Image Quality as a Problem of Computational Vision

Elena A. Fedorovskaya
Eastman Kodak Company
Rochester, New York

Abstract

The present paper describes the framework and results of experiments aiming at developing computational methods to predict perceived image quality and perceived attributes, based on digital image data. Subjective judgments of overall contrast, sharpness, lightness, colorfulness, and quality were collected for 126 images presented on the monitor screen. For the analysis, digital images were represented in CIELAB color space, which was used as an approximation of the perceptually uniform color space. We utilized a computational vision approach and showed that all global attributes, including quality, can be predicted based on the combination of features (image properties) that could be considered as relevant elements at different levels of mental representation. The dependence of the global attributes upon the same features explains their mutual correlations, a known empirical fact. The regression results look very promising but require further refinement in terms of the feature assessment.

Introduction

Rapidly growing numbers of applications and services related to digital imaging and digital photography emphasize the output image quality as one of their most important product attributes. Delivery of these applications and services require the development of elaborate image processing techniques to maintain and, if necessary, enhance the perceived image quality of each individual image. An intelligent image processing system, that can selectively apply specific image processing or image enhancement techniques to any given image can potentially save processing time associated with the application of image processing algorithms, and can improve customer satisfaction by enhancing the perceived quality of pictures that will benefit from this processing. Predicting perceived quality from image data in an automatic fashion, and determining a particular image characteristic or attribute that may need improvement, therefore becomes an increasingly important part of such an intelligent imaging system.

Approaches to Image Quality

There currently exist several directions in pictorial image quality research that were initially developed for specific purposes. These approaches are usually implemented in conjunction with certain experimental methods and empirical models used to explain the data. Among them we can differentiate:

Psychophysical or Psychometric Approach

Within this approach, subjective image quality judgments and their biases became the main subject of investigation. This work was pioneered Eastman Kodak Company researchers: MacAdam[1] and Bartleson.[2] and further developed by other researchers.[3,4] A central premise of this research is that perceived attributes can be measured using perceptual scaling techniques and that a relationship may be derived to explain the connection between a physical or system parameter and the perceptual scale. Image quality is then understood as an aggregate of perceived image attributes.

System-Based Approach

The goal of this approach is to establish a set of specific system parameters to attain reproduction aims. According to this approach parameters that describe the performance of a capture or a reproduction imaging system are set to produce images according to the defined specification. Image quality is then regarded as a certain quantifiable level of such performance. Usually this approach considers a single aim or a few aims that would satisfy a broad range of images, and thus characterizes the system performance in terms of the perceived quality of an average or representative set of images. The approach has proven to be a useful tool for optimization of various system parameters. The main challenges of this approach are: a scene dependency, because aims for the images of different scenes could be different (example, a portrait image compared to a scenic image), and a system dependency: a new reproduction method or a system that incorporates novel technical solutions, brings new parameters that may require the reassessment of the system performance. For example, a tone-reproduction curve, as a system parameter relevant to the system contrast, could be replaced by the scene-dependent tone scale. This process produces results difficult to predict, based on the concept of the tone scale as a smooth monotonic function and related concept of an input image as an average image for the on-aim photography.

Signal Processing Based Approach

This approach originated from the need to measure and compare physical or technical properties (e.g., resolution, bit depth, noise, histogram, compression rate, quantization error, etc.) of digital images and image transformations (filtering, sampling, interpolation, etc.) with respect to output quality. Because these properties are not directly related to the observer's "psychological reality," their estimation originally did not include observer's perception. This methodology is a broadening of the signal processing approach. An image is thus considered to be a complex signal, and image quality is treated as a measure of this signal. From this, such metrics as signal-to-noise ratio, root-mean-square error, etc., were derived and used with respect to image quality. Within the context of this approach, a human observer is implicitly understood as another technical device that captures and registers physical complex signals. Although useful in specific instances, this paradigm is limited in its capability for generalization.

