Optimizing color reproduction of natural images

Huib de Ridder
Delft University of Technology

Optimizing color reproduction of natural images

S.N. Yendrikhovskij, F.J.J. Blommaert, H. de Ridder
IPO, Center for Research on User-System Interaction
Eindhoven, The Netherlands

Abstract

The paper elaborates on understanding, measuring and optimizing perceived color quality of natural images. We introduce a model for optimal color reproduction of natural scenes which is based on the assumption that color quality of natural images is constrained by perceived naturalness and colorfulness of these images. To verify the model, a few experiments were carried out where subjects estimated the perceived 'naturalness', 'colorfulness' and 'quality' of images of natural scenes. The judgments were related to statistical parameters of the color point distribution across the images in the CIELUV color space. It was found that the perceptually optimal color reproduction can be derived from this statistic within the framework of our model. We specify naturalness, colorfulness and quality indices, describing the observer's judgments. Finally, an algorithm for optimizing perceived quality of color reproduction of natural scenes is discussed.

1. Introduction

Within the recent past, the *quantity* of images produced by color devices (e.g. displays, cameras, printers, etc.) has been increasing rapidly. Yet, the perceived *quality* of images has been improving more slowly and still is often far from perfect. There usually is no problem in obtaining a desirable image, but there often is a problem in obtaining the image of a desirable quality. Conceptually, the basic idea in obtaining images of optimal quality is simple: image quality has to be measured by the technical devices in a way similar to observer's judgments, and (if necessary) optimized before presenting the image to the observer. The implementation of this idea is much more difficult, and requires elaboration upon an objective criterion correlating with subjective strengths of image quality, and a robust algorithm for image quality enhancement. But first and foremost, it requires a fundamental understanding of the visuo-cognitive process that underlies an evaluation of image quality. This paper subsequently focuses on a) understanding, b) measuring, and c) optimizing perceived color quality of natural images.

The block diagram in Figure 1 specifies the process underlying color appreciation of natural images. Natural images refer to any hard or soft-copy representing natural, i.e. real-life, scenes by means of a color device such as CRT displays, printers, etc. The images are perceived in terms of color attributes (e.g. hue, colorfulness, and lightness). The perceived color attributes are assumed to be compared with color attributes recalled from memory in association with similar categories of objects. The object categories are extracted from the population of apparent object colors seen in the past and can be represented by the prototype, or the most typical category member. Similarity of the perceived and recalled attributes determines perceived naturalness of reproduced colors.

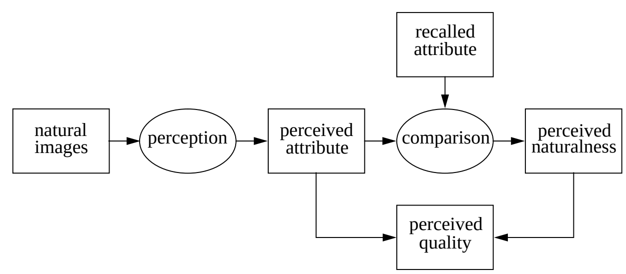

Figure 1: Diagram of the process that underlies the subjective impression of color image quality. Rectangles denote representation of the input and output of the visuo-cognitive processing; ellipses denote transformation of the input to the output.

Perceived naturalness of color images is assumed to be the first principal constraint imposed on perceptual image quality. However, a picture is unlikely to be judged of high quality if the items it contains are not distinguishable. We argue that the high degree of discrimination is the second principal constraint imposed on perceptual image quality. Recent studies [1−5] have shown small but systematic differences between perceived naturalness and quality of images when chroma varies: in general, subjects prefer slightly more colorful images, despite the fact that they realize that these images look somewhat unnatural. In line with these findings, we assume that when people look at images containing familiar categories of objects, two primary factors shape their subjective impression of how optimal colors are reproduced: colors should be perceived both realistically (naturalness constraint), and distinctly (discrimination constraint). Since the two constraints might conflict with each other, perceived color quality is suggested to be modeled as a compromise between them.

