Diana Kalenova

COLOR AND SPECTRAL IMAGE ASSESSMENT USING NOVEL QUALITY AND FIDELITY TECHNIQUES

Thesis for the degree of Doctor of Science (Technology) to be presented with due permission for public examination and criticism in the Auditorium of the Student Union House at Lappeenranta University of Technology, Lappeenranta, Finland on the 12th of December, 2009, at noon.

Acta Universitatis
Lappeenrantaensis
366

Supervisor Professor Heikki Kälviäinen
Professor Pekka Toivanen
Adjunct Professor Vladimir Bochko

Laboratory of Machine Vision and Pattern Recognition
Department of Information Technology
Faculty of Technology Management
Lappeenranta University of Technology
Finland

Reviewers Associate Professor Pavel Zemcik
Department of Computer Graphics and Multimedia
Faculty of Information Technology
Brno University of Technology
Czech Republic

Dr. Birgitta Martinkauppi
InFotonics Center
University of Joensuu
Finland

Opponent Associate Professor Pavel Zemcik
Department of Computer Graphics and Multimedia
Faculty of Information Technology
Brno University of Technology
Czech Republic

Adjunct Professor Markku Hauta-Kasari
InFotonics Center
University of Joensuu
Finland

ISBN 978-952-214-857-5
ISBN 978-952-214-858-2 (PDF)
ISSN 1456-4491

Lappeenrannan teknillinen yliopisto
Digipaino 2009

Preface

The work presented in this thesis has been carried out at the Laboratory of Machine Vision and Pattern Recognition in the Department of Information Technology in the Faculty of Technology Management of Lappeenranta University of Technology, Finland, between 2003 and 2009.

I would like to express my deep gratitude to my supervisors, Professor Pekka Toivanen and Dr. Vladimir Bochko, and especially Professor Heikki Kälviäinen for their guidance, encouragement, and fruitful cooperation throughout the work.

I thank the staff of the Laboratory of Machine Vision and Pattern Recognition in the Department of Information Technology in the Faculty of Technology Management at Lappeenranta University of Technology for a very stimulating atmosphere, Professors Jussi Parkkinen and Timo Jääskeläinen, and Dr. Arto Kaarna for being great co-authors in several articles.

I gratefully acknowledge the help of the reviewers Prof. Zemcik and Dr. Martinkauppi for their criticism and valuable comments on the work done.

I also wish to thank all of my friends, especially my very dear Olga and Zhanna for their constant support, patience and help. Without these people, life would have been a dread.

My deepest affection belongs to my parents Margarita and Guennadiy for their support and help - without them this study would not have been possible.

Lappeenranta, December 2009

Diana Kalenova

Abstract

Diana Kalenova
COLOR AND SPECTRAL IMAGE ASSESSMENT USING NOVEL QUALITY AND FIDELITY TECHNIQUES
Lappeenranta, 2009
136 p.

Acta Universitatis Lappeenrantaensis 366
Diss. Lappeenranta University of Technology

ISBN 978-952-214-857-5
ISBN 978-952-214-858-2 (PDF)
ISSN 1456-4491

Keywords: image quality, image fidelity, color and spectral image analysis, perceptual image quality, digital image processing, machine vision

UDC 004.932.2 : 004.932

The ongoing development of the digital media has brought a new set of challenges with it. As images containing more than three wavelength bands, often called spectral images, are becoming a more integral part of everyday life, problems in the quality of the RGB reproduction from the spectral images have turned into an important area of research. The notion of image quality is often thought to comprise two distinctive areas - image quality itself and image fidelity, both dealing with similar questions, image quality being the degree of excellence of the image, and image fidelity the measure of the match of the image under study to the original.

In this thesis, both image fidelity and image quality are considered, with an emphasis on the influence of color and spectral image features on both. There are very few works dedicated to the quality and fidelity of spectral images. Several novel image fidelity measures were developed in this study, which include kernel similarity measures and 3D-SSIM (structural similarity index). The kernel measures incorporate the polynomial, Gaussian radial basis function (RBF) and sigmoid kernels. The 3D-SSIM is an extension of a traditional gray-scale SSIM measure developed to incorporate spectral data. The novel image quality model presented in this study is based on the assumption that the statistical parameters of the spectra of an image influence the overall appearance. The spectral image quality model comprises three parameters of quality: colorfulness, vividness and naturalness. The quality prediction is done by modeling the preference function expressed in JNDs (just noticeable difference). Both image fidelity measures and the image quality model have proven to be effective in the respective experiments.

3D	Three dimensional
BDMM	Blockwise Distortion Measure for Multispectral Images
BFD($\ell : c$)	Bradford University Color Difference Equation
CIE	Commission International d'Eclairage
CMC($\ell : c$)	Colour Measurement Committee Color Difference Equation
CNI	Color Naturalness Index
DQE	Detective Quantum Efficiency
FUN	fidelity, usefulness, naturalness
HVS	Human Visual System
IQC	Image Quality Circle
JND	Just Noticeable Difference
LCD	Leeds Colour Difference
LSF	Line Spectral Frequencies
MSE	Mean Squared Error
MOS	Mean Opinion Score
MTF	Modulation Transfer Function
OTF	Optical Transfer Function
PCA	Principal Component Analysis
PIDM	Perceptual Image Distortion Map
PSF	Point Spread Function
PSNR	Peak Signal to Noise Ratio
RBF	Radial Basis Function

SIDM	Spectral Image Distortion Map
SSIM	Structural Similarity Index
SVM	Support Vector Machines
UQI	Universal Quality Index

I. Kalenova, D., Botchko, V., Parkkinen, J., Jääskeläinen, T., "Spectral Color Appearance Modeling", *Proceedings of The Digital Photography Conference: Image Processing, Image Quality and Image Capture Systems (PICS)*, Rochester, New York, USA, May 13-16, 2003, pages 381–385.

II. Kalenova, D., Toivanen, P., Botchko, V., "Color Differences in a Spectral Space", *Proceedings of Color, Graphics, Imaging and Vision (CGIV), 2nd European Conference*, Aachen, Germany, April 5-8, 2004, pages 368–371.

III. Kalenova, D., Toivanen, P., Botchko, V., "Spectral Image Distortion Map", *Proceedings of Pattern Recognition (ICPR), 17th International Conference*, Cambridge, UK, August 23-26, 2004, pages 668–671.

IV. Kalenova, D., Toivanen, P., Bochko, V., "Color Differences in a Spectral Space", *Imaging Science and Technology*, Vol. 49, No. 4, 2005, pages 404–409.

V. Kalenova, D., Toivanen, P., Bochko, V., "Probabilistic Spectral Image Quality Model", *Proceedings of International Color Association (AIC), 10th Congress*, Granada, Spain, May 8-13, 2005, pages 1641–1645.

VI. Kalenova, D., Toivanen, P., Bochko, V., "Preferential Spectral Image Quality Model", *Proceedings of Image Analysis (SCIA), 14th Scandinavian Conference*, Joensuu, Finland, June 19-22, 2005, pages 389–398.

VII. Kalenova, D., Dochev, D., Bochko, V., Toivanen, P., Kaarna, A., "A Novel Technique of Spectral Image Quality Assessment Based on Structural Similarity Measure", *Proceedings of Color, Graphics, Imaging and Vision (CGIV), 3rd European Conference*, Leeds, UK, June 19-22, 2006, pages 499–502.

In this thesis these publications are referred to as *Publication I*, *Publication II*, *Publication III*, *Publication IV*, *Publication V*, *Publication VI* and *Publication VII*.

1 Introduction **11**
1.1 Background . . . 11
1.2 Research Problem . . . 14
1.3 Overview and Aims of the Thesis . . . 14
1.4 Summary of Publications . . . 15

2 Image Fidelity **17**
2.1 Pixel Difference Based Measures . . . 19
2.1.1 CIELAB Based Measures . . . 19
2.2 Measures Accounting for Image Structure . . . 23
2.2.1 Structural Similarity Index . . . 24
2.2.2 Blockwise Distortion Measure for Multispectral Images . . . 27
2.3 Correlation Based Measures . . . 29
2.3.1 Conventional Color Similarity Measures . . . 29
2.3.2 Kernel Similarity Measures . . . 31
2.4 Summary and Discussion . . . 32

3 Image Quality **33**
3.1 Observer Perception Attributes . . . 34
3.2 Image Quality Models . . . 35
3.2.1 FUN Model . . . 36
3.2.2 Computational Image Quality Model . . . 36
3.2.3 Kayargadde Image Quality Theory . . . 37
3.2.4 Spectral Color Appearance Model . . . 38
3.3 Summary and Discussion . . . 43

4 Subjective Image Quality and Fidelity Evaluation **45**
4.1 Direct Scaling . . . 46
4.2 Threshold Method . . . 46
4.3 Pairwise Comparison . . . 46
4.4 Psychophysical Image Quality Measurement Standard . . . 47
4.4.1 Quality Ruler Method . . . 48
4.5 Perceptual Image Distortion Map . . . 48
4.6 Summary and Discussion . . . 49

5 Experiments **51**
5.1 Spectral Databases Used in This Thesis . . . 51
5.2 Image Quality Experiments . . . 53
5.3 Image Fidelity Experiments . . . 61
5.3.1 Image Fidelity Experiments Using Kernel Similarity Measures . . . 61
5.3.2 Image Fidelity Evaluation Using 3D-SSIM . . . 66
5.4 Summary and Discussion . . . 67

6 Discussion **73**

Bibliography **77**

A Publications **85**

Chapter I

Introduction

The nature and the scope of imaging have been undergoing dramatic changes in the recent years. The latest trend is the appearance of multiprimary displays and printers that reproduce images with a closer spectral match to the original scene captured [46, 86]. The appearance of such devices gives rise to a problem already existing for conventional tools - the assessment of perceptual image quality given only the physical parameters of the image. The demand for a quantitative analysis of image quality has dramatically increased.

1.1 Background

The invention of the first optical instruments, the optical telescope and microscope (1600-1620), created the notion of "the quality of the image" in science. This concept reoccured with an application to photography in 1860-1930, was further developed with the appearance of cinematography and television in 1935-1955, and keeps developing today [32]. Image quality research has traditionally been associated with the detection of visible distortions in an image, such as blockiness, noise or any other artefacts introduced by imaging or transmitting systems [25]. This approach led to a common misconception of the term "image quality" and that, in turn, brought about the problem of the definition of this notion [94]. Another source of ambiguities is the fact that the term image quality is an intuitive concept that has vague boundaries closely connected with human perception [55]. In order to be able to give a definition for this notion, it is necessary to look into its constituents. First of all, it is necessary to define what is meant by quality, what images are, what the end users are, and hence what the requirements imposed upon image displays are.

The main areas of application of image quality research lie in the design of imaging chains and their components, those that can further be used in medical, forensic and many other scientific and industrial applications [17]. At the end of every imaging chain, there usually is a human observer who makes a judgment of whether the image received is of quality good enough for the intended application [31]. Having human observers as

end-users of image quality assessment sets significant restrictions. A number of visual effects imposed by the human visual system should be taken into account.

One more constraint on the term image quality is connected with the part of the imaging chain to which the quality of the images is connected, i.e. image acquisition, recording, transmission and reproduction systems. Each of these parts puts forward specific demands, requirements and tasks associated with them [76].

Considering all of the above, we can move on to the definition of the term image quality itself. According to Merriam-Webster's dictionary, the notion of quality can be defined as [11]:

- a peculiar and essential character;
- an inherent feature;
- the degree of excellence;
- a distinguishing attribute;
- the attribute of an elementary sensation that makes it fundamentally unlike any other sensation.

Thus, the degree of excellence of the image can be put as the basis of the working definition of image quality. To limit the scope of this work, the assumption has to be made that by an image in this work we assume the digital form of images. A **gray-level image** in general, as defined by Gonzalez and Woods [45], is a two-dimensional function $f(x, y)$, where the values of the function are gray-level intensities of the image. **Color images**, e.g. RGB, CIELAB or CIELUV color images consist of three planes representing red, green and blue color intensity values, respectively, for RGB images or luminance and two chrominance color components [47, 95]. Thus, the color image is a three-dimensional matrix, with the third being the color or so called spectral dimension. **Spectral images**, being a particular class of images, are also presented as three dimensional matrices. The only difference is that the third dimension can contain more than three components [21]. At the same time, a digital image, be it gray-level or spectral, can be considered from several points of view. On the one hand, they are two or three-dimensional, in the case of spectral images, signals or functions; on the other hand, images are carriers of visual information. These two viewpoints produce two different definitions and consequently two approaches to image quality: image fidelity and image quality itself. Furthermore, visuo-cognitive processing is an essential stage of human interaction with the environment. Thus, image quality should be described in terms of adequacy of the image for the given task rather than of the visibility of distortions [56]. Given all of the above, a definition given by Brian Keelan [64] presents the most adequate interpretation of the term "image quality" for our research:

Image quality - "an impression of [image] merit or excellence, as perceived by an observer neither associated with the act of ... [acquisition], nor closely involved with the subject matter depicted" [64].

A few words should be said about the choice of the observer. As stated in the definition of image quality, the observer should be neither the one taking the image nor the one being the subject of the image. Both of the parties are closely connected with the image content and have requirements that are not readily quantifiable and hard to model. For photographers, such attributes would include lighting quality or composition, and for the subject of the image that would be e.g. "preserving a cherished memory or conveying a subject's essence" [64]. Thus, research concentrating on the field of image quality should limit the observers participating in the experiments [26, 64].

Image fidelity is a complementary notion to image quality. It is important not to confuse these terms [12, 41, 89] since these notions are related, but are often negatively correlated. The two are sometimes used interchangeably. Image fidelity can be defined as the visibility of a distortion or information loss, while image quality refers primarily to a degree of preference of the image [89]. The term image quality is harder to quantify and predict since the perception of the quality of the image is a multidimensional sensation, meaning that the judgment depends on a number of factors in complex combinations [38]. Images with significant distortions can often be perceived as having higher quality. Engeldrum [32] distinguishes image quality and image impairment approaches to image appearance evaluation. The quality approach, according to Engeldrum, models the judgment of image quality directly, independent of the reference, while the impairment approach quantifies the degradation of the quality from a certain ideal or point of reference. According to Farell [38], image fidelity can be expressed in terms of the probability of detection of a distortion, often called threshold judgments. Image quality, in turn, is quantified by suprathreshold judgments - preference or rank ordering. These two dissimilar approaches originate from different areas of application of the resulting images. On the one hand, television, image compression, and optics have as the origin an "original" or "ideal" image, which is impossible to capture or transmit due to some system limitations. An alternative viewpoint comes from photography and digital imaging systems, where the original image does not exist, and only a mental image can serve as such [32]. The most general definition of the term is given by Yendrikhovskij [97]:

Image fidelity - "the degree of apparent match of the reproduced image with the external reference, i.e. original" [97].

The term "original" here comprises a number of media: a natural scene, an image from another device (display, printer, projector, etc.) or any other unprocessed image [97]. From the historical point of view, the concept of a reference or original image, widely used in image fidelity research, first appeared in the work of George B. Airy in 1834, when he formulated a diffraction pattern of a clear circular aperture - the "Airy disk". This physical limit became the measure of ultimate image quality [32]. Optical Image Quality deviations appeared in 1902 in the famous work by K. Stehl - the Stehl intensity ratio [68], first defined as an "image quality measure" [32, 68]. These measures continued with "image fidelity" defined as the mean-squared-error difference, "relative structural content" and "correlation quality" [68] that lay the groundwork for future research efforts [32].

A natural classification of both image quality and image fidelity research originates from the degree of involvement of the observer in the assessment process. Based upon this

attribute, image quality research can be divided into subjective and objective methods [53].

Objective [94] image assessment methods concentrate on finding quantitative measures that can automatically predict perceived image quality and image fidelity. Techniques that belong to this area of research are based upon some physical measures or some image characteristics that somehow describe the appearance of the image [32].

Subjective methods of image quality and fidelity [87] evaluation require performing human visual test assessments that would yield the evaluation of the image. One of the most common examples of such tests is the mean opinion score (MOS). A human observer is used in such kind of research as the measurement instrument [32].

The two approaches form different kinds of tasks as the origin: subjective assessment methods attempt *to measure* the quality or fidelity, while objective methods seek *to predict* these, i.e. to estimate the overall impression of a given image based upon its inherent features [55].

1.2 Research Problem

In this work, the problem of the quality and fidelity of color and spectral images is considered. There are very few works dedicated to spectral image assessment [15, 62, 61]. This thesis deals primarily with issues of image quality and image fidelity in the spectral domain, ignoring specific spatial distortions. Combining both spatial and spectral approaches is the topic of future investigation. It has been touched upon in *Publication VII* and thus will not be considered in detail in this thesis.

Image quality research began in an attempt to find a relationship between spectral image appearance and the statistical characteristics of the image under study. The main question that was asked at the time was whether it is possible to predict and affect image appearance and consequently image quality using statistical image attributes, and what the preference function is in this case. This idea later evolved into the Spectral Appearance Model that would allow the assessment and prediction of spectral image quality.

Image fidelity research originated in a study of the possibility of applying conventional color image fidelity measures to spectral images, and led to the development of the novel kernel similarity measures. The main research question was to find image fidelity measures that would allow the quantification of specific spectral distortions introduced into the image, and at the same time would model human perception of the fidelity of spectral images.

1.3 Overview and Aims of the Thesis

This study is dedicated to color and spectral image quality and fidelity, and the structure of the thesis reflects the main research problem of this work. This thesis is divided into six chapters. Chapter 1 introduces the research field, research problem and objectives of the thesis.

Following the introduction, Chapter 2 presents the notion of image fidelity; the most prominent measures developed in the field are given here. Novel objective image fidelity metrics developed in this thesis are also described. The measures include kernel similarity measures and 3D-SSIM.

Chapter 3 presents the second part of the research - image quality. The most significant objective image quality models existing in this area of research are described, and the novel model developed for spectral images is given on top of that.

Chapter 4 completes the theoretical basis of image quality and image fidelity description by presenting some of the works in the field of subjective image quality and image fidelity.

Chapter 5 presents the practical results of both parts of the thesis, i.e. image fidelity and image quality. Preliminary conclusions are given in this chapter. Image datasets used in the thesis are described.

Chapter 6 contains discussion, conclusions and possible future research directions. The thesis is concluded with an appendix containing the publications. An overview of the publications is given in Section 1.4.

1.4 Summary of Publications

This thesis contains seven publications: one journal article, which has been published in an international journal and six conference papers. The publications can be divided into two broad topic areas. *Publication I*, *Publication VI*, and *Publication V* are dedicated to the topic of image quality, while *Publication II*, *Publication III*, *Publication IV*, and *Publication VII* deal with image fidelity.

Publication I introduces a Spectral Color Appearance model based on the statistical image model that sets a relationship between the parameters of the spectral and color images, and the overall appearance of the image. A set of tests on the capability of the model to evaluate image quality and predict observer judgments is included. The author of this thesis developed the model based on Vladimir Botchko's idea, performed the experiments, and wrote the article.

Publication II introduces a set of color similarity metrics in a spectral space. These are based on a popular pattern recognition technique - kernels. Three measures are proposed: the polynomial, Gaussian radial basis function (RBF) and sigmoid kernels. These are tested against a Munsell Matte spectral dataset [3], and compared with twelve conventional measures from [49]. *Publication II* is based on ideas of the author and Pekka Toivanen, the implementation and experiments were performed by the author. The author of the thesis was also the principal author of the publication.

The measures proposed in *Publication II* are used in *Publication III* to create a Spectral Image Distortion Map - a technique of spectral image fidelity evaluation. As a result, a gray-scale spectral distortion image is obtained, where the intensity of each of the pixels is a difference between the original image and the distorted one. The author proposed the idea, developed the algorithm, performed the experiments and was the principal author of the publication.

Publication IV is an extended journal version of *Publication II*, and most of the previously presented material is repeated. The color similarity measures and the experiments are

described in more detail. In addition, results of the tests on the performance of the measures against a database of metameric colors are included. The author developed and implemented the metrics, and wrote the article.

Publication V is an extension of *Publication I*, where the Spectral Color Appearance Model is further tested in a set of paired comparison experiments. The experiments are performed in a manner yielding assessments calibrated in JNDs of overall quality. A mean quality loss function is presented over all of the scenes and observers. The author of the thesis performed the experiments, gathered and analyzed the results, and wrote the article.

Publication VI is an extension of the previously published *Publication I*, where the Spectral Color Appearance Model is improved by adding the parameter Naturalness. A set of subjective tests on the performance of the model as a whole is also included. The author produced the idea, created the tests and gathered expert data, generalized the results obtained, and wrote the article.

Publication VII introduces an extension of the conventional gray-scale image based technique SSIM (structural similarity index) [94]. The novel 3D-SSIM is used as a spectral image fidelity measure. The performance of the 3D-SSIM is compared with the measures proposed in *Publication II* and a subjective quality measure - the Perceptual Image Distortion Map [101]. The author performed the tests, generalized the ideas and wrote the article.

Chapter II

Image Fidelity

Image fidelity according to Yendrikhovskij [97] can be defined as the degree of an apparent match of the image under study with an external reference also called the original. Thus image fidelity measures can be considered as perceptual error measures. The original can be a natural scene, an image from another device (display, printer, projector, etc.) or any other unprocessed image [97]. By image in this chapter we assume three-dimensional signals or functions [45]. The purpose of the fidelity assessment, in turn, is improvement of the images to be used by human observers.

Several assumptions, which rule out significant ambiguities, have to be made before engaging in image fidelity research. According to de Ridder [25] one of the most important implicit considerations is that the original image is always the one with the higher degree of quality and consequently preference. Another restriction is that image fidelity can not be determined without a direct comparison between the original and the processed images [25]. By image fidelity in this case we assume objective image fidelity methods, i.e. qualitative measures that can automatically predict perceived image fidelity [32].

The main task of any objective image fidelity research is to find a measure that would allow modeling of the human visual system (HVS) response in the task of image fidelity evaluation, based on some physical or statistical characteristics of the images under study [54]. The wide range of existing measures vary in the number and type of physical image parameters and statistical image characteristics that constitute them.

Many of the physical attributes used in image fidelity research originated from analogue systems. Some of the most popular physical parameters are given in Table 2.1 [54]. The first column lists some of the main attributes assessed, and the second column presents consequent physical measures that give the estimate of image attributes.

Here PSF stands for the Point Spread Function, OTF for the Optical Transfer Function, LSF for the Line Spectral Frequencies, MTF for the Modulation Transfer Function, and DQE for the Detective Quantum Efficiency [54].

Although the list is far from being complete, measures based upon these are widely used in research and industrial applications.

Table 2.1: Physical parameters of image quality [54]

Attribute	Physical measure
Color	Spectral data, chromaticity diagrams, color spaces
Tone (contrast)	Density, pixel value, characteristic curve, tone reproduction curve, gamma, histogram
Resolution (detail)	Resolving power, l/mm, dpi, ppi
Sharpness (edges)	Acutance, PSF, OTF, LSF, MTF
Noise	Granularity, noise power (Wiener) spectra, autocorrelation function
Information	Entropy, information capacity, DQE

In this thesis, we deal primarily with color and spectral image fidelity measures, and measures that can be applied to both gray-scale and color images. According to Hild [48], the criteria for a successful color fidelity measure are as follows:

- the similarity measure should account for the perceptual differences in the color attributes in a balanced way;
- the functional relation between a single color attribute and a similarity measure should be one-to-one;
- the functional relation between the values of the fidelity measure and the color features of the image should be monotonous;
- the whole range of color characteristics of the image should map into the full range of the similarity measure;
- there should be a possibility of adjusting the sensitivity of the measures.

The measures considered in this thesis comply with the principles mentioned above, and moreover, these were put in the basis of the development process of the image fidelity research performed.

At the moment, there are various classification methods for image fidelity measures. One of the most comprehensive overviews of these metrics is presented by Avcibas [16]. The categories proposed are [16]:

- pixel difference-based measures (such as the mean squared error and the formulae derived from it);
- correlation-based measures, i.e. correlation between pixels and vector angular directions (e.g. mean angle similarity, normalized cross-correlation [14]);

- edge-based measures, edge displacements and precision (e.g. the Pratt edge measure [84], edge stability [22]);
- spectral distance-based measures, i.e. Fourier magnitude or phase spectral discrepancy (e.g. the spectral phase error, the block spectral magnitude error [69, 77]);
- context-based measures, i.e. distortion measures based on multidimensional context probability (e.g. the Hellinger distance, the rate distortion measure [83]);
- HVS-based measures, i.e. HVS-weighed spectral distortion measures or browsing similarity (e.g. the HVS absolute norm, DCTune [42]).

This categorization was chosen as the basis of the classification of the measures given in this thesis.

2.1 Pixel Difference Based Measures

According to Avcibas [16], these measures calculate the distortions between the images on the basis of the pixelwise differences or the moments of the error images.

One of the most widely used image fidelity measures that falls into this category is the mean squared error (MSE), computed as a mean of the squared difference between the original and modified images [55].

$$MSE(I, I') = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} |I'(i,j) - I(i,j)|^2, \tag{2.1}$$

where $I(i,j)$ and $I'(i,j)$ are luminances of the original and distorted images I and I' of the pixels with coordinates i and j, respectively, and the image size is $M \times N$. The MSE is abbreviated as $MSE(I, I')$ for simplicity, with i and j omitted.

An extension of the MSE often used in literature is the Peak Signal-to-Noise Ratio (PSNR). The PSNR is, in fact, a normalized version of the MSE [55].

$$PSNR = 10 \log_{10} \frac{R^2}{MSE}, \tag{2.2}$$

where R is the luminance range of the display media. These measures and the formulae derived from them have the benefit of simplicity in understanding and realization. However, these have a weak correlation with the perceived visual quality [44].

2.1.1 CIELAB Based Measures

CIE ΔE L*a*b* [4] was recommended by the Commission International d'Eclairage (CIE) in 1976. L*a*b* is a perceptually uniform color space, where L* is the lightness scale, a* is the red-green scale, and b* is yellow-blue. CIE ΔE L*a*b* computes a Euclidean distance between corresponding pixels in the original and modified images in the L*a*b* color space [4]:

$$\Delta E = \sqrt{(\Delta L*)^2 + (\Delta a*)^2 + (\Delta b*)^2}. \tag{2.3}$$

One ΔE unit is equivalent to a threshold detection perceptual color difference. Calculating the dissimilarity in the L*a*b* color space ensures that equal ΔE values correspond to equal perceptual distances. However, this measure suffers the drawback of being perceptually non-uniform [9].

In an attempt to improve CIE ΔE L*a*b*, a number of measures have been introduced: CIE94 (CIE ΔE_{94} L*a*b*), the Bradford University color difference equation (BFD(ℓ : c)) [71, 72], the Colour Measurement Committee (CMC(ℓ : c)) [23] equation and the Leeds Colour Difference (LCD) [66, 78] equation [70]. The measures have proven to perform better than the conventional CIE ΔE L*a*b*. However, these are still not the perfect color differencing solution at the same time. All four can be reduced to a generic equation representing the following formulae [70]:

$$\Delta E = \sqrt{\left(\frac{\Delta L*}{k_L S_L}\right)^2 + \left(\frac{\Delta C*}{k_C S_C}\right)^2 + \left(\frac{\Delta H*}{k_H S_H}\right)^2 + \Delta R}, \tag{2.4}$$

where S_L, S_C and S_H are weighting functions for lightness, chroma and hue components. $\Delta L*$, $\Delta C*$, $\Delta H*$ are CIE L*a*b* lightness, chroma and hue differences, and S_L is equal to one in CIE94. k_L, k_C and k_H are parametric factors adjusted in accordance with the viewing conditions (texture, background, separation). ΔR is an interactive term between the chroma and hue difference expressed as [70]:

$$\Delta R = R_T f(\Delta C * \Delta H*). \tag{2.5}$$

For CMC(ℓ : c) [23] and CIE94 ΔR is equal to zero, and R_T is slightly different for LCD formula [66, 78]. The chroma is calculated, in turn, as [70]

$$C^*_{ab} = \sqrt{a*^2 + b*^2}. \tag{2.6}$$

Hue is defined as [70]

$$H^*_{ab} = \tan^{-1}(b^*/a^*). \tag{2.7}$$

Despite the seeming similarity, the equations have significant discrepancies in lightness and hue estimation [70].

Further development of the famous ΔE equation led to the emergence of the CIEDE2000 [9] equation, which would be able to account for the characteristics of the HVS in the task of image fidelity evaluation. There have been several attempts to perfect the conventional measure to be able to achieve perceptual uniformity. The final equation is formed in the following way: first L*, a*, b* and C* are computed as [70]

$$\begin{aligned} L^* &= 116 f(Y/Y_n) - 16 \\ a^* &= 500[f(X/X_n) - f(Y/Y_n)] \\ b^* &= 200[f(Y/Y_n) - f(Z/Z_n)] \\ C^*_{ab} &= \sqrt{a*^2 + b*^2}, \end{aligned} \tag{2.8}$$

where X, Y and Z are CIE 1931 XYZ tristimulus values [2] of a given color, X_n, Y_n and Z_n are the tristimulus values of the reference white point, and conversion function $f(I)$ is defined as [70]

$$f(I) = \begin{cases} I^{1/3} \; for \; I > 0.008856 \\ f(I) = 7.7871 + 16/116 \quad otherwise. \end{cases} \tag{2.9}$$

Then all of thus received L*, a*, b* and C* are rescaled to account for the human perception features as [70]

$$\begin{aligned} L' &= L^* \\ a' &= (1 + G)a^* \\ b' &= b^* \\ C' &= \sqrt{a'^2 + b'^2} \\ h' &= \tan^{-1}(b'/a'), \end{aligned} \tag{2.10}$$

where rescaling function G is defined as [70]

$$G = 0.5\left(1 - \sqrt{\frac{\overline{C*_{ab}}^7}{\overline{C*_{ab}}^7 + 25^7}}\right), \tag{2.11}$$

where $\overline{C*_{ab}}$ is the arithmetic mean of $C*_{ab}$ for a pair of samples.

$\Delta L'$, $\Delta C'$ and $\Delta H'$ are calculated then, where the first two are just differences between the values in the original and modified images, and $\Delta H'$ has a slightly different form [70]:

$$\Delta H' = 2\sqrt{C'_b C'_s}\sin\left(\frac{\Delta h'}{2}\right), \tag{2.12}$$

where $\Delta h'$ is computed similarly to $\Delta L'$, $\Delta C'$. The hue difference thus takes into account the angle between the chroma vectors in original C'_b and modified C'_s images. The overall CIEDE2000 has the following form [9, 70]:

$$\Delta E = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T\left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)}, \tag{2.13}$$

where the constituents of the equation have similar interpretations as those in Eq. (2.4). The weighting function for the lightness is expressed as [70]

$$S_L = 1 + \frac{0.015(\overline{L'} - 50)^2}{\sqrt{20 + (\overline{L'} - 50)^2}}, \tag{2.14}$$

where $\overline{L'}$ is the mean of the lightness for a pair of samples. The weighting function for chroma is defined as [70]

$$S_C = 1 + 0.045\overline{C'}, \tag{2.15}$$

where $\overline{C'}$ is the mean of the chroma for a pair of samples. The weighting function for the hue proposed in [19] in turn is defined as

$$S_H = 1 + 0.015\overline{C'}T, \tag{2.16}$$

where T is meant to account for the angle between hue vectors [70]:

$$T = 1 - 0.17\cos(\overline{h'} - 30^\circ) + 0.24\cos(2\overline{h'}) + 0.32\cos(3\overline{h'} + 6^\circ) - 0.2\cos(4\overline{h'} - 63^\circ). \tag{2.17}$$

Function R_T, which was created to be able to account for the purpose of improving the performance of the color difference equations in the blue region, is expressed as [70]

$$R_T = -\sin(2\Delta\theta)R_C, \tag{2.18}$$

where $\Delta\theta$ is given as [70]

$$\Delta\theta = 30\exp\{-[(\overline{h'} - 275^\circ)/25]^2\}. \tag{2.19}$$

The R_T function given in such a way is similar to that in LCD [66, 78] and BFD$(\ell : c)$ [71, 72] equations, with R_C being slightly modified to improve the performance of the equation on neutral colors and high chroma colors [70]:

$$R_C = 2\sqrt{\frac{\overline{C'}^7}{\overline{C'}^7 + 25^7}}. \tag{2.20}$$

These measures were designed for calculating differences on large uniform color patches. However, the sensitivity of the human eye to color variations depends on spatial patterns in the image [18, 82], ignored in CIE ΔE L*a*b* based measures.

2.2 Measures Accounting for Image Structure

As Fairchild points out [36], significant research efforts concentrate on the issues of color differences and color appearance ignoring the spatial properties of the human vision. Moreover, works concentrating on spatial distortions, in turn, neglect the influence of these artefacts on the color appearance with a few exceptions to this rule [60, 80, 94, 102].

In an attempt to create a measure that would be able to account for spatial patterns in the task of color image fidelity evaluation, a spatial extension of the conventional measure S-CIELAB was introduced [102]. The idea behind S-CIELAB is that prior to the calculation of CIE ΔE L*a*b*, spatial preprocessing is realized. The purpose of this stage is to imitate spatial blurring occurring in the human eye [102].

Given all of the above, S-CIELAB is computed as follows [102]:

1. Convert the images given into an opponent color space. Bäuml, Poirson and Wandell [18, 82] in a series of psychophysical experiments have determined the optimal transformation from the CIE 1931 XYZ [2] into three opponent color planes, consequently representing luminance, red-green and blue-yellow channels [102]:

$$\begin{aligned} O_1 &= 0.279X + 0.72Y - 0.107Z \\ O_2 &= -0.449X + 0.29Y - 0.077Z \\ O_3 &= 0.086X - 0.59Y + 0.501Z. \end{aligned} \tag{2.21}$$

   The CIE 1931 XYZ color space [2] is one of the first mathematically defined color spaces, X, Y and Z in this case are the tristimulus values of red, green and blue. This color space was built on direct measurements of the human eye and serves nowadays as the basis for many color spaces (e.g. RGB, CIE L*a*b*) [74].

2. Each of the received images is filtered with a two-dimensional spatial kernel of the following form [102]:

$$f = k\Sigma_i w_i E_i, \tag{2.22}$$

   where k is a scale factor, and w_i are weights varying for each plane separately. E_i is calculated as follows [102]:

$$E_i = k_i \exp[-(x^2 + y^2)/\sigma_i^2], \tag{2.23}$$

   where k_i scale factors are chosen in such a way that E_i sums to 1. σ_i is the spread expressed in degrees of the visual angle. Parameters w_i and σ_i vary for each of the opponent color planes separately. The values of these are presented in Table 2.2.

3. The final stage is to perform the inverse transformation of the image after filtering it into CIE XYZ color coordinates, and then into the CIE L*a*b* color space. Thus received images are subject to the CIE ΔE L*a*b* calculation.

