presents the average judgments of one subject who preferred significantly less colorful images of the STADHUIS scene than did the other subjects. Yet, the same nonlinearity between quality and naturalness can be seen as in Fig. 4.

The results of the present study leave open the question as to why there is a discrepancy between quality and naturalness. The preference for more colorful images hints at a need of higher color contrast. This may be advantageous for improving object recognition.[2] Another possibility is the influence preferences and emotions may have on quality judgments.[14,15] At the moment, explanations like these are mere speculations. It is evident that further research is required before the experimentally established difference between naturalness and quality can be better understood.

Conclusions

In this study, the relationship between perceptual image quality and naturalness was investigated by varying the average saturation of color images of natural scenes at various lightness levels. Both naturalness and image quality were found to be nonmonotonically related to average saturation over a large range of lightness levels. The relation between the quality naturalness judgments and average saturation could always be described by an inverted U-shaped function. At the original lightness level, the inverted U-shaped function for the quality judgments was always found to be shifted to higher saturation values relative to the function for the naturalness judgments. This result suggests that subjects tend to prefer more colorful images, although they realize that these images look somewhat unnatural. This difference between naturalness and quality diminished with decreasing lightness. This is mainly caused by the naturalness function's shifting toward higher saturation levels with decreasing lightness, whereas the quality function is hardly affected by such lightness changes.