At present, the effect of the discrimination constraint is confined to the impact of colorfulness, and the quality model is tested for displayed images. The model can be adopted for other domains (e.g. lightness, hue) and color devices (photocameras, printers, etc.). In the following research, we conducted several experiments in which subjects estimated perceived naturalness, colorfulness and quality of natural images. Subsequently, we specify naturalness and colorfulness indices, describing these assessments. The two indices are used to obtain the color quality index and optimize color reproduction of natural images.

2. Naturalness index

2.1. Basic assumptions

The essential problem in developing a naturalness index for CRT displays can be formulated as follows: *Given the monitor's RGB measurements of an image only, how can the perceived naturalness of this image be computed?* To solve the problem, we introduce here three main assumptions that can be used for computation of a naturalness index.

Assumption 1: Digitized images of natural scenes are represented as color statistics, i.e. color point distributions, in some perceptual and approximately uniform color space.

The consequence of this assumption is that a) spatial information is disregarded; b) RGB values of the displayed images are transformed into a psychometrical color space, e.g. the CIELUV color space; and c) images are considered as an ensemble of color points in this space, where each point represents one pixel of a corresponding image. Although the CIELUV space is not an appearance space, the basic statistical parameters (e.g. mean, spread) of the color point distribution across the images in the CIELUV color space correlate positively with apparent colorfulness[3], lightness[2] and hue[1] of images of natural scenes.

Assumption 2: Different segments in the CIELUV color space, particularly in the CIE u'v' chromaticity diagram, have different probabilities of being occupied by the color point distribution of images that are perceived to be natural.

The validity of this assumption was investigated [5] on the basis of color distribution of 78 images from two Photo CDs. It was shown that the distribution is not spread uniformly in the CIE u'v' plane and can be roughly divided into four regions: achromatic, orange-yellow, yellow-green, and blue parts of the diagram. The areas have no distinct borders and form a sort of triangle with the orange, green and blue colors in the corners. For the present, we divide the CIE u'v' diagram into three crucial sections (see Figure 2), for convenience referred to as 'skin', 'grass', and 'sky' segments, that were determined from a research on naturalness judgments of these colors[6]. The segments are crucial, or privileged, in the sense that they cover areas of possible variation for the three most familiar and frequently reproduced categories of colors. Human faces, grass, and sky can be found in an absolute majority of images of natural scenes. However, the names of the segments should not be understood literally, because colors of other object categories (e.g. orange, foliage, water) fall within their borders as well. A consequence of dividing the CIE u'v' diagram into the privileged segments is that the perceived naturalness of any color reproduction can be estimated locally within these segments.

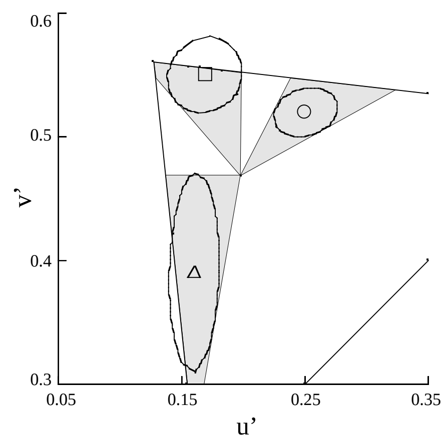

Figure 2: *The 'skin' (circle), 'grass' (square), and 'sky' (triangle) segments for calculation of the naturalness index. Ellipses: the best-fitting Gaussian σ-contour ellipses representing the naturalness judgments of these colors.*

Assumption 3: The differences between average and so-called 'prototypical' saturation values within the privileged segments are represented by means of a Gaussian function.