Table 2.2: Filter parameters for S-CIELAB [102]

Plane	Weights w_i	Spreads σ_i
Luminance	0.921	0.0283
	0.105	0.133
	-0.108	4.336
Red-green	0.531	0.0392
	0.330	0.494
Blue-yellow	0.488	0.0536
	0.371	0.386

The S-CIELAB model proposed by Wandell [102] is based on the assumption that the spatial characteristics of the HVS are isotropic. As it was shown in [52], the MTF of the human eye depends on the directivity of the spatial frequency response, i.e. the sensitivity of the human eye in the vertical and horizontal directions is greater than in the diagonal direction. A two-dimensional MTF was proposed by Miyake et al. in [75]. The MTF of the human eye can be presented as [75]:

$$M(u,v) = M_0(\omega)[1 - \{1 - \gamma(\omega)\}|\sin^b 2\phi|], \tag{2.24}$$

where u and v are spatial frequencies, b is a coefficient, $M_0(\omega)$ is the MTF of the horizontal direction, and ω and ϕ are defined by the following equation

$$\begin{cases} \omega = \sqrt{u^2 + v^2} \\ \phi = \arctan u/v. \end{cases} \tag{2.25}$$

The modified S-CIELAB is then calculated as follows: the first step is similar to the conventional S-CIELAB; each of the images under consideration is transformed into the opponent color space. After that, the MTF of the human eye is calculated with Eq. (2.24), and thus received spectra are convolved. Then, the spatial frequency filtering of the original S-CIELAB is applied. The resulting difference is calculated using the Euclidean distance [75].

Since the ΔE calculation is the step not connected directly to the overall S-CIELAB calculation, some other image fidelity measure can also be used as the last step of the whole calculation.

2.2.1 Structural Similarity Index

In an attempt to create a measure that would allow us to quantify image fidelity both in the spectral and spatial domains, a 3D-SSIM measure was proposed in *Publication VII*. The measure is an extension of the previously published structural similarity index (SSIM) [94] that was created for gray-scale images. The SSIM generalizes the universal quality index (UQI) proposed by Wang [92, 93]. The idea behind the measure according to Wang [94] is that the human visual system (HVS) is highly adapted to extracting structural information from the image. Wang [94] defines the structural information as

inherent attributes of the image that characterize the structure of the objects in the scene, independent of the local luminance and contrast. Thus each of the three, i.e. luminance, contrast and structure, should be considered separately [94].

The SSIM assesses image fidelity using three previously described criteria: luminance, contrast and structure, all being relatively independent, meaning that a change in any of them does not affect the rest. The overall similarity measure should comply with the following conditions [94]:

1. Symmetry: $f(x, y) = f(y, x)$;
2. Boundedness: $f(x, y) \leq 1$;
3. Unique maximum: $f(x, y) = 1$ if and only if x=y,

where $f(x, y)$ is a general form of the separate components of the SSIM.

Given all of the above, luminance $l(x_i, x_j)$ can be defined based on the mean intensities of the signals x_i and x_j [94] as

$$l(x_i, x_j) = \frac{2\mu_{x_i}\mu_{x_j} + C_1}{\mu_{x_i}^2 + \mu_{x_j}^2 + C_1}, \tag{2.26}$$

where μ_{x_i} and μ_{x_j} are means of the input vectors x_i and x_j, respectively.

To estimate the contrast component $c(x_i, x_j)$, the mean intensity is subtracted from the signal. The contrast is then computed as the standard deviation of the resulting zero-mean signal [94]

$$c(x_i, x_j) = \frac{2\sigma_{x_i}\sigma_{x_j} + C_2}{\sigma_{x_i}^2 + \sigma_{x_j}^2 + C_2}, \tag{2.27}$$

where σ_{x_i} and σ_{x_j} are consequent standard deviations.

To be able to compute the structural component $s(x_i, x_j)$, the signal should be zero-mean and have unit standard deviation. The correlation between thus normalized vectors defines the structure comparison [94]:

$$s(x_i, x_j) = \frac{2\sigma_{x_i x_j} + C_3}{\sigma_{x_i}\sigma_{x_j} + C_3}, \tag{2.28}$$

where $\sigma_{x_i x_j}$ is the correlation coefficient of x_i and x_j, and σ_{x_i}, σ_{x_j} are taken from the contrast calculation.

Terms C_1, C_2, and C_3 introduced in Eqs. (2.26 - 2.28) are small constants, included to avoid instability when the denominators are close to zero, and defined as [94]

$$C_1 = (K_1 L)^2,\ C_2 = (K_2 L)^2 \ and\ C_3 = C_2/2, \tag{2.29}$$

where L is the dynamic range of the pixel values ($L = 255$ for 8 bits/pixel gray-scale images), and $K_1 \gg 1$ and $K_2 \gg 1$ are two scalar constants.

The SSIM is thus formed in the following way [94]:

$$SSIM(x_i, x_j) = [l(x_i, x_j)]^{\alpha}[c(x_i, x_j)]^{\beta}[s(x_i, x_j)]^{\gamma}, \tag{2.30}$$

where α, β and γ are variable positive sensitivity parameters that define the relative importance of the three components. If all of the components of the SSIM are equally important, as in this work, each of the sensitivity parameters is set to 1.

Any image fidelity measure can be computed either locally or globally relative to the image [94]. The term globally means that the measures are computed on a pixelwise basis, while locally implies that a windowing approach is used. The use of either depends on the metric used [94]. In the case of the SSIM it was proven in [93] that the index should be applied locally due to a number of reasons according to [94]:

- image statistics are spatially non-stationary;
- image distortions are often space-variant;
- at a typical viewing distance only local areas of the image can be perceived with high resolution by a standard human observer;
- localized measures result in a spatially varying fidelity map, providing more information on image impairments.

In the case of the SSIM, an 11×11 circular-symmetric Gaussian weighting function normalized to the unit sum is used, with a standard deviation of 1.5 samples. Local statistics μ_x, σ_x and $\sigma_{x_i x_j}$ are computed using this windowing approach. The overall SSIM is computed via averaging over the entire image [94].

The technique presented in *Publication VII* is an extension of the SSIM. Conventional SSIM is applied to gray-scale images, while 3D-SSIM primarily to spectral images. However, we can assume that the strong correlation between the spectra and the color reproduced from that spectrum would allow the assessment of color images computed from the spectral images. The idea was to create a measure that would allow assessment of both spatial and spectral distortions.

A similar idea was proposed for color images in [20]. To be able to adapt the SSIM to color image assessment, the SSIM was applied to the images converted into the IPT color space [37] to each of the channels separately, and the results were combined using a geometrical mean [20]. IPT is a novel uniform color space. I in this color space is lightness, the red-green dimension is denoted as P, and yellow-blue as T. The IPT color space is designed in such a way that it accurately models the constant perceived hue [37].

In the framework of this research, two possible solutions were proposed in [28]:

1. apply the conventional two-dimensional SSIM to every band and average the result over the whole image;

2. extend SSIM to three dimensions (3D-SSIM).

The first approach is a straightforward solution, while the second requires explanation. The idea behind the 3D-SSIM calculation is to apply a three-dimensional window instead of a previously mentioned two-dimensional Gaussian weighting function for the calculation of the local statistics μ_x, σ_x and $\sigma_{x_i x_j}$, used in Eqs. (2.26 - 2.28). The weighting function proposed in *Publication VII* is an $11 \times 11 \times 11$ sliding window expressed as [45]

$$h(x, y, z) = \sqrt{2\pi}\sigma A \exp(-2\pi^2\sigma^2(x^2 + y^2 + z^2)), \tag{2.31}$$

where A and σ are constants adjusting the windowing function, σ is usually chosen to be close to the standard deviation of the image, and x, y and z are the coordinates of the appropriate weight, which vary from 1 to the maximal size of the window in the appropriate direction.

Thus the algorithm of the 3D-SSIM calculation is realized as in the conventional SSIM, except for the use of the windowing function. The reason for using a three-dimensional window is the fact that adjacent spectral bands in an image are highly correlated, and localized statistics contain more information about the image structure [28].

Both of the approaches of spectral image evaluation have been tested in [28]. It was shown that the novel 3D-SSIM presented in *Publication VII* outperforms the conventional SSIM applied band by band.

2.2.2 Blockwise Distortion Measure for Multispectral Images

A spectral image fidelity measure proposed by Kaarna et al. [62, 61], a blockwise distortion measure for lossy compression of multispectral images (BDMM), is given here as one of the few works existing in the field of image fidelity of spectral images. The measure is based on a popular technique, two-dimensional blockwise distortion, modified to calculate the difference between the original spectral image and the compressed/reconstructed spectral image. The proposed measure exhibits the following properties [61]:

- BDMM takes into account both relative and absolute errors;
- comparisons between different sets of images are possible;
- independency of the coding methods;
- BDMM is based on essential characteristics of the image;
- according to [61] visual inspection of the images agrees with the results obtained by BDMM.

The modified blockwise distortion measure is computed as follows: a sliding cube of size $3 \times 3 \times 3$ is used for computation of the characteristics of the image. The contrast, spatial and spectral structure, and number of different gray-levels are computed for each pixel [61].

Contrast e_c characterizes the way each pixel differs from the background. It can be defined as a local brightness change, and can be computed using standard deviation as [90]

$$e_c = \frac{(\sigma^o - \sigma^{cr})^2}{\max(1, \sigma^o)}, \tag{2.32}$$

where σ^o is the standard deviation of the original image and σ^{cr} is the standard deviation of the compressed/reconstructed image. Normalization by σ^o is done in order to account for the effect of the human eye being more sensitive to changes in low contrast regions than in high contrast regions.

The spatial structure e_s accounts for blurring and jaggedness artefacts in an image, and it is the response to edge detection operation in a block. The spatial structure error is computed as a sum of all edge-detection operations normalized by the contrast value of the block [90].

$$e_s = \frac{(|G_x^o - G_x^{cr}| + |G_y^o - G_y^{cr}| + |G_z^o - G_z^{cr}|)}{3\max(1, \sigma^o)}, \tag{2.33}$$

where G_x, G_y, G_z are Laplacian edge-detection filters extended to a three-dimensional case for the original and the compressed/reconstructed images.

Quantization error e_q gives an estimate of the blockiness in the image. It is defined as the number of different gray-levels Q in the block in an image [90]

$$e_q = (Q^o - Q^{cr})^2, \tag{2.34}$$

where Q^o is the number of different gray-levels in a block in the original image, and Q^{cr} in the reconstructed/compressed image.

The overall BDMM is obtained by averaging each of the error components over the whole image [61]:

$$E_c = \frac{1}{s}\sum_{i=1}^{S} e_{ci},\ E_s = \frac{1}{s}\sum_{i=1}^{S} e_{si},\ E_q = \frac{1}{s}\sum_{i=1}^{S} e_{qi}, \tag{2.35}$$

where $S = N^2M$, with N^2 being the number of pixels in the image and M the number of bands in the image.

The overall BDMM is computed as follows [61]:

$$E_{lin} = w_c f(E_c) + w_s f(E_s) + w_q f(E_q), \tag{2.36}$$

where w_c, w_s and w_q are consequent importance weights, and f is a scaling function defined as

$$f(E_i) = 1 - \min(1, \frac{E_i}{k_i}),\ i = c, s, g, \tag{2.37}$$

where k_i is a scaling coefficient that scales BDMM to the range $[0, 1]$, and at the same time serves as a threshold for $E_i > k_i$, $f(E_i) = 0$ [61].

2.3 Correlation Based Measures

An alternative approach is presented by correlation based measures. It was shown in [35] that image fidelity can be quantified in terms of the correlation function. These measures incorporate correlation between pixels and vector angular directions [16]. These measures estimate the similarity between two images under study, and therefore in this sense they are complementary to the difference-based measures [15].

2.3.1 Conventional Color Similarity Measures

Twelve conventional color similarity measures were proposed in color-difference literature [67, 81] and tested in [48, 49]. These measures can be attributed to the category of the correlation based measures. All of the measures take two p-dimensional input color vectors x_i and x_j varying in the value range $[0, 1]$, and compute a real number in the range $[0, 1]$ as an output. The measures seek to account for the angle between the color vectors, which in itself bears important information [48, 49]. The measures presented in [67, 81] are as follows:

Metric 1

$$S_1 = \frac{x_i x_j^t}{|x_i||x_j|} = \cos\theta, \tag{2.38}$$

where θ is the angle between vectors x_i and x_j.

Metric 2

$$S_2 = \left(\frac{x_i x_j^t}{|x_i||x_j|}\right)\left(1 - \frac{\big||x_i| - |x_j|\big|}{\max\big(|x_i|, |x_j|\big)}\right). \tag{2.39}$$

Metric 3

$$S_3 = \frac{|x_i|\cos\theta + |x_j|\cos\theta}{\big(|x_i|^2 + |x_j|^2 + 2|x_i||x_j|\cos\theta\big)^{1/2}}. \tag{2.40}$$

Metric 4

$$S_4 = \frac{\big(|x_i|^2 + |x_j|^2 + 2|x_i||x_j|\cos\theta\big)^{1/2}}{|x_i| + |x_j|}. \tag{2.41}$$

Metric 5

$$S_5 = \frac{\big(|x_i|^2 + |x_j|^2 - 2|x_i||x_j|\cos\theta\big)^{1/2}}{\big(|x_i|^2 + |x_j|^2 + 2|x_i||x_j|\cos\theta\big)^{1/2}}. \tag{2.42}$$

Metric 6. Correlation coefficient method

$$S_6 = \frac{\sum_{k=1}^{p} |x_{ik} - \bar{x}_i||x_{jk} - \bar{x}_j|}{\left(\sum_{k=1}^{p}(x_{ik} - \bar{x}_i)^2\right)^{1/2}\left(\sum_{k=1}^{p}(x_{jk} - \bar{x}_j)^2\right)^{1/2}}, \tag{2.43}$$

where $\bar{x}_i = \frac{1}{p}\sum_{k=1}^{p} x_{ik}$

Metric 7. Exponential similarity method

$$S_7 = \frac{1}{p}\sum_{k=1}^{p} \exp\left(-\frac{3}{4}\frac{(x_{ik} - x_{jk})^2}{\beta_k^2}\right), \tag{2.44}$$

where $\beta_k^2 > 0$ is determined experimentally.

Metric 8. Absolute-value exponent method

$$S_8 = \exp\left(-\beta\sum_{k=1}^{p} |x_{ik} - x_{jk}|\right), \tag{2.45}$$

where $\beta > 0$

Metric 9. Absolute-value reciprocal method

$$S_9 = 1 - \beta\sum_{k=1}^{p} |x_{ik} - x_{jk}|, \tag{2.46}$$

where β is determined empirically.

Metric 10. Maximum-minimum method

$$S_{10} = \frac{\sum_{k=1}^{p} \min(x_{ik}, x_{jk})}{\sum_{k=1}^{p} \max(x_{ik}, x_{jk})}. \tag{2.47}$$

Metric 11. Arithmetic-mean minimum method

$$S_{11} = \frac{\sum_{k=1}^{p} \min(x_{ik}, x_{jk})}{\frac{1}{2}\sum_{k=1}^{p}(x_{ik} + x_{jk})}. \tag{2.48}$$

Metric 12. Geometric-mean minimum method

$$S_{12} = \frac{\sum_{k=1}^{p} \min(x_{ik}, x_{jk})}{\sum_{k=1}^{p}(x_{ik}x_{jk})^{1/2}}. \tag{2.49}$$

These measures were designed for the purpose of estimating the image fidelity of RGB and L*a*b* images, but they can also be applied to spectral images. It was shown in [48] that the Exponential similarity method and the Absolute-value exponent method have shown the most promising results for color images.

2.3.2 Kernel Similarity Measures

A novel image fidelity approach was proposed in *Publication II* and *Publication IV*. The measures presented are based upon a well-known pattern recognition technique - kernel support vector machines (SVM) [88]. The ideas lying in the basis of kernel machines were proposed by Vapnik in [91] and later evolved into a broad field. SVMs incorporate supervised learning methods for classification and regression of data [88]. A kernel in this case can be considered an extension of a conventional dot product computed in a certain feature space. Thus SVMs map the input data into a higher dimensional space, where the data can be separated more easily [88].

The measures proposed in *Publication II* and *Publication IV* include the Gaussian radial basis function, polynomial and sigmoid kernels.

The polynomial kernel color similarity measure, modified to account for the nature of human perception, can be presented as follows [88]:

$$S_{polynomial} = (y_i, y_j)^d, \tag{2.50}$$

where henceforth $y_{i,j} = RVx_{i,j}$, d is a variable sensitivity parameter, x_i and x_j are input p-dimensional spectral vectors, V is a spectral luminous efficiency function for photopic vision [6], and R is the spectral radiance function of a certain light source [5]. The similarity functions have a general form of $S(x_i, x_j)$, the arguments are omitted for simplicity in the formulae shown in this work; (,) in the formulae given here is a notation for a dot product between vectors [88].

The sigmoid kernel based similarity metric can be presented as follows [88]:

$$S_{sigmoid} = \tanh(\langle y_i, y_j \rangle k + \vartheta), \tag{2.51}$$

where k and ϑ are variable parameters, and the dot product is denoted as $\langle$, $\rangle$ for simplicity of notation.

The Gaussian radial basis function kernel then has the following form [88]:

$$S_{Gaussian} = \exp\left(- \frac{\|y_i - y_j\|^2}{2\sigma^2} \right), \tag{2.52}$$

where $\sigma > 0$, σ is a parameter of the sensitivity of the function.

This approach was further developed in *Publication III*. A Spectral Image Distortion Map (SIDM) was proposed. What is calculated in the SIDM, in fact, is a pixelwise spectral distortion. Kernel similarity measures Eqs. (2.50 - 2.52) are applied to calculate a global image fidelity measure. As a result a grayscale spectral distortion image is obtained, where the intensity of each of the pixels is a difference between the original image and the distorted one obtained via kernel similarity measures on a $[0, 1]$ scale, from "not similar at all" to "identical". From the point of view of probability theory, the SIDM represents a map of probabilities of the subjects identifying a certain pixel as similar.

2.4 Summary and Discussion

In this chapter, we considered image fidelity, which can be defined as the degree of an apparent match of the image under study with an external reference also called the original [97]. The most prominent measures developed in the field and the novel objective image fidelity metrics developed in this thesis were presented. These primarily include color and spectral image fidelity measures, and measures that can be applied to both gray-scale and color images. All of the metrics have been divided into three broad categories based on the classification proposed by Avcibas [16]: pixel difference based measures, measures accounting for image structure and correlation based measures.

Pixel based measures included some of the most popular ones existing in the field: the MSE, PSNR, CIE ΔE L*a*b*, and extensions of these measures BFD($\ell : c$) [71, 72], CMC($\ell : c$) [23], LCD [66, 78], CIE94 ΔE and the most recent CIEDE2000 [70].

Measures accounting for image structure included an extension of the conventional CIE ΔE L*a*b* - S-CIELAB [102], a measure created for the assessment of spectral images - BDMM [61], the SSIM, and an extension proposed in this thesis for the 3D-SSIM proposed in *Publication VII*.

Correlation based measures included novel kernel similarity measures proposed in *Publication II*, *Publication IV* and further developed into a Spectral Image Distortion Map (SIDM) in *Publication III*. In addition to these, twelve conventional image fidelity measures proposed in [67, 81] were presented.

Chapter III

Image Quality

Image quality is defined in this thesis as "an impression of [image] merit or excellence, as perceived by an observer neither associated with the act of ... [acquisition], nor closely involved with the subject matter depicted" [64]. Note that by image quality here we assume objective image quality. An objective measure can be defined as a quantitative measure correlated with the perceived attributes of image appearance, and accounting for the viewing conditions and the properties of the HVS [64]. By an image in image quality assessment we assume a carrier of the visual information, and the end user, thus a human observer. Moreover, it should be emphasized that in this and the following chapter we are dealing with color images reproduced from spectral images, and the effect of the spectral image statistics on the color image appearance.

The Image Quality Circle (IQC) [30] proposed by Engeldrum and presented in Figure 3.1 suggests a framework for image quality research in general. The main problem considered in this area of studies is how to relate the technology variables and customer quality preference. Technology variables according to Engeldrum [32] are the inherent elements or qualitative parameters of the imaging system (whatever the display media be), which are manipulated in order to change image quality, e.g. the toner size, resolution or paper parameters. Customer quality preference in this case is the overall image quality impression that can be quantified in a number of ways (from 0 to 1, or "bad" to "excellent", etc.) [32].

These two parts of IQC (technology variables and customer quality preference) can be directly connected with subjective image quality methods. However, such methods are time and human resource consuming [32] (these methods are the subject of Chapter 4 of this thesis). On the other hand, not suffering from this drawback, objective image quality given as the second half of the circle can be divided into two main constituents: physical image parameters and customer perceptions. Physical image parameters are quantitative functions or parameters used in image quality description, such as Wiener spectra, MTF or saturation. They are connected to technology variables via system models that comprise analytical models predicting the relationship between the two. At the same time, customer perceptions, or as Engeldrum [32] defines them, the "nesses",

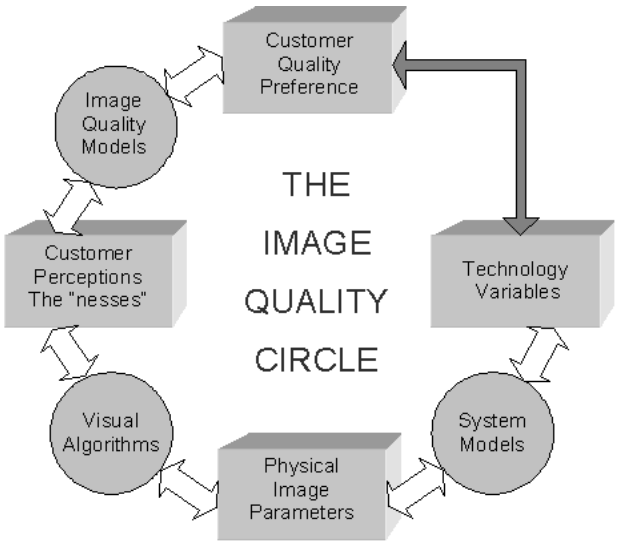


Figure 3.1: Image Quality Circle [32]

emphasizing the perceptual character of these as opposed to the physical, are perceptual attributes of image quality, such as sharpness, graininess or colorfulness. Customer perceptions are connected to customer quality preference via mathematical functions describing the tradeoff that the observer makes when judging the image. The "nesses" are varied through visual algorithms relating the physical image parameters to customer perceptions [32]. One of the main questions facing image quality research is the choice of customer perception attributes - the "nesses" - and image quality models.

This part of the thesis concentrates on the observer perception attributes or "nesses" and image quality models connecting these to customer quality preference.

3.1 Observer Perception Attributes

There is a limited number of observer perception attributes or "nesses" appearing in literature, most of which are bound to the ones originating from applications, e.g. photography, image compression, image coding, etc. [32].

Observer perception attributes, or as they are often called in literature, image quality attributes, can be categorized in a number of ways. One of the possible classifications was proposed by Keelan [64]:

Artefactual image quality attributes deal with artefacts introduced by various parts of the imaging chain. The appearance of such artefacts always leads to the degradation of image quality. Such attributes include blockiness, graininess, redeye, and various other digital artefacts. Such attributes have a certain detection threshold beyond which the quality of the reproduction decreases.

Preferential image attributes are nearly always visible in the image. With the increase of the value of the attributes, quality also increases up to a certain optimum point, and further change of the attributes leads to the loss of quality. The optimum is observer dependent. Such attributes include, among others, color balance, contrast, colorfulness (saturation), memory color reproduction, etc.

Aesthetic attributes are more subjective in nature and thus more difficult to quantify. Examples of such attributes include the lighting quality or composition. These attributes depend on the observer. A photographer and the subject of the image can have differing views of the aesthetic nature of the image, as does an independent observer.

Personal attributes are also subjective and highly observer dependent. To clarify the contents of such attributes, the following examples can be mentioned: preserving a cherished memory, conveying a subject's essence. The difference between these and the previous category is that such attributes involve only the parties involved in the act of photography.

A significant problem connected with the choice of the "nesses" is the interdependence between the customer perception attributes, meaning that the change of one of the perceptual attributes of the image affects the others. Thus the "nesses" should be chosen in such a way that such dependencies are taken into account or are relatively independent [64].

3.2 Image Quality Models

Image quality models are designed in such a way that they would connect customer quality preference to customer perceptions [30]. Image quality models can take various forms, however. The most common one appearing in literature can be summarized as follows [33]:

$$IQ = (a_1 ness_1^p + a_2 ness_2^p)^{1/p}, \tag{3.1}$$

where IQ is the image quality judgment value, p is the exponent or the power of the model, a_1 and a_2 are predefined coefficients or weights, and $ness_1$ and $ness_2$ are the values of the customer perception attributes or "nesses" [33].

This model, basically, describes the process of the formation of the observer image quality judgment. Thus, this equation can be thought of as a combination rule of the perception attributes forming a resulting quality estimate. The p power of the model defines the amount of perceptual quality parameters used. If a user is asked to assess a set of images with only one "ness" varied within the set, according to Eq. (3.1) the relationship between a separate "ness" and image quality judgment is linear or monotonous [33]. It is often true for some image parameters mostly belonging to the category of artefactual observer perception attributes. However, there are various "nesses" that do not follow this pattern [39, 40], e.g. colorfulness or contrast.

The number and the variety of image quality models appearing in literature is significant. We will consider several solutions that influenced the model proposed in this thesis.

3.2.1 FUN Model

FUN stands for fidelity, usefulness and naturalness. The model was proposed by Yendrikhovkij in [100]. The overall image quality assessment here can be given as a weighted sum of these three attributes, with each of the weights being in the range from 0 to 1, in such a way that their sum is equal to 1. Such a tradeoff allows the modeling of the relative importance of each. The weights depend primarily on the area of application of certain images and their contents [97]. Thus, the model complies with Eq. (3.1) describing a typical quality model.

Fidelity

Fidelity according to Yendrikhovskij [97] is "the degree of apparent match of the reproduced image with the external reference, i.e. original". The degree of fidelity represents how close the image under study is to the original. The higher the fidelity is, the closer is the match. Such an attribute is especially important for fine art or various catalogue reproductions [97].

Usefulness

Usefulness is, in turn, "the degree of apparent suitability of the reproduced image to satisfy the corresponding task" [97]. The degree of usefulness depends on the application of the image, or as it can also be described, the needs of the viewer, i.e. what is expected of the image in order for it to be useful. An image with the highest degree of usefulness ideally presents the maximum number of distinguishable features. A mathematical definition of usefulness as a term of the number of discriminable items in an image was proposed by Janssen and Blomaert in [58, 59]. The usefulness term is affected by contrast and color saturation. This index is often negatively correlated with fidelity due to the fact that original images might not be of high enough discriminability. An example of typical usefulness sensitive applications would be medical images, military night-vision images, remote sensing images, etc.

Naturalness

Naturalness is defined by Yendrikhovskij [97] as "the degree of apparent match between the reproduced image and internal references, e.g. memory prototypes". A number of studies exploring the connection between color variations and naturalness exist [98, 99, 96, 57, 58, 40, 24]. These attempt to develop a color naturalness index (CNI) that would allow predicting the observer evaluation of the degree of naturalness of colors reproduced. CNI is basically a distance measure between the colors of the objects present in the image and the prototypical colors of the objects. Simply put, it describes the degree of the similarity of the color of an object in the scene with the prototypical color for that category [97].

3.2.2 Computational Image Quality Model

An approach closely connected with the FUN model [97] is presented by Janssen in [55]. Quality is regarded here as the degree of the ability of the visual and cognitive systems to use the information presented within the image [55].

Janssen [55] defines image quality in his work as the adequacy of a given image as an input to visual perception, whereas the adequacy term is described by discriminability and identifiability of the items given in the image.

According to Janssen, the process of the human perception of an image consists of three main stages [55]:

1. construction of the internal representation of the image;
2. matching the thus obtained representation with the memorized reality;
3. semantic processing of the information.

All of the steps of the perception of an image have to be successful in order for the image to be perceived of good enough quality for any given task. To be able to achieve that several conditions have to be met. For one, the internal representation has to be precise enough, and the degree of similarity between this representation and the memorized reality has to be close enough. This means that for a successful interpretation of the image, which is assumed to be the quality perception of the image, it should create the best match between the internal representation and knowledge of reality [55].

Given the conditions stated, discriminability and identifiability are regarded as the perceptual image attributes that can effectively describe image quality. Discriminability is described as the average difference between the colors presented in the image, and thus is closely related to colorfulness. Identifiability, in turn, is the difference between the image and the memory object colors, making identifiability closely related to naturalness. The main focus of the model lies in describing the process of image quality judgment formation rather than modeling or predicting the outcome of the quality perception [57].

This model describes the quality of the image in terms of usefulness and naturalness, where usefulness can be defined as "precision of the internal representation of the image", and naturalness as "the degree of correspondence between the internal representation of the image and knowledge of the reality as stored in memory". Thus quality is described as the tradeoff of the degree of naturalness and usefulness. The two terms can often be negatively related, meaning that some distortions might be introduced into the image in order for it to be more useful, which would affect the degree of naturalness, e.g. increasing the contrast or saturation of the image [55].

The relationship between naturalness and usefulness is defined as [57]:

$$IQ = \lambda D + (1 - \lambda)m, \tag{3.2}$$

where IQ is the image quality value, λ is a naturalness vs. usefulness coefficient varying in the range 0 to 1, D is the strength of the discriminability, and m is the naturalness.

3.2.3 Kayargadde Image Quality Theory

The Kayargadde theory of image quality [63] uses two perceptual attributes: blur (unsharpness) and noisiness (speckle). According to Kayargadde, blur and noise are inherent to the process of acquisition.

Blur estimation consists of two steps: first, the regions of the local one-dimensional edges are determined by finding the local maxima in the gradients, and two-dimensional structures are removed by excluding points with high energy. The second step includes the edge parameters amplitude, sigma, offset, displacement, and angle of the edge, which are determined using polynomial transforms. Thus, Kayargadde gives the following mathematical definition of blur [63]:

$$S_b = 1 - \frac{1}{[1 + (\sigma_{bi}/\sigma_{b0})^2]^{1/4}}, \tag{3.3}$$

where σ_{bi} is the average spread of image blurring, and σ_{b0} is the intrinsic blur of the human early visual pathway. The value used in the experiments by Kayargadde is 0.65 arc min. [27] and is determined as [63]

$$\sigma_{b0} = \sqrt{\sigma_0^2 + (C_{ab}d)^2}, \tag{3.4}$$

where σ_0 is a blur constant equal to 0.5 arc min., C_{ab} is a constant equal to 0.08 arc min./mm, and d is the pupil size in mm [63].

The objective estimation of noise is based upon the assumption that images have areas where the luminance distribution is locally zero-dimensional, i.e. homogeneous areas found by selecting the regions with a small gradient energy. Noise is determined according to Kayargadde in a similar manner as blur [63]:

$$S_n = 1 - \frac{1}{[1 + (\sigma_{ni}/\sigma_{n0})^2]^{1/4}}, \tag{3.5}$$

where σ_{ni} is the standard deviation of the noise in the image, and σ_{n0} is the standard deviation of the noise in the early visual pathway. The value of σ_{n0} is determined empirically [63].

Kayargadde [63] determined that S_n and S_b noisiness and blur are not totally independent, and that the combination of these objective image quality measures presents a good measure of the perceived image quality. However, the method suffers from the significant drawback that images with a speckled texture would have low quality according to this technique, despite being perceived as being of a high quality [27].

3.2.4 Spectral Color Appearance Model

The spectral color appearance model proposed in *Publication I*, *Publication V* and *Publication VI* describes quality in terms of preferential image attributes. The model is intended to create a paradigm that would allow the description of the quality of spectral images using objectively measurable parameters of spectral images in connection with the subjective quality metrics. The model per se assesses the quality of color images reproduced through the spectral images. Quality here is said to depend upon three parameters: colorfulness, vividness and naturalness. We assume here is that quality means preferential spectral quality attribute based evaluations.

An underlying hypothesis of the spectral color appearance model is that variations of the statistical characteristics of spectral images affect the overall appearance. Kurtosis impacts image highlight reproduction, and standard deviation, in turn, affects image contrast and color saturation.

To be able to affect image appearance through the use of the parameters of spectral images, a generalized statistical image model which describes the behavior of the statistical characteristics of natural images is used. The statistical model represents any given spectral image as

$$\mathbf{f}(\mathbf{x}) = \boldsymbol{\mu} + \mathbf{D}\mathbf{g}(\mathbf{x}), \tag{3.6}$$

where $\mathbf{f}(\mathbf{x})$ is a natural spectral image presented as an n-dimensional vector random field, $\mathbf{x}$ is a vector with each element being a spatial dimension, $\boldsymbol{\mu}$ is a vector of means of each of the spectral bands, $\mathbf{g}(\mathbf{x})$ is a normalized vector image with a zero mean and unit standard deviation for each component, $\mathbf{D} = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$, where σ_i is the standard deviation of the i-th component.

Based on this model, colorfulness and vividness can be varied using the three parameters α, β and k_{max}. One of the assumptions set as the basis of the statistical image model is that standard deviation variations affect the overall image colorfulness appearance. Vector $\boldsymbol{\sigma}$ can be presented as follows:

$$\boldsymbol{\sigma} = \alpha\beta\boldsymbol{\sigma}_v + (1-\alpha)\boldsymbol{\sigma}_c, \tag{3.7}$$

where $\alpha = (\sigma_{max} - \sigma_{min})/\sigma_{max}$ is a constant characterizing the relationship between constant σ_c and variable σ_v parts of the standard deviation $\boldsymbol{\sigma}$, and β is a contrast variation coefficient. Thus by varying α and β, colorfulness can be affected. Values of $\alpha < 1$ with $\beta = 1$ significantly reduce colorfulness, while $\beta > 1$ with $\alpha = 1$ increase the colorfulness appearance.

Another quality attribute, vividness, according to the statistical image model, is influenced by kurtosis variations. A gamma-Charlier histogram transform function of the normalized vector image with a zero mean and unit standard deviation $f_s(\mathbf{x})$ and a kurtosis vector $\mathbf{k}$ allows us to affect the vividness of the image as follows:

$$\mathbf{g}(\mathbf{x}) = \mathbf{H}(f_s(\mathbf{x}), \mathbf{k}). \tag{3.8}$$

To affect image appearance through histogram transform, all kurtosis elements are proportionally modified according to the given maximum of the kurtosis value k_{max}.

Quality manipulation is a complicated task that requires significant computational resources. Spectral images contain large amounts of information which have to be manipulated in order to influence the overall impression of the display. Usually, some implicit assumptions are made in order to limit the amount of computations. The assumption underlying this study is that only global variations are taken into account, which, in turn, originates from the fact that all parts of the image have been captured under the

same illuminant or belong to the same object. Thus, the same modifications are applied to all pixels of the image irrespective of the content [40].

So far, we have discussed the terms colorfulness and vividness without giving a proper definition for the image attributes serving as the base of the model.

