Image quality semantics

JOURNAL OF IMAGING SCIENCE AND TECHNOLOGY • Volume 41, Number 5, September/October 1997

Image Quality Semantics

T. J. W. M. Janssen and F. J. J. Blommaert
IPO, Center for Research on User-System Interaction, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

In this contribution we discuss image quality in the context of the visuo-cognitive system as an information-processing system. To this end, we subdivide the information-processing as performed by the visuo-cognitive system into three distinct processes: (1) the construction of a visual representation of the image, (2) the interpretation of this representation by means of a confrontation with memory, and (3) task-directed semantic processing of the interpreted scene to formulate a proper response.

A successful completion of these processes can be ensured only when two main requirements are satisfied: (1) the visual representation of the image should be sufficiently precise, and (2) the degree of correspondence between the visual representation and "knowledge of reality" as stored in memory should be high.

We then relate these requirements to the attributes "usefulness" and "naturalness" of the image and give a functional description of image quality in terms of naturalness and usefulness. To conclude, experimental results supporting this description of image quality will be discussed.

Journal of Imaging Science and Technology 41: 555–560 (1997)

Introduction

A major part of research activity in the field of image quality is directed toward the development of reliable, widely applicable, instrumental image quality measures. There are two important reasons for developing instrumental measures: (1) quality evaluation by means of subjective assessment tests is quite expensive and time consuming, and (2) a posteriori assessing of the image quality of a given design does not allow for an a priori optimization of this design, thus condemning the design of image (re-)production systems to remain an iterative procedure.

At present, much of the research concerning instrumental image quality measures is based on an approach that can be characterized as a "signal evaluation" approach. In this approach, the image is regarded as a complex signal that deviates more or less from the complex signal that represents the "ideal" or "original" image. Images are defined in the physical or perceptual domain—in the latter case using models of the earliest stages of visual perception—and quality measures are defined as distances in an appropriate function space, e.g., Euclidian distance between actual and original image.

In contrast to this approach, we regard the processing of images by the visuo-cognitive system not as the evaluation of complex signals but instead as the *processing of visual information*. Realizing that this information processing is an essential part of an observer's interaction with his environment, we characterize the quality of an image in a more meaningful manner as the degree to which the image can be successfully exploited by the observer. We therefore consider the visuo-cognitive system (1) as an information-processing system[1–4], and (2) as an integral part of an observer's interaction with his environment.[5–8]

Original manuscript received December 27th, 1996.

e-mail: jansf@ipo.tue.nl

Quality and Information Processing

Understanding Information-Processing Systems. We adopt the general viewpoint in computational cognition[9,10] that (1) the visuo-cognitive system can be considered to be an information-processing system, and (2) information-processing systems can be completely understood only when they are understood at three distinct levels. These levels are (1) the semantic level, i.e., the level describing the system in terms of computational goals and strategies, (2) the algorithmic level, i.e., the level describing the implementation of the computational theory into algorithms and associated representations, and (3) the level of physical implementation, i.e., the level describing the physical implementation of these algorithms and representations.

As stated above, any information-processing system can be completely understood only when it can be appropriately described at *all three* levels, which for reasons of simplicity may be designated as the levels of "what and why," "how," and "where." The approach we choose to pursue here is strictly top-down, i.e., we first try to gain a fundamental understanding of the "what and why" of the processing of images by the visuo-cognitive system to arrive at an understanding of what image quality is, and then proceed with the "how" and "where." Our present purposes therefore are served best with a description of visuo-cognitive information-processing at the semantic level.

Quality of Information. At a semantic level, the interaction of an observer with his environment can be described by a cycle consisting of three activities: (1) perception, i.e., acquiring information from the environment and constructing an internal representation from it, (2) cognition, i.e., interpreting the obtained internal representation, and (3) action, i.e., responding appropriately to this interpretation.