The color statistic of image i containing n_{ij} pixels in a privileged segment j has a certain average saturation, S_{ij}. A statistic of S_{ij} for images that are perceived to be natural can be represented by a probability density function with the mean, denoted as μ_j, and the standard deviation, denoted as σ_j. The mean, μ_j, is defined as the expected value of S_{ij}, and can be understood as some 'prototypical' color, or the most typical saturation value, within a privileged segment j. It was shown[6] that the perceived naturalness of a reproduced object color can be characterized by a monotonically decreasing function (e.g. a Gaussian density function) of the distance in the CIELUV space between the apparent color of the object and the prototypical color of that object category. Similarly, we propose that the naturalness index N_{ij} for an image i within a privileged segment j can be defined by a Gaussian density function of the differences between S_{ij} and μ_j values:

$$N_{ij} = exp(-0.5(\frac{S_{ij} - \mu_j}{\sigma_j})^2) \quad (1)$$

Then the total naturalness index N_i for an image i can be defined as a sum of the naturalness indices per segment weighted with the number of pixels n_{ij} within a privileged segment j:

$$N_i = \frac{\sum n_{ij} N_{ij}}{\sum n_{ij}} \quad (2)$$

Note that N_i varies from 0 (the most unnatural image) to 1 (the most natural image).

2.2. Verification: Experiment 1

The aim of this experimental study was to examine the perceived naturalness of 30 images of real-life scenes by multiplying CIE saturation values of every pixel in the images with a scaling factor.

2.2.1. Method

Eight subjects, 4 females and 4 males, with no deficiencies in color vision took part in the experiment. Their age varied between 20 and 27. They were students at Eindhoven University of Technology and had little or no knowledge of color reproduction technology.

Thirty digitized pictures from two Photo CDs were chosen for the experiment. The pictures represented typical categories of photographs: portraits, landscapes, people in nature, natural objects (flowers, animals etc.), and artificial objects (planes, boats, etc.). The images were described by their color point distributions in the CIELUV color space. The transformation into the CIELUV color space was made by using standard formulae[7]. A new set of images was prepared by changing the CIELUV saturation value for each pixel while the CIELUV lightness and hue-angle were kept constant. Every picture was manipulated twice. The saturation values of the master (unprocessed) images were multiplied by two randomly taken scaling factors: the first constant ranging from 0.0 to 0.9, and the second constant ranging from 1.1 to 2.0. The resulting 60 images plus 12 of the 30 original images composed a total set of 72 pictures. The pictures were displayed by an Image Sequence Processor ISP500 of a Digital Video System on a high-resolution 50 Hz non-interlaced BARCO color monitor driven by a SUN-3/260 workstation.

The experiments were run in a dark room with a white dimly lit 2.5 cd/m^2 background behind the monitor. The subjects were seated approximately 1.5 m in front of the screen. The images were displayed with a flexible timing until the subjects gave a response. During a 5-min period of adaptation to the room illumination and screen luminance, the subjects studied instructions and then did a training series of 10 stimuli. The observers estimated the naturalness of the displayed images, defined as 'the degree of correspondence with reality'. The observers were asked to make their judgments using an eleven-point numerical category scale ranging from 0 (the lowest degree of perceived naturalness) to 10 (the highest degree of perceived naturalness). All 72 images were randomly shown two times during the experiment.

2.2.2. Results

The naturalness index was derived by fitting the naturalness judgments with Equation 1 and Equation 2 using a least-squares fitting procedure. The correlation between the obtained naturalness index and naturalness scores reported by observers is $r = 0.87$ $(F = 212, p < 0.001)$ and explains the major variability in the experimental data. As can be seen in Figure 3, the naturalness of some images is predicted better than that of others. In general, the prediction deteriorates for images containing objects, whose saturation values might deviate considerably from the prototypical mean values without destroying the impression of naturalness. For example, some building materials, an animal's hide, and the color of the sun may fall in the 'skin' segment, have a very low or high saturation value, and still appear to be 'natural'. But considering that even for these objects the prediction is not extremely bad, and the fact that we only used a first-order model, the performance of the naturalness index is quite good.

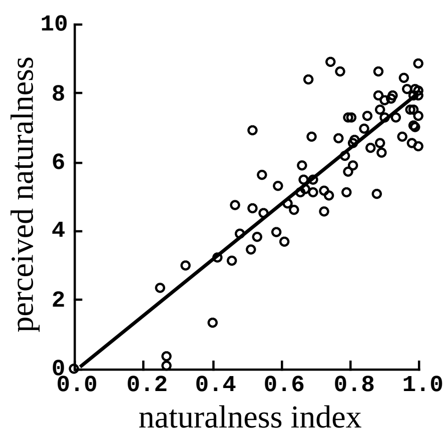

Figure 3: Naturalness judgments derived from Experiment 1 versus the naturalness index.