Colorfulness

Fedorovskaya et al. [40] define colorfulness as the presence and vividness of colors in an image. The definition adapted in this work states that colorfulness can be described as the contrast and saturation of the colors in the image under study. An assumption made in this work is that colorfulness belongs to a class of preferential image quality attributes. This means that with the increase of this parameter, the quality of the image also increases up to a certain point, after which it starts to decrease. Moreover, the location of the optimum point is heavily viewer dependent.

An example of the change in the colorfulness parameter in a spectral image is shown in Figure 3.2.

Figure 3.2: Color reproduction of image inlab2 [79]; original ($\alpha = 1$, $\beta = 1$) (left), and processed ($\alpha = 1$, $\beta = 1.6$) (right)

Here, the image on the left is the color reproduction of the original spectral image, and the image on the right is the modified image with an increased colorfulness value ($\alpha = 1$, $\beta = 1.6$). Looking at Figure 3.2, we can state that with the increase of the colorfulness parameter, the overall contrast and saturation of the colors in the image increase.

Fedorovskaya and de Ridder have previously invested significant efforts in colorfulness assessment. A number of experimental works in this area [39, 40, 24] have shown that saturation, chroma and colorfulness do not have a monotonous functional connection with image quality. The three terms used in these papers constitute three aspects of colorfulness, as we see in our research.

Vividness

Vividness can be defined as highlight and shadow intensity. The actual term was derived in the course of the experiments, during which the subjects, when asked to describe the impression of the images with low values of this attribute, described them as "lacking

life" and "not vivid enough". From these descriptions, we can draw the preliminary conclusion that vividness has a positive correlation with naturalness.

The effect of applying the vividness parameter change to a spectral image is presented in Figure 3.3.

Figure 3.3: Color reproduction of image inlab5 [79]; original ($k_{max} = 45$) (left), and processed ($k_{max} = 80$) (right)

Here, the image on the left is the color reproduction of the original, and the image on right hand-side is the color reproduction of the spectral image with the increased vividness ($k_{max} = 80$). The areas of the highlight and shadow have stronger intensities, while the rest of the image remains relatively similar.

Another aspect of the vividness parameter is its connection with colorfulness. Some of the observers at the experimental stage perceived the vividness change to be similar to the colorfulness change. To illustrate the difference, let us look at a sample of a quality ruler for both of the image attributes presented in Figure 3.4.

Here the top row represents the change of the vividness attribute and the bottom that of colorfulness. The cut-out parts of the flower show the difference in the change of colors when applying either of the attributes. Looking at Figure 3.4, we can state that although the effect of vividness and colorfulness on image appearance is visually similar, the difference between the two specifically in the highlight and shadow regions is clear.

An assumption made in this work is that both vividness and colorfulness belong to a class of preferential image quality attributes, meaning that with the increase of this parameter the quality of the image also increases up to a certain point, after which it starts to decrease.

Naturalness

Image quality perception depends on both internal and external factors. Viewing conditions, physical properties and characteristics of the stimuli, and the context in which the image is to be used are among the most important external factors, while internal ones include prior experience and conditioning, the memory of scenes and similar images, points of interest in the images, emotions etc. [38].

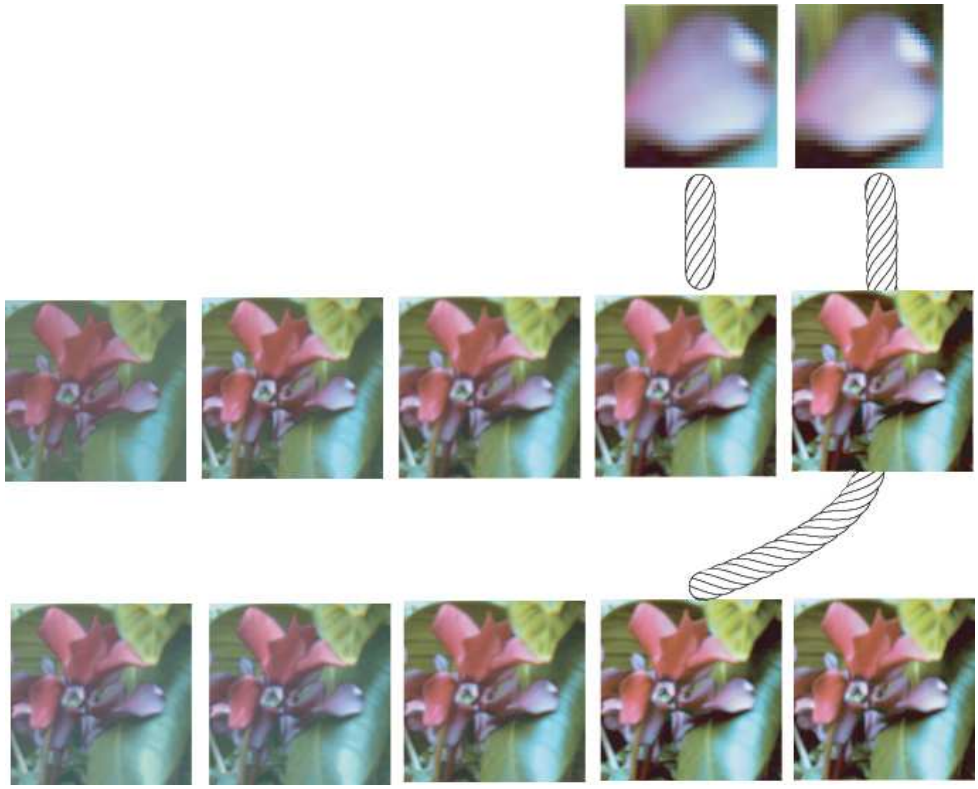


Figure 3.4: Quality ruler example inlab1 of vividness (top row) and colorfulness (bottom row)

Naturalness, according to Yendrikhovskij [96], is the degree of correspondence of the image in question and an internal representation or the memorized image of it [96, 40]. The connection between image quality and image naturalness can be expressed in the intuitive assumption that in order for the image to be of high quality it should at least be perceived as natural [40]. Moreover, in this work we have also adapted the assumption that the visuo-cognitive process should be regarded from the information processing point of view rather than from the signal processing one. Janssen and Blommaert [57] describe the process of image perception as processing of the visual information received, which has to be reconstructed into a visual form and is further perceived with regard to the memory representation.

One of the goals of this work was to find a relationship between the naturalness parameter and the two parameters proposed in this study, i.e. colorfulness and vividness, and, in turn, to establish the influence of naturalness on the overall appearance of the image under study. An initial hypothesis is based upon the assumption adapted from works of Janssen and Blommaert [57], stating that naturalness is a necessary, but not a sufficient parameter to define the quality of an image. Some images may appear unnatural, but

still be perceived to be of hight quality. Thus, the principle of discriminability should also be considered, represented in this work by colorfulness and vividness. However, as we have previously said both of the parameters have a certain threshold of increase of quality, after which the image still remains discriminable. However, the image becomes so unnatural that the quality perception drops. Thus, quality can be considered to be a trade-off between colorfulness and vividness on the one hand, and naturalness on the other.

3.3 Summary and Discussion

In this chapter, we considered objective image quality, which is defined here as "an impression of [image] merit or excellence, as perceived by an observer neither associated with the act of ... [acquisition], nor closely involved with the subject matter depicted" [64]. Quality is considered here from the point of view of observer perception attributes and image quality models connecting them to customer quality preference.

Preferential and artefactual attributes are the main focus of this part of the work. They include blur, noisiness, fidelity, naturalness, colorfulness, detectability and vividness.

The models described in this part include the ones that have significantly influenced the spectral color appearance model presented in the framework of this research and could be included in further developments. The computational image model [55], the FUN model [97], and the Kayargadde image quality theory [63] were presented.

The spectral color appearance model that sets a relationship between objectively measured parameters of spectral images and subjective quality metrics was proposed in this thesis. The model assumes that image appearance can be affected using the statistical characteristics of the image. Three perceptual parameters are used to describe quality: colorfulness, vividness and naturalness. Colorfulness is the contrast and saturation of colors in the image. Vividness is defined as the highlight and shadow intensity, and naturalness is the degree of correspondence of the image to the internal representation of it. An assumption was made in this chapter that naturalness is a necessary, but not a sufficient condition for the quality of an image.

Chapter IV

Subjective Image Quality and Fidelity Evaluation

Up to this point, we have been dealing with objective image assessment techniques, i.e. the measures and methods seeking to automatically predict the perceived image appearance [32]. This chapter is dedicated, in turn, to subjective methods of image assessment. They require human visual test assessments [87]. Human observers act as the measurement instruments in the subjective experiments [32]. The goal of this chapter is to describe subjective methods and techniques used in the experimental part of this thesis and in *Publications I-VII* both for image quality and image fidelity.

The first standard of subjective image quality evaluation was published in 1974, i.e. ITU-R BT.500-7 "Methodology for the subjective assessment of the quality of television pictures" [7]. This standard establishes a certain reference image against which the impairment factors that affect the overall appearance are measured, as well as some of the earlier approaches to subjective television picture assessment [13].

Observer evaluations in subjective experiments in general can be performed in two different ways: absolute and comparative [34]. Absolute subjective image quality evaluation is performed on each image under study separately; the image is assigned some rating on a given category scale. Comparative image assessment, on the other hand, as the name suggests, presents the observers a number of images to compare and rate either from best to worse or by assigning some rating based upon image comparison [34]. Comparative image assessment falls into both the fidelity and quality evaluation categories, while absolute assessment is primarily a subjective image quality technique.

There are various subjective image assessment methods, such as direct scaling [85], the rank order method [64], the threshold method [43, 89, 65], pairwise comparison [38, 64], categorical sort [64], the perceptual image distortion map (PIDM) [101], etc. The number of subjective techniques and their methodologies vary greatly, and the list given above is far from complete. The methods used in this work and the methods closely related to them are described in more detail.

4.1 Direct Scaling

The most commonly applied method in subjective image quality and fidelity evaluation is the mean opinion score (MOS), which is often called direct scaling [85]. The method involves quantifying the quality of the image under study on a certain scale, which can be expressed as either an integer (from 1 to 5, or 1 to 10) or a term describing the quality of the reproduction (not acceptable to excellent, etc.). To obtain the overall quality judgment, individual responses over all of the subjects are averaged, which reduces variability in the data.

Here, as in most of subjective assessment techniques, both expert and non-expert observers can be used, depending on the goal of the experiment. In this case, by experts we mean subjects educated in the area of the experiments, be it compression, halftoning or some other image modification concerned in the experiments [38].

One of the most significant drawbacks of this method is the subjectiveness of the assignment of the scale to a certain image ordering. Another issue is that the results of the tests are unitless, which complicates further analysis and comparisons with other experiments [38].

4.2 Threshold Method

A method lacking the drawbacks of the direct scaling technique is the threshold method [43, 89, 65]. The idea behind it is that the user should make a certain judgment, i.e. place a threshold of the visibility of a certain artefact, difference between the visual stimulus, or preference of a certain stimuli over the others. The main advantage of such a method is that the results obtained using this method are more reliable compared to the above-mentioned technique, the task is less vague, and when the observer is provided with feedback the variance of the results obtained using the threshold method is minimized. On the other hand, such a method does not allow the investigation of a connection between a single factor influencing the image appearance and the overall impression, given that image attributes often have a strong correlation. It is hard to single out a separate image characteristic contributing to the judgment. Moreover, this method does not allow predicting image fidelity judgments [38].

4.3 Pairwise Comparison

The pairwise comparison method, or as it is often called paired, the comparison technique [38, 64] unlike the previous two methods produces reliable and informative data on the subjective image quality above the threshold of visibility or preference. The idea behind the method is that the subjects are presented with pairs of images with a single attribute or a series of varying attributes, and the subjects are to identify the image with the highest quality. This method allows us to estimate the factors that influence the suprathreshold judgments. Moreover, this method allows us to predict the perceived image quality judgments. This technique is designed for measuring small quality preferences. However, larger sample differences are significantly more difficult to quantify using this method [64].

The selection of either of the above-mentioned psychometric methods is often a difficult task. Moreover, none of the methods produce the results calibrated against an existing numerical scale, making the comparison of the results practically impossible [64].

4.4 Psychophysical Image Quality Measurement Standard

A recently proposed standard in the area of subjective image quality assessment is ISO 20462-1:2005 "A psychophysical image quality measurement standard" [65] that sets a number of rules for measuring image quality, the interpretation of the results and comparison between the experiments.

The standard introduces the notions of a triplet comparison technique and that of the quality ruler, both of which yield assessments calibrated in just noticeable difference (JND) units [64].

Most quality research is based upon the assumption that the human perception of image quality and fidelity is probabilistic in its nature, i.e. the scale of the distortions introduced into the image has a certain probability of detection [64]. This, in turn, means that the distortions can be related to a certain function or interval scale.

According to Keelan [64], a particular instance of the judgment is drawn from a probability distribution of possible perceptions. One JND can be regarded as representing a just significant difference of quality. In other words, the JND is a stimulus difference that yields a 75%:25% proportion in a forced-choiced paired comparison, i.e. falling half way between a random guess (50%:50%) and a certainty (100%:0%) [64]. Therefore, JND is the smallest possible difference between two stimuli, noticeable by observers [10]. To be able to estimate the JND interval scale, the probability observed in a paired comparison experiment is used in an approximation function. Usually a normal perceptual response distribution is assumed, and the standard introduces the angular function [64]

$$JNDs = \frac{12}{\pi} \sin^{-1}(\sqrt{p}) - 3, \tag{4.1}$$

where p is a forced-choice paired comparison probability. The possible values of JND computed using Eq. (4.1) lie within the interval $[-3, 3]$. This function lacks the most significant drawback of the Gaussian function, i.e. the extended tails of the latter do not always match the real perceptual responses [64].

The standard [10] distinguishes two types of JND units. An attribute and quality JNDs. Attribute JND is a measure of the detectabilty of the appearance of differences. In our study, we have used the quality JND, which is a measure of the significance of perceived differences on overall image quality [64].

The choice of the JND as the measure of image quality in this research is not incidental. As Keelan points out, JNDs could serve as an excellent basis for the construction and calibration of numerical scales of preference. Having the judgments expressed in JNDs gives an opportunity for comparing the results obtained with some reference system. Moreover, the results could also be generalized over a number of attributes to produce the overall judgment function [64].

The standard also defines the conditions for the experiments. To be able to produce relative JND values, at least ten observers and three scenes have to be involved in the experiments. The observers have to be tested for normal color vision and visual acuity. A number of rules of the physical experimental setup that have to be taken into account, i.e. the general lighting, duration of the experiments, background color etc., are also defined [65].

4.4.1 Quality Ruler Method

The quality ruler technique, as the name suggests, presents the user with a number of images of the same scene with variations in a separate attribute, with a previously set in JND units quality or fidelity difference between the samples [65].

The ruler is constructed according to two main requirements: it should be easy to identify the samples closest in quality to the one under study, second requirement is that the stimulus under test and the closest in appearance image should be inspected under similar conditions. The quality ruler can be both hardcopy and softcopy [65]. Hardcopy here means that the images are presented in a printed format, while softcopy assumes a digital form of the images under study.

Various attributes can be used in this method, but it is generally not recommended to use the preference image attributes, such as color and tone characteristics, which are individual and depend on personal likes and dislikes, because of the instability of the final ranking [65]. This method is not currently applied in our work, but will be used in future fidelity research involving SSIM and other measures involving the assessment of spatial image distortions.

4.5 Perceptual Image Distortion Map

One of the subjective techniques of image fidelity assessment used in this thesis is Perceptual Image Distortion Map (PIDM) proposed by Zhang et al. [101]. The idea is to be able to predict the fidelity of the reproduction at different locations within the image, i.e. establish the locations of the errors perceived by subjects in the reproduced image, and the severity of the distortions [101].

This is a full-reference, comparative method, i.e. to be able to produce a PIDM, both the original and modified images are used. To obtain a distortion map, subjects are presented as two images - the original and an image with the distortions introduced. Test subjects are asked to identify the regions within the image where the two appear to be different by using a digital marker. Any shape and size of the marker can be used for this purpose, taking into account that the chosen option limits the spatial resolution of the measured image distortion map, and the marker should also be taken into account when generalizing the results. The actual image distortion map can be considered as a map of the probabilities of a certain region in the modified image to be marked as varying from the original. Thus, having a number of subjective empirical maps, an overall PIDM is constructed by constructing the overall probability. To illustrate the areas of difference within the image, the PIDM is normally presented as a gray-scale image with gray level intensity at each pixel corresponding to the probability of a mark covering the pixel at

the same location in the modified image, from white - regions marked most often, to black - not marked at all. The PIDM can be used separately as an image fidelity measure or as a reference point for the assessment of the performance of objective measures [101].

4.6 Summary and Discussion

In this chapter, we briefly considered a number of methods of subjective image quality and fidelity assessment. Subjective techniques require performing human visual test assessments to be able to make an evaluation, where observers act as the measurement instruments [87]. This chapter was dedicated to methods used in the experimental part of this work and some of the techniques closely related to them.

All of the methods given can be divided into two broad categories: absolute and comparative [34]. Subjective methods considered here include direct scaling [85], the rank order method [64], the threshold method [43, 89, 65], pairwise comparison [38, 64], categorical sort [64], and the perceptual image distortion map (PIDM) [101].

Two standards of subjective image quality have also been considered, i.e. ISO20462-1:2005 "A psychophysical image quality measurement standard" [65], and ITU-R BT.500-7 "Methodology for the subjective assessment of the quality of television pictures" [7]. The standards set the rules for measuring subjective image quality, define the conditions necessary for the experiments, and provide a basis for further analysis of the results.

Chapter V

Experiments

In this chapter we will consider experiments performed in the framework of this thesis, intended to evaluate the theoretical assumptions presented in previous chapters.

Experiments are divided into two broad parts: objective image quality and objective image fidelity. Image quality experiments are based on the research presented in *Publication I*, *Publication V*, and *Publication VI*, while image fidelity tests use material published in *Publication II*, *Publication III*, *Publication IV*, and *Publication VII*. All of the experiments include methods and models of subjective image quality and fidelity assessment.

MacAdam [73] describes the process of experimental image quality research as an approach consisting of five basic stages:

1. vary images in a systematic manner using certain image attributes;
2. perform visual judgments;
3. analyze the results and find correlation with the varied physical parameters;
4. find optima;
5. improve the underlying model.

Such a procedure is true both for image quality and fidelity research and is applied with variations in this work.

5.1 Spectral Databases Used in This Thesis

In this work, the following publicly available databases were used:

1. Both image fidelity and image quality parts use a spectral database of natural scenes from [79], the whole database consists of 29 images. Five images - *inlab1*, *inlab2*, *inlab5*, *jan13am* and *rleaves* were selected for the purpose of the experiments. Each image has the following dimensions: 256x256 in the spatial dimension and 31 components in the spectral dimension as 8 bits/pixel format. The images were captured by a CCD (charge coupled device) camera with narrow band interference filters in a 400-700 nm wavelength range at 10 nm intervals. The images selected were taken indoors (in a controlled environment, i.e. dark-lab or glass-house), in order to avoid unnecessary wind effects. The quality of the images is described in the database as: "Best" and "Medium: slight corrections" [79].

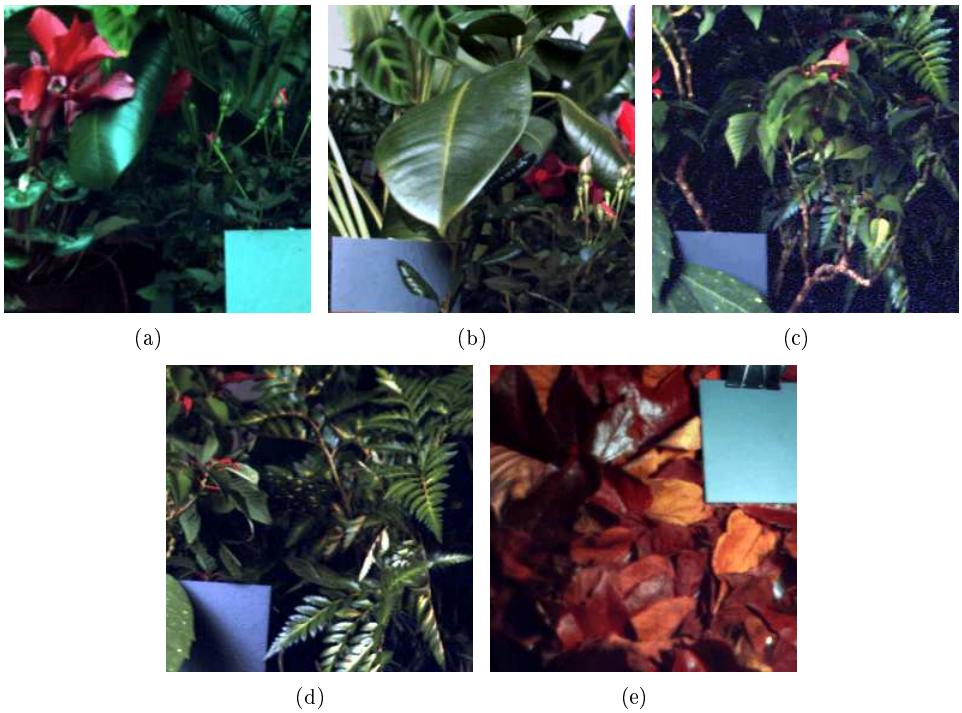


Figure 5.1: Color reproduction of spectral images from [79] used in this thesis. a)*inlab1*; b) *inlab2*; c) *inlab5*; d) *jan13am*; e) *rleaves*

2. Another set of spectral data measured in the University of Joensuu [1] is a set of spectra of 1269 Munsell Colors Matte color chips [3]. This dataset is used in image fidelity experiments. The Munsell color system is one of the most widely used and is one of the most appropriate data sets for the purpose of the perceptual scaling experiments. The reflectance spectrum was taken with a Perkin-Elmer Lambda 9 UV/VIS/NIR spectrophotometer in the 380 - 800 nm interval with a 1 nm wavelength resolution producing a total of 421 spectral components [1]. The dataset is used in this thesis in image fidelity experiments.

3. A set of tests concerning the possibility of discrimination of metameric colors was performed on a dataset of metamers, containing a total of three assortments of different colors. The reflectance spectrum of these was measured at a 380 - 780 nm interval with a 2 nm wavelength resolution under illuminants D65 and A, resulting in 201 spectral components. The data set consists of a red metameric sheet, Macbeth metameric sheet and reflectance spectrum of a part of a passport. The dataset was obtained from Joensuu University, the Department of Computer Science, Finland. The metameric color discrimination experiments were performed in the framework of image fidelity experiments.

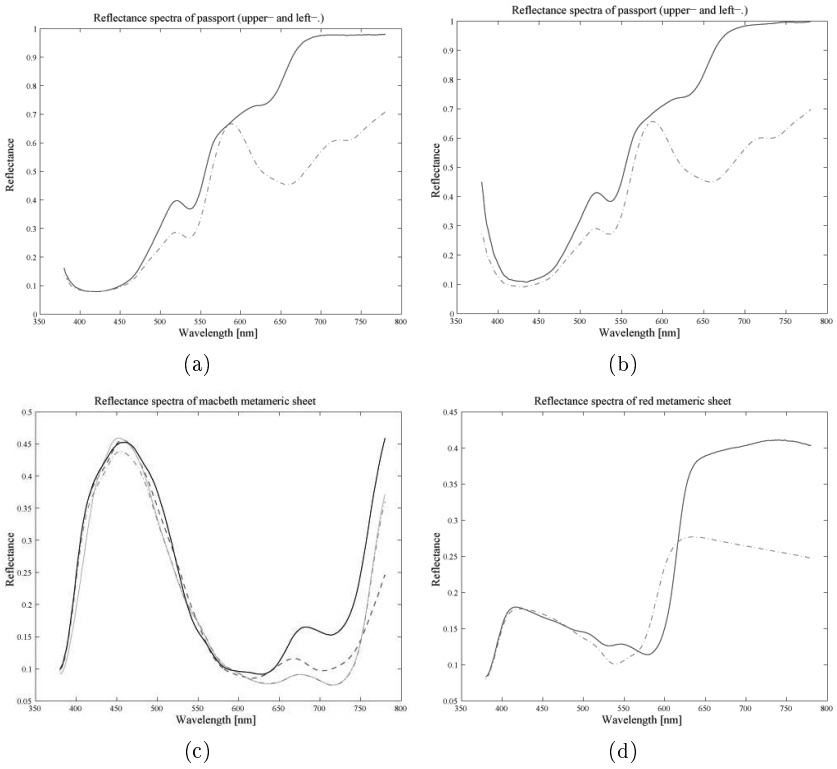


Figure 5.2: Reflectance spectra of metameric samples. a) passport upper and left corners under D65 illuminant; b) passport upper and left corners under A illuminant; c) Macbeth metameric sheet under D65; d) red metameric sheet under D65

5.2 Image Quality Experiments

Image quality experiments concentrate on applying the Statistical Image Model described in Chapter 3 to images of natural scenes. The tests performed consist of three parts corresponding to the three image attributes described in the model: colorfulness, vividness

and naturalness. The main idea behind the experiments is to ascertain the existence of the functional relation between these image attributes and perceptual image assessments, and estimate the form of the relation.

Experimental settings, which include the number and the criteria for the selection of observers, test stimuli and observer instructions were chosen to comply with the ISO "Psychophysical image quality measurement standard" [65]. According to this standard, relative quality judgments should be obtained from at least ten observers and three scenes, and all of the observers have to be tested for normal vision and visual acuity. Given that, the experimental settings included twenty observers with normal or corrected-to-normal vision without color deficiencies (tested using the Ishihara Color Tests [51]). Among the observers ten were so-called expert observers, meaning that they were aware of the area of the research, and the other ten were non-expert observers. The observers included staff and students of the Department of Information Technology of Lappeenranta University of Technology. To prevent the loss of the quality of judgments due to fatigue, the duration of the experimental sessions was limited to one hour. If more time was needed, the experiments continued after a break. Viewing conditions followed the requirements given in [8], i.e. general room lighting was shaded and set so that it neither directly nor indirectly influenced the viewing surface. The color temperature of the room lighting was 5600K, which is close to daylight lighting. This was taken into account when RGB representation of the images under study were produced from the spectral data, by setting the appropriate value of the reference white. When viewing the images, the frames were darkened to a 10% brightness level for a width of 75 mm. The images were presented on a CRT display. However, the limitations of the color gamut on the CRT displays were not taken into account in the experiments. Neither were the viewing angles integrated into the experimental setup assumptions. Although original images were presented, it should be emphasized that they were not explicitly identified to observers as such. A significant part of the setup are instructions given to test subjects. An important issue here is the definition of the artefact or attribute being assessed in the series of tests. According to Keelan [64] instructions should explicitly describe the origin and the appearance of the effect under study. Consequently, the instructions explained each of the three image quality attributes as follows:

Colorfulness: "You will be presented a series of images. Your task is to assess the colorfulness of the images using an integer from one to ten, with one corresponding to the lowest degree of colorfulness and ten to the highest. Colorfulness consists of contrast and color saturation. Therefore, pay special attention to the edge contours, reproduction of the details, saturation of colors, sharpness, contrast, and brightness."

Vividness: "You will be presented a series of images. Your task is to assess the vividness of the images using an integer from one to ten, with one corresponding to the lowest degree of vividness and ten to the highest. Vividness can be defined as the highlight reproduction in an image. Thus, pay special attention to the reproduction of the lightest and darkest areas."

Naturalness: "You will be presented a series of images. Your task is to assess the naturalness of the images using an integer from one to ten, with one corresponding

to the lowest degree of naturalness and ten to the highest. Naturalness is defined in this case as the degree of correspondence between the image reproduced on the screen and reality, i.e. the original scene as you picture it."

Experiments were performed on spectral images in the first dataset described in Section 5.1. First, a set of test data was produced using the colorfulness parameter, varied through standard deviation, using Eq. (3.7). By changing the α and β coefficients, it was possible to receive new values for constant and variable parts of the standard deviation. This procedure was applied to images with values of (α,β) equal to (0.55,1), (0.75,1), (1,1.3), (1,1.6). The intervals were chosen to be uneven for the purpose of further analysis, and also to exclude the possibility of observers seeing a certain trend in image appearance and adjusting the quality judgments accordingly. Observers were presented RGB reproductions of the original and modified spectral images. To produce these color reproductions spectrum for each of the points in the images was converted into CIE 1931 XYZ color space [2] using the color-matching functions. Thus received XYZ values were then converted into RGB representation. The results of the assessments are summarized in Table 5.1. It should be emphasized that the general room lighting conditions were taken into account by selecting the appropriate value of the reference white.

Table 5.1: Mean values of colorfulness evaluation scores

Quality	Image 1	Image 2	Image 3	Image 4	Image 5
inlab1	2.67	4.50	6.00	6.00	7.17
inlab2	2.67	4.83	7.00	7.83	8.67
inlab5	3.00	4.50	5.67	6.83	7.67
jan13am	3.33	4.83	7.00	7.67	8.17
rleaves	2.50	3.50	4.83	4.83	5.83

Each cell in Table 5.1 corresponds to an averaged colorfulness evaluation score, with the highest and lowest scores excluded from calculations. Every number represents the value pairs of α and β described above, applied to each of the images given in the first column. Image 3 here is the original image. Average variance within the judgments of the observers is equal to 1.2, which shows that the spread of opinions within the observers was not vast. Looking at Table 5.1, it can be stated that with the increase in the colorfulness, the quality of the image under study increases.

The second set of test data was produced through variation of the vividness parameter. Test images were produced with the help of Eq. (3.8) (with k_{max} equal to 5, 10, 30, 60). Table 5.2 presents the results of the assessments. Again, the intervals between the values of the parameter were chosen to be uneven for the same reason as in the colorfulness case.

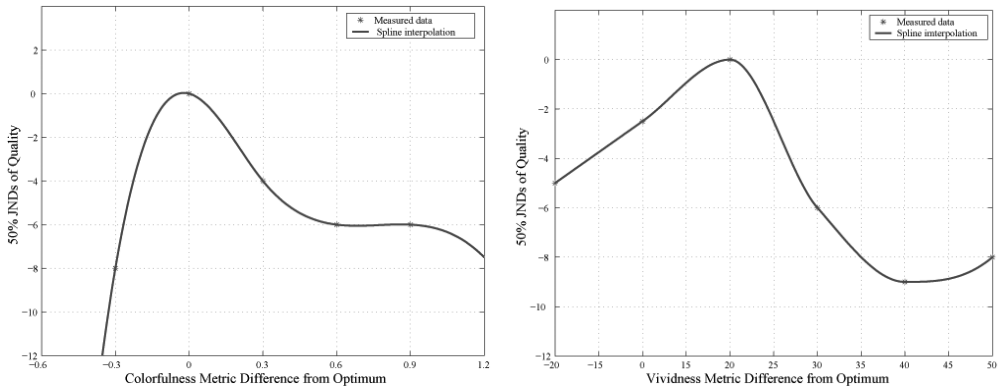
Each cell in Table 5.2 corresponds to an average over all of the observers' vividness evaluation scores, with the highest and lowest scores excluded from calculations. Here again, Image 3 is the original image. Average variance within the judgments of the observers is equal to 1.58, which shows that the spread of opinions within the observers was higher than in colorfulness evaluation, but still not high. Looking at Table 5.2, it

Table 5.2: Mean values of vividness evaluation scores

Quality	Image 1	Image 2	Image 3	Image 4	Image 5
inlab1	2.56	4.00	6.44	7.19	8.19
inlab2	5.88	6.12	7.94	4.29	3.76
inlab5	3.18	4.71	5.76	7.29	8.12
jan13am	4.29	5.88	6.71	7.41	5.59
rleaves	4.65	5.82	5.94	4.76	5.00

can be stated that with the increase in the value of vividness, quality also increases up to a certain point - the optimum - and further changes lead to a loss of quality.

To present and analyze the results, the notion of the just noticeable difference (JND) is used. To construct the quality preference function, the value of each of the attributes of the highest rated image was identified for each scene and observer, and the fraction of times each one of the images was chosen was computed. To convert each fraction to a probability density, it was divided by the width of an attribute interval including the sample position. The resulting preference distribution function is presented in Figure 5.3. The circles represent the calculated JND values and the line a spline fit estimate of the function computed using Matlab.



(a) (b)

Figure 5.3: Measured quality loss functions of (a) colorfulness and (b) vividness averaged over the collection of scenes and observers

Although it has been found that both colorfulness and vividness are observer and scene susceptible, the quality loss function for each observer and scene can be regarded as a member of a family, and could thus be averaged to produce a mean quality loss function for the collection of observers and scenes tested [64].

Figure 5.3 (a) presents the quality loss function of the colorfulness and (b) vividness image attributes. Here, the x-axis represents the difference from the optimal position of

each of the attributes and the y-axis is the JND variation. The series of experiments both in vividness and colorfulness yielded an image that was preferred by most of the observers. The statistical characteristics of that image gave us the point of the optimum. Any changes in the values of vividness and colorfulness lead to the drop in the preference of the images. As the attributes deviate the quality drops off. The value of the drop off in JNDs is given on the y-axis. And the optimum lies at the point (0,0).

Looking at both plots, it can be stated that quality has an optimum, so that with the increase of either of the image attributes quality increases to a certain point, after which it starts to decrease. Moreover, the peak of preference lies almost within the original, with both vividness and colorfulness being slightly higher. A preliminary conclusion that can be drawn is that observers prefer slightly more saturated images with deeper shadows and higher contrast.

Both colorfulness and vividness image sets were further used in naturalness assessment experiments. The goal of this part of the experiments was to find out the influence of colorfulness and vividness on the appearance of image naturalness and the influence of image naturalness on the impression of image quality. The results of the experiment are summarized in Table 5.3.

Table 5.3: Mean values of naturalness evaluation scores

Quality	**Image 1**		**Image 2**		**Image 3**		**Image 4**		**Image 5**	
	1	2	1	2	1	2	1	2	1	2
inlab1	2.37	5.41	5.83	6.34	8.12	8.34	9.67	9.83	8.15	9.12
inlab2	2.25	6.03	6.93	7.98	8.93	9.67	9.87	5.56	8.41	3.98
inlab5	3.84	7.34	6.55	7.53	8.34	9.12	9.85	9.53	8.56	9.34
jan13am	3.10	6.56	7.12	8.96	8.67	9.87	9.73	6.76	8.17	4.87
rleaves	3.32	6.17	6.86	9.10	8.16	8.17	9.54	7.65	9.50	5.35

Each cell in Table 5.3 corresponds to an averaged naturalness evaluation score, with maximum and minimum values excluded from consideration. Columns denoted with 1 present results of the tests produced through the colorfulness change, and these with 2 through the vividness change. *Image* 1 and *Image* 2 have the parameters (α, β, k_{max}): (0.55,1,5), (0.75,1,10), *Image* 3 is the original with parameters equal to (1,1,k_{orig}), where k_{orig} is the maximal value of kurtosis in the original image, *Image* 4 and *Image* 5 have parameters (α, β, k_{max}) equal to (1,1.3,30), (1,1.6,60), respectively. Note that either (α, β) (for colorfulness change) or (k_{max}) (for vividness change) were varied, while the rest of the parameters were kept constant.