At this point, we may already infer that to ensure a proper response to occurrences in the outside world, certain requirements must be imposed upon the information

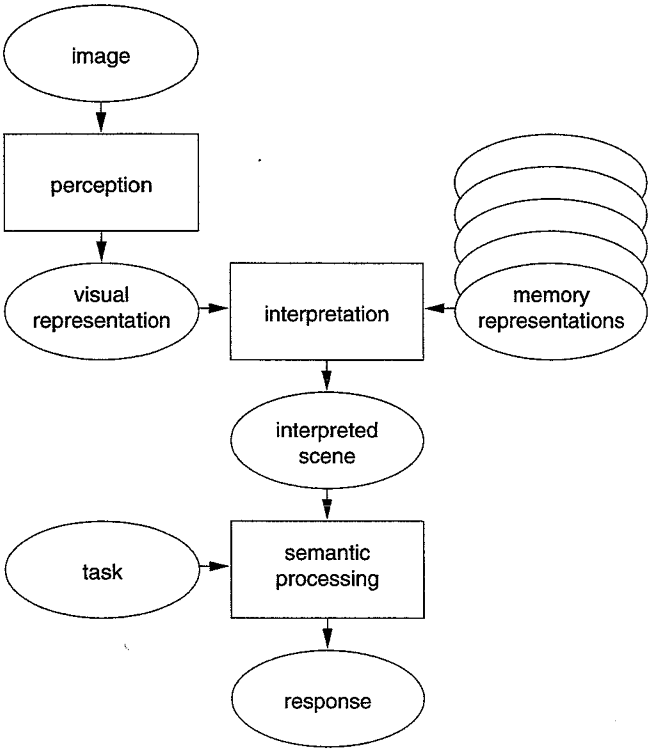

Figure 1. A diagrammatical depiction of visuo-cognitive processing of images. In this diagram, ellipses denote representations of information, and rectangles denote processes transforming one representation into another. Note that we consider the "response" to be a formulated sequence of actions, *not* its manifestation in terms of motoric events.

that is acquired from the environment. When the quality of information is considered to be the degree to which these requirements are satisfied, we arrive at two important conclusions: (1) the quality of information can be defined only within the context of an observer interacting with his environment, and (2) the quality of information refers to the appropriateness of this information as a basis for a proper response to outside world occurrences.

Quality of Images. Realizing that images are the medium for visual information, we now focus our attention on the question of what requirements should be imposed upon an image. At first thought, the requirements that a "good" image should satisfy would seem to be precision and reliability. The above outlined ideas, however, lead to a somewhat more complicated requirement: for an image to be of "good" quality, the observer's *interpretation* of this image should be successful, i.e., should with high probability be correct. The imposition of this requirement is justified by realizing that *it is of vital importance that there is no discrepancy between "what really is there" and what the observer assumes is there*. The next section will focus on the question of how this optimally can be secured.

Image Quality Semantics

Image Processing by the Visuo-Cognitive System. We subdivide the processing of images by the visuo-cognitive system in three distinct processes (see, e.g., Refs. 1 and 2): (1) perception, i.e., the construction of a visual representation of the image using primarily low-level knowledge of the visual world, (2) interpretation, i.e., the confrontation ("matching") of this visual representation with memory representations, and (3) task-directed semantic processing of the interpreted scene to formulate a response. A diagrammatical depiction of the processing of images by the visuo-cognitive system is shown in Fig. 1.

The collection of memory representations of the outside world, which from this point onward we refer to as "knowledge of reality," we regard as well defined but nevertheless fuzzy. As an example of this, consider that although observers "know" what the prototypical characteristics of a certain object are, they usually are unable to define a clear distinction between prototypical and nonprototypical object characteristics. Knowledge of reality therefore may be thought of as accumulated knowledge of (the behavior of) outside world statistics. Instead of referring to a match between the visual representation and a memory representation, henceforth we more appropriately refer to a match between the visual representation and knowledge of reality.