3. Colorfulness index

3.1. Basic assumptions

The problem in developing a colorfulness index for CRT displays can be formulated as follows: *Given the monitor's RGB measurements of an image only, how can perceived colorfulness of this image be computed?*

Colorfulness has been defined as the attribute of a visual sensation according to which the perceived color of an area appears to be more or less chromatic[7]. Not much is known about what determines perceived colorfulness of displayed images of natural scenes. Recently, some progress has been made in modeling the relation between colorfulness judgments of displayed images and statistical parameters of the CIELUV chroma distributions across the images[3,8]. It was proposed that colorfulness depends on two factors. The first factor is the distance of the image colors from a neutral gray; this factor might be modeled as the mean of the distribution of the chroma values of all individual image colors. The second factor is the distance between individual image colors; this factor might be modeled as the spread of the average chroma expressed in, for instance, the standard deviation of the distribution of chroma values. Indeed, experimental data have shown that direct-scaling and difference-scaling of the perceived strength of colorfulness is linearly related in good approximation to the weighted sum of the average chroma and the standard deviation of the chroma when those parameters are varied[3,8].

At present, we ignore the lightness differences between images. In this case, chroma can be replaced by saturation, since for the CIELUV space saturation equals chroma divided by lightness. We assume that the average saturation S_i of image i and its standard deviation σ_i are linearly related with the same weight to colorfulness judgments. Then, the colorfulness index C_i is assumed to be defined by the expression:

$$C_i = S_i + \sigma_i \tag{3}$$

Note that C_i varies from 0 (achromatic image) to C_{max} (most colorful image).

3.2. Verification: Experiment 2

3.2.1. Method

Eight subjects estimated the colorfulness of the displayed images, defined as 'presence and vividness of colors in the image'. The observers were asked to make their judgments using an eleven-point numerical category scale ranging from 0 (the lowest degree of perceived colorfulness) to 10 (the highest degree of perceived colorfulness). Images and experimental conditions were identical to the ones of Experiment 1.

3.2.2. Results

The colorfulness judgments as a function of the colorfulness index are shown in Figure 4. Every point represents averaged scores over 16 trials (2 repetitions $*$ 8 subjects). A linear regression analysis demonstrates that the correlation between these two variables is high $(r = 0.91, F = 327, p < 0.001)$.

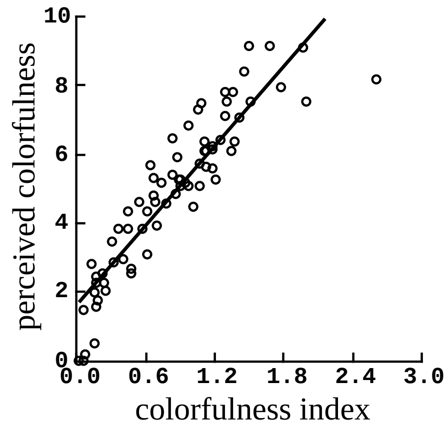

Figure 4: *Colorfulness judgments derived from Experiment 2 versus the colorfulness index.*

4. Color quality index

4.1. Basic assumptions

The problem in developing a color quality index for CRT displays can be formulated as follows: *Given the monitor's RGB measurements of an image only, how can the perceived quality of this image be calculated?*

Results of previous investigations[4,5,9] have shown that the quality judgments of images of natural scenes can be modeled as a linear combination of the colorfulness and naturalness judgments of these images when saturation varies. Therefore, the color quality index Q_i for image i may be calculated as a linear combination of the colorfulness and naturalness indices:

$$Q_i = w_q N_i + (1 - w_q)(C_i / C_{max}) \tag{4}$$

where C_i / C_{max} is the normalized colorfulness index. The weighting parameter, w, describes a compromise between the naturalness and colorfulness constraints: when $w_q = 1$, perceived quality is exclusively determined by perceived naturalness of the image, when $w_q = 0$, quality is determined by perceived colorfulness of the image only. Note that Q_i varies from 0 (image of the worst color quality) to 1 (image of the best color quality).