Looking at Table 5.3, we can state that the peaks of the naturalness judgments do not lie within the original image area, which, in turn, brings us to the conclusion that users generally prefer slightly modified images.

To further analyze the results, let us look at the dependency between the naturalness observer perception attribute and the statistical parameters of the spectral images varied in the experiments (colorfulness and vividness). Figure 5.4 presents a plot of the relationship between naturalness, vividness and colorfulness.

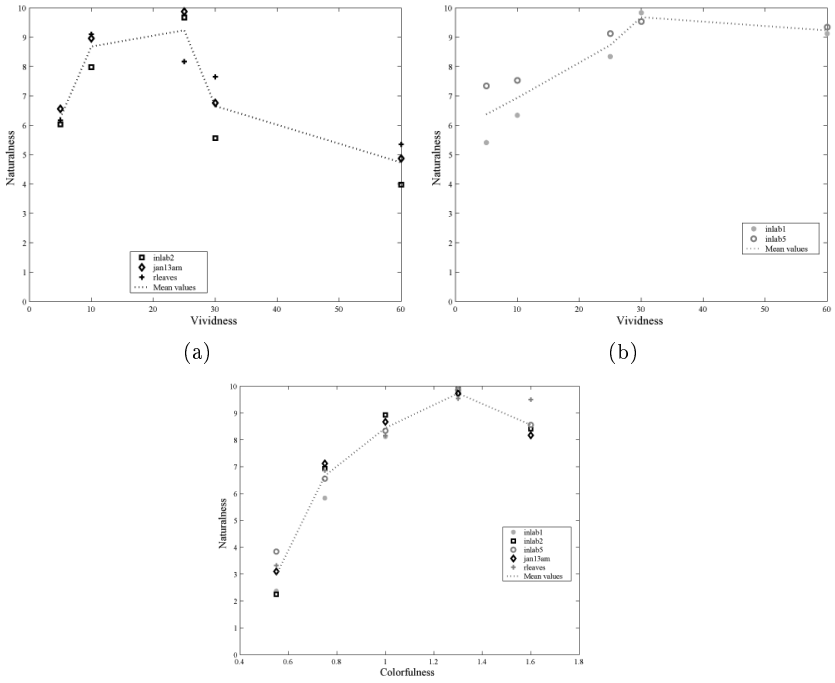


Figure 5.4: Averaged naturalness estimations vs. statistical parameters of spectral images. Vividness (a) of inlab2, jan13am, rleaves; vividness (b) of inlab1 and inlab5; colorfulness (c) of all of the images

Looking at Figure 5.4, we can state that a positive correlation exists between the vividness, the colorfulness of spectral images and the naturalness perceptual image attributes. Figure 5.4(a) and (b) present the relation between the vividness spectral image attribute and naturalness judgments. It is clearly visible that the naturalness maximum lies at points close to the original image. Figure 5.4(a) contains a plot of the inlab2, jan13am, rleaves image judgments, and Figure 5.4(b) - inlab1 and inlab5. We have separated the plots into two parts due to a different form of dependency between the attributes. In the first case images exhibit a sharper drop in the naturalness judgments than in the second one. In fact, in the second case the decrease in naturalness is such that the naturalness remains approximately close to the maximal value. Such a phenomenon can be attributed to the fact that images inlab1 and inlab5 contain objects that attract the most of the observers' attention compared to the objects situated in the background. Moreover, these objects lie in the red area of the spectrum, which means that observers are not susceptible to minor variations in these areas due to the properties of the human visual system. Thus, the drop in quality and in naturalness is less distinct.

Figure 5.4(c) demonstrates a connection between the colorfulness parameter and the naturalness constraint. It can be stated that observers perceive images slightly more colorful than the original ones as being the most natural. Moreover, considering the fact that observers have previously rated images with higher colorfulness values as being of higher quality, we can state that memory color reproduction influences the preferred color reproduction of the objects.

A plot of the quality judgments versus naturalness values for both vividness (red circles) and colorfulness (blue asterisk) test sets (Figure 5.5) was constructed to be able to analyze the correlation of naturalness and quality perception. The data was taken from Tables 5.1- 5.3. Such a comparison is possible due to the fact that experimental settings (number of observers, number and contents of scenes, viewing conditions, etc.) are similar in both of the experiments. Moreover, the algorithm of modification and values of the statistical parameters of spectral images are the same.

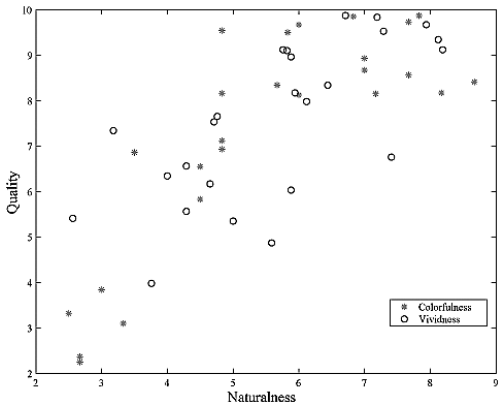


Figure 5.5: Averaged naturalness estimations vs. perceptual quality estimations of images with the vividness (circles) and colorfulness (asterisk) change

Looking at Figure 5.5, we can state that there is a positive dependency between the quality judgments and the naturalness constraint. This means that with the increase of naturalness the quality also increases, which proves the preliminary assumption that in order for the image to be of good quality it should at least be perceived natural. To further illustrate the dependency, a correlation coefficient between the quality judgments and naturalness estimations computed over all of the scenes equals to 0.8196 for colorfulness, 0.7018 in the case of vividness test sets, and the overall correlation is equal to 0.7752. The spread in the plot can be attributed to a lack of test images and a rough scale of spectral image attributes adopted in this study.

Even though we can see from the plot in Figure 5.5 that there is a connection between the naturalness constraint and the quality judgments of the users, it is relatively difficult to predict what would be the effect of the naturalness change on image quality, and how fast the quality decreases with the decrease in naturalness, which in turn can be varied through variation of any of the attributes of spectral images. Accordingly, naturalness could serve as a universal image attribute that would allow modeling both image quality and the joint effect of attributes of spectral images on the overall perception of the image reproduction. In order to model the effect of naturalness on quality, a preference distribution function of naturalness expressed in terms of JNDs (Figure 5.6) has been constructed in a similar manner as with colorfulness and vividness.

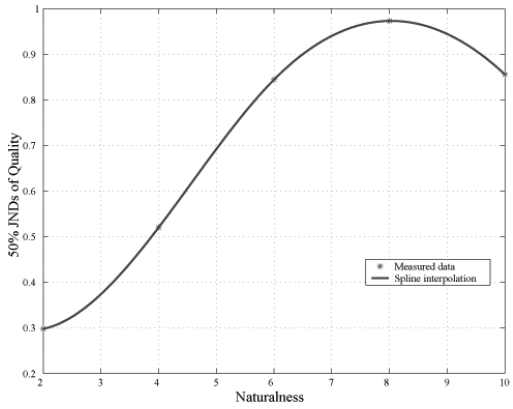


Figure 5.6: Averaged preference distribution function of naturalness

Looking at Figure 5.6 and the results of the previous experiments, the conclusion can be drawn that with the increase of naturalness, the quality of the images increases. However, at a certain point, naturalness starts to decrease. Thus, a high degree of naturalness is a necessary, but not a sufficient condition for the quality perception. A usefulness condition has to be satisfied as well, which in turn leads to a discriminability principle. In other words, even if the image gives an impression of being unnatural, it might be perceived as being of high quality, due to the fact that the information in the image is easily discriminable. Therefore, naturalness has a strong connection with the statistical characteristics of spectral images and image quality perception on the whole, being a necessary, but not a determinative factor.

5.3 Image Fidelity Experiments

The main purpose of image fidelity experiments in this study was to choose the metric that would be able to model human observer perception in the task of color discrimination, given that the measure should work both on color and spectral images.

5.3.1 Image Fidelity Experiments Using Kernel Similarity Measures

Experiments conducted in the framework of image fidelity research using kernel similarity measures consist of three consequent phases, evolving from color differencing to actual image fidelity:

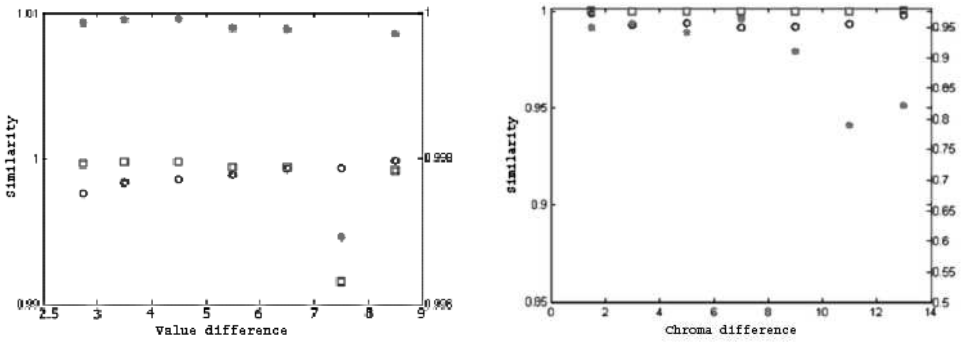
- experiments performed on the Munsell Colors Matte Dataset; [3]
- experiments performed on a set of metameric color samples;
- experiments performed on natural images.

The choice of the dataset of Munsell Colors is not incidental. The Munsell color system is known to be perceptually uniform, meaning that each color is separated from its closest neighbors by an equal perceptual distance [74]. Given the purpose of the experiments, measurements produced against this dataset should produce as even a response as possible. To this end, a series of spectra of Munsell colors with constant values of Hue and Chroma and adjacent values of Value were input into the measures given in Eqs. (2.38-2.52). The primary goal of this part of experiments is to find a metric that would give comparable values of differences for perceptually equally disparate colors, but would also account for the full spectrum of changes in each of the three color dimensions (Hue, Chroma, Value) with a full range of values.

The first set of color samples out of the dataset of Munsell Colors for the experiments was chosen to cover the full range of the Value dimension. The Hue was selected to be 5R, 5B and 5G, the Chroma equal to 1, and the Value in the range of 2.5 to 9. The second set was similar to the previous one, except that the Value and Hue were kept constant while the Chroma was varied. Thus, the Hue was set to 5R, 5B and 5G, the Value to 6, and Chroma in the range of 1 to 14.

Values of the sensitivity parameters of the image fidelity measures in the experiments were set as follows: in Eq. (2.45) and (2.46) $\beta = 0.00012$; in Eq. (2.50) $d = 7$; in Eq. (2.51) $\sigma = 0.6$; and in Eq. (2.52) $k = 1$, $\vartheta = -0.1$. These values are set empirically and will be subject of further experiments.

Based upon the experimental settings described above, diagrams of the functional relations between similarity measures, the Value and Chroma have been obtained (see Figure 5.7). The diagrams show the three best metrics from the point of view of the linearity of the response given to the inputs, meaning that the measures gave close to linear (to a certain extent) responses to the changes in Value and Chroma. The metrics that generated the smoothest responses are Metric 1 (Eq. (2.38)), Metric 8 (Eq. (2.45)) and the Gaussian radial basis function (Eq. (2.52)), presented in Figure 5.7.



(a) (b)

Figure 5.7: Functional relations between similarity measures(a) Similarity versus Value, (b) Similarity versus Chroma (Metric 1 - asterisk (second scale); Metric 8 - Absolute Value Reciprocal Method - circles; Gaussian RBF metric - squares) for 5R Hue (sensitivity parameters set to $\beta = 0.00012$; $d = 7$; $\sigma = 0.6$, k = 1 and ϑ=-0.1)

In Figure 5.7, the asterisks represent Metric 1, circles Metric 8, and squares the Gaussian RBF measure. A surge in the value of the similarity measures observed both with the changes in Value and Chroma at the point of the Chroma difference 10 and 12, and Value difference 7 and 8, can be explained by the effect of metamerism. The spectra of the colors differ by a greater value compared to the visual perception of the difference.

An important aspect of color differencing considered here is the ability to model the human perception of metameric colors. Metamerism is a phenomenon describing a state where two colors with differing spectral power distributions appear to be identical - in other words, different spectra producing equivalent receptor responses [50, 95]. The sources of the metamerism can be different, i.e. illuminant, geometric, observer, and field size [36]. In this work, we consider only the illuminant induced metameric effect since it is easier to model and closer to the ongoing experiments. The assumption in this study is that a successful metric should be able to show the existence of dissimilarities between colors under certain illuminants, at the same time exhibiting similarities under other illuminants. That way the closer the result obtained using the similarity measures is to 1 the better the given measures model metameric color perception. For this purpose, a test was run in *Publication IV* using measures presented in Eqs. (2.38-2.52) against a dataset of metameric colors presented is Section 5.1. The results of the tests are given in Table 5.4.

Each cell in Table 5.4 presents the results of the comparison of the metameric colors in each of the four sets (upper row), measured using the metrics in Eqs. (2.38-2.52) (left side column). Here, the term Set means a sample of metameric colors. Set 1 is the reflectance spectrum of the red metameric sheet, Set 2 of Macbeth metameric sheet, and Sets 3 and 3(A) are reflectance spectra of a part of a passport. Sets 1-3 are taken under illuminant D65, under which the differences between the color patches are not visible, and Set 3(A)

Table 5.4: Values of similarities in paired comparison experiments on metameric colors (sensitivity parameters set to $\beta = 0.00012$; $d = 7$; $\sigma = 0.6$; $k = 1$ and $\vartheta = -0.1$)

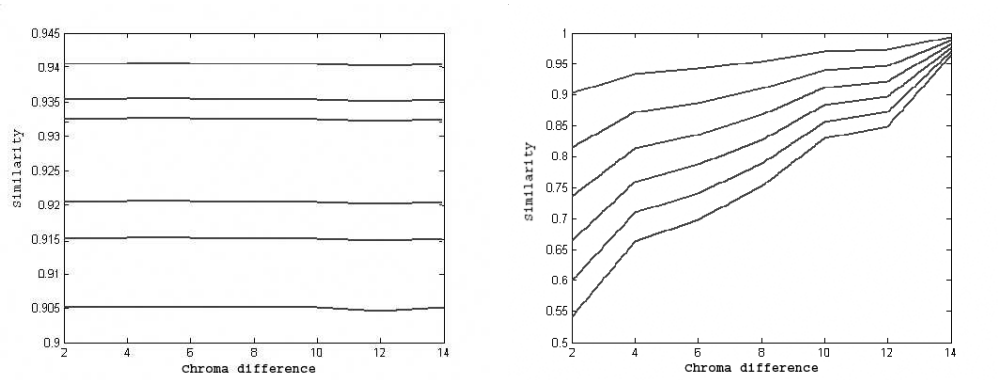
	Set 1	**Set 2**	**Set 3**	**Set 3(A)**
Metric 1	1.0000	1.0000	1.0000	1.0000
Metric 2	0.8108	0.8853	0.8160	0.8153
Metric 3	1.0000	1.0000	1.0000	1.0000
Metric 4	1.0000	1.0000	1.0000	1.0000
Metric 5	0.8882	0.9365	0.8870	0.8860
Metric 6	0.8670	0.8930	0.8635	0.8621
Metric 7	0.6076	0.7680	0.7039	0.7017
Metric 8	1.0000	1.0000	1.0000	1.0000
Metric 9	1.0000	1.0000	1.0000	1.0000
Metric 10	0.8108	0.8853	0.8160	0.8132
Metric 11	0.8882	0.9365	0.8870	0.8854
Metric 12	0.8967	0.9396	0.8973	0.8970
Polynomial	1.0000	1.0000	1.0000	0.9532
Gaussian RBF	1.0000	1.0000	1.0000	0.9643
Sigmoid	0.9989	0.9990	0.9988	0.9843

under illuminant A, under which the difference between color patches becomes visible.

Looking at Table 5.4 and considering the purpose of the experiments, several metrics can be singled out as conforming with the requirement of modeling human observer behavior in image fidelity assessment tasks, meaning the possibility of discrimination of metameric colors under certain illuminants (Set 3(A)) and similarity of colors under other illuminants (Sets 1-3). The most promising results were produced using the Polynomial and Gaussian RBF kernel similarity measures, with the Sigmoid kernel measure performing slightly worse.

Another important requirement imposed on image fidelity measures is the possibility of adjustment of the sensitivity of measurements. Looking at Eqs. (2.38-2.52), it can be stated that the kernel methods (Eqs. (2.50-2.52)), the exponential similarity measure (Eq. (2.44)), the absolute value exponential measure (Eq. (2.45)), and the absolute value reciprocal measure (Eq. (2.46)), possess similarity terms in them. However, the kernel methods provide significantly better control over the whole equation. The results of testing of the sensitivity parameters in the measures for the Gaussian RBF and Metric 8 (absolute value exponent) are presented in Figure 5.8. The choice of the Gaussian RBF and Metric 8 functions for testing is not incidental. Both of these metrics have shown the best results for the smoothness of the response of the sensitivity function requirement, and were among the best in the metamerism experiments compared to the rest of the measure exhibiting sensitivity terms. Thus, it would seem reasonable to continue comparing them.

Functional relations between the similarity values of Gaussian RBF and Metric 8 for different sensitivity values are shown in Figure 5.8, where both β and σ have been varied in the range of $[0.005, 0.03]$ with evenly spaced intervals. Looking at the resulting plots,



(a) (b)

Figure 5.8: Sensitivity of (a) Gaussian RBF (b) Metric 8 (absolute-value exponent) for 5R Hue, Value 6 (β and σ varying in the range 0.005 - 0.03)

it is clear that the sensitivity of the measures can be varied through the use of the special terms in a certain range. However, the absolute value exponential method tends to produce nonlinearities with the growth of the sensitivity value at higher values of the Chroma and Value, which is an undesirable effect in this case.

The parameters of kernel similarity measures were set manually, but they are far from optimal. The goal of the following experiment was to find the coefficients in kernel similarity measures formulae that would give equal proportions between a set of differences, independent of the color space the colors are given in [29]. For that purpose, Eqs. (2.50-2.52) were refactored into the following form [29]:

$$S_{polynomial} = (ay_i, by_j)^d, \tag{5.1}$$

where a, b and d are parameters of the equation to be optimized. The resulting values after optimization using the Matlab optimization toolbox are $a = b = 1$; $d = 7$.

The sigmoid kernel similarity measure was transformed into

$$S_{sigmoid} = \tanh(\langle y_i, y_j \rangle k + \vartheta), \tag{5.2}$$

where k and ϑ are parameters for optimization. Their values are $k = 1$ and $\vartheta = 0.7$ after optimization.

The Gaussian RBF kernel measure was modified into the following form:

$$S_{Gaussian} = \exp\left(- a \frac{\|y_i - y_j\|^b}{2\sigma^2} \right), \tag{5.3}$$

where a, b and σ are optimization parameters, equal to $a = 1$, $b = 2$, and $\sigma = 0.707$ as a result.

The optimization has been performed on a Munsell dataset [3] against CIE L*a*b*, CIE L*u*v* and RGB color spaces.

The next part of the image fidelity experiments consisted of applying kernel similarity measures and twelve conventional measures to images of natural scenes described in Section 5.1. Kernel similarity measures were used here to produce a Spectral Image Distortion Map (SIDM). What is calculated here, in fact, is a pixelwise spectral distortion. As a result, a gray-scale spectral distortion image is obtained, where the intensity of each of the pixels is a difference between the original image and the distorted one.

To imitate color distortions all of the images were compressed using the principal component analysis (PCA) algorithm down to two principal components. PCA is a technique that allows extracting smaller sets of variables with less redundancy from a higher dimensional data. Color reproductions of original images and reconstructions after compression are given in Figure 5.9.

The areas of color difference are clearly visible in the images, and concentrate primarily in red and brown regions. Given that the overall coloring of the image is green the first two eigenvectors contain the spectral information of the main colors. By adding the third the image would have retained a color scheme very close to the original.

The images thus obtained were used to produce SIDM using each of the three kernel similarity measures. Figure 5.10 (a, d) present the result of the application of the polynomial kernel, (b,e) the Gaussian radial basis function, and (c,f) the sigmoidal kernel. The level of intensity in the maps corresponds to the similarity scale: from black "not similar at all" to white "identical".

Comparing the maps obtained, it can be stated that they present a similar result to a certain extent. Moreover, comparing the original and the compressed image it is possible to select the areas that are different in the images, and these are similar to the results obtained with the use of the image distortion maps.

To assess the accuracy of the SIDM, the subjective image quality technique the PIDM described in Chapter 4 is used. Five subjects were presented two sets of images, consisting of an original and a compressed image. The users were asked to mark the regions that appeared different with a rectangular digital marker of the size 4 by 4 pixels with different levels of gray-level intensity. Black means "not similar at all" and white "identical". The subjects were instructed to mark the whole image area. Figure 5.11 presents the mean of all subject maps.

Figure 5.11 clearly indicates that the PIDM presents a practically excellent fit to the SIDM calculated through the use of the Gaussian RBF. Nevertheless, certain errors exist, which can be attributed to the fact that the marker size and shape caused several inaccuracies in stamping identical regions several times. To further illustrate the accuracy of the SIDM, the differences between the mean value of the PIDM and each of the three SIDM maps have been computed (see Table 5.5). The last column presents the value of the deviation of the normalized SCIELAB error image from the PIDM.

The PIDM presents a full map of empirical distortion data, which can be used in the task of evaluation of the accuracy of the metrics presented. Thus, looking at Table 5.5 it

Table 5.5: Comparison of SIDM, SCIELAB and PIDM

	Polynomial	**Gaussian RBF**	**Sigmoidal**	**SCIELAB**
inlab 2	0.0499	0.0395	0.0551	0.0606
inlab 5	0.0395	0.0291	0.0581	0.0609

can be concluded that the most accurate evaluation of the human response in the quality estimation task is obtained through the use of the Gaussian RBF kernel and the least accurate with SCIELAB, although the deviation between the two is not large.

From the point of view of the probability theory it can be stated that the SIDM presents the probability of the subject identifying a certain pixel as similar, which allows avoiding time and money consuming expert surveys, and gives the possibility of computing the distortion values dynamically as the information is discarded from the image, as for example in a lossy compression task.

5.3.2 Image Fidelity Evaluation Using 3D-SSIM

The second part of the image fidelity experiments is dedicated to applying the SSIM (see Section 2.2) to spectral image assessment.

The goal of the experiments was to find the best way of extending the conventional SSIM (Eq. (2.30)). The first approach assumes that conventional SSIM is applied band by band and the result is averaged over the whole image. The second option is to apply the three-dimensional window given in Eq. (2.31).

For the purpose of this experiment, original and modified images from the previous experiment were used (see Figure 5.9). The result of applying conventional and 3D-SSIM is shown in Figure 5.12.

Looking at Figure 5.12, it is clear that 3D-SSIM presents a clearer picture of the distortions introduced into the image. However, the visual inspection of the results of the tests is not enough to establish a clear picture of the efficiency of either of the modifications of SSIM. For that reason, the PIDM out of the previous experiments was applied. Again an average over the whole PIDM was computed and subtracted from the average over the 3D-SSIM and SSIM maps. The results of the previous experiment and the SSIM experiment are given in Table 5.6.

Table 5.6: Comparison of SIDM, SSIM, 3D-SSIM and PIDM

	Polynomial	**Gaussian RBF**	**Sigmoidal**	**3D-SSIM**	**SSIM**
inlab 2	0.0499	0.0395	0.0551	0.1636	0.2374
inlab 5	0.0395	0.0291	0.0581	0.0986	0.1745

Each cell in Table 5.6 represents the differences of the averages of kernel measures and both SSIM based measures from the average computed over PIDMs of the appropriate images. Thus, the closer the value in each of the cells to 0 the closer the measure is to PIDM. The range of possible values in this table is from 0 to 1. Looking at Table

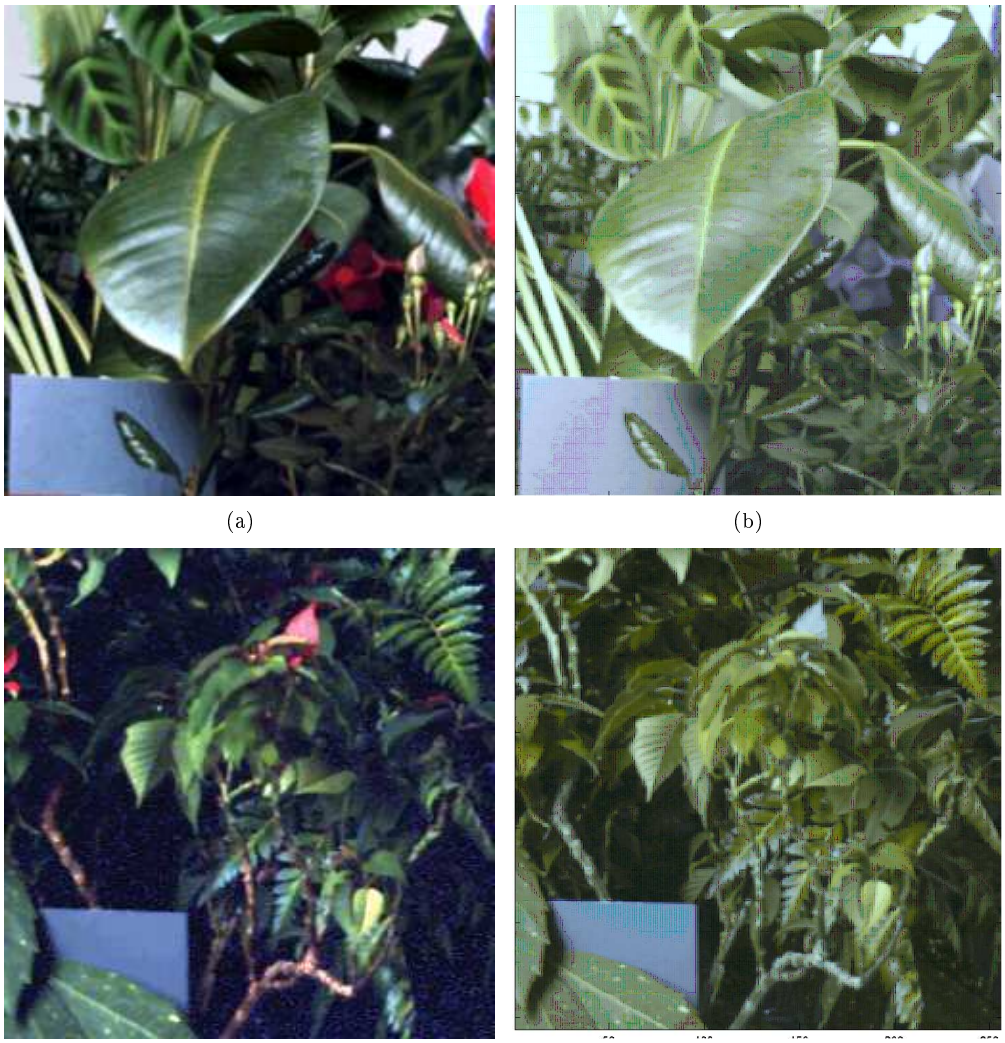
5.6, it can be stated that the 3D-SSIM performs significantly better compared with the conventional SSIM applied band by band. However, it is still worse than kernel similarity measures, and specifically the Gaussian RBF metric.

5.4 Summary and Discussion

In this chapter, we described the experiments performed in the framework of this thesis, intended to evaluate the theoretical assumptions presented in previous chapters. The experiments performed can be divided into two broad parts: objective image quality and image fidelity.

Image quality experiments concentrate on applying the Statistical Image Model to images of natural scenes. The tests performed correspond to the three image attributes described in the model: colorfulness, vividness, and naturalness. Several conclusions can be drawn based on the results of the experiments. Primarily, the results of the tests indicate that vividness, colorfulness and naturalness are positively correlated. Furthermore, the peaks of naturalness in vividness experiments lie close to the original images, while with colorfulness slightly modified images are perceived as more natural. Moreover, the experiments show that observers generally prefer slightly more saturated images with deeper shadows, meaning that higher values of colorfulness and vividness result in higher quality judgments. Furthermore, with the increase of naturalness the perceived quality increases. However, at a certain point naturalness starts to decrease. Thus, a high degree of naturalness is a necessary, but not a sufficient condition for a high quality perception. A usefulness condition has to be satisfied, as well, which in turn leads to a discriminability principle. Meaning that, even if the image gives an impression of being unnatural, it might be perceived as being of high quality due to the fact that the information in the image is easily discriminable. Therefore, naturalness has a strong connection with the statistical characteristics of spectral images and image quality perception on the whole, being a necessary, but not a determinative factor.

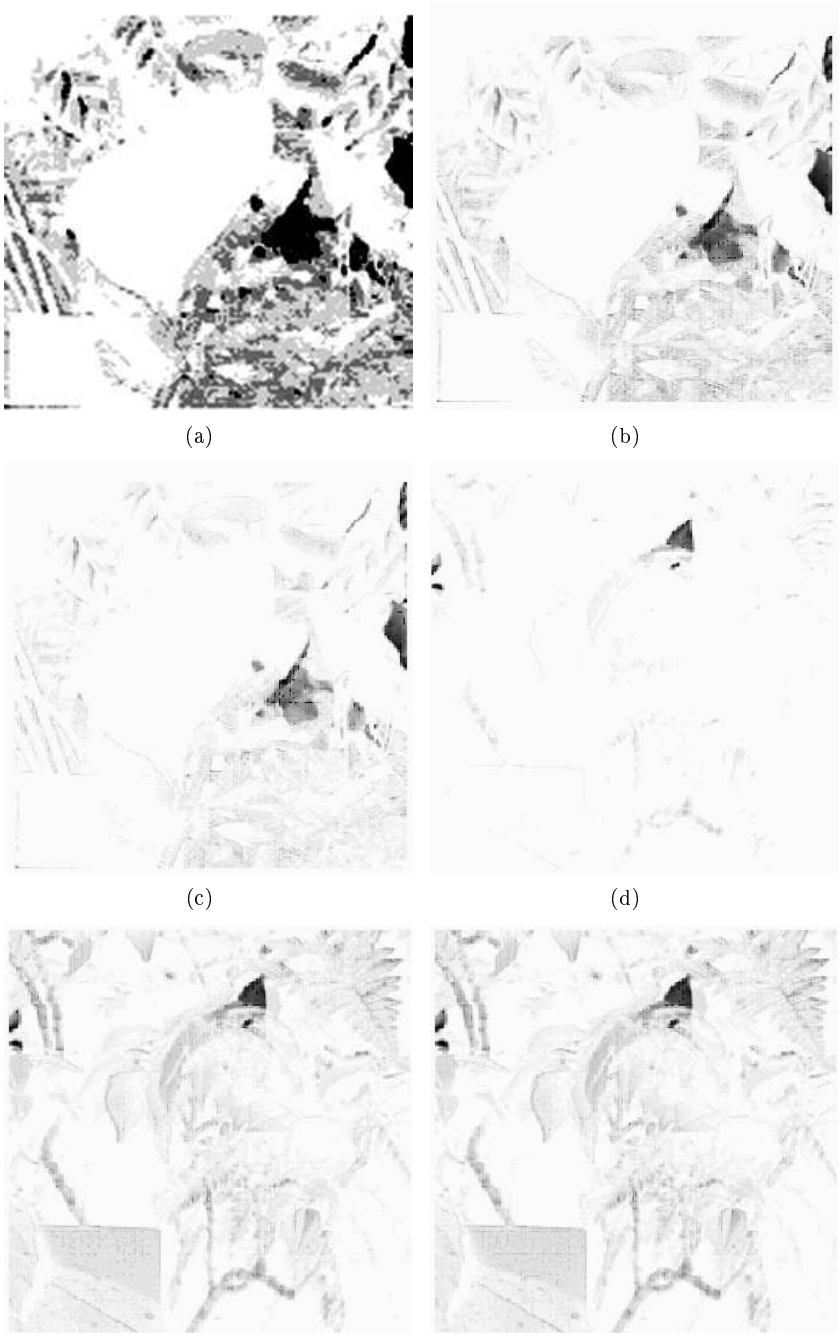
The goal of image fidelity experiments is to find a measure that would model the human observer behavior in the task of color discrimination and would also apply to both color and spectral images. The criteria for the choice of the measure was that the measure should give comparable values of difference for perceptually equally disparate colors, and would also account for the full spectrum of change in the color. The optimal measure should also have the possibility of adjusting the sensitivity of the assessment. Another important aspect taken into account was to model the human observer in the task of discrimination of metameric colors. The measures used are twelve conventional color similarity measures (Eqs. (2.38-2.49)), SCIELAB (2.23), kernel similarity measures (Eqs. (2.50-2.52)) and the novel measure 3D-SSIM the latter two proposed in this thesis. Based on the results obtained, it can be concluded that the most accurate evaluation of the human response in the quality estimation task is obtained through the use of the Gaussian RBF kernel and the least accurate with SCIELAB, although the deviation between them is not large.



(a) (b)

(c) (d)

Figure 5.9: Color reproduction of spectral images inlab2 (a) original, (b) reconstruction after compression (PCA 2); inlab5 (c) original, (d) reconstruction after compression (PCA 2)



(a) (b)

(c) (d)

(e) (f)

Figure 5.10: Spectral image distortion maps. inlab2 (a,b,c); inlab5 (d,e,f). (a,d) polynomial kernel; (b,e) Gaussian radial basis function and (c,f) sigmoidal kernel

(a) (b)

Figure 5.11: Perceptual image distortion maps. Inlab2 (a); inlab5 (b)

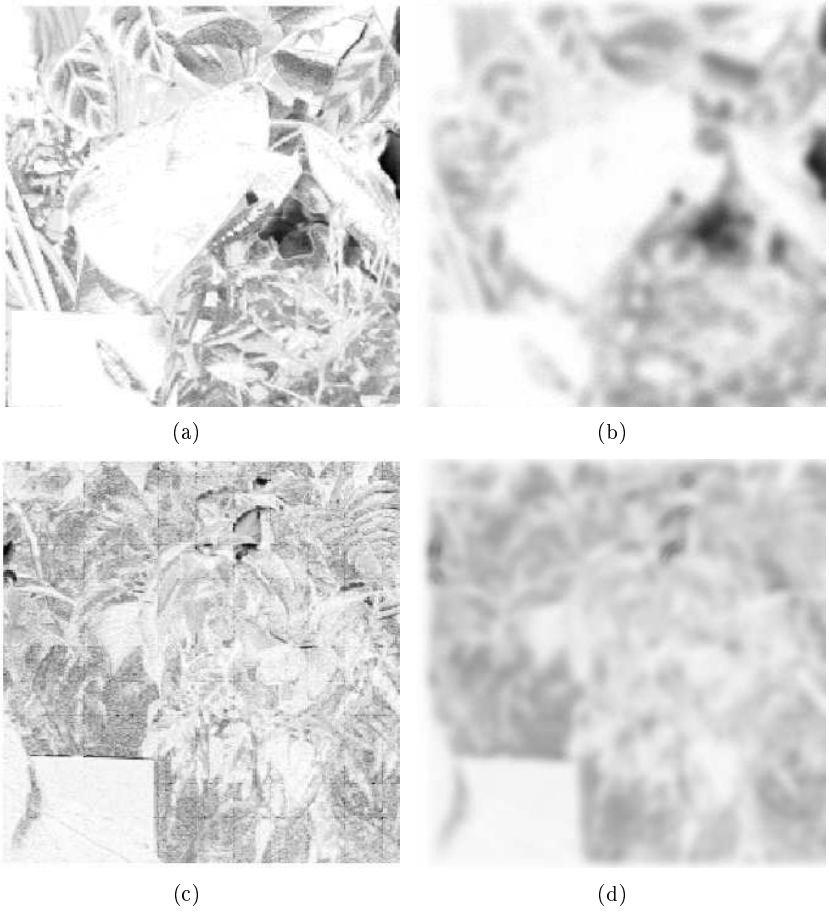


Figure 5.12: SSIM difference maps. inlab2 (a,c) and inlab5 (b,d). 3D-SSIM (a,b) and averaged SSIM (c,d)

Chapter VI

Discussion

The main objective of this research was to try to find a solution to the problem of quantifying quality and fidelity of color and spectral images. The primary focus of this thesis lies in spectral domain distortions, not taking into consideration spatial artefacts. Evolving from a study of a relationship between spectral image appearance and the statistical characteristics of images under study, a Spectral Appearance Model that allows the assessment and prediction of spectral image quality, has been developed. This direction of research was later on supplemented with a study of the possibility of applying conventional color image fidelity measures to spectral images, which, in turn, eventually led to the development of novel kernel based similarity measures. This work can be divided into two broad parts, stemming from the above-mentioned directions of the research.