A Low Vision-Based Approach

This approach, commonly referred to as vision modeling was pioneered during the early 1990's[5,6] and has received significant attention since.[7] It originated from the need to assess threshold level changes in an image due to processing techniques, such as image compression, filtering, sampling, etc. This approach can be considered as an integration of low-level vision data obtained in psychophysics to enable differential response for complex images. The output of the visual model-based algorithm could be described as a measure of visual difference between a reference image and an image under consideration. Although the vision model-based approach has demonstrated important successes in simulating low level visual processes,[8] its main challenges, with respect to image quality, are associated with difficulties in formulating a combination metric that allows for the integration of spatially localized difference maps into a single visually adequate global measure of visual difference, and the absence of a clear relationship between absolute visual difference and image quality. Because it is impossible to infer which image from a pair of images has higher quality considering only a visual difference map, the image quality concept within this approach is reduced to the concept of image fidelity with a crucial role of the reference (original) image, frequently understood as having optimal quality. In addition, higher-level visual processes, which have not yet been incorporated into these models (e.g., constancy, attention, subjective importance of different regions of the image), may significantly influence the judgment of the visual difference and, consequently, the predictive power of the model.

Information Processing Approach

A very novel and powerful approach was recently proposed in an attempt to explain experimentally observed discrepancy between judgments of naturalness and quality preferences.[9-11] This approach emphasizes visual information processing in understanding and modeling of perceived image quality.[12] Assuming that visual processing of images is a goal-directed process, and stressing its active nature from the observer's stand point, the suggested approach formulates then image quality as the degree of the adequacy of the image as input to the vision stage of the interaction process. Two requirements are proposed in considering this adequacy: discriminability and identifiability of the image content. The fruitfulness of this paradigm was demonstrated by developing computational methods to define colorfulness, naturalness and quality for images subjected to global variations along perceptual dimensions in color space.[11] It was not clear, however, how the particular implementation of this approach if strictly followed could lead to predicting quality of an arbitrary image from an arbitrary source.

The provided classification is relative, in a sense that often there is a combination of the approaches in any given investigation. However, they still can be recognized by their primary focus and applied methods. A very useful way to visualize their specialty could be derived from the diagram of Image Quality Circle suggested by Engeldrum,[13] if one could imagine a short link to quality determination from the appropriate blocks in the circle that denote components of the chain that relate technology variables (system parameters) of the imaging system to resulting customer quality preferences.

Below we describe the paradigm that could be utilized to integrate advantages and unique knowledge obtained through various approaches. The paradigm is very closely related to the information processing approach suggested by Janssen.[12] However, rather then considering information processing in general, we would like to make a main emphasis on the understanding of vision process as a structure of multiple levels of mental representation, a notion that allows us to more fully explore perceptual properties of images with respect to image quality.

Computational Vision as a Paradigm Toward Image Quality

Computational vision is a multidisciplinary field that integrates a number of disciplines: neurophysiology, psychology, and artificial intelligence, which considers vision as a computational process and "emphasizes information, knowledge, representation, constraints, and processes, rather than details of mechanisms."[14] Within the computational approach to vision that is described in a number of papers,[14, 15] a vision system is often structured as a succession of levels of representation. The initial levels are constrained by what is possible to compute directly from the image, while higher levels are dictated by the information required to support the ultimate goal. In between, the order of representation is constrained by what information is available at preceding levels and what is required by succeeding levels. For example, the sensing process, which takes place in early vision, enables conversion from light flux incident on a photosensitive receptor array to a brightness measurement by the photosensing mechanism that often involves spatial quantization. The next stage of processing attempts to detect spatial and temporal changes such as discontinuities in brightness or brightness gradient, line ending or local anomalies in a homogeneous field, in the image and to make them explicit. The output representation is an array with feature description recorded at each location. Marr's raw primal sketch is an example of a suitable representation for image features.[15] The raw primal sketch uses three kinds of primitives to describe intensity changes: various types of edge, lines or thin bars, and blobs. Each is characterized in terms of orientation, size (length and width, or a blob's diameter), contrast, position, and termination points.