Naturalness, Usefulness, and Quality. Given the visuo-cognitive processes outlined above, we now ask how a successful interpretation of an image best can be secured. Returning to Fig. 1, we readily may conclude that for a successful interpretation of an image, the interpretation process should result in a satisfactory match between the visual representation and knowledge of reality. We therefore quite directly arrive at two principal requirements that an image of "good" quality should satisfy: (1) the visual representation of the image should be sufficiently precise, and (2) the degree of correspondence between the visual representation and knowledge of reality as stored in memory should be high.

We now are able to formalize the preceding discussion by defining (1) the *usefulness* of an image to be the precision of the visual representation of the image, and (2) the *naturalness* of an image to be the degree of correspondence between the visual representation of the image and knowledge of reality as stored in memory. Using these definitions, we define the quality of an image to be the degree to which the image is both useful and natural.

The sets of requirements that one needs to impose on an image to maximize its usefulness or its naturalness in general will not coincide; e.g., detection or discrimination of objects in an image may require "exaggeration" of certain features of this image, resulting in a less natural reproduction of the image. Therefore, given the above definition of image quality, we postulate that the quality of an image will be given by a compromise resulting from simultaneously evaluating to what degree the image satisfies the sets of requirements that lead to maximizing the usefulness or the naturalness of the image.

Experiments

Experiment 1: Influences of Naturalness and Usefulness on Image Quality. ***Aim.*** To test the semantic description of image quality as given above, we performed an experiment allowing us to measure the influences of naturalness and usefulness on image quality. To this end, we selected two kinds of manipulation, i.e., varying the color temperature of the reference white* and varying chroma,[†,11]

* Defined as "the temperature of a Planckian radiator whose radiation has the same chromaticity as that of a given stimulus." Unit: kelvin (K).

† Defined as "the colorfulness of an area judged as a proportion of the brightness of a similarly illuminated area that appears to be white or highly transmitting." In the experiments C^*_{uv}, a correlate of chroma in the CIELUV color space, was scaled.

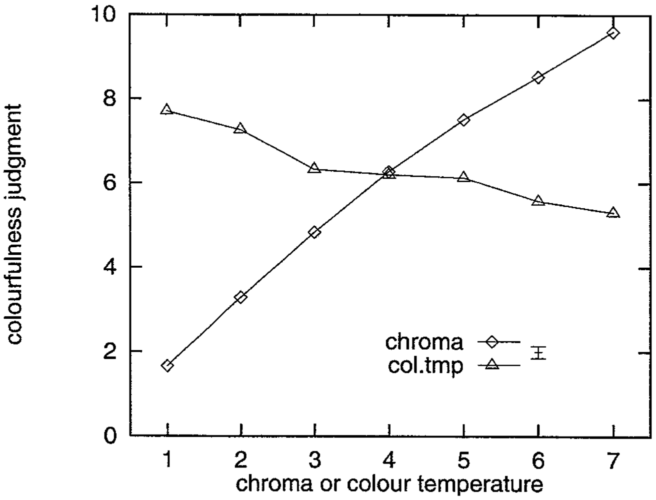

Figure 2. Colorfulness judgments (averaged over subjects and scenes) for the conditions chroma (diamonds) and color temperature of the reference white (triangles). The error bar denotes a distance of two average standard errors in the mean. The numbers 1–7 on the horizontal axis denote, for chroma, scaling by, respectively, 0.50, 0.63, 0.79, 1.00, 1.26, 1.59, and 2.00, and for the reference white a color temperature of, respectively, 4650K, 5150K, 5800K, 6500K, 7400K, 8650K, and 10,300 K.