The first part of this thesis is dedicated to image fidelity. Fidelity is defined in this work as the degree of an apparent match of the image under study with an external reference, also called the original [97]. A number of conventional image fidelity measures served as the basis for this research, forming the theoretical basis for the novel measures proposed in this work. Conventional measures primarily include color and spectral image fidelity measures, and metrics that can be applied to both gray-scale and color images. In the framework of this thesis, all of these were divided into three broad categories, taking classification proposed by Avcibas [16] as an underlying principle. The resulting categories are: pixel difference based measures, measures accounting for image structure, and correlation based measures.

Pixel based measures, as the name implies, comprise the measures calculated based on the pixelwise differences between the images [16]. These include almost all of the most popular fidelity metrics existing in literature: the MSE, PSNR [55], CIE ΔE L*a*b* [4], and extensions of these measures - BFD($\ell : c$) [71, 72], CMC($\ell : c$) [23], LCD [66, 78], CIE94 ΔE [70], and the recently proposed CIEDE2000 [70].

Another category of image fidelity measures considered in this thesis are correlation based measures, which include twelve conventional image fidelity metrics proposed in [81, 67]. These are based on the assumption that correlation between images is related to overall

image fidelity. The measures, including among others the Exponential similarity method, the Correlation Coefficient method and the Absolute-value reciprocal method, were designed and applied primarily to color images. The possibility of applying these metrics to the task of spectral image assessment has been tested in *Publication II* and *Publication IV*. Furthermore, a set of novel kernel similarity measures has been proposed in these publications and further developed into a Spectral Image Distortion Map in *Publication III*. The novel metrics include the Gaussian RBF, polynomial and sigmoid kernels. These measures are based upon a well-known pattern recognition technique - kernel support vector machines [88]. A preliminary result was obtained in *Publication II* and *Publication IV* stating that novel kernel similarity measures outperform conventional correlation based measures in the task of spectral image fidelity assessment. Furthermore, the Gaussian RBF produced the most accurate assessments considering traditional color similarity methods.

Most of the methods of image fidelity evaluation given above and described in literature concentrate on color artefacts. A significant number of measures exist in the field of spatial distortions. However, the influence of spatial properties of images on color reproduction is largely ignored [36], with a few exceptions to this rule [60, 80, 94, 102]. Measures accounting for image structure include an extension of the conventional CIE ΔE L*a*b* - S-CIELAB [102], a measure created for the assessment of spectral images - the BDMM [61], SSIM [94], and 3D-SSIM, an extension of the measures described in this thesis, initially proposed in *Publication VII*. For further development of the fidelity research, we have proposed to either combine the existing measures proposed in *Publication IV* with a spatial assessment method or extend the latter to a three-dimensional case to be able to incorporate spectral data. A 3D-SSIM method that calculates conventional SSIM [94] using a three-dimensional window was obtained as a result.

A number of experiments performed on the fidelity measures have proven that kernel similarity measures proposed in this thesis, and designed to quantify specific spectral distortions, provide a closer match to the human observer behavior in the task of color discrimination and fidelity assessment compared to the rest of the tested measures. However, the results also indicate that further improvement of the measures is possible.

The second part of the research is dedicated to image quality. This notion can be approached from various angles which results in a number of various definitions [64, 89]. The source of the ambiguity of the term stems from the fact that quality is an intuitive concept, closely connected with human perception [94]. The definition chosen for this thesis was given in [64] and states that quality is "an impression of [image] merit or excellence, as perceived by an observer neither associated with the act of ... [acquisition], nor closely involved with the subject matter depicted" [64], with several restrictions imposed, i.e. by images in this case we assume their digital form, and the end users are considered to be human observers.

The main question facing any image quality research is whether the quality could be readily quantified, and if so, what the attributes are that characterize it [30]. Both of the issues are addressed in this work. This study began by addressing the question of whether spectral image appearance can be affected by statistical image attribute variation. This research was further developed into the study of the assessment and prediction of spectral image quality.

An excellent framework for any image quality research was proposed by Engeldrum [30], i.e. the image quality circle. This concept connects technology variables with customer quality preference via customer perception attributes and image quality models [30]. This thesis concentrates on finding these two parts and the connection to customer quality preference.

One of the most significant problems of image quality research is the choice of the attributes of observer perception, or so called "nesses" [32]. Preferential and artefactual image attributes are the main focus of this thesis. These include blur, noisiness, fidelity, detectability, naturalness, colorfulness and vividness.

The second part of the circle - the models considered in this work are the FUN model [97], computational image model [55], Kayargadde model [63], and spectral color appearance model proposed in this thesis.

The spectral color appearance model proposed in this work describes quality in terms of preferential image attributes, i.e. colorfulness, vividness, and naturalness. The model is based upon the assumption that statistical characteristics of spectral images affect the overall appearance. Kurtosis variation affects image highlight and shadow intensity, called vividness in this work. And standard deviation impacts image contrast and saturation, i.e. colorfulness. The third parameter of the spectral color appearance model stems from the intuitive assumption that in order for the image to be of high quality it should at least be perceived as natural [57].

The results of the experiments performed in the framework of this work indicate that the three parameters of the spectral color appearance model, i.e. colorfulness, vividness and naturalness, are positively correlated. Moreover, with the increase of naturalness the quality of the images increases. However, at a certain point, naturalness starts to decrease. Thus, a high degree of naturalness is a necessary, but not a sufficient condition for quality perception. A usefulness condition has to be satisfied as well, which in turn leads to the discriminability principle. In other words, even if the image gives the impression of being unnatural, it might be perceived as being of high quality due to the fact that the information in the image is easily discriminable. Therefore, naturalness has a strong connection with the statistical characteristics of spectral images and image quality perception on the whole, being a necessary, but not a determinative factor. This means that although an image might be unnatural, as long as it is discriminable it might be considered of high quality. The trade-off of these two conditions produces the most optimal result. Overall, the Spectral Appearance Model proposed in this thesis has proven to be capable of predicting and affecting image appearance and image quality, with the help of the statistical image attributes.

The future research direction in the case of both image fidelity and image quality is connected to the problem of incorporating spatial data into the assessments. Preliminary results obtained in this area show promising results. Four centuries of the history of image quality and image fidelity research could have brought a solution to the problem of effective image assessment. However, we are as far from a complete solution as we were back then [32].

[1] *Joensuu spectral database of Munsell colors.* University of Joensuu Color Group, http://spectral.joensuu.fi/. Referred on 12.2004.

[2] CIE(1932). Commission Internationale de l'Eclairage proceedings. Cambridge University Press, 1931.

[3] *Munsell Book of Color - Matte Finish Collection.* Munsell Color, 1976.

[4] CIE 1978 Recommendations on uniform color spaces, color difference equations, psychometric color terms. Supplement No. 2 to CIE publication No. 15 (E.-1.3.1) 1971/(TC-1.3), 1978.

[5] CIE 15.2-1986 Part 2: CIE standard illuminant D65 relative spectral power distribution. CIE Publication 15.2, 1986.

[6] CIE 1988 2 degree spectral luminous efficiency function for photopic vision. CIE Publication 086, 1990.

[7] ITU-R BT.500-7 methodology for the subjective assessment of the quality of television pictures, 1995.

[8] ISO 3664:2000 viewing conditions - graphic technology and photography. International Organization for Standardization, 2000.

[9] CIE improvement to industrial colour-difference evaluation. CIE Publication No. 142-2001, 2001.

[10] ISO 20462-1:2005, photography – psychophysical experimental method for estimating image quality – part 1: Overview of psychophysical elements. International Organization for Standardization, 2003.

[11] *Merriam-Webster's Collegiate Dictionary.* Merriam-Webster, Inc., 2003.

[12] AHUMADA, A. J., AND NULL, C. H. Image quality: a multidimensional problem. In *Digital Images and Human Vision*, A. B. Watson, Ed. MIT Press, 1993, ch. 11, pp. 141–148.

[13] ALLNATT, J. W., AND LEWIS, N. W. Subjective quality of television pictures with multiple impairments. *Electronics Letters 1*, 7 (1965), 187–188.

[14] Androutos, D., Platanitis, S. K. N., and Venetsanopoulos, A. N. Distance measures for color image retrieval. In *IEEE International Conference on Image Processing* (Chicago, IL, USA, 1998), vol. 2, IEEE, pp. 770–774.

[15] Avcibaş, I., and Sankur, B. Statistical analysis of image quality measures. In *10th European Signal Processing Coneference (EUSIPCO'2000)* (Tampere, Finland, 2000), pp. 2181–2184.

[16] Avcibaş, I., Sankur, B., and Sayood, K. Statistical evaluation of image quality measures. *Journal of Electronic Imaging 11*, 2 (2002), 206–223.

[17] Barten, P. G. J. *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality.* SPIE Press, 1999.

[18] Bäuml, K. H., and Wandell, B. A. The color appearance of mixture gratings. *Vision Research 36*, 18 (1996), 2849–2864.

[19] Berns, R. S. Derivation of a hue-angle dependent, hue-difference weighting function for CIEDE2000. In *9th Congress of the International Colour Association* (Rochester, USA, 2001), vol. 4421, SPIE, pp. 638–641.

[20] Bonnier, N., Schmitt, F., Brettel, H., and Berche, S. Evaluation of spatial gamut mapping algorithms. In *14th Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications (CIC'06)* (Scottsdale, USA, 2006), pp. 56–61.

[21] Bovik, A. C. *Handbook of Image and Video Processing.* Academic Press, 2000.

[22] Carevic, D., and Caelli, T. Region based coding of color images using Karhunen-Loeuve transform. *Graphical models and image processing 59, Issue 1* (1997), 27–38.

[23] Clarke, F. J. J., McDonald, R., and Rigg, B. Modification to the JPC79 colour-difference formula. *Journal of the Society of Dyers and Colourists 100, Issue 4* (1984), 128–132.

[24] de Ridder, H. Naturalness and image quality saturation and lightness variation in color images. *The Journal of Imaging Science and Technology 40*, 6 (1996), 487–493.

[25] de Ridder, H. Image processing and the problem of quantifying image quality. In *IEEE International Conference on Image Processing (ICIP'01)* (Thessaloniki, Greece, 2001), vol. 2, IEEE, pp. 3–6.

[26] Deffner, G., Yuasa, M., McKeon, M., and Arndt, D. Evaluation of display-image quality: experts vs. non-experts. *SID Symposium Digest 25* (1994), 475–478.

[27] Dijk, J. *In search of an objective measure for the perceptual quality of printed images.* PhD thesis, Delft Technical University, Delft, Holland, 2004.

[28] Dochev, D., Bochko, V., Kalenova, D., Toivanen, P., and Kaarna, A. 3D similarity index for evaluating quality of lossy compressed spectral images. In *Third European Conference on Color in Graphics, Imaging and Vision (CGIV'06)* (Leeds, UK, 2006), Society for Imaging Sciences and Technology, pp. 199–204.

[29] Dochev, D., Bochko, V., Kalenova, D., Toivanen, P., and Kaarna, A. Defining optimized formulas for equal differences in color space and in spectral space. In *The second international workshop on image media quality and its applications* (Chiba, Japan, 2007), pp. 113–119.

[30] Engeldrum, P. G. Measuring customer perception of print quality. *TAPPI journal 73*, 3 (1990), 161–164.

[31] Engeldrum, P. G. A framework for image quality models. *Journal of Imaging Science and Technology 39*, 4 (1995), 312–318.

[32] Engeldrum, P. G. Image quality modeling: Where are we? In *Image Processing, Image Quality and Image Capture Systems (PICS'99)* (Savannah, USA, 1999), Society for Imaging Sciences and Technology, pp. 251–255.

[33] Engeldrum, P. G. Extending image quality models. In *Image Processing, Image Quality and Image Capture Systems (PICS'02)* (Portland, USA, 2002), vol. 5, IS&T, pp. 65–69.

[34] Eskicioglu, A. M. Quality measurement for monochrome compressed images in the past 25 years. In *2000 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Istanbul, Turkey, 2000), vol. 4, pp. 1907–1910.

[35] Eskicioglu, A. M., and Fisher, P. S. Image quality measures and their performance. *IEEE Transactions on communications 43*, 12 (1995), 2959–2965.

[36] Fairchild, M. Modelling colour appearance, spatial vision, and image quality. In *Color image science: exploiting digital media*, L. W. MacDonald and M. R. Luo, Eds. Wiley, 2002, ch. 17, pp. 357–370.

[37] Fairchild, M. D., and Ebner, F. Development and testing of a color space (IPT) with improved hue uniformity. In *Color Imaging Conference: Color science, systems and applications (CIC'98)* (Scottsdale, USA, 1998), vol. 6, IS&T, SID, pp. 8–13.

[38] Farell, J. E. Image quality evaluation. In *Color Imaging: Vision and Technology*, L. W. MacDonald and M. R. Luo, Eds. Wiley, 1999, ch. 15, pp. 285–314.

[39] Fedorovskaya, E. A., Blommaert, F. J. J., and de Ridder, H. Perceptual quality of color images of natural scenes transformed into CIELUV color space. In *Color Imaging Conference: Color science, systems and applications (CIC'93)* (Scottsdale, Arizona, USA, 1993), vol. 1, Society for Imaging Sciences and Technology, SID, pp. 37–40.

[40] Fedorovskaya, E. A., de Ridder, H., and Blommaert, F. J. J. Chroma variations and perceived quality of color images of natural scenes. *Color research and applications 22*, 2 (1997), 96–110.

[41] Ford, A. M. Determination of compressed image quality. In *Colour Imaging: Vision and Technology*, L. W. MacDonald and M. R. Luo, Eds. Wiley, 1999, ch. 16, pp. 315–338.

[42] Frese, T., Bouman, C. A., and Allebach, J. P. Methodology for designing image similarity metrics based on human visual system models. In *Conference on Human Vision and Electronic Imaging II* (San Jose, USA, 1997), B. E. Rogowitz and T. N. Pappas, Eds., vol. 3016, pp. 472–483.

[43] Georgeson, M. A., and Sullivan, G. D. Contrast constancy: deblurring in human vision by spatial frequency channels. *Journal of Physiology*, 252 (1975), 627–656.

[44] Girod, B. What's wrong with mean-squared error. In *Digital images and human vision*, A. B. Watson, Ed. MIT press, 1993, ch. 15, pp. 207–220.

[45] Gonzalez, R. C., and Woods, R. E. *Digital Image Processing*, 2nd ed. Prentice Hall, 2002.

[46] Hardeberg, J. Y., and Gerhardt, J. Characterization of an eight colorant inkjet system for spectral color reproduction. In *Second European Conference on Colour in Graphics, Imaging and Vision (CGIV'04)* (Aachen, Germany, 2004), Society for Imaging Sciences and Technology, pp. 263–267.

[47] Hauta-Kasari, M. *Computational techniques for spectral image analysis*. PhD thesis, Lappeenranta University of Technology, Lappenranta, Finland, 1999.

[48] Hild, M. On the effectiveness of color similarity measures in background-frame differencing applications. In *First European Conference on Colour in Graphics, Imaging and Vision (CGIV'02)* (Poitier, France, 2002), IS&T, pp. 304–309.

[49] Hild, M., and Emura, T. Which color similarity measure is most effective for background-frame differencing? In *Color Imaging Conference: Color science, systems and applications (CIC'01)* (Scottsdale, AZ, USA, 2001), Society for Imaging Sciences and Technology, SID, pp. 168–173.

[50] Hunt, R. W. G. *The Reproduction of Color*. Wiley, 2004.

[51] Ishihara, S. *Tests for colour-blindness*. Handaya Tokyo, Hongo Harukicho, 1917.

[52] Ishihara, T., Oishi, K., Tsumura, N., and Miyake, Y. Dependence of directivity in spatial frequency response of the human eye (1), (2). *Journal of the Society of Photographic Science and Technology of Japan 65*, 2 (2002), 121–127, 128–133.

[53] Ivkovic, G., and Sankar, R. An algorithm for image quality assessment. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)* (Quebec, Canada, 2004), vol. 3, IEEE, pp. 713–716.

[54] Jacobson, R., and Triantaphillidou, S. Metric approaches to image quality. In *Color image science: exploiting digital media*, L. W. MacDonald and M. R. Luo, Eds. Wiley, 2002, ch. 18, pp. 371–392.

[55] Janssen, R. *Computational image quality.* SPIE Press, 2001.

[56] Janssen, T. J. W. M. Understanding image quality. In *International Conference on Image Processing (ICIP'01)* (Thessaloniki, Greece, 2001), vol. 2, IEEE, p. 7.

[57] Janssen, T. J. W. M., and Blommaert, F. J. J. Image quality semantics. *Journal of Imaging Science and Technology 41*, 5 (1997), 555–560.

[58] Janssen, T. J. W. M., and Blommaert, F. J. J. Predicting the usefulness and naturalness of color reproductions. *Journal of Imaging Science and Technology 44*, 2 (2000), 93–104.

[59] Janssen, T. J. W. M., and Blommaert, F. J. J. Visual metrics: discriminative power through flexibility. *Perception 29, Issue 8* (2000), 965–980.

[60] Jin, E. W., Feng, X. F., and Newell, J. The development of a color visual difference model (CVDM). In *Image Processing, Image Quality, Image Capture, Systems Conference (PICS'98)* (Portland, USA, 1998), The Society for Imaging Sciences and Technology, pp. 154–158.

[61] Kaarna, A., and Parkkinen, J. Blockwise distortion measure for lossy compression of multispectral images. In *X European Signal Processing Conference (EUSIPCO 2000)* (Tampere, Finland, 2000), pp. 2197–2200.

[62] Kaarna, A., and Parkkinen, J. Quality measure for multispectral image compression. In *First International Conference in Graphics and Image Processing (CGIP'2000)* (Saint-Etienne, France, 2000), pp. 274–279.

[63] Kayargadde, V. *Feature extraction for image quality prediction.* PhD thesis, Eindhoven Technical University, Eindhoven, Holland, 1995.

[64] Keelan, B. W. *Handbook of image quality: characterization and prediction.* Marcel Dekker, Inc., 2002.

[65] Keelan, B. W., and Urabe, H. ISO 20462, a psychophysical image quality measurement standard. In *Image Quality and System Performance* (San Jose, USA, 2004), Y. Miyake and D. R. Rasmussen, Eds., vol. 5294, SPIE–IS&T, pp. 181–189.

[66] Kim, D. H., and Nobbs, J. H. New weighting functions for the weighted CIELAB colour difference formula. In *Proceedings of 8th Congress of International Colour Association (AIC Colour)* (Kyoto, Japan, 1997), vol. 1, pp. 446–449.

[67] Leung, Y. *Spatial Analysis and Planning under Imprecision.* North-Holland Publishers, 1988.

[68] Linfoot, E. H. *Fourier methods in optical image evaluation*, revised ed. The Focal Press, 1964.

[69] Lohmann, A. W., Mendlovic, D., and Shabtay, G. Significance of phase and amplitude in the Fourier domain. *Journal of Optical Society of America A*, 14, Issue 11 (1997), 2901–2904.

[70] Luo, M. R., Cui, G., and Rigg, B. The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color research and application 26*, 5 (2001), 340–350.

[71] Luo, M. R., and Rigg, B. BFD(l:c) colour-difference formula, Part 1- development of the formula. *Journal of the Society of Dyers and Colourists*, Issue 2, 103 (1987), 86–94.

[72] Luo, M. R., and Rigg, B. BFD(l:c) colour-difference formula, Part 2- performance of the formula. *Journal of the Society of Dyers and Colourists*, 103, Issue 3 (1987), 126–132.

[73] MacAdam, D. L. Quality of color reproduction. In *Proceeding of the Institute of Radio Engineers* (1951), vol. 39, Issue 5, pp. 468–485.

[74] Malacara, D. *Color Vision and Colorimetry: Theory and Applications.* SPIE Press, 2002.

[75] Miyake, Y., and Nakagawa, S. Evaluation of image quality based on the human visual characteristics. *Transactions of the Institute of Electronics, Information and Communication Engineers. A J89-A, Issue 11* (2006), 858–865.

[76] Myake, Y., and Bai, J. Application of human visual characteristics to evaluation and designing of imaging system. *Journal of the Imaging Society of Japan 41*, 4 (2002), 333–342.

[77] Nill, N. B., and Bouzas, B. H. Objective image quality measure derived from digital image power spectra. *Optical Engineering 31, Issue 4* (1992), 813–825.

[78] Noriega, L. A., Heptinstall, A. R., Luo, M. R., and Westland, S. The perception of achromatic differences. In *Colour Science 98, Colour Physics Symposium* (Leeds, UK, 1998), vol. 3, pp. 195–204.

[79] Párraga, A., Brelstaff, G., Troscianko, T., and Moorhead, I. Color and luminance information in natural scenes. *Journal of the Optical Society of America A 15*, 3 (1998), 563–569.

[80] Pattanaik, S. N., Fairchild, M. D., Ferwerda, J. A., and Greenberg, D. P. Multiscale model of adaptation, spatial vision, and color appearance. In *Color Imaging Conference: Color science, systems and applications (CIC'98)* (Scottsdale,vUSA, 1998), The Society for Imaging Sciences and Technology, SID, pp. 2–7.

[81] Plataniotis, K., N., and Venetsanopoulos, A., N. *Color Image Processing and Applications.* Springer-Verlag, 2000.

[82] Poirson, A. B., and Wandell, B. A. Pattern-color separable pathways predict sensitivity to simple colored patterns. *Vision Research 36*, 4 (1996), 515–526.

[83] Popat, K., and Picard, R. W. Cluster based probability model and its application to image and texture processing. *IEEE transactions on image processing Issue 2, 6* (1997), 268–284.

[84] Pratt, W. K. *Digital Image Processing.* Wiley, 1978.

[85] Riskey, D. R. Use and abuses of category scales in sensory measurement. *Journal of Sensory Studies Issue 3-4, 1* (1986), 217–236.

[86] Rosen, M. R., Hattenberger, E., and Ohta, N. Spectral redundancy in a 6-ink inkjet printer. In *Image Processing, Image Quality, Image Capture Systems Conference (PICS'03)* (Rochester, USA, 2003), Society for Imaging Sciences and Technology, pp. 236–243.

[87] Roufs, J. A. J. Perceptual image quality: concept and measurement. *Philips Journal of Research 47*, 1 (1992), 35–62.

[88] Schölkopf, B., and Smola, A. J. *Learning with kernels: support vector machines, regularization, optimization and beyond.* MIT Press, 2002.

[89] Silverstein, D. A., and Farell, J. E. The relationship between image fidelity and image quality. In *IEEE International Conference on Image Processing (ICIP'96)* (Lausanne, Switzerland, 1996), vol. 1, IEEE, pp. 881–884.

[90] Sonka, M., Hlavac, V., and Boyle, R. *Image Processing, Analysis, and Machine Vision*, 2nd ed. CL-Engineering, 1998.

[91] Vapnik, V. *Estimation of dependencies based on empirical data*, 2nd ed. Springer-Verlag, 2006.

[92] Wang, Z. *Rate scalable foveated image and video communications.* PhD thesis, The University of Texas, Austin, Texas, 2001.

[93] Wang, Z., and Bovik, A. C. A universal image quality index. *IEEE Signal Processing Letters 9*, 3 (2002), 81–84.

[94] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4 (2004), 600–612.

[95] Wyszecki, G., and Stiles, W. S. *Color science: concepts and methods, quantitative data and formulae.* Wiley-Interscience, 2000.

[96] Yendrikhovskij, S. N. *Color reproduction and the naturalness constraint.* PhD thesis, Eindhoven Technical University, Eindhoven, Holland, 1998.

[97] Yendrikhovskij, S. N. Image quality and color categorisation. In *Color image science: exploiting digital media*, L. W. MacDonald and M. R. Luo, Eds. Wiley, 2002, ch. 19, pp. 393–420.

[98] Yendrikhovskij, S. N., Blommaert, F. J. J., and de Ridder, H. Color reproduction and the naturalness constraint. *Color research and applications 24, Issue 1* (1999), 52–67.

[99] Yendrikhovskij, S. N., Blommaert, F. J. J., and de Ridder, H. Towards perceptually optimal colour reproduction of natural scenes. In *Colour Imaging: Vision and Technology*, L. W. MacDonald and M. R. Luo, Eds. Wiley, 1999, ch. 18, pp. 363–382.

[100] Yendrikhovskij, S. N., de Ridder, H., Fedorovskaya, E. A., and Blommaert, F. J. J. Colourfulness judgments of natural scenes. *Acta Psychologica Issue 1, 97* (1997), 79–94.

[101] Zhang, X., Setiawan, E., and Wandell, B. A. Image distortion maps. In *5th Color Imaging Conference: Color Science, Systems, and Applications (CIC'97)* (Scottsdale, AZ, USA, 1997), vol. 5, Society for Imaging Sciences and Technology, SID, pp. 120–125.

[102] Zhang, X., and Wandell, B. A. A spatial extension of CIELAB for digital color image reproduction. *SID Symposium Digest*, 27 (1996), 731–734.

Appendix I

Publications

Publication I

KALENOVA, D., BOTCHKO, V., PARKKINEN, J., JÄÄSKELÄINEN, T.,
Spectral Color Appearance Modeling

Proceedings of The Digital Photography Conference: Image Processing, Image Quality and Image Capture Systems (PICS), Rochester, New York, USA, May 13-16, 2003, pages 381-385.

Spectral Color Appearance Modeling

Diana Kalenova, Vladimir Botchko, Jussi Parkkinen, and Timo Jaaskelainen**
Department of Information Technology, Lappeenranta University of Technology, P.O. Box 20, FIN-53851 Lappeenranta, Finland
diana.kalenova@lut.fi, vladimir.botchko@lut.fi
**Department of Computer Science, Department of Physics, University of Joensuu, P.O. Box 111, FIN-80101 Joensuu, Finland*
jussi.parkkinen@cs.joensuu.fi, timo.jaaskelainen@joensuu.fi

Abstract

In this paper, a spectral color appearance model is considered. This model is used for color image quality estimation when a color image is reproduced through a spectral image. The model is based on the statistical image model that sets a relationship between the parameters of the spectral images and color images, and the overall appearance of the image. This was used to create a set of test color images that were evaluated by twenty color-normal test persons. The results of their evaluation were used for synthesis of a Fuzzy Logic System. Based on the test persons estimation two-dimensional colorfulness-vividness quality evaluation surfaces were built for each image and for image groups. The study shows that a spectral-color relationship and Fuzzy Logic Inference are efficient in color image quality evaluation.

Introduction

In this paper, the spectral color appearance model for quality of color and spectral image estimation is considered. The model can be used in a number of different applications, from digital archiving to communications and industrial applications.

Most of the works on color appearance models are dedicated to the problem of reproduction of colors under different illuminants so that the overall appearance remains unchanged and independent of the viewing conditions, and the device used to represent the image [2]. The model, described in this paper, is used for color image quality estimation when a color image is reproduced through a spectral image. Two criteria are devised for this purpose: colorfulness and vividness. These, in turn, are based on a statistical spectral image model [1], developed on the basis of the statistical characteristics study of natural images. According to the model, statistical characteristics of the image, like kurtosis and standard deviation, affect the appearance of the image. Kurtosis impacts image highlight reproduction, and standard deviation affects image contrast and color saturation. A set of test images was produced using the statistical model.

In order to prove the feasibility of the given spectral appearance model twenty color-normal persons graded the images and the results were input into a Mamdani Fuzzy Inference System, that models the behavior of the human expert and predicts the quality estimate depending on the level of colorfulness and vividness. For the latter, a set of tests was performed on the Fuzzy Logic System, where all images except one were used as a training data set and the image left was used as a testing image. The tests have proven that the prediction error is low in an average case and the Fuzzy Expert System is an efficient tool in predicting the human observer behavior in a quality estimation task. Similar work, dealing with discrimination of the performance of the human observer in a visual detection task in gray level images, is presented in [6].

For spectral images, a visual quality estimation method using a component gray level image for quality evaluation is proposed in [5].

Statistical Model

A generalized statistical spectral model, which copies the behavior of the statistical characteristics of natural images, is used in this work, as follows [1]:

$$\mathbf{f}(\mathbf{x}) = \boldsymbol{\mu} + \mathbf{D}\mathbf{g}(\mathbf{x}) \qquad (1)$$

where $\mathbf{f}(\mathbf{x})$ is a n-dimensional vector random field, and $\mathbf{x}$ is a vector with each element being a spatial dimension; $\boldsymbol{\mu}$ is a mean vector, $\mathbf{g}(\mathbf{x})$ is a normalized vector image with zero mean and unit standard deviation for each component, $\mathbf{D}=\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, where σ_i is a standard deviation in the component.

Eq.1 is usually used for the texture images. However, in this study, this model is used in images that are homogeneous or consist of several homogeneous regions. In the last case, the model is used for each region segmented before the modeling.

To provide the spectral and color parameter change that affect the image appearance, the following relationships are used. The vector $\boldsymbol{\sigma}$ is presented as follows:

$$\boldsymbol{\sigma}=\alpha\beta\boldsymbol{\sigma}_v+(1-\alpha)\boldsymbol{\sigma}_c \quad (2)$$

where $\alpha=(\sigma_{max}-\sigma_{min})/\sigma_{max}$ is the relationship between constant and variable parts of standard deviation and β is a contrast variation coefficient, $\boldsymbol{\sigma}_v$ is a variable component vector of $\boldsymbol{\sigma}$, $\boldsymbol{\sigma}_c$ is a constant component vector of $\boldsymbol{\sigma}$.

Finally, $\mathbf{g}(\mathbf{x})$ is centered by substracting its mean as follows:

$$\boldsymbol{g(x)}=\boldsymbol{g(x)}-E(\boldsymbol{g(x)}) \quad (3)$$

value of the right-hand side is computed and substituted in $\mathbf{g}(\mathbf{x})$, $E(\mathbf{g}(\mathbf{x}))$ is the expectation value of $\mathbf{g}(\mathbf{x})$. $\mathbf{g}(\mathbf{x})$ is defined through gamma-Charlier histogram transform of $\mathbf{f}_s(\mathbf{x})$ and a kurtosis vector $\mathbf{k}$ as follows [1]:

$$\boldsymbol{g(x)}=H(\boldsymbol{f}_s(\boldsymbol{x}),\boldsymbol{k}) \quad (4)$$

where $\mathbf{f}_s(\mathbf{x})$ is a normalized image of $\mathbf{f}(\mathbf{x})$, with zero mean and unit standard deviation for each component, the elements of $\mathbf{k}$ are kurtosis, calculated in components. To affect the image appearance through histogram transform, the kurtosis change is used as follows: all kurtosis elements are proportionally modified according to the given maximum of the kurtosis value

α, β and $\mathbf{k}$ are used for modifying the colorfulness and vividness change in the tests.

Experiment

Experiments were performed on spectral images from [7]. Five images – *inlab1, inlab2, inlab5, jan13am* and *rleaves* were selected. Each image has the following dimensions: 256x256 in the spatial dimension and 31 components in the spectral dimension. The fragments with a size 128x128 and 31 components of these images were used in the test.

Colorfulness

First, a set of test images was produced using the colorfulness parameter. The term includes both contrast and color saturation. Thus, colorfulness was varied through standard deviation, using Eq.2. By changing the α and β coefficients it was possible to receive new values for constant and variable parts of the standard deviation. This procedure was applied to the images with values of (α,β) equal to (0.55,1), (0.75,1), (1,1.3), (1,1.6). Fig.1 shows how the colorfulness change affects the color reproduction of a spectral image.

Figure 1. inlab2. A color reproduction of the original image (left) and processed image (α=1, β=1.6) (right)

Vividness

The second set of tests was produced through variation of the vividness parameter, closely related to highlight reproduction in an image. As the highlight was modified through kurtosis change, the test images were produced with the help of Eq.3 (with k_{max} equal to 5, 10, 30, 60). Fig.2 shows the effect of the vividness change in color reproduction of a spectral image.

Figure 2. inlab 5. A c*olor reproduction of the original image (left) and modified image (k_{max}=80) (right)*

Both test sets were evaluated by 20 color-normal persons using a 1 to 10 scale. Where 1 was assumed as the lowest quality and 10 the highest. The means of all of the scores for both colorfulness and vividness were computed. The results are presented in Tables 1 and 2. In Table 1 Image 1 and Image 2 have parameters (α,β): (0.55,1), (0.75,1), Image 3 is the original, Image 4 and Image 5 have parameters (α,β): (1,1.3), (1,1.6). In Table 2 Image 1 and Image 2 have the following values of k_{max}: 5, 10, respectively. Image 3 is the original, Images 4 and 5 have respectively, k_{max} equal to 30, 60.

Table 1. Mean values of quality judgments in colorfulness change

Quality	Image **1**	Image **2**	Image **3**	Image **4**	Image **5**
Inlab1	2.67	4.50	6.00	6.00	7.17
Inlab2	2.67	4.83	7.00	7.83	8.67
Inlab5	3.00	4.50	5.67	6.83	7.67
Jan13AM	3.33	4.83	7.00	7.67	8.17
Rleaves	2.50	3.50	4.83	4.83	5.83

Table 2. Mean values of quality judgments in vividness change

Quality	Image **1**	Image **2**	Image **3**	Image **4**	Image **5**
Inlab1	2.56	4.00	6.44	7.19	8.19
Inlab2	5.88	6.12	7.94	4.29	3.76
Inlab5	3.18	4.71	5.76	7.29	8.12
Jan13AM	4.29	5.88	6.71	7.41	5.59
Rleaves	4.65	5.82	5.94	4.76	5.00

The experiment showed that there is a close relationship between the quality judgment and the parameters of the spectral color appearance model.