The local edge and blob description in the raw primal sketch must be organized into spatially coherent units (e.g., boundaries and regions) for subsequent analysis. Some basic grouping processes occur at this stage; organizing elements into straight lines and smooth curves, and to cluster elements into regions of textures. This stage corresponds to what Marr denoted as a full primal sketch.

Palmer,[16] using Marr's theoretical framework as a foundation, decomposed visual perception at the algorithmic level into four major stages beyond the retinal image itself. Each stage is defined by a different kind of output representation and the processes that are required to compute it from the input representation. Applying the labeling scheme from the information processing approach, each stage is named according to the kind of information it represents explicitly: the image based, surface-based, object-based, and category-based stages of perception.

The computational vision approach seems to be a powerful paradigm for understanding and modeling perceived image quality. The fruitfulness of the information processing approach to computationally define image quality and naturalness has already been demonstrated by Endrikhovski and Janssen.[11,12] Our previous research to predict perceived overall contrast from digital image information using a computational vision paradigm yielded promising results.[17]

We believe that a fully implemented computational vision framework will allow us to overcome major limitations of the existing approaches to image quality, namely, algorithm dependency, system dependency, and scene dependency, by considering image properties relevant to appropriate levels of mental representations of a scene, as well as expressing and measuring the scene in terms of their adequacy to human visual and cognitive processes. Therefore, image quality can be understood as a measure of a multi-level interaction between the user and an image.

We propose the following assumptions:

1) Judgments of perceived image quality, as well as overall attributes (e.g., overall contrast, overall lightness, colorfulness, etc.), are derived from certain mental representations and, therefore, can be analyzed with respect to their corresponding elements.
2) Perceived image quality and perceived overall attributes, as related to the integral impression from the image, may involve elements from multiple levels (or stages) of representation.
3) The elements relevant to the perceived attributes and quality can be computationally evaluated and, therefore, the prediction of perceived attributes and quality can be calculated.
4) In perceived quality and attribute predictions, the elements measured can be combined.

We think that allowing multiple representations and plurality of perceptually relevant elements to contribute to perceived image quality assessment might help to solve the existing problems of scene, algorithm, and system dependency, with respect to image quality modeling. In principle, all perceptually, and broadly, psychologically-defined representations can be considered as giving rise to quality judgments. Already mentioned are: sensory array (retinal image, input image, intensity image, or color array are different terms used by different authors), image-based stage, surface-based stage, object-based stage, and category-based stage.[16] Additionally, other kinds of representations can potentially be included, e.g., emotionally related, symbol-, or metaphor-related. However, the difficulty lies in defining relevant scene entities for those representations and assessing them in a systematic manner.

Examples of relevant elements that characterize events at each stage of visual processing for different representations are array elements for the sensory array representation, considered both in a spatial and frequency domain. Edges, lines, or regions for the image-based representation; local patches of 2D surface at some slant, located at some distance, could be considered as the elements for the surface-based representation; while volumetric primitives are the entities for object-based representation, etc. Elements can be associated with a number of descriptors, for example, size, orientation, contrast, position, as well as statistics. We will call these descriptors as features.

There are a variety of possible combination rules that can be applied to combine the features, for example, Minkowski metric with different exponents. The linear summation is the simplest combination rule, which we will consider in this paper.

In this case, if $\mathcal{F} = F_1,\ldots, F_n$ is the set of measurements performed for the corresponding n features, and $\mathcal{A} = A_1,\ldots, A_m$ – a set of m overall image attributes, then $\mathcal{A} = \mathcal{F} \cdot B$, where B is the nxm- matrix of weights, assigned to those features.