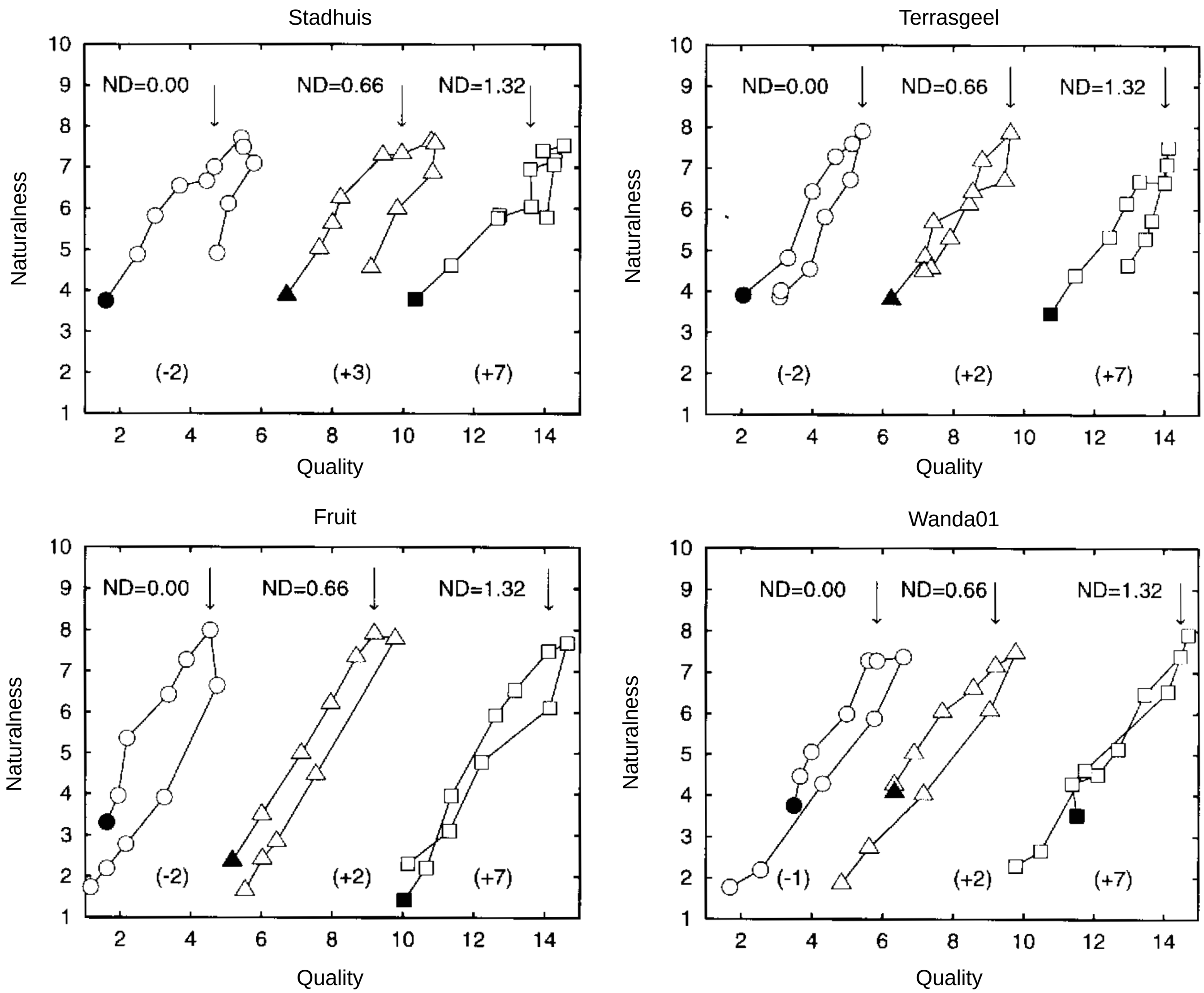

Figure 4. Naturalness estimates versus quality judgments per lightness level (ND = 0.00, 0.66, 1.32) and scene. The filled symbols denote the images with the lowest average saturation. Solid lines link images at neighboring saturation levels. The numbers in parentheses indicate how much the naturalness–quality functions have been shifted along the quality axis.

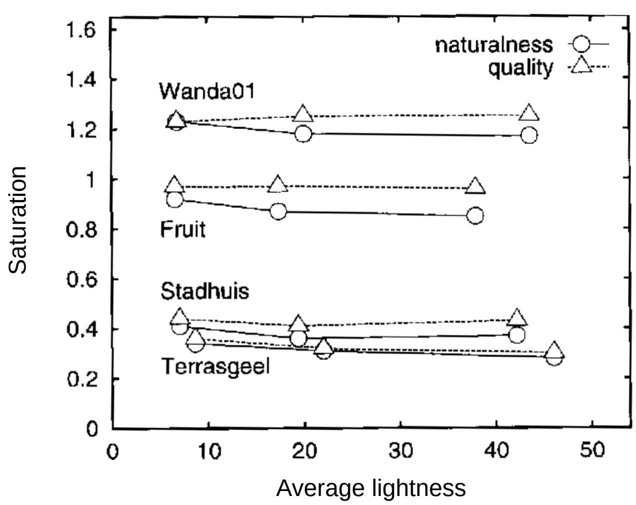

Figure 5. Saturation levels at which polynomials fitted to the quality and naturalness judgments shown in Figs. 2 and 3, respectively, are at their maximum values. Note that, in general, the qualitatively optimal images are more colorful than the images considered to be the most natural ones.

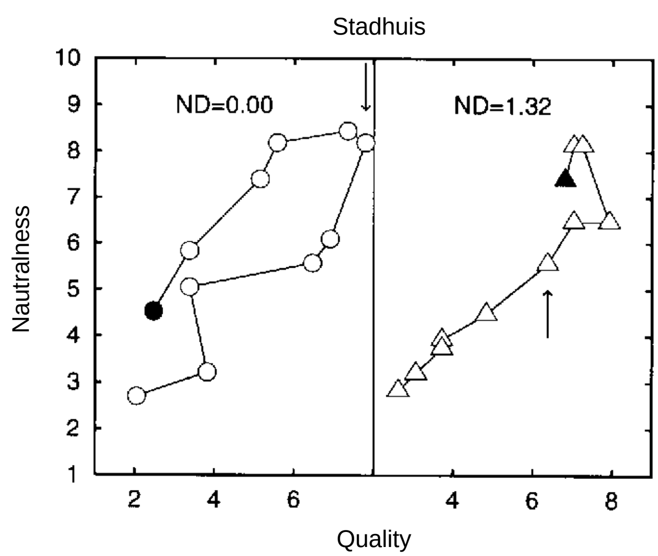

Figure 6. Naturalness judgments versus scaled quality at two lightness levels. Subject: IK. For further details, see Fig. 4.