Analysis showed that the test persons prefer the enhanced color images to the images with less colorfulness. This corresponds to the results obtained by other studies [4]. However, a novel result was that the highlight (vividness) parameter has a maximum at some point.

Fuzzy Logic System

On the basis of the results obtained during the experiments, a Mamdani Fuzzy Inference System (FIS) using *min* and *max* for *T-norm* and *T-conorm operators* was synthesized [3]. It has two crisp inputs: Colorfulness and Vividness. These are presented by values αβ and k, respectively. Output is Quality on a 1 to 10 scale. Since a crisp output value is needed, the obtained results are deffuzified. For the defuzzification method the centroid of area z_{COA} is used as follows:

$$z_{COA} = \frac{\int_z \mu_A(z) z dz}{\int_z \mu_A(z) dz} \quad (5)$$

where $\mu_A(z)$ is the aggregated output membership function. The system is based on nine rules, which help to model the test persons behavior.

For the tuning of the system several shapes of membership functions for both input and output variables were tested. The experimental results have proven that 'bell' membership function yielded better results in an average case for the input variables and trapezoidal for the case of the output variable 'Quality'. Membership function plots are shown in Fig.3.

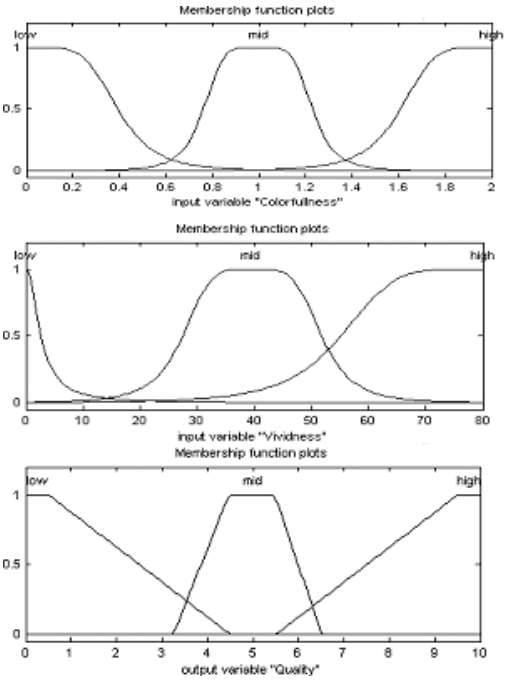


Figure 3. Membership function plots. Input variable Colorfulness, Vividness and Output variable Quality (from top to bottom)

These membership functions led to a three-dimensional overall input-output surface presented in Fig.4. At that point we have received interesting results that correspond to those received in section Experiment. On the one hand, colorfulness parameter behaved as was expected, increasing monotonically. On the other hand, vividness produced a clear maximum on the output surface in the case of the images with small details of different colors (inlab1, inlab5).

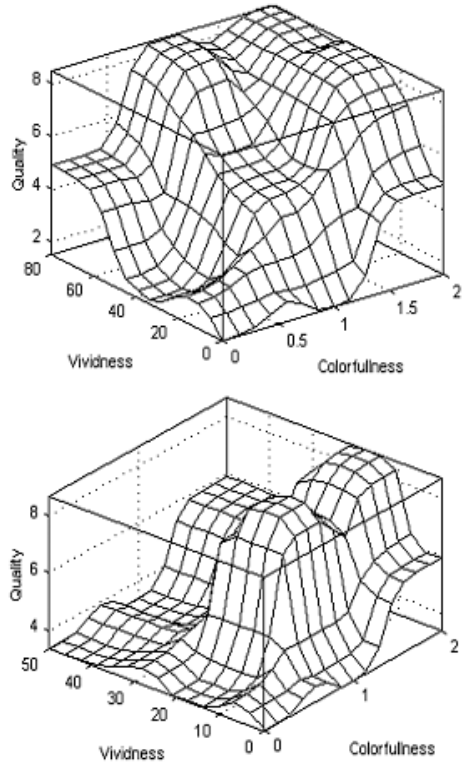


Figure 4. Overall input-output surfaces. inlab2,jan13am,rleaves (top), inalb1,inalb5 (bottom)

An experiment similar to the one presented is given in [6], where a Fuzzy Expert System to facilitate the discrimination of the human observer qualities for the purpose of visual detection in gray level images was created. It was shown that the system behaved as expected for an average case.

Next, a test for the expert evaluation model using the designed system is discussed.

Experimental results

A set of tests was performed to prove that the described Fuzzy Logic System is capable of modeling the behavior of an expert in a quality estimation task.

The tests were done in the following manner: the FIS was rebuilt to incorporate the expert's judgments for all of the images except one. Thus, the new system evaluated four images out of five. Then, the parameters of the image, which was not left out, were input into the new system. The judgment predicted was substracted from the judgment, received using a full set of the original five images. Then a similar operation was performed on the four other. The obtained values are given in Table 3. Thus, each image one by one was used as a test image in a Fuzzy Logic System trained with the rest of the images. The procedure was performed for each of the parameters separately.

Table 3. Results of the Fuzzy Logic System testing

Colorfulness	**1**	**2**	**3**	**4**	**5**
Inlab 1	0.67	0.33	1.00	1.67	1.00
Inlab 2	0.25	0.67	1.08	1.58	1.67
Inlab 5	0.67	0.33	1.00	1.67	1.00
Jan13AM	0.75	0.67	1.08	1.33	0.92
Rleaves	0.50	1.33	2.17	2.92	2.58
Vividness	**1**	**2**	**3**	**4**	**5**
Inlab 1	0.80	0.74	0.70	0.00	0.30
Inlab 2	1.33	0.26	1.66	1.60	1.29
Inlab 5	0.80	0.74	0.70	0.00	0.30
Jan13AM	0.94	0.05	0.32	2.66	1.08
Rleaves	0.39	0.21	1.34	1.05	0.21

Each cell of Table 3 contains the value of the difference between the expert judgment and the judgments modeled in the test. Table 3 is divided into two parts. The upper part in Table 3 gives the results for the case when the vividness parameter remained constant and the colorfulness was varied. The lower part of Table 3 has a constant colorfulness and a variable vividness. In the first column, the name of the test image is given, and in the first row, the number of the image with modified parameters (see section Experiment) is given.

The results show that the proposed Fuzzy Logic System models the behavior of the human expert in a quality estimation task. The average error did not exceed ten percent of the judgments in an average case. To get a more accurate prediction of the image quality a larger training set is needed.

Conclusion

In this paper, a spectral color appearance model was considered. The model was used in color image quality estimation when a color image was reproduced from a spectral image.

A statistical spectral image model was used in the study. According to this model the highlight appearance in the image was closely related to the kurtosis value, and contrast and saturation were connected by the standard deviation. Varying the statistical parameters of the image a set of test images was obtained. The images were evaluated by twenty color-normal test persons. And the evaluation results were input into a Fuzzy Logic System.

To conclude, it is possible to say that test persons prefer more enhanced images to those with little colorfulness. However, in the case of highlights the behavior is not so straightforward. The quality values have a clear maximum at some point in the vividness variation. This means, that in human evaluation the optimal values of highlight in the image are neither the high nor the low.

The study has shown that the spectral-color relationship and Fuzzy Logic Inference were efficient in color image quality evaluation.

Acknowledgments

We would like to acknowledge the help of all of the test persons who participated in the experiments in this study.

References

1. Vladimir Botchko, Heikki Kälviäinen, and Jussi Parkkinen, Highlight Reproduction Using Multispectral Texture Statistics, Proc. The Third International Conference on Multispectral Color Science, pg. 61. (2001).
2. Robert Hunt, J. Color Res. Appl., 19 (1994).
3. Jyh-Shing Roger Jang, Chuen-Tsai Sun, Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence, Prentice Hall, 1997, pg. 423.
4. Ruud Janssen, Computational Image Quality, SPIE press, 2001, pg. 65
5. Arto Kaarna, Jussi Parkkinen, J. of the Imaging Society of Japan, 41, 4 (2002).
6. Jose Padilla-Medina, and Francisco Sanchez-Marin, J. of Fuzzy Sets and Systems, 132, 3 (2002).
7. Alejandro Parraga, Gavin Brelstaff, Tom Troscianko, J. Opt. Soc. Am. A, 15, 3 (1998).

Biography

Diana Kalenova received the engineer degree in Information Technology in Economics from St. Petersburg State University of Telecommunications in 2001 and is currently doing her M.Sc. in Information Technology study at Lappeenranta University of Technology, Finland, in the framework of the International Master's Program in Information Technology. Her work is primarily focused on spectral image analysis, color appearance models, statistical characteristics of images and image quality issues.

Vladimir Botchko is a senior assistant in the Information Technology Department at Lappeenranta University of Technology. Dr. Botchko received his Master's degree at the Electrotechnical Telecommunication University of St. Petersburg, and Ph.D degree at Electrotechnical University of St. Petersburg. His interests lie in the field of computer vision, computer graphics, pattern recognition, spectral and color image analysis. Dr. Botchko is a member of Finnish Pattern recognition Society.

Jussi Parkkinen received an M.Sc. degree in medical physics in 1982, and a Ph.D. degree in mathematics in 1989 from University of Kuopio, Finland. Since 1999 he has been professor of Computer Science at the University of Joensuu, Finland. His research interests are in color and image analysis, pattern recognition, and molecular computing. He is a member of the Finnish Pattern recognition Society.

Timo Jaaskelainen received a Ph.D. degree in 1981 in physics from University of Joensuu, Finland. In 1991, he joined the Department of Physics, University of Joensuu, Finland, as an associate professor. Since 1994 he has been a professor in the same department. He is also the head of department. His research interests include optical material research, optical metrology, color research and information optics.

Publication II

KALENOVA, D., TOIVANEN, P., BOTCHKO, V.,
Color Differences in a Spectral Space

Proceedings of Color, Graphics, Imaging and Vision (CGIV), 2nd European Conference, Aachen, Germany, April 5-8, 2004, pages 368-371.

Color Differences in a Spectral Space

Diana Kalenova, Pekka Toivanen, Vladimir Botchko
Department of Information Technology, Lappeenranta University of Technology,
P.O. Box 20, FIN-53851 Lappeenranta, Finland
diana.kalenova@lut.fi, pekka.toivanen@lut.fi, vladimir.botchko@lut.fi

Abstract

In this study, color similarity metrics in a spectral space are considered. The paper gives a brief overview of several existing measures and presents a novel approach based on kernel methods. New similarity measures for spectral images based upon kernel methods include polynomial, Gaussian radial basis function and sigmoid kernels. The performance of the methods is tested against the Munsell Matte spectral dataset. Kernel methods are compared to twelve well-known similarity metrics, i.e. Correlation Coefficient, Exponential Similarity, Maximum-Minimum methods, etc. Spectral differences of constant Hue, adjacent Values and Chromas have been evaluated using these metrics. The tests show that the proposed Gaussian radial basis function kernel metric performs significantly better, compared to the rest of the measures.

Introduction

In this paper, a novel kernel based approach to color similarity estimation problem is proposed. The methods used can be employed in a number of different applications, including image compression, electronic commerce, archiving etc.

One of the most popular color similarity metrics is Euclidean distance defined in the CIE L*a*b* color space [1]. It has an advantage of simplicity in understanding and realization, however the metric is not efficient. Euclidean distance calculates the difference between colors not taking into account the angle between color vectors, which produces a significant divergence for RGB image reproduction.

Another set of color similarity metrics was proposed in [2,3]. The set contains twelve well-known color similarity measures, created upon the assumption that an optimal color similarity metric should take into account both the distance and the angle between color vectors. The measures are based upon popular distance functions, such as Mahalanobis or Hamming distances. All of the metrics have been tested in [4]. It has been shown that the absolute-value exponential method was the most effective in the task of color differencing [4].

All of the above-mentioned measures have been applied to standard RGB or CIE L*a*b* color spaces. They can be extended to incorporate spectral data. An approach to spectral color differencing has been proposed in [11]. The measures are, in this case: N-dimensional Euclidean distance between two radiance spectra and Euclidean distance in this space. These measures are defined based on two spectral databases: Munsell Glossy [8] and NCS [11].

This measure has been applied to spectral differences of constant chroma, adjacent hues and adjacent values in Munsell and NCS-databases. Then a three-dimensional conical color-space with first three PCA-eigenvectors of NCS- and Munsell data as basis vectors has been defined and analyzed. The Euclidean distance has been computed in this color space. The metric performed significantly better than standard CIE L*a*b* DeltaE, CIE94 and CIEDE2000 formulae. However, as most of the measures built upon Euclidean distance, it suffers from a serious drawback of neglecting the angle between color vectors, which carries significant information in itself [11].

In this paper, a novel approach of color difference estimation, incorporating spectral data, is proposed. The color similarity metrics introduced are based on a well-known pattern recognition technique – kernel methods [5]. A kernel can be considered an extension of a canonical dot product, which in turn, is one of the techniques described in [2,3]. Kernel methods applied in this study include: polynomial, Gaussian radial basis function (RBF) and sigmoid kernels.

Spectral differences between colors have been measured using a Munsell Matte Collection spectral dataset [6], against constant Hues and adjacent values of Chromas and Values. The results of the measurements produced using the kernel-based methods [5] have been compared to twelve well-known color metrics [2,3] extended to incorporate spectral data.

Color Similarity Measures

Color similarity measures generally take two spectral vectors as an input and produce an output on a 0 to 1 scale. Where 0 means that the colors are “not similar at all” and 1 means “identical” [7].

In this paper, a kernel-based approach to color similarity is proposed. The performance of these measures is compared with the performance of twelve widely known metrics [2,3]. Kernels can be assumed to be dot products of vectors in a certain feature space, meaning that if we have two vectors x_i and x_j in the input domain X, we can produce a mapping:

$$\Phi : X \to H \qquad x \quad x := \Phi(x) \tag{1}$$

So the dot product is computed in thus induced feature space [5]. The kernel similarity measures considered in this paper include polynomial, Gaussian radial basis function and sigmoid kernels.

Polynomial kernel similarity measure can be presented as follows:

$$S_{polynomial} = (x_i, x_j)^d \tag{2}$$

where d is the parameter of the sensitivity of the measure, x_i and x_j are input p-dimensional color vectors. The similarity functions have a general form of $S(x_i,x_j)$, the arguments are skipped for simplicity in the formulae shown in this paper.

The Gaussian radial basis function kernel has the following form:

$$S_{Gaussian} = \exp\left(-\frac{\|x_i - x_j\|^2}{2s^2}\right) \tag{3}$$

where $\sigma > 0$, σ is the parameter of the sensitivity of the function.

And the sigmoid kernel based similarity metric can be presented as follows:

$$S_{sigmoid} = \tanh(k*(x_i, x_j) + J) \tag{4}$$

where k an ϑ are variable parameters.

The twelve other metrics are as follows [7]:

Metric 1

$$S_1 = \frac{x_i x_j^t}{|x_i||x_j|} = \cos(q) \tag{5}$$

where θ is the angle between vectors x_i and x_j

Metric 2

$$S_2 = \left(\frac{x_i x_j^t}{|x_i||x_j|}\right)\left(1 - \frac{||x_i| - |x_j||}{\max(|x_i|, |x_j|)}\right) \tag{6}$$

Metric 3

$$S_3 = \frac{|x_i|\cos(q) + |x_j|\cos(q)}{\left(|x_i|^2 + |x_j|^2 + 2|x_i||x_j|\cos(q)\right)^{\frac{1}{2}}} \tag{7}$$

Metric 4

$$S_4 = \frac{\cos(q)\left(|x_i|^2 + |x_j|^2 + 2|x_i||x_j|\cos(q)\right)^{\frac{1}{2}}}{|x_i| + |x_j|} \tag{8}$$

Metric 5

$$S_5 = 1 - \frac{\left(|x_i|^2 + |x_j|^2 - 2|x_i||x_j|\cos(q)\right)^{\frac{1}{2}}}{\left(|x_i|^2 + |x_j|^2 + 2|x_i||x_j|\cos(q)\right)^{\frac{1}{2}}} \tag{9}$$

Metric 6: Correlation Coefficient Method

$$S_6 = \frac{\sum_{k=1}^{p} |x_{ik} - \bar{x}_i||x_{jk} - \bar{x}_j|}{\left(\sum_{k=1}^{p}(x_{ik} - \bar{x}_i)^2\right)^{1/2}\left(\sum_{k=1}^{p}(x_{jk} - \bar{x}_j)^2\right)^{1/2}} \tag{10}$$

where $\bar{x}_i = \frac{1}{p}\sum_{k=1}^{p} x_{ik}$

Metric 7: Exponential Similarity Method

$$S_7 = \frac{1}{p}\sum_{k=1}^{p}\exp\left(-\frac{3}{4} * \frac{(x_{ik} - x_{jk})^2}{b_k^2}\right) \tag{11}$$

where $b_k^2 > 0$ is a parameter that is determined experimentally.

Metric 8: Absolute-Value Exponent Method

$$S_8 = \exp\left(-b\sum_{k=1}^{p}|x_{ik} - x_{jk}|\right) \tag{12}$$

where $\beta > 0$.

Metric 9: Absolute-Value Reciprocal Method

$$S_9 = 1 - b\sum_{k=1}^{p}|x_{ik} - x_{jk}| \tag{13}$$

β is determined empirically.

Metric 10: Maximum-Minimum Method

$$S_{10} = \frac{\sum_{k=1}^{p}\min(x_{ik}, x_{jk})}{\sum_{k=1}^{p}\max(x_{ik}, x_{jk})} \tag{14}$$

Metric 11:Arithmetic-Mean Minimum Method

$$S_{11} = \frac{\sum_{k=1}^{p}\min(x_{ik}, x_{jk})}{\frac{1}{2}\sum_{k=1}^{p}(x_{ik} + x_{jk})} \tag{15}$$

Metric 12: Geometric-Mean Minimum Method

$$S_{12} = \frac{\sum_{k=1}^{p}\min(x_{ik}, x_{jk})}{\sum_{k=1}^{p}(x_{ik}x_{jk})^{1/2}} \tag{16}$$

Experiment

The tests of the viability of color similarity metrics were performed on Munsell Colors Matte dataset [6] (1269 matt Munsell color chips). The reflectance spectra had been measured by a Perkin-Elmer lambda 9 UV/VIS/NIR spectrophotometer in the 380 nm - 800 nm interval with 1 nm wavelength resolution [8].

The purpose of the experiment was to choose a metric that would give comparable values of differences for perceptually equally disparate colors, and at the same time would account for the changes in Hue, Value and Chroma with the whole range of values. Another requirement imposed on the measures was the possibility of adjustment of the sensitivity of similarity measurements.

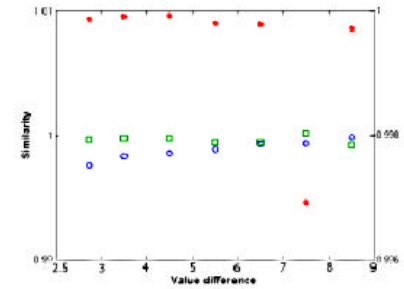
The first set of experiments concerned the possibility of color discrimination based on spectral information of Munsell colors with constant values of Hue and Chroma

and adjacent values of Value. The set was chosen so that it covers the entire range of Values. Hue was selected to be 5R, 5B and 5G, Chroma equal to 1, and Values ranging from 2.5 to 9.

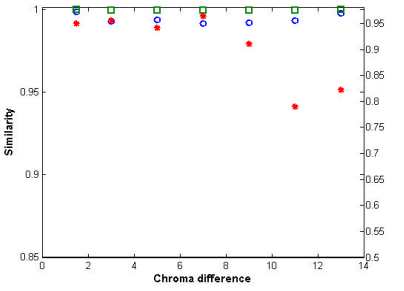
The second set of tests was similar to the previous one, except that the Values and Hues remained constant, while Chroma was varied. Thus, Hue was set to 5R, 5B and 5G, Value to 6, and Chroma ranged from 1 to 14.

In order to account for the visual uniformity of the Munsell color dataset, the input data was multiplied by Spectral Luminous Efficiency Function for photopic vision [9] and illumination factor [10].

Considering the purpose of the whole experiment, the results obtained can be divided into several categories. First, would come the metrics that gave comparable values of differences for perceptually equally disparate colors. For that purpose, based upon the experimental settings described above, diagrams of the functional relations between similarity measures, Value and Chroma have been obtained (see Fig. 1). The diagrams show the three best metrics from the point of view of the linearity of the response given to the inputs. Meaning that the measures gave close to linear (to a certain extent) responses to the changes in Value and Chroma. The metrics that gave the smoothest responses are Metric 1, Metric 8 and Gaussian radial basis function, presented in Fig.1.



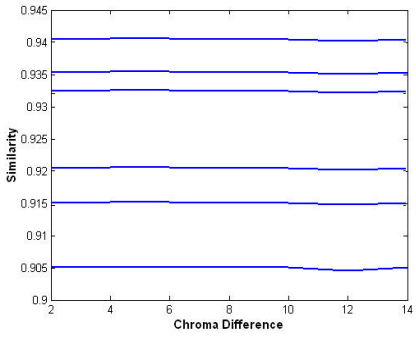
(a)



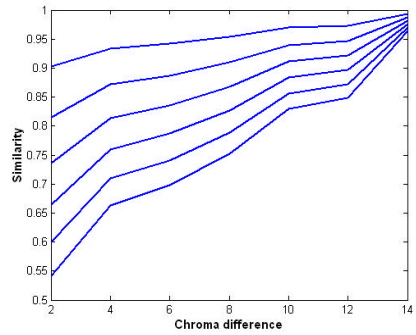
(b)

Figure 1. (a) Similarity vs. Value, (b) Similarity vs. Chroma (Metric 1 – red asterisk (second scale); Metric 8 – Absolute-Value Reciprocal Method blue circles, RBF – green squares) for 5R Hue

The second requirement set upon the metrics has been the possibility of adjustment of the sensitivity of similarity measurements. Looking back at the Eq. 2-16, it can be stated that the kernel methods (polynomial, Gaussian RBF and sigmoid), exponential similarity, absolute-value exponential, absolute-value reciprocal methods poses similarity terms, which allow adjustment of the measurements. However, the kernel methods provide a significantly better control over the whole equation. The results of testing of the sensitivity parameters in the measures for Gaussian RBF and Metric 8 (absolute-value exponent) are presented in Fig. 2.



(a)



(b)

Figure 2. Sensitivity of (a) Gaussian RBF (b) Metric8 (absolute-value exponent) for 5R Hue, Value 6

A functional relation between the similarity values of Gaussian RBF and Metric 8 for different sensitivity are shown in Fig. 2, where both β and σ have been varied in the range of 0.005 and 0.03. The resulting graphs show that the sensitivity of the measures can be varied through the use of the special terms in a certain range. However, the absolute value exponential method tends to produce non-linearities with the growth of the sensitivity value. The choice of the Gaussian RBF and Metric 8 functions for testing is not accidental. Both of these metrics have shown the best results for the smoothness of the response

of the sensitivity function requirement, thus it would seem reasonable to continue comparing them.

And the final requirement upon the similarity measures that has been set is the possibility of accounting for the changes in Hue, Value and Chroma with the whole range of values. This requirement implies that the sensitivity of the metric should be such that it would account for the just noticeable by a human eye differences in the color. The best results have been produced by kernel metrics, Metric7 (exponential similarity) and Metric 8 (absolute-value exponential).

Considering the results obtained, several metrics could be considered as the most promising. The best results have been obtained using the Gaussian RBF and absolute-value exponential metrics. Both of these measures have produced approximately linear responses to perceptually uniformly distributed values of Value and Chroma (with Gaussian RBF producing the smoothest response (see Fig.1)), at the same time, accounting for the change of color vectors with the whole range of values. The sensitivity of the function could be adjusted. However control over the sensitivity in the Gaussian RBF metric is significantly better due to a special term σ introduced into the formula. All of that brought us to a conclusion of the efficiency of the Gaussian RBF measure in the task of color discrimination.

Another result obtained from the experiments performed, is that the response of the similarity functions became smoother with the introduction of the efficiency curve [9] and the illumination factor [10].

Conclusion

In this paper, color similarity metrics in spectral space have been considered. The measures include twelve well-known metrics created upon well-known distance functions, such as Mahalanobis or Hamming distances, and a set of novel kernel-based color similarity measures. The performance of all of these measures has been tested against Munsell Matte spectral dataset [8]. The purpose of the experiments has been to find a metric that gives comparable values of differences for equally disparate color, at the same time accounting for the values of change in Hue, Chroma and Value with the whole range of values. Based on the results presented above, it can be concluded that the kernel-based approach to color differencing in spectral space is efficient. The performance of the kernel-based metrics gave results comparable with the traditional color similarity measures. Moreover, Gaussian radial basis function kernel performed more effectively considering the traditional color similarity methods. The response of the measure was smoother in the case of both the Value and the Chroma change, and the sensitivity of the function was greater and could be varied in a more efficient way (through the use of a special sensitivity term). Furthermore, weighting the input data by the efficiency curve [9] and illumination factor [10] produced a smoother output of the similarity functions.

Acknowledgments

I would like to acknowledge the help and valuable advice of Ms. Irina Mironova, who helped greatly during the research.

References

1. Recommendations on Uniform Color Spaces, Color Difference Equations, Psychometric Color Terms, Publication CIE 15 (E.-1.3.1), Suppl.No.2, Bureau Central de la CIE, Vienna, 1971.
2. Konstantinos Plataniotis, Anastasios Venetsanopoulus, Color Image Processing and Applications, Springer-Verlag, Berlin, 2000, pg. 70.
3. Yee Leung, Spatial Analysis and Planning under Imprecision, North-Holland Publishers, Netherlands, 1988, pg 116.
4. Michael Hild, Takashi Emura, Which Color Similarity Measure is Most Effective for Background-Frame Differencing?, Proc. 9th CIC 2001, pg. 168-173. (2001).
5. Bernard Schölkopf, Alex Smola, Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, Cambridge, MA, 2002, pg. 45.
6. Munsell Book of Color - Matte Finish Collection Munsell Color, Baltimore, Md., 1976.
7. Michael Hild, On the Effectiveness of Color Similarity Measures in Background-Frame Differencing Applications, Proc. CGIV, pg. 306. (2002).
8. Color Spectra Database http://www.it.lut.fi/ip/research/color/database/database.html (cited 10.08.2003).
9. Spectral Luminous Efficiency Function for Photopic Vision, Publication CIE 86-1990.
10. CIE Standard Illuminant D65 Relative Spectral Power Distribution, CIE 15.2-1986.
11. Joni Orava, Timo Jaaskelainen, Jussi Parkkinen, Color Differences in a Spectral Space, Proc. PICS 2003, pg. 205-209, (2003).

Biography

Diana Kalenova received the engineer degree in Information Technology in Economics from St. Petersburg State University of Telecommunications in 2001 and M.Sc. in Information Technology at Lappeenranta University of Technology, Finland, in 2003. Since 2003 she has been working at Lappeenranta University of Technology, Finland, as an assistant. Her work is primarily focused on spectral image analysis, color appearance models, statistical characteristics of images and image quality issues. Ms. Kalenova is a member of Pattern Recognition Society of Finland.

Professor, Doctor of Technology Pekka Toivanen has graduated from Helsinki University of Technology in 1989. Since 1991 he has been working at the Lappeenranta University of Technology, at the Department of Information Technology. He is at the moment a head of the laboratory of Information Processing of the Department of Information Technology

at Lappeenranta University of Technology. Pekka Toivanen is a member in SPIE and IEEE. His main research interests are multispectral image processing, color, and distance transforms.

Vladimir Botchko is a senior assistant in the Information Technology Department at Lappeenranta University of Technology. Dr. Botchko received his Master's degree at the Electrotechnical Telecommunication University of St. Petersburg, and Ph.D degree at Electrotechnical University of St. Petersburg. His interests lie in the field of computer vision, computer graphics, pattern recognition, spectral and color image analysis. Dr. Botchko is a member of Finnish Pattern Recognition Society.

Publication III

KALENOVA, D., TOIVANEN, P., BOTCHKO, V.,
Spectral Image Distortion Map

Proceedings of Pattern Recognition (ICPR), 17th International Conference, Cambridge, UK, August 23-26, 2004, pages 668–671.

Spectral Image Distortion Map

Diana Kalenova, Pekka Toivanen, Vladimir Botchko
Department of Information Technology, Lappeenranta University of Technology
diana.kalenova@lut.fi, pekka.toivanen@lut.fi, vladimir.botchko@lut.fi

Abstract

In this paper a novel technique of spectral image quality evaluation using Spectral Image Distortion Map (SIDM) is proposed. The method is based on a recent approach to evaluation of color differences in a spectral space. What is calculated here, in fact, is a pixelwise spectral distortion. As the measure of the dissimilarity a novel kernel based similarity measure is used. The metric produces comparable values of differences for perceptually equally disparate colors. As a result a gray-scale spectral distortion image is obtained, where the intensity of each of the pixels is a difference between the original image and the distorted one. A Perceptual Image Distortion Map (PIDM) has also been constructed to show the accuracy of SIDM. A comparison of PIDM and SIDM shows that the latter provides an excellent fit to the response of the human visual system.

1. Introduction

In this paper, a novel technique of spectral image quality estimation is proposed. The method can be employed in a number of different applications, including image compression, Internet museums, electronic commerce etc.

The main objective for a creation of such measures is primarily lossy compression applications. A quality measure should be established with a possibility of computing the distortion value dynamically as the information is discarded from the image. This kind of metric should also be able to account for the characteristics of the human visual system.

Several approaches to spectral image distortion measurements exist at the moment [1,2,3]. Most of the quality measures have emerged from gray-scale image metrics: mean-squared error, signal to noise ratio, percentage maximum absolute distortion etc. Nevertheless, none of these measures accounts for the characteristics of the human visual system. A Blockwise Distortion Measure for Multispectral images (BDMM) has been suggested in [2]. The measure seeks to compute a quality measure that would correspond with the human evaluation; however, it deals with the artifacts in the spatial direction and does not account for the specific spectral distortions.

The algorithm, described in this paper, suggests that the distortion measure can be built as a pixelwise difference between the original and the distorted (compressed) image. As the difference metric a novel kernel based similarity metric [4] is used. The kernel methods applied in this study include: polynomial, Gaussian radial basis function (RBF) and sigmoid kernels. It was shown in [4] that these similarity metrics correspond well with the properties of the human visual system and produce comparable values of differences for perceptually equally disparate colors. As a result a gray-scale spectral distortion map is obtained, which shows the areas where the visible distortions are in the image, and how large the distortions are.

2. Color similarity measures

One of the most popular color similarity metrics so far has been Euclidean distance [5]. It has an advantage of simplicity in understanding and realization, however the metric is not optimal. Euclidean distance calculates the difference between colors not taking into account the angle between color vectors, which produces a significant divergence for RGB image reproduction [4].

Another set of color similarity metrics was proposed in [6, 7]. The set contains twelve well-known color similarity measures, created upon the assumption that an optimal color similarity metric should take into account both the distance and the angle between color vectors. All of the metrics have been tested in [8]. It has been shown that the absolute-value exponential method was the most effective in the task of color differencing [4]. A spectral discrimination possibility has been tested in [4]. The performance of the methods has been tested against the Munsell Matte spectral dataset [9]. The performance of the conventional color discrimination measures has been compared with the performance of a newly introduced kernel based similarity measures. It was shown [4] that the metrics produce comparable values of differences for perceptually equally disparate colors [4].

0-7695-2128-2/04 $20.00 (C) 2004 IEEE

A set of kernel based similarity measures suggested in [4] is used in this paper. The kernel similarity measures considered in this paper include polynomial, Gaussian radial basis (RBF), and sigmoid.

The color similarity measures, commonly, have two spectral vectors as inputs and produce an output on a 0 to 1 scale, where 0 means that the colors are "not similar at all" and 1 means "identical" [10].

Kernels can be assumed to be dot products of vectors in a certain feature space, meaning that if we have two vectors x_i and x_j in the input domain X, we can produce a mapping [11]:

$$\Phi : \mathrm{X} \to \mathrm{H}$$
$$x \mapsto x := \Phi(x) \tag{1}$$

Polynomial kernel similarity measure can be presented as follows [11]:

$$S_{polynomial} = \left(x_i, x_j\right)^d \tag{2}$$

where d is a parameter of the sensitivity of the measure, x_i and x_j are input p-dimensional vectors. In order to account for the characteristics of the human visual system the input data is multiplied by Spectral Luminous Efficiency Function for photopic vision [12] and illumination factor [13].

The Gaussian RBF kernel has the following form [6]:

$$S_{Gaussian} = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{3}$$

where $\sigma > 0$, σ is the parameter of the sensitivity of the function.

And the sigmoid kernel based similarity can be presented as follows [6]:

$$S_{sigmoid} = \tanh(k * (x_i, x_j) + \vartheta) \tag{4}$$

where k and ϑ are variable sensitivity parameters.

3. Experiments

Experiments were performed on spectral images from [14]. Two images – inlab2 and inlab5 were selected. Each image has the following dimensions: 256x256 in the spatial dimension and 31 components in the spectral dimension. The images were captured by a CCD (charge coupled device) camera in a 400-700 nm wavelength range at 10 nm intervals. The images selected were taken indoor (in a controlled environment, i.e. dark-lab or glass-house).

First, both of the images were compressed using PCA down to two principal components. The color reproductions of original images and reconstructed after compression are given in Fig.1.

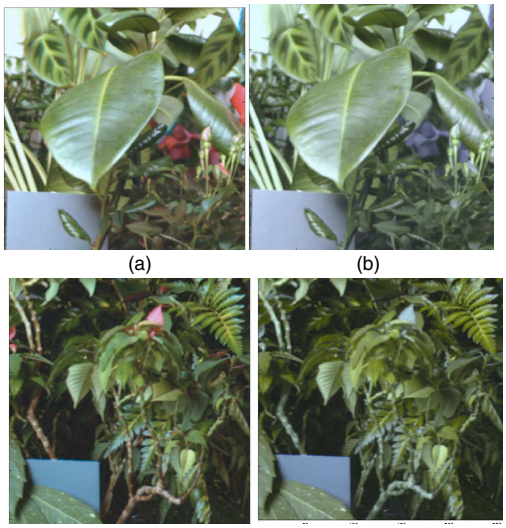


Figure 1. Color reproduction of spectral images *inlab2* (a) original, (b) reconstruction after compression (PCA 2); *inlab5* (c) original, (d) reconstruction after compression (PCA 2)

The areas of color difference are clearly visible in the images, and concentrate primarily in red and brown regions.

The next step after compression was to multiply both of the image data sets by Spectral Luminous Efficiency function for photopic vision [12] and illumination factor [13].