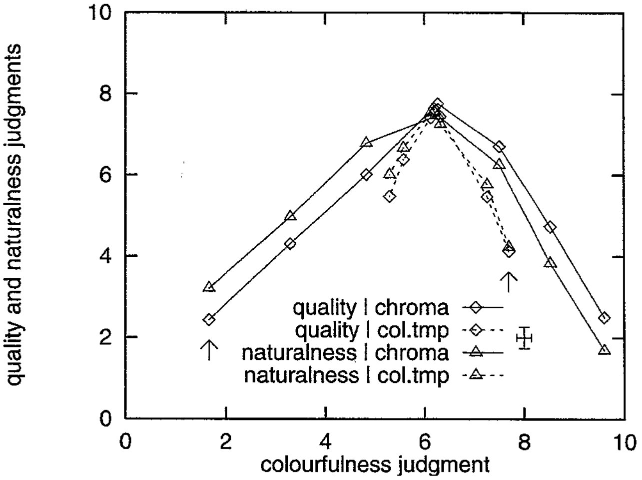

Figure 3. Quality judgments (diamonds) and naturalness judgments (triangles) versus colorfulness judgments (all averaged over subjects and scenes) for the conditions chroma (solid lines) and color temperature of the reference white (dashed curves). The error cross denotes a distance of two average standard errors in the mean, and the arrows denote the images with lowest chroma and the lowest color temperature of the reference white.

which we expected to affect naturalness and usefulness in distinctly different ways. The first manipulation was expected to influence only the naturalness of the image, whereas the second manipulation was expected to influence both the naturalness and the usefulness of the image.

Description. The experiment, similar to experiments described in Refs 12, 13, and 14, was performed using four color images of natural scenes taken from a Kodak Photo CD test disk. The color temperature of the reference white was varied between 4650 K and 10,300 K, 6500 K being the original, in seven steps of perceptually equal size, and chroma was scaled by a constant ranging from 0.5 to 2.0 in seven steps.

Seven subjects participated in the experiment. In three separate sessions they were shown on a CRT the complete set of images, in random order, with three replications. In the first session, the subjects' task was to judge the quality[‡] of the images, in the second session to judge the colorfulness[§] of the images, and in the third session to judge the naturalness[¶] of the images. Subjects were instructed to use an 11-point numerical scale ranging from 0 ("bad" or "weak") to 10 ("excellent" or "strong").

Results. Figure 2 shows colorfulness judgments (averaged over subjects and scenes) versus chroma (diamonds) and color temperature of the reference white (triangles). The effect of scaling chroma on judged colorfulness is clearly visible. Less clear, although still significant, is the influence of color temperature of the reference white on judged colorfulness; images are judged less colorful for higher temperatures of the reference white.

Figure 3 shows quality judgments (diamonds) and naturalness judgments (triangles) versus colorfulness judgments, all averaged over subjects and scenes, for the conditions chroma (solid curves) and color temperature of the reference white (dashed curves). The figure shows that, as expected, quality correlates well with naturalness, although for the condition chroma the curve for quality is shifted with respect to the curve for naturalness toward higher values of colorfulness (and hence toward higher values of chroma). Note that a similar shift does *not* occur for the condition color temperature of the reference white. This seems to dismiss the possibility of a straightforward "preference" for images with higher colorfulness.

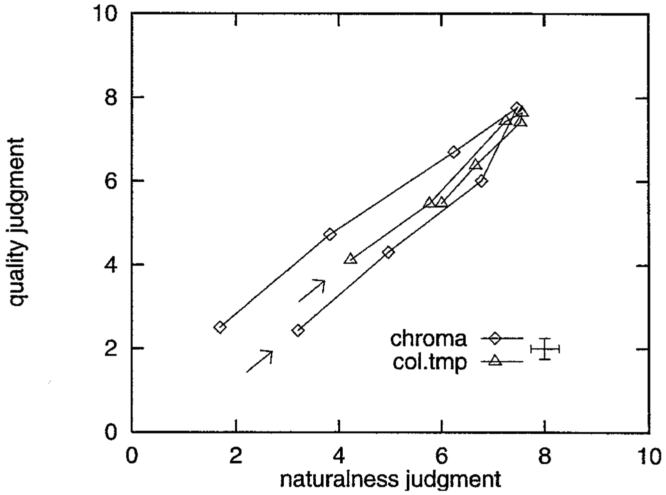

Figure 4. Quality judgments (averaged over subjects and scenes) versus naturalness judgments (also averaged) for the conditions chroma (diamonds) and color temperature of the reference white (triangles). The error cross denotes a distance of two average standard errors in the mean, and the arrows denote the images with lowest chroma and the lowest color temperature of the reference white.