4.2. Verification: Experiment 3

4.2.1. Method

Eight subjects estimated the color quality of the displayed images, defined as 'the degree to which you like the reproduction of colors in the image'. The observers were asked to make their judgments using an eleven-point numerical category scale ranging from 0 (the lowest degree of perceived quality) to 10 (the highest degree of perceived quality). Images and experimental conditions were identical to the ones of Experiment 1.

4.2.2. Results

The quality judgments as a function of the color quality index are shown in Figure 5. Every point represents averaged scores over 16 trials (2 repetitions $*$ 8 subjects). Apparently, the proposed color quality index evaluates the perceived quality of the images in close accordance with the observer's assessments $(r = 0.90, F = 290, p < 0.001)$. Through the least-squares fit of the quality judgments using Equation 4 we have found that the weighting parameter w describes a compromise between the naturalness and colorfulness constraints equals 0.75.

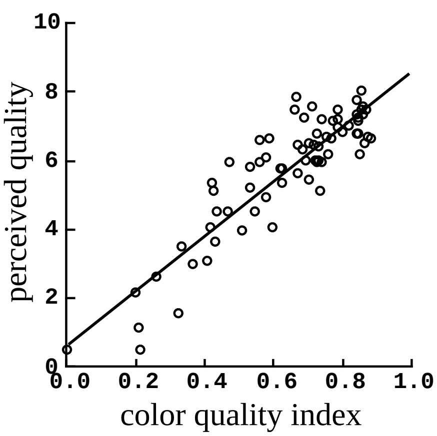

Figure 5: *Quality judgments derived from Experiment 3 versus the color quality index.*

5. An algorithm for optimizing color quality of natural images

The proposed quality index is an attempt to develop a perceptual criterion to estimate whether the reproduction is optimal. This criterion can further be used to optimize the quality of this reproduction. Basically, the optimization should be performed in two stages: 1) computation of the 'naturalness-index' of the image under investigation, and 2) computation of the direction and magnitude for the colorfulness enhancement. Below we present an example of such an algorithm.

The first two steps of the algorithm are the same as for calculating the naturalness index. The third step is based on a comparison of the average, S_{ij}, and prototypical, μ_j, saturation values:

$$k_{ij} = \frac{\mu_j}{S_{ij}} \tag{5}$$

The results of the comparison are weighted per segment:

$$k_i = \frac{\sum n_{ij} k_{ij}}{\sum n_{ij}} \tag{6}$$

where n_{ij} denotes a number of pixels within the segments. The 'naturalness' constraint is implemented as following:

$$S_i' = k_i S_i \tag{7}$$

where S_i is the saturation value of the master image, and S_i' is the new saturation value after the processing. Due to possible discrepancies between perceived naturalness and values of the naturalness index, it is necessary to restrict multiplications with very small/big constants k_i that might cause prohibitive color artifacts. For this, we propose to use an additional 'original' constraint:

$$S_i'' = w_o S_i' + (1 - w_o) S_i \tag{8}$$

where S_i'' is the new saturation value, and

$$w_o = \frac{1}{1 + |1 - k_i|} \tag{9}$$

A colorfulness enhancement is performed in the final stage of the algorithm. The enhancement is based on color transformations that slightly increase colorfulness of processed images, for instance by multiplication of the images saturation values by constant 1.2 (see research[9] for details).

5.1. Verification: Experiment 4

5.1.1. Method

The proposed algorithm was tested for images from the TV-net. The pictures represented typical categories of scenes: out and in-door footage, advertising, movies, news, sport, and music clips. New images were prepared by changing the CIELUV saturation values for each pixel in accordance with the algorithm of the color quality optimization. The resulting 20 processed and 20 master images were assessed in quality twice by 10 observers. The subjects judged perceived quality of the displayed images, using a scale from 0 to 10. The quality description and other experimental conditions were similar to the ones of Experiment 3.