The results of the multiplication were used in the experiment of SIDM evaluation. Three such maps were computed through the use of Eq. 2,3,4. The similarity between images was computed on a pixelwise basis. Fig. 2 (a, d) present the result of application of polynomial kernel; (b,e) Gaussian radial basis function and (c,f) sigmoidal kernel. The level of the intensity in the maps corresponds to the similarity scale: from black "not similar at all" to white "identical". Comparing the maps obtained it can be stated that they present a similar to a certain extent result. Moreover, comparing the original and the compressed image it is possible to select the areas that are different in the images, and these are similar to the results obtained with the use of the image distortion maps.

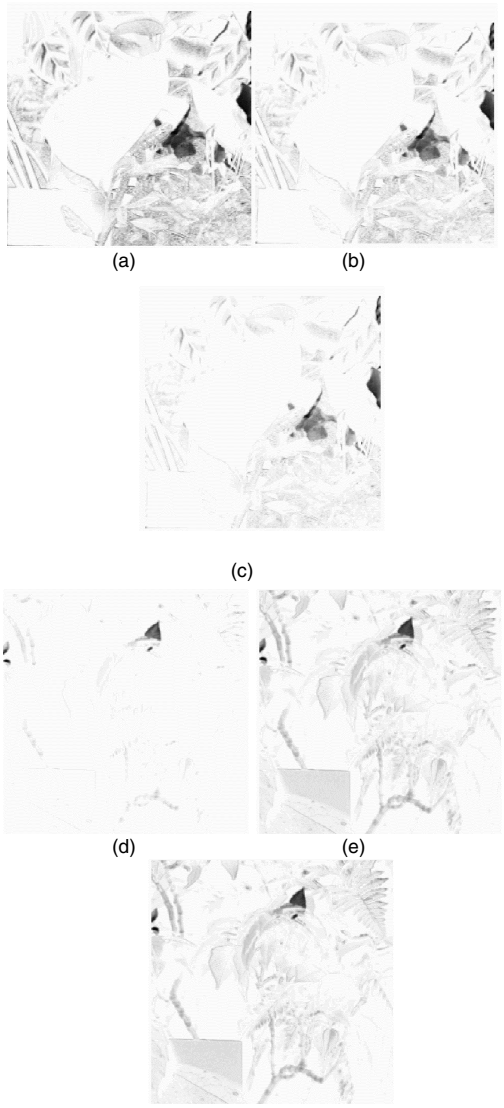


Figure 2. Spectral image distortion maps. Inlab2 (a,b,c); inlab5 (d,e,f). (a,d) polynomial kernel; (b,e) Gaussian radial basis function and (c,f) sigmoidal kernel

5. Experimental results

The accuracy of the SIDM has been tested using Perceptual Image Distortion Map (PIDM) and SCIELAB [15]. Five subjects have been presented two sets of images, consisting of an original and a compressed image (Fig. 1). The users have been asked to mark the regions that appeared different with a rectangular digital marker of size 4 by 4 pixels with different levels of gray-level intensity. Where black means ”not similar at all” and white ”identical”. The subjects were instructed to mark the whole image area. Fig. 3 presents the mean of all subject maps.

Figure 3. Perceptual Image Distortion Map for images: (a) *inlab2*; (b) *inlab5*

Fig. 3 clearly indicates that PIDM presents a practically excellent fit to the SIDM calculated through the use of Gaussian RBF. Nevertheless, certain errors exist, which can be attributed to the fact that the marker size and shape caused several inaccuracies in stamping identical regions several times.

The values of the errors of the SIDM are given in Table 1, where each of the cells in first three columns present the mean deviation of the SIDMs, obtained through the use of polynomial, Gaussian RBF and sigmoidal kernels for each of the images, from the values of PIDM. Last column presents the value of the deviation of the normalized SCIELAB [15] error image from the PIDM.

Table 1. Comparison of SIDM, SCIELAB and PIDM

	Polynomial	Gaussian RBF	Sigmoidal	SCIELAB
Inlab 2	0.0499	0.0395	0.0551	0.0606
Inlab 5	0.0395	0.0291	0.0581	0.0609

SCIELAB is a spatial extension of a standard CIELAB color metric, which allows measuring color reproduction errors in digital images [5]. It is a widely excepted industry standard. This metric is based on empirical data of color discrimination. SCIELAB error measure has been computed from the color reproduction of the inlab2 and inlab5 spectral images. This metric can be considered applicable in this case, since the subjects in the PIDM experiments dealt with color reproductions only.

PIDM presents a full map of empirical distortion data, which can be used in the task of evaluation of the accuracy of the metrics presented. Thus, looking at Table 1 it can be concluded that the most accurate evaluation of the human response in the quality estimation task is obtained through the use of Gaussian RBF kernel, while the worst with SCIELAB, although the deviation between these is not large.

Taking into consideration all of the above it can be stated that SIDM is a quality evaluation technique that accurately predicts the response of a human visual system in a distortion evaluation task. The values of the SIDM vary in the range from 0 to 1, representing the difference values from "not similar at all" to "identical". From the point of view of the probability theory it can be stated that SIDM presents a probability of the subject identifying a certain pixel as similar, which allows avoiding time and money consuming procedure of expert survey, and gives the possibility of computing the distortion values dynamically as the information is discarded from the image, as for example in a lossy compression task.

5. Conclusion

In this paper a novel technique of spectral image quality evaluation using SIDM has been proposed. The method is based on a recently published spectral discrimination study [4]. The algorithm, given in this paper, computes a pixelwise difference between the original and the distorted images. A gray-scale image distortion map is obtained as a result, where the intensity of each of the pixels corresponds to the value of the similarity between them, which, in turn, shows the areas where the visible distortions are in the image, and how large the distortions are. As the similarity measure a novel kernel based technique is used, which has proven to be applicable in the task of color discrimination analysis [4]. Several parameters can be obtained using SIDM: mean, median, and standard deviation of the spectral differences, which would allow estimation of the overall image distortion. The measures have been applied to several spectral images with spectral distortions introduced. The performance of SIDM has been tested against a Perceptual Image Distortion Map [15] and SCIELAB. PIDM presents a full map of empirical distortion data, while SCIELAB is a widely accepted spatial extension of CIELAB industrial standard of measuring color reproduction errors in digital images [15]. Comparing the results obtained, it can be stated that PIDM presents a practically excellent fit to the SIDM calculated through the use of Gaussian RBF. Therefore it can be concluded that SIDM mimics closely the response of the human visual system in a task of quality evaluation and can be considered among the best approaches to evaluation of spectral distortions introduced into the spectral images.

10. References

[1] A. Kaarna, J. Parkkinen, "Quality Metric for Multispectral Image Compression", Journal of the Imaging Society of Japan, vol. 41, no. 4, 2002, pp. 379-391.

[2] A. Kaarna, J. Parkkinen, "Blockwise Distortion Measure for Lossy Compression of Multispectral Images", Proc. X European Signal Processing Conference, EUSIPCO 2000, Tampere, Finland, September 4-8, 2000, pp. 2197-2200.

[3] M. J. Ryan, J. F. Arnold, "A Suitable Distortion Measure for the Lossy Compression of Hyperspectral Data", Geoscience and Remote Sensing Symposium Procs., IGARSS '98. 1998 IEEE International, 6-10 July 1998, vol4., pp. 2056-2058.

[4] D. Kalenova, P. Toivanen, V. Botchko, "Color Differences in a Spectral Space", Proc. 2nd European Conference on Color in Graphics, Imaging and Vision, CGIV 2004, in April 5 - April 8, 2004, pp. 368-371.

[5] Recommendations on uniform color spaces, color difference equations, psychometric color terms, Publication CIE 15 (E.-1.3.1), Suppl.No.2, Bureau Central de la CIE, Vienna, 1971.

[6] Plataniotis, K. N, Venetsanopoulus, A. N., *Color Image Processing and Applications,* Springer-Verlag, Berlin, 2000.

[7] Leung, Y., *Spatial Analysis and Planning under Imprecision*, North-Holland Publishers, The Netherlands, 1988.

[8] M. Hild, T. Emura, "Which color similarity measure is most effective for background-frame differencing?", Proc. 9th Color Imaging Conference, 2001, pg. 168-173.

[9] *Munsell Book of Color* - Matte Finish Collection Munsell Color, Baltimore, Md., 1976.

[10] M. Hild, "On the Effectiveness of Color Similarity Measures in Background-Frame Differencing Applications", Proc. European Conference on Color in Graphics, Imaging and Vision, CGIV 2002, pp. 306.

[11] Scholkopf B., Smola A., *Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond,* MIT Press, Cambridge, MA, 2002.

[12] *Spectral Luminous Efficiency Function for Photopic Vision*, Publication CIE 86-1990.

[13] *CIE Standard Illuminant D65 Relative Spectral Power DiStribution*, CIE 15.2-1986.

[14] A. Parraga, G. Brelstaff, T. Troscianko, "Color and Luminance Information in Natural Scenes", J. of Optical Society of America A, 15, 3 (1998), pp. 563-569.

[15] X. Zhang, E. Setiawan, B. A. Wandell, "Image Distortion Maps", Proc. 5th Color Imaging Conference. Color Science, Systems, and Applications, Scottsdale, AZ, USA, Nov. 1997, pp. 120-125.

0-7695-2128-2/04 $20.00 (C) 2004 IEEE

Publication IV

KALENOVA, D., TOIVANEN, P., BOCHKO, V.,
Color Differences in a Spectral Space

For information on reprints or reproduction contact
Donna Smith
Managing Editor
The Journal of Imaging Science and Technology
Society for Imaging Science and Technology
7003 Kilworth Lane
Springfield, Virginia 22151 USA
703/642-9090 extension 17
703/642-9094 (fax)
donna.smith@imaging.org
www.imaging.org

Color Differences in a Spectral Space

Diana Kalenova, Pekka Toivanen and Vladimir Bochko
Department of Information Technology, Lappeenranta University of Technology, Lappeenranta, FINLAND

In this work, a novel approach to color differencing in a spectral space is presented. The algorithm is based on well known pattern recognition technique-kernel methods, which include polynomial, Gaussian radial basis function (RBF) and sigmoid kernels. This article starts with a brief overview of several existing methods created both for color and spectral data. The performance of novel measures is tested against the Munsell Matte spectral dataset, and a spectral database of metameres. The results of tests obtained for kernel methods are compared with those produced by twelve conventional similarity metrics, i.e., Correlation Coefficient, Exponential Similarity, Maximum–Minimum methods, etc. The assumption behind experiments is that methods should model the behavior of human observers in the task of color differencing. The tests show that the proposed Gaussian RBF kernel metric performs significantly better, compared to the rest of the measures.

Journal of Imaging Science and Technology 49: 404–409 (2005)

Introduction

The idea of reproduction of spectral images on printed media has recently gained popularity, resulting in a number of publications.[1,2] Appearance of such devices extends the scope of traditional areas of application of spectral-image-based algorithms beyond remote sensing and archiving to more observer oriented ones. As a consequence, a necessity in methods of modeling of human perception of the quality of spectral images increases. A particular area of research in this field is estimation of color differences in spectral images. In this article, a novel approach to color similarity estimation is presented. The measures given are based on well known kernel techniques. The methods used can be employed in a number of different applications, including image compression, electronic commerce, imaging, industrial applications, like textile color differencing, etc.

Although it is assumed that much more research needs to be done in the field of color differences, one of the most popular conventional color similarity metrics Euclidean distance defined in the CIE L*a*b* color space is extensively used.[3] It has an advantage of simplicity in understanding and realization, however, the metric is not efficient. Euclidean distance calculates the difference between colors not taking into account the angle between color vectors, which produces a significant divergence for RGB image reproduction.[4] Several modifications have been introduced into this formula, intended to improve the perceptual uniformity of CIE L*a*b* ΔE, resulted in CIE ΔE_{94} and CIEDE2000 equations.

Original manuscript received July 19, 2004

Corresponding Author: Diana Kalenova, Diana.Kalenova@lut.fi

Another set of color similarity metrics was proposed in Refs. 5 and 6. The set contains twelve well-known color similarity measures, created upon the assumption that an optimal color similarity metric should take into account both the distance and the angle between color vectors, thus eliminating the drawbacks of conventional methods. The measures are based upon popular distance functions, such as City Block or Cosine distances. All of the metrics have been tested in Ref. 7. It has been shown that the absolute value exponential method can be considered to be the most effective measures in the task of color differencing.[4,7]

All of the above-mentioned metrics have been applied to standard RGB or CIE L*a*b* color spaces. Nevertheless, many of these can be extended to incorporate spectral data. An approach directly dealing with spectral color differencing has been proposed in Ref. 8. The measures are, in this case: N-dimensional Euclidean distance between two radiance spectra, which is an extension of the conventional ΔE metric. The second measure is also a Euclidean distance, defined in a three-dimensional conical color-space with first three PCA-eigenvectors of NCS- and Munsell data as basis vectors. These measures have been applied to spectral differences of constant chroma, adjacent hues and adjacent values in Munsell Glossy[9] and NCS[8] databases. Both of the metrics perform significantly better than standard CIE L*a*b* ΔE, CIE ΔE_{94} and CIEDE2000 formulae. However, as most of the measures built upon Euclidean distance, it suffers from a serious drawback of neglecting the angle between color vectors, which carries significant information in itself.[7,8]

In this article, a novel approach to color difference estimation, incorporating spectral data, is proposed. The color similarity metrics introduced are based on a well-known pattern recognition technique-kernel methods.[10] A kernel can be considered an extension of a canonical dot product, which in turn, is one of the techniques

described in Refs. 5 and 6. Kernel methods applied in this study include: polynomial, Gaussian radial basis function (RBF) and sigmoid kernels. Spectral differences between colors have been measured using a Munsell Matte Collection spectral dataset,[11] against constant Hues and adjacent values of Chromas and Values. The perceptual phenomenon of metamerism has been tested using a dataset of spectra of metameric colors. The results of the measurements produced using the kernel-based methods[10] have been compared to twelve well-known color metrics,[5,6] extended to incorporate spectral data.[4]

Color Similarity Measures

The main task facing any color differencing research is given two color vectors or patterns to determine the degree of similarity between these, returning a certain real number. Thus, the output of a color similarity function can either be an exact difference in intensities of the pixels, or a probability of the two vectors being alike. In our case the color similarity measures presented take two spectral vectors, varying in the range [0,1] (with the difference between vectors thus lying in the range of [0,1]), as an input and produce an output on a 0 to 1 scale, where 0 means that the colors are "not similar at all" and 1 means "identical".[12]

In this study several color similarity measures are considered. Most of these measures are based on various distance functions, like City Block or Cosine distance metrics. A novel kernel-based approach to color similarity proposed in this article is a well-known machine learning and pattern recognition technique. Kernels can be assumed to be dot products of vectors in a certain feature space, meaning that if we have two vectors x_i and x_j in the input domain X, we can produce a mapping:

$$\begin{aligned} &\Phi: X \to H \\ &x \mapsto x := \Phi(x) \end{aligned} \tag{1}$$

So the dot product is computed in thus induced feature space. A geometrical interpretation of the dot product is, in turn, a cosine of the angle between normalized vectors, which brings us to the notion of the Cosine distance.[10]

The kernel similarity measures considered in this article include polynomial, Gaussian radial basis function and sigmoid kernels. In order to account for the visual uniformity of the Munsell color dataset, the input data was multiplied by Spectral Luminous Efficiency Function for photopic vision,[13] denoted as V, and spectral radiance function of a certain light source,[14] denoted as R.

Polynomial kernel color similarity measure, modified to account for the nature of the human perception, can be presented as follows:

$$S_{polynomil} = (y_i, y_j)^d \tag{2}$$

where henceforth $y_{i,j} = R^*V^*x_{i,j}$, d is the parameter of the sensitivity of the measure, x_i and x_j are input p-dimensional color vectors, where the values of vector differences may vary significantly, however in this study, they remain relatively small. The similarity functions have a general form of $S(x_i,x_j)$, the arguments are skipped for simplicity in the formulae shown in this article; (,) in the formulae given here is a dot product between vectors.

The Gaussian radial basis function kernel then has the following form:

$$S_{Gaussian} = \exp\left(-\frac{\|y_i - y_j\|^2}{2\sigma^2}\right) \tag{3}$$

where $\sigma > 0$, σ is the parameter of the sensitivity of the function.

The sigmoid kernel based similarity metric can be presented as follows:

$$S_{sigmoid} = \tanh\left(\langle y_i, y_j \rangle k + \vartheta\right) \tag{4}$$

where k and ϑ are variable parameters, and dot product is denoted as $\langle\ ,\ \rangle$.

The performance of these measures is compared with the performance of twelve widely known metrics presented in Refs. 5 and 6.

The twelve other metrics are as follows[7]:

Metric 1

$$S_1 = \frac{x_i x_j^t}{|x_i||x_j|} = \cos(\theta) \tag{5}$$

where θ is the angle between vectors x_i and x_j

Metric 2

$$S_2 = \left(\frac{x_i x_j^t}{|x_i||x_j|}\right)\left(1 - \frac{||x_i| - |x_j||}{\max(|x_i|, |x_j|)}\right) \tag{6}$$

Metric 3

$$S_3 = \frac{|x_i|\cos(\theta) + |x_j|\cos(\theta)}{\left(|x_i|^2 + |x_j|^2 + 2|x_i||x_j|\cos(\theta)\right)^{\frac{1}{2}}} \tag{7}$$

Metric 4

$$S_4 = \frac{\cos(\theta)\left(|x_i|^2 + |x_j|^2 + 2|x_i||x_j|\cos(\theta)\right)^{\frac{1}{2}}}{|x_i| + |x_j|} \tag{8}$$

Metric 5

$$S_5 = 1 - \frac{\left(|x_i|^2 + |x_j|^2 - 2|x_i||x_j|\cos(\theta)\right)^{\frac{1}{2}}}{\left(|x_i|^2 + |x_j|^2 + 2|x_i||x_j|\cos(\theta)\right)^{\frac{1}{2}}} \tag{9}$$

Metric 6: Correlation Coefficient Method

$$S_6 = \frac{\sum_{k=1}^{p} |x_{ik} - \bar{x}_i||x_{jk} - \bar{x}_j|}{\left(\sum_{k=1}^{p}(x_{ik} - \bar{x}_i)^2\right)^{1/2}\left(\sum_{k=1}^{p}(x_{jk} - \bar{x}_j)^2\right)^{1/2}} \tag{10}$$

where $\bar{x}_i = \frac{1}{p}\sum_{k=1}^{p} x_{ik}$

Metric 7: Exponential Similarity Method

$$S_7 = \frac{1}{p}\sum_{k=1}^{p}\exp\left(-\frac{3}{4}*\frac{\left(x_{ik}-x_{jk}\right)^2}{\beta_k^2}\right) \tag{11}$$

where $\beta^2_k > 0$ is a parameter that is determined experimentally.

Metric 8: Absolute Value Exponent Method

$$S_8 = \exp\left(-\beta\sum_{k=1}^{p}\left|x_{ik}-x_{jk}\right|\right) \tag{12}$$

where $\beta > 0$.

Metric 9: Absolute Value Reciprocal Method

$$S_9 = 1-\beta\sum_{k=1}^{p}\left|x_{ik}-x_{jk}\right| \tag{13}$$

β is determined empirically.

Metric 10: Maximum-Minimum Method

$$S_{10} = \frac{\sum_{k=1}^{p}\min(x_{ik},x_{jk})}{\sum_{k=1}^{p}\max(x_{ik},x_{jk})} \tag{14}$$

Metric 11: Arithmetic Mean Minimum Method

$$S_{11} = \frac{\sum_{k=1}^{p}\min(x_{ik},x_{jk})}{\frac{1}{2}\sum_{k=1}^{p}(x_{ik}+x_{jk})} \tag{15}$$

Metric 12: Geometric Mean Minimum Method

$$S_{12} = \frac{\sum_{k=1}^{p}\min(x_{ik},x_{jk})}{\sum_{k=1}^{p}(x_{ik}x_{jk})^{1/2}} \tag{16}$$

Sensitivity parameters β, d, σ, k and ϑ have been manually set to certain values in this study, according to the criteria of the experimental settings, which will be discussed in the sections to follow. A series of experiments have been made with different values of these parameters and the most suitable have been given in this article.

Experiment

The tests of the viability of color similarity metrics were performed on Munsell Colors Matte dataset[9] (1269 matte Munsell color chips). The reflectance spectra had been measured by a Perkin–Elmer Lambda 9 UV/VIS/NIR spectrophotometer in the 380 – 800 nm interval with 1 nm wavelength resolution producing a total of 421 spectral components.[9] A set of tests concerning the possibility of discrimination of metameric colors was performed on a dataset of metamers, containing a total of three assortments of different colors. The reflectance spectra of these had been measured in a 380 – 780 nm interval with 2 nm wavelength resolution under illuminants D65 and A, resulting in 201 spectral components.

The main purpose of the experiments performed was to choose the metric that would be able to model human observer perception in the task of color discrimination. That means, first of all, a metric that should give comparable values of differences for perceptually equally disparate colors, and at the same time would account for the changes in Hue, Value and Chroma with the whole range of values. In this respect, one of the most important aspects of human perception in this area, the problem of metamerism, has to be considered. Metamerism is observed when two color samples appear to match under a particular light source, and then do not match under a light source with a different spectral power distribution. Metamerism is the result of the intrinsic property of color, which is a sensation of the human observer rather than a property of a certain colored object.[15] Thus, when modeling perceptual color differences, the metrics should be able to account for this effect. Another requirement imposed on the measures was the possibility of adjustment of the sensitivity of similarity measurements.

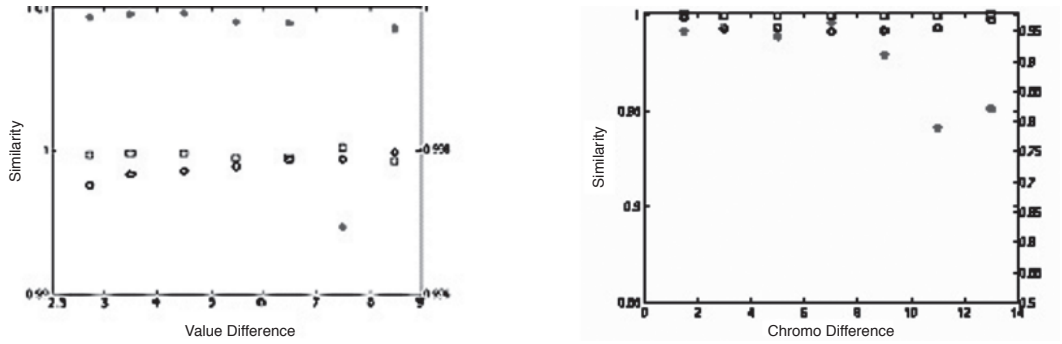
The first set of experiments concerned the possibility of color discrimination based on spectral information of Munsell colors with constant values of Hue and Chroma and adjacent values of Value. The Munsell color system is known to be practically uniform, meaning that each color is separated from its closest neighbor by equal perceptual distance.[15] Thus the measures considered were each expected to produce close to identical responses in each of the paired comparisons.

The first set was chosen so that it covers the entire range of Value. Hue was selected to be 5R, 5B and 5G, Chroma equal to 1, and Value ranging from 2.5 to 9.

The second set of tests was similar to the previous one, except that the Value and Hue remained constant, while Chroma was varied. Thus, Hue was set to 5R, 5B and 5G, Value to 6, and Chroma ranged from 1 to 14. Measures, defined in Eqs. (2)–(16), have been used in these experiments. The values of the sensitivity parameters set for the purpose of the experiments are as follows: in Eqs. (12) and (13) $\beta = 0.00012$; in Eq. (2) $d = 7$; Eq. (3) $\sigma = 0.6$ and in Eq. (4) $k = 1$, $\vartheta = -0.1$.

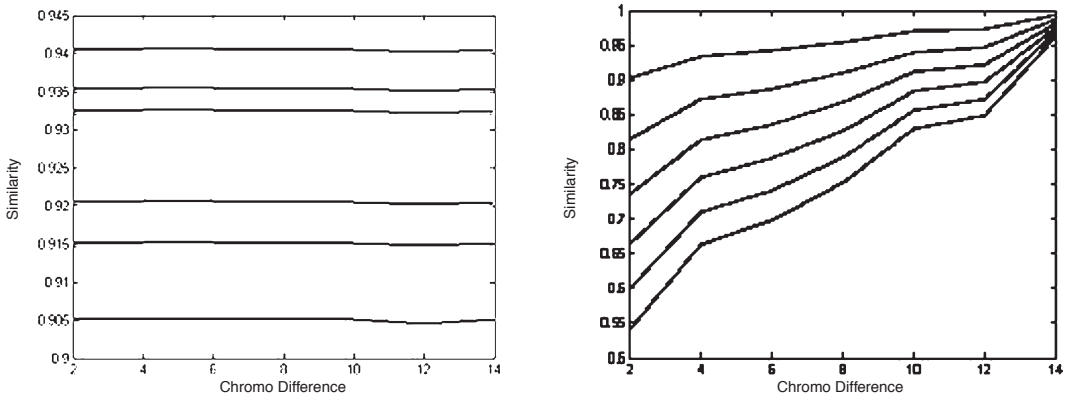
Considering the purpose of the whole experiment, we can divide the results obtained into several categories. First, the metrics that gave comparable values of differences for perceptually equally disparate colors should be referred. For that purpose, based upon the experimental settings described above, diagrams of the functional relations between similarity measures, Value and Chroma have been obtained (see Fig. 1). The diagrams show the three best metrics from the point of view of the linearity of the response given to the inputs, meaning that the measures gave close to linear (to a certain extent) responses to the changes in Value and Chroma. The metrics that generated the smoothest responses are Metric 1 (Eq. (5)), Metric 8 (Eq. (12)) and Gaussian radial basis function (Eq. (3)), presented in Fig. 1. A surge in the value of the similarity measures observed both with the changes in Value and Chroma at the point of Chroma difference 10–12, and Value difference 7–8, can be explained by the effect of metamerism. The spectra of the colors differ by a greater value as compared to the visual perception of the difference. The effect of the metamerism and responses of the measures to metameric pairs comparison will further be considered in a separate test.[4]

The second requirement set upon the metrics has been the possibility of adjustment of the sensitivity of similarity measurements. Considering Eqs. (2)–(16), it can be stated that the kernel methods Eqs. (2)–(4)



(a) (b)

Figure 1. (a) Similarity versus Value, (b) Similarity versus Chroma (Metric 1 – red asterisk (second scale); Metric 8 – Absolute-Value Reciprocal Method blue circles; RBF metric – green squares) for 5R Hue (sensitivity parameters set to $\beta = 0.00012$; $d = 7$; $\sigma = 0.6$, $k = 1$ and $\vartheta = -0.1$)



(a) (b)

Figure 2. Sensitivity of (a) Gaussian RBF (b) Metric8 (absolute-value exponent) for 5R Hue, Value 6 (β and σ varying in the range 0.005–0.03)

(polynomial, Gaussian RBF and sigmoid), exponential similarity, Eq. (11), absolute value exponential, Eq. (12), absolute value reciprocal Eq. (13) methods poses similarity terms, which allow adjustment of the measurements. However, the kernel methods provide a significantly better control over the whole equation. The results of testing of the sensitivity parameters in the measures for Gaussian RBF and Metric 8 (absolute value exponent) are presented in Fig. 2.[4]

A functional relation between the similarity values of Gaussian RBF and Metric 8 for different sensitivity term values are shown in Fig. 2, where both β and σ have been varied in the range of 0.005 – 0.03 with evenly spaced intervals. The resulting diagrams show that the sensitivity of the measures can be varied through the use of the special terms in a certain range. However, the absolute value exponential method tends to produce nonlinearities with the growth of the sensitivity value at higher values of Chroma and Value, which is an undesirable in this case effect. The choice of the Gaussian RBF and Metric 8 functions for testing is not accidental. Both of these metrics have shown the best results for the smoothness of the response of the sensitivity function requirement, thus it would seem reasonable to continue comparing them.

The next requirement set upon the similarity measures is the possibility of accounting for the changes in Hue, Value and Chroma with the whole range of values. This requirement implies that the sensitivity of the metric should be such that it would account for the differences in the color just noticeable by a human eye. The best results have been produced by kernel metrics Eqs. (2)–(4), Metric 7 (exponential similarity, Eq. (11)) and Metric 8 (absolute value exponential Eq. (12)).

The next set of tests concerns the possibility of discrimination of metameric colors. The main purpose

TABLE I. Values of similarities in paired comparison experiments on metameric colors (sensitivity parameters set to β = 0.00012; d = 7; σ = 0.6, k = 1 and ϑ = –0.1)

	Set 1	Set 2	Set 3	Set 3 (A)
Metric 1	1.0000	1.0000	1.0000	1.0000
Metric 2	0.8108	0.8853	0.8160	0.8153
Metric 3	1.0000	1.0000	1.0000	1.0000
Metric 4	1.0000	1.0000	1.0000	1.0000
Metric 5	0.8882	0.9365	0.8870	0.8860
Metric 6	0.8670	0.8930	0.8635	0.8621
Metric 7	0.6076	0.7680	0.7039	0.7017
Metric 8	1.0000	1.0000	1.0000	1.0000
Metric 9	1.0000	1.0000	1.0000	1.0000
Metric 10	0.8108	0.8853	0.8160	0.8132
Metric 11	0.8882	0.9365	0.8870	0.8854
Metric 12	0.8967	0.9396	0.8973	0.8970
Polynomial	1.0000	1.0000	1.0000	0.9532
Gaussian RBF	1.0000	1.0000	1.0000	0.9643
Sigmoid	0.9989	0.9990	0.9988	0.9843

of this study on the whole is to find a metric that would model human perception in the most accurate way. Thus although the spectra of the metameric assortments differ, the metrics should show very little or no differences under certain illuminants, and presence of differences under those illuminants that enable human observers to see the dissimilarities between colors. The responses of the measures presented by Eqs. (2)–(16) to metameric pairs comparison are given in Table I.

Each cell in Table I presents the result of the input of the sets of metameric colors (upper row) into the measures considered in this article (left side column). In the experiments, Sets 1–3 are taken under illuminant D65 with the differences between colors being invisible, and Set 3 (A) under illuminant A, under which the difference between the colors becomes visible. Considering the results presented in Table I we can state that there are several metrics that satisfy the requirements imposed upon them, meaning that the possibility of discrimination of metameric colors under certain illuminants and similarity of colors under other illuminants. We are interested in the measures that produced the best results in previous experiments. From that point of view all of the kernel metrics, Gaussian RBF function, and Polynomial could be considered among the most promising, while Metric 7 recognized the differences between metameric colors in all of the cases. Metric 8 at the same time was incapable of determining the differences in all of the sets.

In light of these results, several metrics could be considered as the most promising. The best results have been obtained using the Gaussian RBF and absolute value exponential (Metric 8) metrics. Both of these measures have produced approximately linear responses to perceptually uniformly distributed values of Value and Chroma (with Gaussian RBF producing the smoothest response (see Fig. 1)), at the same time, accounting for the change of color vectors with the whole range of values. The sensitivity of the function could be adjusted. However control over the sensitivity in the Gaussian RBF metric is significantly better due to a special term σ introduced into the formula.[4]

The last experiment concerning the possibility of discrimination of metameric colors shows that among the two best metrics only Gaussian RBF function was capable of modeling the behavior of the human expert in the task of color differencing, meaning that the metric was capable of detecting the dissimilarities under certain illuminants and showing the presence of the similarities between colors in the case of illuminant D65. All of that brought us to a conclusion of the efficiency of the Gaussian RBF measure in the task of color discrimination. Another result obtained from the experiments performed, is that the response of the similarity functions became smoother with the introduction of the efficiency curve[13] and the illumination factor.[14]

Conclusions

In this article, color similarity metrics in spectral space have been considered. The measures include twelve well-known metrics created upon well-known distance functions, such as City Block or Cosine distances, and a set of novel kernel-based color similarity measures. The performance of all of these measures has been tested against Munsell Matte spectral dataset[9] and a dataset of metameric colors.

The purpose of the experiments has been to find a metric that would model the behavior of the human observer in the task of color differencing. This task implies, first of all, that the metrics considered should give comparable values of differences for equally disparate color, at the same time accounting for the values of change in Hue, Chroma and Value with the whole range of values.

Based upon the results presented above, it can be concluded that the kernel-based approach to color differencing in spectral space is efficient. The performance of the kernel-based metrics gave results comparable with the traditional color similarity measures. Moreover, the Gaussian RBF kernel performed more effectively considering the traditional color similarity methods. The response of the measure was smoother in the case of both the Value and the Chroma change, and the sensitivity of the function was greater and could be varied in a more efficient way (through the use of a special sensitivity term). Furthermore, weighting the input data by the Spectral Luminous Efficiency Function for photopic vision[13] and spectral radiance function of a certain light source[14] produced a smoother output of the similarity functions.

Another important aspect considered in this article is the possibility of discrimination of metameric colors. In order to model the human perception in the most accurate way the metrics should be able to show the dissimilarities between the metameric colors under certain illuminants, at the same time exhibiting similarities in the case of other illuminants. From this point of view it can be stated that Gaussian RBF function performed in the most rigorous way. On the whole, it is necessary to emphasize that in consideration of the overall purpose of the research, i.e., the possibility to model the behavior of the human expert in the task of color differentiation, the Gaussian RBF function can be considered the most efficient.

Acknowledgment. The spectra in the Munsell database and in the database of metameric colors were obtained from the Department of Computer Science, University of Joensuu, Finland.

References

1. J. Hardeberg and J. Gerhardt, "Characterization of an Eight Colorant Inkjet System for Spectral Color Reproduction", *Proc. Second European Conference on Colour Graphics, Imaging and Vision*, (IS&T, Springfield, VA, 2004) p. 263.

2. M. Rosen, E. Hattenberger and N. Ohta, “Spectral Redundancy in a 6-inkInk-jet Printer”, *Proc. The Digital Photography Conference 2003,* (IS&T, Springfield, VA,, 2003) p. 236.
3. *Recommendations on Uniform Color Spaces, Color Difference Equations, Psychometric Color Terms*, Publication CIE 15 (E.–1.3.1), Suppl.No.2, Bureau Central de la CIE, Vienna, 1971.
4. D. Kalenova, P. Toivanen and V. Botchko, “Color Differences in a Spectral Space”, *Proc. CGIV 2004*, (IS&T, Springfield, VA, 2004), p. 368.
5. K. Plataniotis and A. Venetsanopoulus, *Color Image Processing and Applications*, (Springer–Verlag, Berlin, 2000), p. 70.
6. Y. Leung, *Spatial Analysis and Planning under Imprecision*, (North-Holland Publishers, Netherlands, 1988), p. 116.
7. M. Hild and T. Emura, “Which Color Similarity Measure is Most Effective for Background-Frame Differencing?”, *Proc. IS&T/SID 9th Color Imaging Conference*, (IS&T, Springfield, VA, 2001), p. 168.
8. J. Orava, T. Jaaskelainen and J. Parkkinen, “Color Differences in a Spectral Space”, *Proc. IS&T's PICS 2003*, (IS&T, Springfield, VA, 2003), pp. 205–209.
9. Color Spectra Database, website http://www.it.lut.fi/ip/research/color/database/database.html (cited 10.08.2003).
10. B. Schölkopf and A. Smola, *Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond*, (MIT Press, Cambridge, MA, 2002), p. 45.
11. *Munsell Book of Color – Matte Finish Collection* (Munsell Color, Baltimore, Md., 1976), p. 150.
12. Michael Hild, “On the Effectiveness of Color Similarity Measures in Background-Frame Differencing Applications”, *Proc. CGIV*, (IS&T, Springfield, VA, 2002), p. 306.
13. *Spectral Luminous Efficiency Function for Photopic Vision*, (CIE Publication 86–1990).
14. *CIE Standard Illuminant D65 Relative Spectral Power Distribution*, (CIE Publication 15.2–1986).
15. D. Malacara, *Color Vision and Colorimetry: Theory and Applications*, (SPIE Press, Bellingham, WA, 2002), pp. 80–112.