Quality judgments versus naturalness judgments, depicted in Fig. 4, again show that the quality–naturalness

‡ Defined as "the degree to which you like the colors in the image."

§ Defined as "the presence and vividness of the colors in the image."

¶ Defined as "the degree to which the colors in the image seem realistic to you."

curve for the condition chroma deviates significantly from a linear relation. The curve has a U shape resulting from the abovementioned shift between quality judgments and naturalness judgments.

Interpretation. The first-order effect, i.e., the high correlation ($r = 0.93$) between naturalness judgments and quality judgments for both conditions, can readily be interpreted in terms of the outlined semantic theory of image quality. To interpret the second-order effect, i.e., the shift between naturalness judgments and quality judgments for the condition chroma, we adopt the CIELUV color space (recommended in 1979 by the Commission Internationale de l' Éclairage) as an appropriate, perceptually uniform color space.

In CIELUV, the image can be thought of as a "cloud of dots," each dot corresponding to one pixel in the image. Scaling chroma can be described as a radial contraction or expansion of the cloud of dots toward or away from the reference white, while changing the color temperature of the reference white can be described by a displacement of the entire cloud along the yellow–blue direction. The relevant difference between the two manipulations follows quite directly from their descriptions in CIELUV: changing chroma results in increased or decreased distances in color space between any pair of dots of which the members do not represent exactly the same color, while changing the color temperature of the reference white has no effect on these distances.

Contrary to manipulations that preserve distances in color space, manipulations that do affect distances in color space also will affect the precision with which the image can be represented internally (since at a presumed, constant level of internal noise, affecting distances in a perceptually uniform space is equivalent to affecting an internal "signal-to-noise ratio"), and hence the usefulness of the image. The above discussion therefore can be concluded as follows: manipulations that do not affect the usefulness of an image, e.g., changing the color temperature of the reference white, will have approximately identical parameter settings for optimizing the naturalness and the quality of the image. Manipulations that do affect the usefulness of an image, e.g., changing chroma, will have different parameter settings for optimizing the naturalness or the quality of an image. In the latter situation, the parameter settings optimizing quality will tend to deviate with respect to those optimizing naturalness toward values that increase the usefulness of the image.

Experiment 2: Image Quality Regarded as a Compromise Between Naturalness and Usefulness. ***Aim.*** To test the description of image quality in terms of a compromise between naturalness and usefulness, we devised an experiment in which we manipulated the brightness contrast of black-and-white images of natural scenes. Assuming that (1) usefulness is linearly related to perceived brightness contrast and (2) the compromise can be adequately described by a linear combination of naturalness and usefulness, we may write image quality (Q) in terms of naturalness (N) and brightness contrast (C) as

$$Q = \lambda_1 N + \lambda_2 C + \lambda_3, \tag{1}$$

and fit the vector $\vec{\lambda}$ to subjects' judgments of quality, naturalness, and contrast as obtained in the experiment.

Description. The experiment was performed using four black-and-white images of natural scenes, obtained by transforming images from a Kodak Photo CD test disk to the CIELUV color space and setting u^* and v^* at zero. We then applied the global, pixel-wise transformation

$$L^{*\prime} = \begin{cases} \left(\dfrac{L^* - L^*_{\min}}{L^*_{\mathrm{ave}} - L^*_{\min}}\right)^{\gamma} \left(L^*_{\mathrm{ave}} - L^*_{\min}\right) + L^*_{\min} & \left(L^*_{\min} \le L^* \le L^*_{\mathrm{ave}}\right) \\ \left(\dfrac{L^*_{\max} - L^*}{L^*_{\max} - L^*_{\mathrm{ave}}}\right)^{\gamma} \left(L^*_{\mathrm{ave}} - L^*_{\max}\right) + L^*_{\max} & \left(L^*_{\mathrm{ave}} < L^* \le L^*_{\max}\right) \end{cases} \tag{2}$$