To test whether this approach is useful for assessing overall perceived image quality, as well as other attributes based on image information, we performed a series of experiments. In these experiments, subjects evaluated overall quality, contrast, sharpness, lightness, and colorfulness of images of natural scenes. Subsequently, their judgments were compared with the measurements of features extracted from the observed images.

Experiments

Images and Viewing Conditions

One hundred twenty six (126) images were chosen for the experiments, representing a variety of scenes, such as indoor and outdoor pictures, different seasons, people in groups and close-up portraits, animals, and images taken under different lighting conditions (e.g., flash, bright sun, twilight, shadows, etc.). The scenes were photographed using color negative film, and digitized with a high-resolution digital scanner. In the stimulus set, 111 pictures were unique color scenes and 15 images were black-and-white versions of some selected color scenes. All images were landscape-oriented. The images varied in their quality and attribute strength as a result of problems at capture (e.g., an indoor picture with an electronic flash) or because of specific scene characteristics (e.g., a backlit scene or misty conditions). High-resolution digital image files were interpolated to obtain a 966 x 654 image size, which was used in the experiments. Stimuli were viewed on a 21" CRT display. White point of the CRT was set to D65. Peak screen luminance was 21 foot-lamberts and the resolution was 1152 x 870. Subjects sat at a distance of 39 inches from the display. Indirect fluorescent lighting produced an ambient illumination level of 20 lux measured at the screen. An adapting neutral field with a luminance of 3.5 foot-lamberts served as a background for image presentation (this level corresponded to an L* value of approximately 50). The visual angle of the stimuli was 15.5 x 10.5 degrees.

Subjects

Seven subjects participated in the experiments. They had normal or corrected-to-normal visual acuity and normal color vision. Their ages ranged from 30 to 50 years. Every subject participated in 6 separate experiments, with 2 sessions per experiment, where they were asked to rate overall quality, overall contrast, overall lightness, overall colorfulness, overall sharpness, and main subject sharpness. The order of the experiments was randomly assigned for each participant.

Attribute Definitions

The following definitions were provided to the subjects:

Perceived Overall Contrast was defined as an integrated impression of differences in lightness, or lightness variation observed within the whole picture.

Overall Sharpness was the overall impression of clarity of edges observed within the whole picture.

Overall Colorfulness was defined as the impression of presence and vividness of colors in the whole picture.

Overall Lightness was defined as the impression of the lightness level that an image produces.

Perceived Quality was defined as the degree of excellence of the reproduction.

Main Subject Sharpness was defined as the overall impression of clarity of edges on the primary subject.

Session Procedure

Stimulus presentation was randomized, with every image presented twice in each experiment. Time presentation for every image was not limited. However, subjects were encouraged to reply promptly by using a slider displayed at the bottom of the screen below the image. The subjects used the slider to identify a position on the scale from 0 to 100, which would reflect the strength of the attribute under consideration. Perceived quality and other attributes were evaluated using a free modulus magnitude estimation technique, where observers are not given a reference image for the attribute; instead, they use their own internal reference. Before the actual session, a short trial session was run for every subject to ensure that subjects understood the instructions and felt confident about the scaling technique. During the trail session 10 to 20 stimuli were presented to assess the degree of confidence and response consistency. Average time for an experimental session was 45 minutes with a break to avoid fatigue.

Image Analysis

We considered representations relevant to the sensory stage and image-based stage as appropriate candidates for feature selection. Although, ideally, a vision model of low level visual processing should be used to transform a physical image, comprised of pixels, into a “sensory array” representation, for the practical purposes of simplification, we approximated this representation by describing image pixels in CIELAB color space, hoping that this approximation would be sufficient in demonstrating the objectives: namely, the applicability of the computational vision approach to image quality.

For the array representation in the spatial domain we selected features that included descriptive statistics for CIELAB pixel distributions of lightness, chroma, and hue angle. Statistical moments: mean, standard deviation, skewness and kurtosis, other statistics, such as maximum and minimum values, were among the features, as well as their combinations. For example, a range contrast feature was calculated using the formula for Michelson contrast, expressed as a ratio of the difference between maximum and minimum lightness values to their sum.