Acknowledgment. The author wishes to thank Els van 't Hullenaar for all her contributions to the research described in this paper.

References

1. R. W. G. Hunt, Visual appreciation, in *The Reproduction of Colour in Photography, Printing & Television,* Fountain Press, Tolworth, UK, 1987, pp. 36–48.
2. M. A. Arbib and A. R. Hanson, Vision, brain, and cooperative computation: an overview, in *Vision, Brain, and Cooperative Computation,* M. A. Arbib, A. R. Hanson, Eds., The MIT Press, Cambridge, MA, 1987, pp. 1–83.
3. S. R. Ellis, Virtual environments and environmental instruments, in *Simulated and Virtual Realities: Elements of Perception,* K. Carr, R. England, Eds., Taylor & Francis, London, 1995, pp. 11–51.
4. Y. Miyake and H. Haneishi, The evolution of nonimpact print—impact on image quality, *Proceedings of the First IS&T/SID Color Imaging Conference: Transforms & Transportability of Color,* Scottsdale, AZ, 1993 pp. 222–227.
5. P. Laihanen, M. Lindholm, S. Rouhiainen, H. Saarelma, and L. Tuuteri, Automatic colour correction, *Proceedings of the Second IS&T/SID Color Imaging Conference: Color Science, Systems and Applications*, Scottsdale, AZ, 1994, pp. 97–102.
6. K. Teunissen and J. H. D. M. Westerink, A multidimensional evaluation of the perceptual quality of television sets, *SMPTE J.* **105**, pp. 31–38 (1996).
7. H. de Ridder, F. J. J. Blommaert, and E. A. Fedorovskaya, Naturalness and image quality: chroma and hue variation in color images of natural scenes, in *Human Vision, Visual Processing, and Digital Display VI,* B. E. Rogowitz, J. P. Allebach, Eds., *Proc. SPIE* **2411**: 51–61 (1995).
8. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae,* John Wiley & Sons, New York, 1982.
9. L. H. Hardy, G. Rand and M. C. Ritter, *H-R-R Pseudoisochromatic Plates for Detecting, Classifying, and Estimating the Degree of Defective Color Vision,* American Optical Company, New York, 1957.
10. R. C. Carter and E. C. Carter, CIE L*u*v* color-difference equations for self-luminous displays, *COLOR Res. and Appl.* **3**: 252–253 (1983).
11. J. Walraven, Color basics for the display designer, in *Color in Electronic Displays,* H. Widdell, D. L. Post, J. D. Grossman, J. Walraven, Eds., Plenum Press, New York, 1992, pp. 3–38.
12. E. A. Fedorovskaya, F. J. J. Blommaert, and H. de Ridder, Perceptual quality of color images of natural scenes transformed in CIELUV color space, *Proceedings of the First IS&T/SID Color Imaging Conference: Transforms & Transportability of Color,* Scottsdale, AZ, 1993, 37–41.
13. H. de Ridder, E. A. Fedorovskaya, and F. J. J. Blommaert, Naturalness and image quality: chroma variation in colour images of natural scenes, *IPO Annual Progress Report* **28**: 89–95 (1993).
14. P. Valdez and A. Mehrabian, Effects of color on emotions, *J. Experimental Psychol. General* **123**: 394–409 (1994).
15. M. P. Hollier and G. Cosier, Assessing human perception, *BT Technol. J.* **14***,* 206–214 (1996).