(where L^* represents the original lightness value of a pixel, $L^{*\prime}$ its new value, and the subscripts "min," "max," and "ave" indicate the minimum, maximum and average lightness values of the original image) on the image using for γ the values 0.25, 0.35, 0.50, 0.71, 1.00, 1.41, 2.00, 2.82, and 4.00. Applying this transformation will, for $\gamma < 1$, decrease and, for $\gamma > 1$, increase the brightness contrast of the image. The minimum and maximum lightness of the image are not affected, while in general the average lightness will remain at approximately the same lightness value.

Eight subjects participated in the experiment. In three separate sessions they were shown on a CRT the complete set of images, in random order, with three replications. In the first session, the subjects' task was to judge the quality[||] of the image, in the second to judge the brightness contrast[#] of the images, and in the third session to judge the naturalness[**] of the images. Subjects were instructed to use an 11-point numerical scale ranging from 0 ("bad" or "low") to 10 ("excellent" or "high").

Results. Contrast judgments (averaged over subjects and scenes) versus the parameter γ are shown in Fig. 5. The figure shows that, as expected, contrast increases for increasing values of γ. Figure 6 shows quality judgments (diamonds, averaged over subjects and scenes) and naturalness judgments (triangles, also averaged) versus

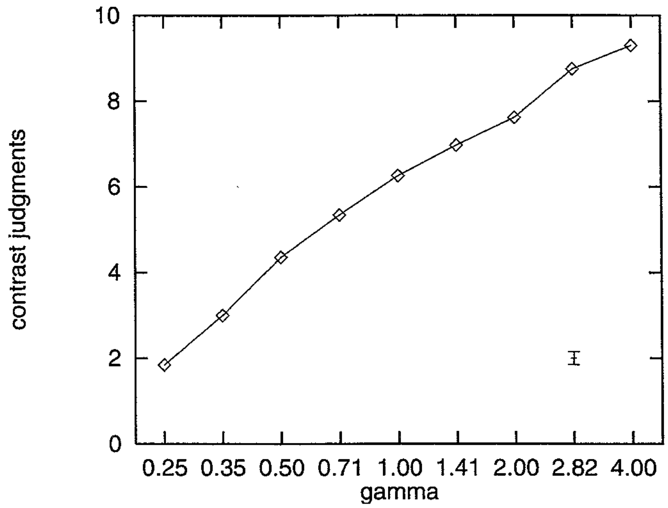

Figure 5. Contrast judgments (averaged over subjects and scenes) versus the parameter γ. The error bar denotes a distance of two average standard errors in the mean.

|| Defined as "the degree to which you like the image."

\# Defined as "apparent light-density differences."

** Defined as "the degree to which the image seems realistic to you."

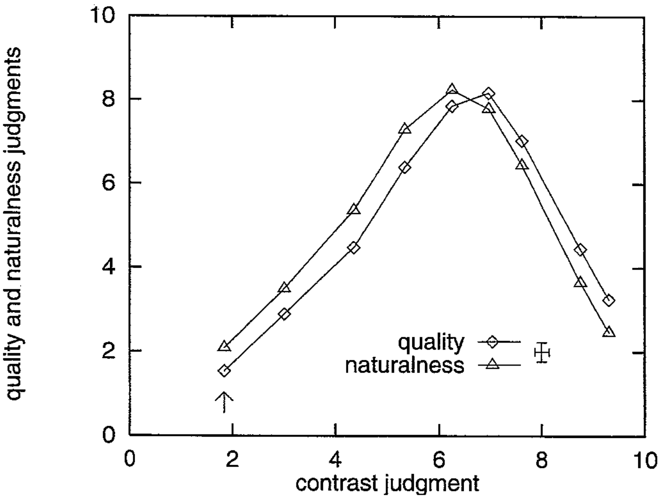

Figure 6. Quality judgments (diamonds) and naturalness judgments (triangles) versus contrast judgments (all averaged over subjects and scenes). The error cross denotes a distance of two average standard errors in the mean, and the arrow denotes the image with lowest γ.