To minimize computation time, all these features were calculated for the low-resolution images obtained by averaging blocks of 8 x 8 pixels and scaling down the image.

For the frequency domain, energy in various frequency bands, as well as ratios were computed using the original, high-resolution image.

For the image-based stage, we used edge-related and region-related descriptors. Edge contrast feature was approximated by the standard deviation of lightness differences between adjacent pixels of the low-resolution image. The color area contrast feature was computed as the average pair-wise difference between mean CIELAB values for the regions multiplied by the number of regions identified in the image.

The regions were obtained by applying an image segmentation algorithm to the low-resolution image.[18]

Other edge characteristics were determined from the full-resolution image. In this case, we used an edge detection method to first identify edges. For all edge pixels edge height, edge gradient and edge width were computed. Mean and maximum values for all of these characteristics were used as features.

The choice of a particular numerical way for the feature assessment was driven by "visual" sense and simplicity considerations. It is obvious, however, that this assessment can be refined and tied to psychophysically established data.

Results

Judgments of image quality and overall perceived attributes were averaged across subjects and used for a subsequent multiple regression analysis to estimate the contribution of all computed features to the perceived attributes. We used a stepwise regression method to choose the best feature combination for every attribute. An additional consideration was to minimize a total number of features participating in the prediction models.

Figure 1 illustrates the results of overall contrast prediction based on the regression analysis. Seven features were identified as significantly contributing to the prediction: maximum lightness, maximum chroma, a distance of the mean image lightness from the background lightness, range contrast, edge contrast, mean edge gradient, and color area contrast. The linear combination of the measures produces a prediction of the overall perceived contrast with the multiple correlation coefficient of approximately 0.77.

Given the simplifications we used with respect to the feature computation, a large variety of color and black-and-white scenes, and the complexity of the subjects' task to produce an estimate of overall perceived contrast, we felt satisfied with the demonstrated ability to predict subjective contrast judgments on an image-to-image basis using a simple linear model.

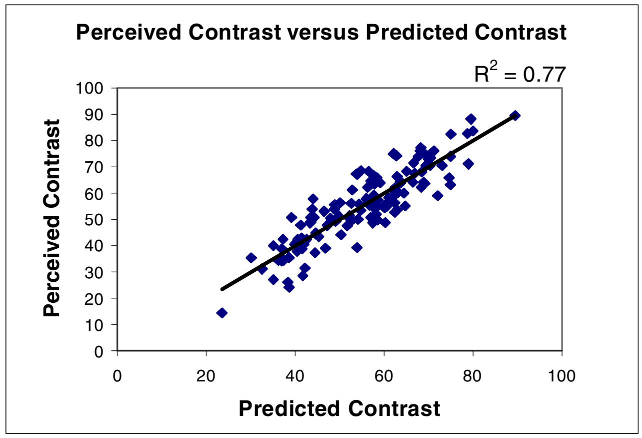

Figure 1. Perceived overall contrast prediction based on a feature combination

We obtained similarly encouraging results for other attributes. The features contributing to perceived quality and attribute predictions are listed in Table 1. As shown in this table, there are total of 13 different features identified as significant. While 4 features contribute to only 1 attribute, e.g., mean lightness for overall lightness prediction or edge width for sharpness, other features systemically appear in the regression equations for many attributes. The overlap in the feature sets predicting different attributes, may be used to explain the empirically known observation that some attributes correlate with each other.

Table 1. Feature Contribution to Perceived Attributes.