5.1.2. Results

Results of the averaged quality judgements are summarized in Figure 6. The algorithm should optimize images of low quality and retain unchanged images of high color quality (solid curve). In general, the differences between quality estimations of images processed by the algorithm and unprocessed images follow this curve. In 16 of the 20 cases the subjects gave higher quality scores to the processed images.

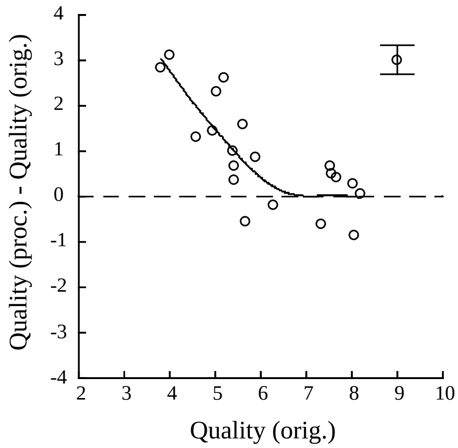

Figure 6: *Relative quality judgements of images processed by the color quality optimization algorithm versus absolute quality judgements of the unprocessed images. The vertical bar indicates twice the average standard error of the mean.*

6. Discussion and conclusions

We present a model for optimal color reproduction of natural scenes and propose naturalness, colorfulness and quality indices that evaluate these perceived attributes in accordance with observer's judgments. The naturalness index can be estimated locally within the 'skin', 'sky', and 'grass' segments in the CIE u'v' chromaticity diagram. The colorfulness index is defined as a sum of the average saturation value and its standard deviation. The color quality index can be calculated as a weighted sum of the colorfulness and naturalness indices.

One of the main consequences of our model is that the original, i.e. 'as it was in nature', reproduction of a real-life scene might not always be the perceptually optimal one. It has long been noted that a colorimetrically precise reproduction of object colors is usually not judged as the most desirable reproduction. The proposed color quality index is an attempt to develop a *perceptual criterion* to estimate whether the reproduction is optimal.

7. References

1. H. de, Ridder, and F. J. J., Blommaert, and E. A., Fedorovskaya, "Naturalness and image quality: chroma and hue variation in color images of natural scenes", *Proc. of the SPIE Conference*, **2411**, pp. 51-61, (1995).
2. H. de, Ridder, "Naturalness and image quality: saturation and lightness variation in color images of natural scenes", *J. Imaging. Sci. and Technol.*, **40**, pp. 487-493, (1996).
3. E. A., Fedorovskaya, H. de, Ridder, and F. J. J., Blommaert, "Chroma variations and perceived quality of color images of natural scenes", *Color Research and Application*, **22**, pp. 96-110, (1997).
4. T.J.W.M., Janssen, and F. J. J., Blommaert, "Image quality semantics", *J. Imaging. Sci. and Technol.*, **41**, pp. 555-560, (1997).
5. S.N., Yendrikhovskij, F. J. J., Blommaert, H. de, Ridder, "Perceptually optimal color reproduction", *SPIE Proc.*, **3299**, (1998).
6. S.N., Yendrikhovskij, F. J. J., Blommaert, H. de, Ridder, S.T.J. van Helvoirt, "Naturalness judgments of object colours", *Proc. of the 8th Congress of the International Color Association in Kyoto, Japan*, pp. 163-166, (1997).
7. G., Wyszecki, and W. S., Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 487, (1982).
8. S.N., Yendrikhovskij, H. de, Ridder, E. A., Fedorovskaya, and F. J. J., Blommaert, "Colourfulness judgments of natural scenes", *Acta Psychologica*, **77**, pp. 79-94, (1997).
9. S.N., Yendrikhovskij, F. J. J., Blommaert, H. de, Ridder, "Towards perceptually optimal color reproduction of natural scenes", *Proc. of the 1st International Conference on Colour Imaging in Multimedia, Derby, UK*, pp. 237-245, (1998).