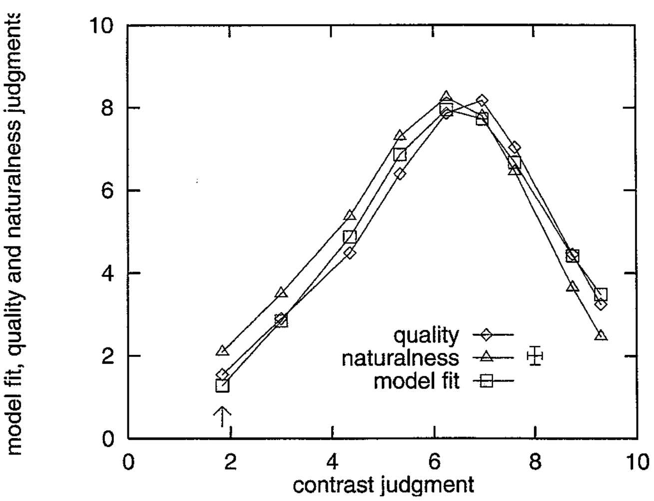

Figure 8. Quality judgments (diamonds) and naturalness judgments (triangles) versus contrast judgments (all averaged over subjects and scenes). The figure also shows the model fit (squares). The error cross denotes a distance of two average standard errors in the mean, and the arrow denotes the image with lowest γ.

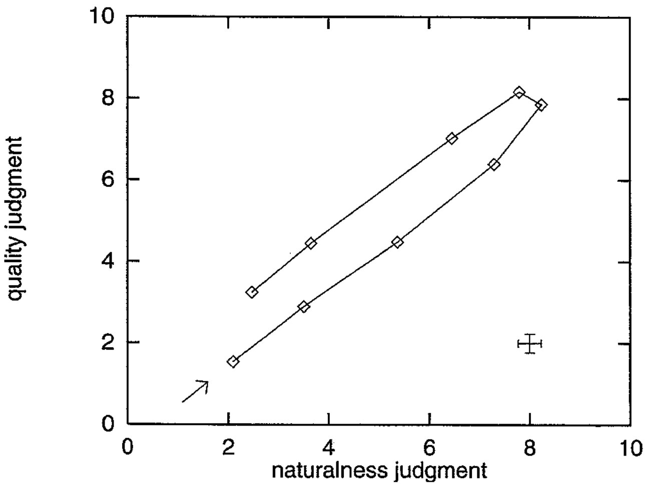

Figure 7. Quality judgments (averaged over subjects and scenes) versus naturalness judgments (also averaged). The error cross denotes a distance of two average standard errors in the mean, and the arrow denotes the image with lowest γ.

contrast judgments (also averaged). The curve for quality is shifted with respect to the curve for naturalness toward higher values of contrast, a result similar to the results for the condition chroma in experiment 1. To conclude, Fig. 7 shows quality judgments versus naturalness judgments. The U shape already found in experiment 1 is clearly visible.

Interpretation. We again find a high ($r = 0.95$) correlation between naturalness judgments and quality judgments, confirming that naturalness is a principal factor constituting image quality. The shift of the curve for quality with respect to the curve for naturalness toward higher values of contrast can readily be interpreted when realizing that higher contrast allows for more accurate detection and localization of edges in the image, and thus for a more precise internal representation of the image.