Attribute / Feature	Con-trast	Light-ness	Sharp-ness	Color-ful-ness	Quality
Maximum lightness	+	+	+	+	+
Maximum chroma	+	+		+	+
Mean lightness to background distance	+		+		+
Mean lightness		+			
Range contrast	+	+	+	+	+
St dev of chroma		+		+	
Spatial frequency band ratio		+	+		+
Edge contrast	+				
Edge hue difference				+	
Max edge gradient		+	+		+
Mean edge gradient	+		+	+	+
Edge width			+		
Color area contrast	+		+		
R squared	0.77	0.72	0.64	0.85	0.57

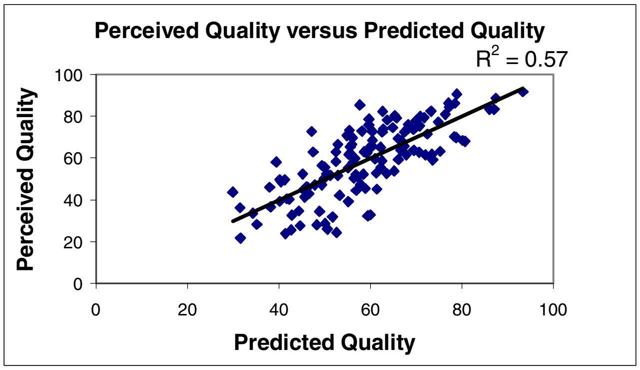

Figure 2. Perceived quality prediction.

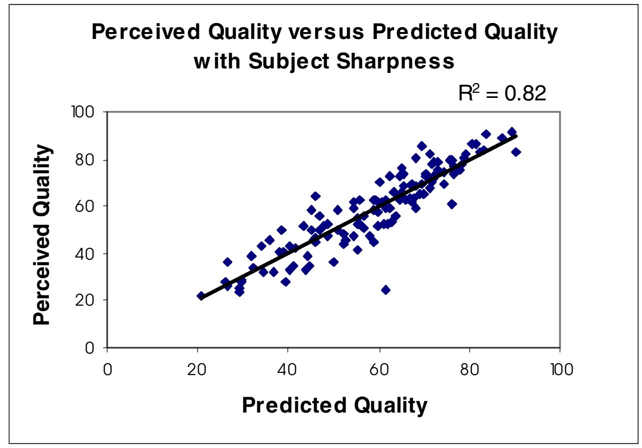

Figure 3. Perceived quality prediction improvement by adding main subject sharpness assessment.

Comparing the coefficient of determination (R^2) measure obtained for different attributes, one can notice that overall sharpness and, especially, quality turned out to be the most difficult to predict attributes: 0.64 for the sharpness prediction and 0.57 for the quality prediction, respectively. Figure 2 shows the regression plot for quality, illustrating that some images were considerably under- and over- predicted. One explanation for this finding, besides the argument about the necessity for better feature extraction methods, could lie in the apparent importance of the characteristics of the main subject rendering for assessing perceived quality and overall sharpness, one of the most critical attributes. The examination of the outliers for those predictions largely supported the assumption. Some of those images had the main subject not in focus (e.g., a close-up of a person's face), while the rest of the picture was sharp. We tested this assumption by allowing the main subject sharpness assessment obtained in the experiments to be a candidate predictor during regression modeling. The main subject sharpness information was found to considerably improve model predictions for several attributes, including sharpness, contrast, and quality. At the same time, the majority of previously identified features were still significant. For example, for overall sharpness, prediction adding main subject sharpness to a predictor list increased the R^2 value from 0.77 to 0.89. At the same time, 7 out of 8 previously determined features retained their significance, while only a maximum edge gradient dropped out. Analogous results were obtained for the quality prediction: the goodness of fit of the linear model improved to reach the R^2 value of 0.82 (Fig. 3), and, yet, 5 out of 7 initially selected features were still present in the resulting predictive combination. This indicates that the main subject sharpness assessment contained important and unique information, which was not directly extracted from the list of computed features designed to represent an entire image.

In the case of quality, the contribution of the main subject sharpness assessment was very substantial: this measure alone accounted for almost 70 percent of the variance. The latter observation points out that the sharpness of the main subject is a prevalent attribute determining quality, and the algorithmic identification of the main subject may be necessary for more precise prediction, especially when parts of an image are out of focus or have other types of problems.

It is worth noting, however, that in many instances, features that are not specifically related to the main subject, but instead, describe the entire image may work well: when an image does not have local problems, or the main subject amounts to the entire image, as in the case of a landscape, etc.

In the current experiments we did not consider higher level representations, e.g., object and category levels, and constraints imposed on image quality. We know, however, from the existing studies, that those characteristics are very important constituents of image quality and need to be included in a complete model. Incorporating the knowledge about higher level visual processing, image quality constraints, as well as further computational refinement of features at the lower levels, could be a next step for this research.

Conclusions

The results of the experiments demonstrate that the computational vision approach appears to be a promising paradigm to further advance the development of image quality modeling.

Perceived image quality and image attributes can be described using a combination of features computationally extracted from the image data.

References

1. D.L. MacAdam, Proc. Institute of Radio Engineers, 36,468 (1951).
2. C.J. Bartleson, J.Photogr.Sci,, 30, 33 (1982).
3. J. Roufs, Philips Journal of Research, 47, 35 (1992).
4. B.W. Keelan, Predicting Multivariate Image Quality from Individual Perceptual Attributes, Proc. PICS, pg.82 (2002).
5. S. Daly, The Visible Differences Predictor: an Algorithm for the Assessment of Image Fidelity, in A. Watson (ed.), Digital Images and Human Vision, MIT Press, Cambridge MA, 1993, pg. 179.
6. J. Lubin, A visual discrimination model for imaging system design and evaluation, in Peli, E. (ed.), Vision Models for Target Detection, World Scientific, Singapore, 1995, pg. 245.
7. A. B. Watson and J. Malo, Video quality measures based on the standard spatial observer, Proc. ICIP, IEEE, pg. III-41 (2002).
8. J. Yang, Proc. IS&T/SPIE, 2002.
9. H. de Ridder, J. Imaging Sci. and Technol., 40, 487 (1996).
10. E.A. Fedorovskaya, H. de Ridder, and F.J.J. Blommaert, Color Res. Appl., 22, 96 (1997).
11. S.N. Yendrikhovskij, F.J.J. Blommaert, and H. de Ridder, Towards perceptually optimal colour reproduction of natural scenes, in L.W. MacDonald and R. Luo (ed.), Colour Imaging: Vision and Technology, 1999, pg. 363.
12. R. Janssen, Computational Image Quality, SPIE Press, Bellingham, Washington USA, 2001.
13. P.G. Engeldrum, Image Quality Modeling: Where Are We?, Proc. PICS, pg. 251 (1999).
14. H. G. Barrow, and J. M. Tenenbaum, Computational approaches to vision, in: K. R. Boff, L Kaufman, and J.P. Thomas, (eds.) Handbook of Perception and Human Performance. Volume II. Cognitive Processes and Performance. John Wiley and Sons, New York, 1986.
15. D. Marr, Vision, San Francisco: Freeman, 1982.
16. S. Palmer, Vision Science: Photons to Phenomenology. The MIT Press, Cambridge, Massachusetts, 1999.
17. E.A. Fedorovskaya, Perceived overall contrast and quality of the tone scale rendering for natural images, Proc. IS&T/SPIE, 2002.
18. J. Luo, R.T. Gray, and H.-C. Lee, Towards physics-based segmentation of photographic color images, Proc. IEEE Int. Conf. Image Process., 1997.

Biography

Elena A. Fedorovskaya received a M.S. in psychology, Ph.D. degree in psychophysiology and a M.S. degree in Applied Mathematics, all from Moscow State University (Russia). She worked as a senior research scientist at Moscow State University. From 1992 to 1993 she was a visiting research fellow at the IPO (the Netherlands). Since 1997 Elena works for Eastman Kodak Company in the area of image psychology studying perceptual, cognitive and emotional aspects of human experience in relation to images.