The least-squares fit of our model to the quality judgments obtained in the experiment is given by

$$Q = \lambda_1 N + \lambda_2 C + \lambda_3, \qquad \vec{\lambda} = (0.90, 0.25, -1.08). \tag{3}$$

Figure 8 shows this fit (squares), together with the quality (diamonds) and naturalness (triangles) judgments. The correlation between the fit and the quality judgments is very high ($r = 0.99$). Considering (1) the primitive nature of the model and (2) the strong nonlinearity of the correlation coefficient as a measure for goodness-of-fit (i.e., goodness-of-fit increases strongly for r approaching one), we may conclude that our description of image quality in terms of a compromise between naturalness and usefulness fits the data very well.

Concluding Remarks

In pursuing a top-down, analytical approach, we have achieved a fundamental interpretation of the processes that play a role in the estimation of the attribute "quality" of an image. We have argued that image quality is, to the observer, a *useful* attribute of an image, expressing how well the observer is able to employ the image as a source of information about the outside world—a view of image quality that is strikingly different from the "perceived distance to the original" philosophy often employed in image quality research.

The results of the experiments discussed in the previous sections support the concept developed here that the quality of an image can be described in terms of a compromise between the naturalness and the usefulness of that image. A logical next step from this point would be to more thoroughly specify the naturalness and usefulness requirements imposed on an image, e.g., by means of formulating algorithms. Implementations of such algorithms will enable (1) the development of instrumental measures for the prediction of image quality; and (2) estimation of parameter settings optimizing the quality of images.

At this point, two generalizations of the ideas discussed here may be interesting to note. First, our description of quality is essentially formulated *independently* of modality, suggesting the possibility of (1) simply applying the same ideas to, for example, the fields of sound or speech quality and (2) of generalizing the current description of image quality to a multimodal semantic description of perceived quality of information presentation. It is highly likely that such a description will prove valuable in the design and evaluation of applications in which a multimodal presentation of information plays a central role.

Second, in our description of image quality we have concentrated on the requirements imposed on the information acquired from the environment. However, requirements ensuring a proper interaction between observer and environment should necessarily include requirements that ensure the ability to adequately respond to the environment. These requirements then, in general, may be imposed on the means the observer is employing to control his environment. Such an approach is likely to result in a general theory of the quality of man–machine interaction.

References

1. D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman, San Francisco 1982.
2. H. G. Barrow and J. M. Tenenbaum, Computational approaches to vision, in *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds., Wiley-Interscience, Chichester, 1986, p. 38.1
3. M. Eimer, Representational content and computation in the human visual system, *Psych. Res.*, **52,** 238, 1990.
4. R. J. Watt, *Understanding Vision,* Academic Press, London, 1991.
5. V. Bruce and P.R. Green, *Visual Perception: Physiology, Psychology and Ecology*, Erlbaum, London, 1985.
6. J. J. Gibson, *The Perception of the Visual World,* Allen and Unwin, London, 1950.
7. J. J. Gibson, *The Senses Considered as Perceptual Systems*, Houghton Mifflin, London, 1966.
8. J. J. Gibson, *The Ecological Approach to Visual Perception,* Houghton Mifflin, London, 1979.
9. A. Newell and H.A. Simon, *Human Problem Solving,* Prentice-Hall, Englewood Cliffs, 1972.
10. A. Newell, *Unified Theories of Cognition,* Harvard University Press, Harvard, 1990.
11. R. W. G. Hunt, *Measuring Colour,* Ellis Horwood, Chichester, 1992.
12. E. A. Fedorovskaya, H. de Ridder, and F. J. J. Blommaert, Chroma variations and perceived quality of color images of natural scenes, *Col. Res. Appl.* **22**, pp. 96–110 (1997).
13. H. de Ridder, E. A. Fedorovskaya, and F. J. J. Blommaert, Naturalness and image quality: chroma variations in color images of natural scenes, *IPO Ann. Prog. Rep.*, **29,** 89 1993.
14. H. de Ridder, Naturalness and image quality: saturation and lightness variation in color images of natural scenes, *J. Imaging Sci. Technol.* **40** (6) pp. 487–493 (1996).