Publication V

KALENOVA, D., TOIVANEN, P., BOCHKO, V.,
Probabilistic Spectral Image Quality Model

Proceedings of International Color Association (AIC), 10th Congress, Granada,
Spain, May 8-13, 2005, pages1641-1645.

Probabilistic spectral image quality model

D. Kalenova, P. Toivanen and V. Bochko
Laboratory of Information Processing, Department of Information Technology, Lappeenranta University of Technology
P.O. Box 20, Lappeenranta, 53850, Finland
Corresponding author: D. Kalenova (Diana.Kalenova@lut.fi)

ABSTRACT

In this paper a novel technique of spectral image quality evaluation - probabilistic spectral image quality model is proposed. This study is based on the statistical image model that sets a relationship between the parameters of the spectral and colour images, and the overall appearance of the image. It has been found that variation of standard deviation of the spectra affects the colourfulness of the image, while kurtosis change influences the highlight reproduction or, so called vividness. The model presented in this study is an extension of a previously published spectral colour appearance model. An original study has been extended to account for the probabilistic nature of perception. A quality loss function, quantifying the decrease in quality in JNDs as a function of deviation of image attributes from the optimum, has been constructed in the framework of the research. The study shows that the presented probabilistic spectral image quality model is efficient in the task of quality of spectral image evaluation and prediction.

1. INTRODUCTION

The idea of reproduction of spectral images on printed media and display has recently gained popularity, resulting in a number of publications[1, 2]. Such multiprimary displays and printers reproduce images with a closer spectral match to the original scene captured. Appearance of such devices gives rise to a problem already existing for conventional tools - assessment of perceptual image quality given only physical parameters of the image. Probabilistic spectral image quality model, presented in this work is intended to create a paradigm that would allow description of the quality of spectral images in terms of objectively measurable parameters of spectral images. The probabilistic spectral image quality model, presented in this paper, can be used for spectral image quality evaluation and prediction in tasks of e.g. imaging devices production and calibration, printing industry and in some other industrial applications.

The model described in this paper is based on the results of a previously published spectral colour appearance model. The model, introduced in[3], has been used for colour image quality estimation, with a colour image, reproduced through the spectral image. Spectral colour appearance model has demonstrated that there is a close connection between the quality judgments of the observers and the parameters of the model, i.e. vividness and colourfulness, which in turn have been proven to depend upon statistical characteristics of the spectral images, in particular standard deviation and kurtosis. This corresponds with the results obtained by other researchers[4]. Fedorovskaya and de Ridder in [4] suggest that scaling of the perceived strength of colourfulness impression is linearly related to the average and standard deviation of the chroma variations in CIELUV colour space.

In every set of images produced through variation of both parameters of the spectral colour appearance model, one image has been found to be of maximal quality, meaning that it had the highest quality judgment given by the observers. Moreover, there has been found a significant difference between the quality judgments of the scenes, meaning that part of the images exhibited a clear maximum at points close to the original scenes, whilst the others have had a significant shift in the highest quality judgment position. This model has already proven to be effective in the task of image quality evaluation, however it lacks universality, in a sense, that units of quality used are artificial and have weak mathematical basis, which, in turn, does not allow comparison with analogous reference systems. Another serious drawback is that modelling of a combined effect of the

parameters of the model on the overall quality impression has been reproduced via Fuzzy Inference System, and has resulted in serious error rates in some cases[3].

The idea underlying the model presented in this paper is that human perception is probabilistic in nature. The units of quality used throughout this work are just noticeable differences (JND), which are natural units with which to perform certain image quality calculations, widely used in the research. JNDs allow constructing calibrated numerical scales that quantify wide ranges of quality[5]. A quality loss function, quantifying the decrease in quality in JNDs as a function of deviation of image attributes from the optimum, has been constructed in the framework of the research on the basis of the results obtained from twenty colour-normal observer's judgments.

In general, the spectral colour appearance model and probabilistic spectral image quality model can be attributed to a class of preferential quality models. A number of publications exist on the topic of preference in colour and tone reproduction in the framework of image quality investigation [6, 7, 8]. All of these works are trying to establish a connection between the preferential attributes of a colour image and the overall perceived image quality.

2. STATISTICAL MODEL

A generalized statistical model, characterizing the behaviour of statistical characteristics of natural spectral images $\mathbf{f}(\mathbf{x})$, presented as n-dimensional vector random field, is described by the following equation[9]:

$$\mathbf{f}(\mathbf{x})=\boldsymbol{\mu}+\mathbf{D}\mathbf{g}(\mathbf{x}) \quad (1)$$

where $\mathbf{x}$ is a vector with each element being spatial dimension; $\boldsymbol{\mu}$ is a mean vector, $\mathbf{g}(\mathbf{x})$ is a normalized vector image with zero mean and unit standard deviation for each component, $\mathbf{D}=\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, where σ_i is standard deviation in the component[3]. The following parameters: α, β and $\mathbf{k}$ are used for modifying the colourfulness and vividness change in the experiment.

Vector $\boldsymbol{\sigma}$ is presented in the following form:

$$\boldsymbol{\sigma}=\alpha\beta\boldsymbol{\sigma}_v+(1-\alpha)\boldsymbol{\sigma}_c \quad (2)$$

where $\alpha=(\sigma_{max}-\sigma_{min})/\sigma_{max}$ is the relationship between constant and variable parts of standard deviation and β is a contrast variation coefficient, $\boldsymbol{\sigma}_v$ is a variable component vector of $\boldsymbol{\sigma}$, $\boldsymbol{\sigma}_c$ is a constant component vector of $\boldsymbol{\sigma}$ [3].

$\mathbf{g}(\mathbf{x})$ is defined through gamma-Charlier histogram transform of $\mathbf{f}_s(\mathbf{x})$ and a kurtosis vector $\mathbf{k}$ as follows[3]:

$$\mathbf{g}(\mathbf{x})=\mathbf{H}(\mathbf{f}_s(\mathbf{x}),\mathbf{k}) \quad (3)$$

where $\mathbf{f}_s(\mathbf{x})$ is a normalized image of $\mathbf{f}(\mathbf{x})$, with zero mean and unit standard deviation for each component. To affect image appearance through histogram transform, all kurtosis elements are proportionally modified according to the given maximum of the kurtosis value reproduction of a spectral image.

The task of quality optimization is a complicated task that requires significant computational resources. Spectral images contain large amounts of information, which have to be manipulated in order to influence the overall impression of the display. Usually, some implicit assumptions are made in order to limit the amount of computations. The assumption underlying this study is that only global variations are taken into account, which, in turn, originates from the fact that all parts of the image have been captured under the same illuminant or belong to the same object. Thus the same modifications are applied to all pixels of the image irrespective of the content[4]. Based upon this principle the generalized statistical model is applied to spectral images in this study.

3. EXPERIMENT

Experiments were performed on spectral images of natural scenes from[10]. Five images – *inlab1*, *inlab2*, *inlab5*, *jan13am* and *rleaves* were selected (see Fig.1). Images have the following dimensions: 256x256 pixels, and 31 spectral components per each pixel. For the purpose of the experiments the area of 128x128 pixels were selected. Images were captured by a CCD (charge coupled device) camera in a 400-700 nm wavelength range at 10 nm intervals. The images selected were taken indoor (in a controlled environment, i.e. dark-lab or glass-house).

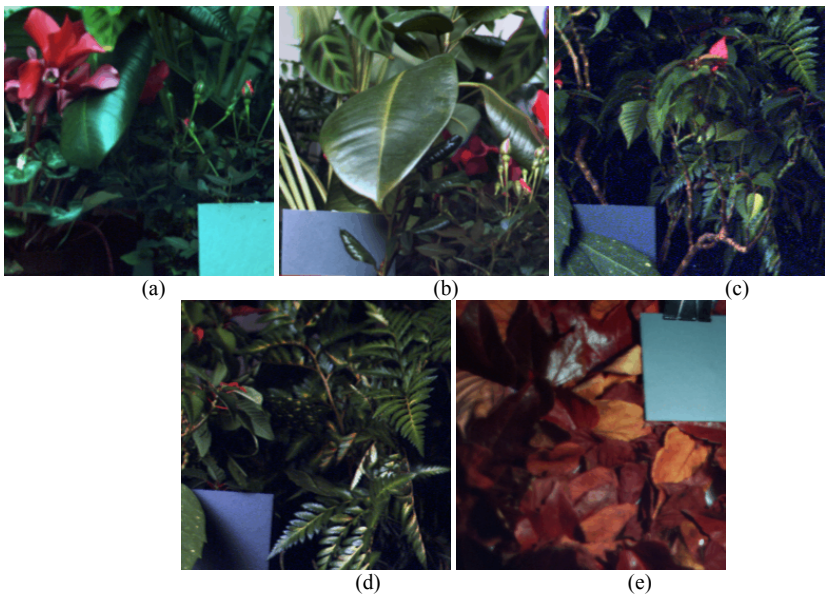
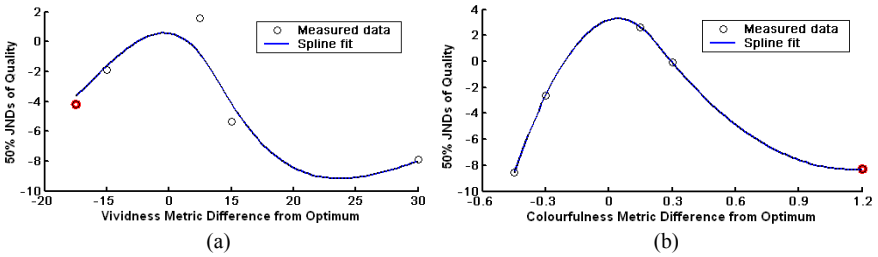


Figure 1: Colour reproduction of original, unprocessed spectral images used in the experiments (a) inlab1, (b) inlab2, (c) inlab5, (d) jan13am, (e) rleaves

First, a set of test images was produced using the colourfulness parameter. The term includes both contrast and colour saturation. Thus, colourfulness was varied through standard deviation, using Eq.2, by changing the α and β coefficients in the ranges of [0.5,1) and [1,1.8] respectively. The second set of tests was produced through variation of the vividness parameter, closely related to highlight reproduction in an image. As the highlight was modified through kurtosis change, the test images were produced with the help of Eq.3 (with k_{max} varied within the interval [5,60]).

Both of the test sets were presented to twenty colour-normal test persons with normal or corrected to normal visual acuity without colour deficiencies. Multiple images with rendition of both of the parameters for each scene have been evaluated using the paired comparison technique, where the subjects had to indicate which of the images is of the highest quality relative to either of the parameters. Thus the experiments have been performed in a manner yielding assessments calibrated in JNDs of overall quality, where one JND can be regarded as representing a just significant difference of quality. Thus, an optimal position for each of the abovementioned parameter values has been obtained, and a mean, over all of the scenes and observers, quality loss function has been obtained (see Fig.2), where x-axis represents the difference from the optimal position of each of the attributes.



(a) (b)

Figure 2. Measured quality loss functions of vividness (a) and colourfulness (b) averaged over the collection of scenes and observers

JND in this study is a so called 50% JND, which means a stimulus difference leading to a 75%:25% proportion in paired comparisons, regardless of the nature of the attribute or the origin of the sample differences, meaning that 75% of users regarded a certain sample to be of higher quality. To construct the quality loss function the value of each of the attributes of the highest rated image was identified for each scene and observer, and the fraction of times each position had been preferred was computed. To convert each fraction to a probability density, it was divided by the width of an attribute interval including the sample position.

Although it has been found that both colourfulness and vividness are observer and scene susceptible, the quality loss function for each observer and scene can be regarded as a member of a family, and thus could be averaged to produce a mean quality loss function for the collection of observers and scenes tested[5]. The experiment showed that there is a close relationship between the quality judgments and parameters of the spectral images.

4. CONCLUSIONS

In this paper, a probabilistic spectral image quality model has been considered. The model sets a relationship between statistical characteristics of spectral images and overall quality judgments. The model described in this paper is an extension of a previously published spectral colour appearance model. The original study has been extended to account for the probabilistic nature of human perception. Varying the statistical parameters of spectral image, i.e. kurtosis and standard deviation representing respectively colourfulness and vividness, a set of test images has been obtained. The images have been evaluated by twenty colour normal subjects in a set of paired comparison experiments. A quality loss function quantifying deviation of image attributes from the optimum in JND units has been constructed. The function can further be used in the task of minimization of deviation from the optimal values of the imaging system, in order to improve the quality of the output. In general, both the probabilistic spectral image quality model and the spectral colour appearance model can be attributed to a class of preferential image quality models and can serve as an efficient tool of image quality characterization and prediction.

References

1. J. Hardeberg, J. Gerhardt, “Characterization of an eight colorant inkjet system for spectral color reproduction”, in Proceedings of the Second European Conference on Colour Graphics, Imaging and Vision (Society for Imaging Science and Technology, Springfield, Va., 2004), pp. 263-267.
2. M. Rosen, E. Hattenberger, N. Ohta, “Spectral redundancy in a 6-ink inkjet printer”, in Proceedings of The Digital Photography Conference 2003 (Society for Imaging Science and Technology, Springfield, Va., 2003), pp. 236-243.
3. D. Kalenova, V. Botchko, T. Jaaskelainen, J. Parkkinen, “Spectral color appearance modeling”, in Proceedings of The Digital Photography Conference 2003 (Society for Imaging Science and Technology, Springfield, Va., 2003), pp. 381-385.
4. E. A. Fedorovskaya, H. de Ridder, F. J. J. Blommaert. “Chroma variations and perceived quality of colour images of natural scenes”, Color research and applications, 22, 96-110 (1997).
5. B. W. Keelan, “Handbook of image quality: Characterization and Prediction”, (Marcel Dekker, Inc., NY, 2002), pp. 35-47.
6. J. D. Buhr, H. D. Franchino, “Color Image reproduction of scenes with preferential tone mapping”, (U.S. Patent #5, 1995),447, 811.
7. H. de Ridder, “Saturation and lightness variation in color images of natural scenes”, J. Imaging Sci. Technol. 40(6), 487-493 (1996).
8. R. Janssen, Computational Image Quality, Society of Photo-Optical Instrumentation Engineers, (Belingham, Washington ,2001), pp. 20-35.
9. V. Botchko, H. Kälviäinen, J. Parkkinen, “Highlight reproduction using multispectral texture statistics”, in Proceedings of the Third International Conference on Multispectral Color Science (Department of Computer Science, University of Joensuu, Joensuu, Finland, 2001), pp. 61-65.
10. A. Parraga, G. Brelstaff, T. Troscianko, “Color and luminance information in natural scenes”, J. Opt. Soc. Am. A 15, 3-5 (1998).

Publication VI

KALENOVA, D., TOIVANEN, P., BOCHKO, V.,
Preferential Spectral Image Quality Model

Proceedings of Image Analysis (SCIA), 14th Scandinavian Conference, Joensuu,
Finland, June 19-22, 2005, pages 389-398.

Preferential Spectral Image Quality Model

D. Kalenova, P. Toivanen, and V. Bochko

Spectral Image Processing and Analysis Group, Laboratory of Information Processing,
Department of Information Technology, Lappeenranta University of Technology,
P.O.Box 20, 53851 Lappeenranta, Finland
{Diana.Kalenova, Pekka.Toivanen, Vladimir.Botchko}@lut.fi

Abstract. In this paper a novel method of spectral image quality characterization and prediction, preferential spectral image quality model is introduced. This study is based on the statistical image model that sets a relationship between the parameters of the spectral and color images, and the overall appearance of the image. It has been found that standard deviation of the spectra affects the colorfulness of the image, while kurtosis influences the highlight reproduction or, so called vividness. The model presented in this study is an extension of a previously published spectral color appearance model. The original model has been extended to account for the naturalness constraint, i.e. the degree of correspondence between the image reproduced and the observer's perception of the reality. The study shows that the presented preferential spectral image quality model is efficient in the task of quality of spectral image evaluation and prediction.

1 Introduction

The nature and the scope of imaging have been undergoing dramatic changes in the recent years. The latest trend is the appearance of multiprimary displays and printers that reproduce images with a closer spectral match to the original scene captured [1,2]. Appearance of such devices gives rise to a problem already existing for conventional tools - assessment of perceptual image quality given only physical parameters of the image. The demand for a quantitative analysis of image quality has dramatically increased. Preferential spectral image quality model, presented in this work is intended to create a paradigm that would allow description of the quality of spectral images in terms of objectively measurable parameters of spectral images in connection with the subjective quality metrics. The preferential spectral image quality model, presented in this paper, can be used for spectral image quality evaluation and prediction in tasks of e.g. imaging device production and calibration, printing industry and in some other industrial applications.

The model described in this paper is based on the results of a previously published spectral color appearance model. The model, introduced in [3], has been used for color image quality estimation, with a color image, reproduced through the spectral image. Spectral color appearance model has demonstrated that there is a close connection between the quality judgments of the observers and the parameters of the model, i.e. vividness and colorfulness, which in turn have been proven to depend upon statistical characteristics of spectral images, in particular, standard deviation and

H. Kalviainen et al. (Eds.): SCIA 2005, LNCS 3540, pp. 389 – 398, 2005.

kurtosis. This corresponds with the results obtained by other researchers [4]. Fedorovskaya and de Ridder in [4] suggest that scaling of the perceived strength of colorfulness impression is linearly related to the average and standard deviation of the chroma variations in CIELUV color space.

In every set of images produced through variation of both parameters of the spectral color appearance model, one image has been found to be of maximal quality, meaning that it had the highest quality judgment given by the observers. Moreover, there has been found a significant difference between the quality judgments of the scenes, meaning that part of the images exhibited a clear maximum at points close to the original scenes, whilst the others have had a significant shift in the highest quality judgment position. This model has already proven to be effective in the task of image quality evaluation, however it lacks universality, in a sense, that units of quality used are artificial and have weak mathematical basis, which, in turn, does not allow comparison with analogous reference systems. Another serious drawback is that modeling of a combined effect of the parameters of the model on the overall quality impression has been reproduced via Fuzzy Inference System, and has resulted in serious error rates in some cases [3].

In general, the spectral color appearance model and the preferential spectral image quality model, introduced in this paper, can be attributed to a class of preferential quality models. A number of publications exist on the topic of preference in color and tone reproduction in the framework of image quality investigation [5, 6, 7]. Among the whole range of preferential characteristics, contrast, saturation and memory color reproduction are the most common ones. These features change is evident in the image and is highly dependent upon the observer and the content of the scene. Normally, such image attributes have an optimal value where the overall impression is most pleasing [7].

Spectral color appearance model has been created upon the assumption that colorfulness and vividness can efficiently describe image quality, with colorfulness including both contrast and saturation. Previous research is extended in this work to account for memory color reproduction or as it will further be called naturalness. The naturalness constraint imposed upon the image quality stems from an intuitive assumption that high quality images should at least be perceived as “natural”. At base, this assumption rests on the psychological model of the quality judgment constitution. Accordingly, the impression of an image is formed as a result of comparison of the output of the process of visual perception and the internal representation of an object, which, in turn, is based on the previous experience (i.e. memory representation or prior knowledge of reality). That is, a high quality image complies with the ideas and expectations of the scene captured in the image at hand. Several works exploring the influence of naturalness on color image quality exist at the moment [4, 8], particularly, in the field of color photography. A direct dependence between the naturalness constraint and the quality judgments has been experimentally found in these, with memory colors being relatively consistent among different observers. However, for the case of colorfulness variation a discrepancy between the naturalness judgments and the perceived quality has been found, i.e. observers perceived more colorful images as being of higher quality, at the same judging these images as unnatural. This phenomenon can be explained from the information-processing point of view, a high degree of naturalness is a necessary, but not a sufficient condition for

quality perception, a usefulness condition has to be satisfied as well, which, in turn, leads to a discriminability principle. In other words, a highly saturated image is perceived to be of high quality, despite being unnatural, due to an increased possibility of discerning certain features in an image [4, 8]. In this study we are trying to establish a connection between quality judgments of spectral images, spectral image attributes and the naturalness constraint with regard to the principles mentioned.

2 Statistical Model

A generalized statistical model, characterizing the behavior of statistical characteristics of natural spectral images $\mathbf{f}(\mathbf{x})$, presented as n-dimensional vector random field, is described by the following equation [9]:

$$\mathbf{f}(\mathbf{x})=\boldsymbol{\mu}+\mathbf{D}\mathbf{g}(\mathbf{x}) \tag{1}$$

where $\mathbf{x}$ is a vector with each element being spatial dimension; $\boldsymbol{\mu}$ is a mean vector, $\mathbf{g}(\mathbf{x})$ is a normalized vector image with zero mean and unit standard deviation for each component, $\mathbf{D}=\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, where σ_i is standard deviation in the component [3]. The following parameters: α, β and k_{max} are used for modifying the colorfulness and vividness change in the experiment.

Vector $\boldsymbol{\sigma}$ is presented in the following form:

$$\boldsymbol{\sigma}=\alpha\beta\boldsymbol{\sigma}_v+(1-\alpha)\boldsymbol{\sigma}_c \tag{2}$$

where $\alpha=(\sigma_{max}-\sigma_{min})/\sigma_{max}$ is the relationship between constant and variable parts of standard deviation, affecting the saturation of colors in an image, and β is a contrast variation coefficient, $\boldsymbol{\sigma}_v$ is a variable component vector of $\boldsymbol{\sigma}$, $\boldsymbol{\sigma}_c$ is a constant component vector of $\boldsymbol{\sigma}$ [3].

$\mathbf{g}(\mathbf{x})$ is defined through gamma-Charlier histogram transform of $\mathbf{f}_s(\mathbf{x})$ and a kurtosis vector $\mathbf{k}$ as follows [3]:

$$\mathbf{g}(\mathbf{x})=\mathbf{H}(\mathbf{f}_s(\mathbf{x}),\mathbf{k}) \tag{3}$$

where $\mathbf{f}_s(\mathbf{x})$ is a normalized image of $\mathbf{f}(\mathbf{x})$, with zero mean and unit standard deviation for each component. To affect the image appearance through histogram transform, all kurtosis elements are proportionally modified according to the given maximum of the kurtosis value k_{max} [3].

The task of quality manipulation is a complicated task that requires significant computational resources. Spectral images contain large amounts of information, which have to be manipulated in order to influence the overall impression of the display. Usually, some implicit assumptions are made in order to limit the amount of computations. The assumption underlying this study is that only global variations are taken into account, which, in turn, originates from the fact that all parts of the image have been captured under the same illuminant or belong to the same object. Thus the same modifications are applied to all pixels of the image irrespective of the content [4]. Based upon this principle the generalized statistical model is applied to spectral images in this study.

3 Experiment

Experiments were performed on spectral images of natural scenes from [10]. Five images – *inlab1*, *inlab2*, *inlab5*, *jan13am* and *rleaves* were selected (see Fig.1). Images have the following dimensions: 256x256 pixels, and 31 spectral components per each pixel. For the purpose of the experiments the area of 128x128 pixels were selected. Images were captured by a CCD (charge coupled device) camera in a 400-700 nm wavelength range at 10 nm intervals. The images selected were taken indoor (in a controlled environment, i.e. dark-lab or glass-house).

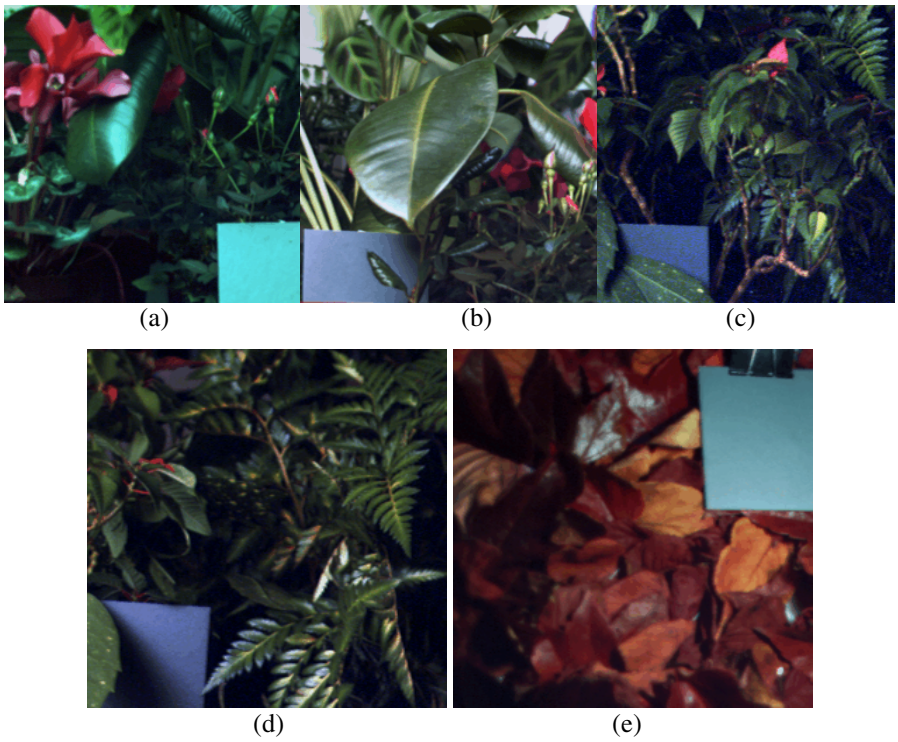


Fig. 1. Color reproduction of original spectral images used in the experiments (a) inlab1, (b) inlab2, (c) inlab5, (d) jan13am, (e) rleaves

Experimental settings, which include the number and the criteria for selection of observers, test stimuli, instructions and viewing conditions requirements, were chosen to comply with [11]. According to this standard relative quality values should be obtained from at least ten observers and three scenes, and all of the observers have to be tested for normal vision and visual acuity. Thus, we chose twenty observers to participate in the experiments. They had normal or corrected-to-normal vision without color deficiencies. To prevent the loss of quality of judgments due to fatigue the duration of the experimental sessions was limited to one hour, in case of more time needed the experiments continued after a break. Viewing conditions followed the

requirements given in [12]. Therefore, the general room lighting was shaded and was set so that it neither directly nor indirectly influenced the viewing surface. When viewing slides, the frames were darkened to a 10% brightness level for a width of 75 mm.

Although original images were presented, it should be emphasized that they were not explicitly identified to observers as such. First, a set of test images was produced using the colorfulness parameter. The term includes both contrast and color saturation. Thus, colorfulness was varied through standard deviation, using Eq.2. By changing α and β coefficients it was possible to receive new values for constant and variable parts of standard deviation. This procedure was applied to the images with values of (α,β) equal to (0.55,1), (0.75,1), (1,1.3), (1,1.6) consequently. The second set of tests was produced through variation of the vividness parameter, closely related to highlight reproduction in an image. As the highlight was modified through kurtosis change, the test images were produced with the help of Eq.3 (with k_{max} equal to 5, 10, 30, 60). The effect of the change of both of parameters on the overall appearance had been shown in [3]. Both test sets were presented to the subjects, who had to rate the naturalness of the images on a ten-point numerical category scale ranging from one (unnatural) to ten (real life). The following instructions for the experiments were given to the observers [4]:

"You will be presented a series of images. Your task is to asses the naturalness of images, using and integer number from one to ten, with one corresponding to the lowest degree of naturalness and ten to the highest. Naturalness is defined in this case as the degree of correspondence between the image reproduced on the screen and reality, i.e. the original scene as you picture it."

The results of the tests are given in Table 1, where each cell corresponds to an averaged naturalness evaluation score, with outliers being excluded from consideration. The columns denoted as 1 present results of the tests produced through the colorfulness change, and 2 with the vividness change respectively. In Table 1 Image 1 and Image 2 have parameters (α,β, k_{max}): (0.55,1,5), (0.75,1,10), Image 3 is the original, Image 4 and Image 5 have respectively parameters (α,β, k_{max}) equal to (1,1.3,30), (1,1.6,60). Note that either (α,β) (for colorfulness change) or k_{max} (for vividness change) were varied, while the rest of the parameters were kept constant.

Table 1. Mean values of naturalness evaluation scores

Quality	Image1		Image 2		Image 3		Image 4		Image 5	
	1	2	1	2	1	2	1	2	1	2
Inlab1	2.37	5.41	5.83	6.34	8.12	8.34	9.67	9.83	8.15	9.12
Inlab2	2.25	6.03	6.93	7.98	8.93	9.67	9.87	5.56	8.41	3.98
Inlab5	3.84	7.34	6.55	7.53	8.34	9.12	9.85	9.53	8.56	9.34
Jan13AM	3.10	6.56	7.12	8.96	8.67	9.87	9.73	6.76	8.17	4.87
Rleaves	3.32	6.17	6.86	9.10	8.16	8.17	9.54	7.65	9.50	5.35

Looking at Table 1 we can state that the peaks of the naturalness judgments do not lie within the original image area, which in turn brings us to a conclusion that users generally prefer slightly modified images. Fig. 2 illustrates the connection between

the naturalness constraint and statistical parameters of the spectral images varied at the experiments.

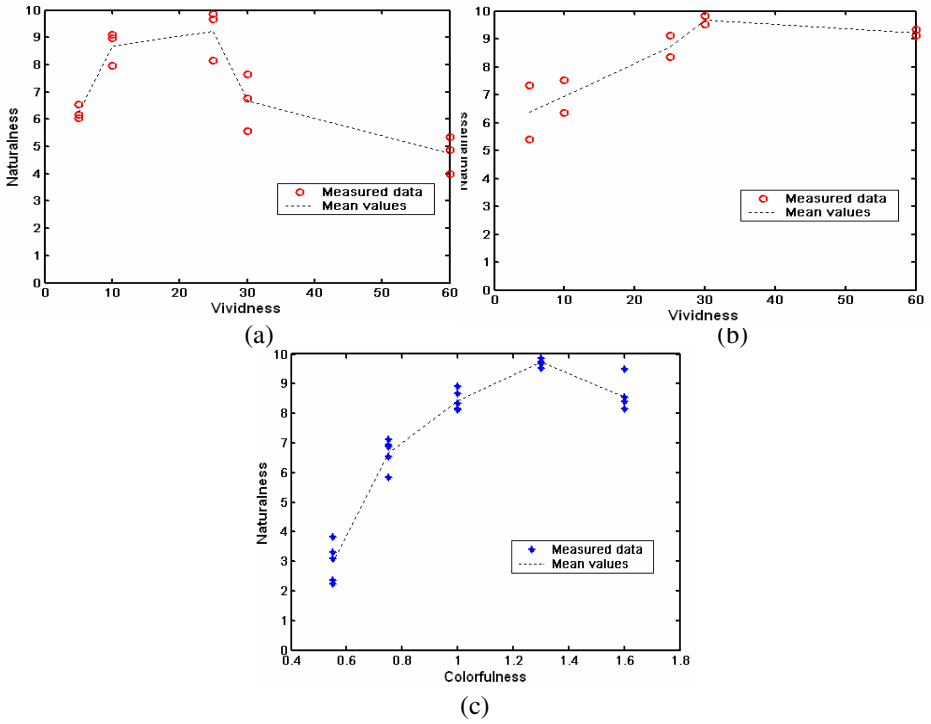


Fig. 2. Averaged naturalness estimations vs. statistical parameters of spectral images. Vividness (a) of inlab2, jan13am, rleaves; vividness (b) of inlab1 and inlab5; colorfulness (c) of all of the images

Fig. 2 demonstrates that there is positive correlation between the attributes of spectral images and the naturalness constraint. Fig. 2 (a) and (b) specifically present the relation between the vividness spectral image attribute and the naturalness constraint. It is clearly visible that the naturalness maximum lies at points close to the original image. We have separated the vividness parameter versus naturalness plots onto two parts due to a different form of dependency between the two. Fig. 2(a) contains a plot of the *inlab2*, *jan13am, rleaves* image judgments, and Fig. 2(b) - *inlab1* and *inlab5*. In the first case images exhibit a sharper drop in the naturalness judgments than in the second one, in fact, in the second case the decrease in naturalness is such that the naturalness remains approximately close to the maximal value. Such discrepancy in the image judgments has also been obtained when evaluating the quality of the images [3]. Both of the phenomena can be attributed to a fact that images *inlab1* and *inlab5* contain objects that attract the most of the observers' attention, compared to the objects situated at the background. Moreover, these objects lie in the red area of the spectrum, which assumes that observers are not

susceptible to minor variations in these areas due to the properties of the human visual system. Thus, the drop in quality and in naturalness is less definitive.

Fig. 2(c) demonstrates the connection between the colorfulness parameter and the naturalness constraint. It can be stated that observers perceive slightly more colorful than original images as being the most natural ones. This effect is consistent with the results obtained in the experiments with color images, stating that there is a tendency for memory colors to be slightly more saturated compared with actual object colors [13]. Moreover, considering the fact that observers have previously rated the images with higher colorfulness values as being of higher quality [3] we can state that memory color reproduction influences the preferred color reproduction of the objects [14].

Another important characteristic of image naturalness is correlation with the quality perception. For this purpose the quality judgment values have been taken from the previous study [3] and plotted against the naturalness obtained in this study. Fig. 3 demonstrates a plot of the quality judgments versus the naturalness constraints for both vividness (red circles) and colorfulness (blue asterisk) test sets. Such a comparison is possible due to the fact that experimental settings (number of observers, number and contents of scenes, viewing conditions, etc.) are similar in both of the experiments, moreover the algorithm of modification and values of the statistical parameters of spectral images are the same.

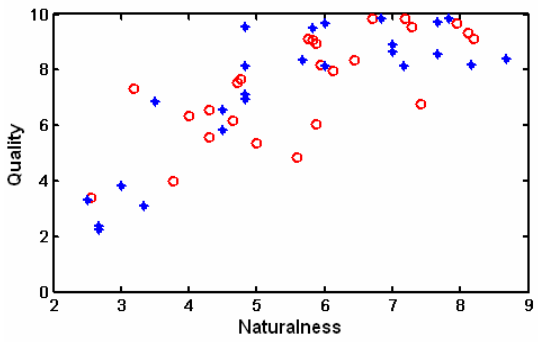


Fig. 3. Averaged naturalness estimations vs. perceptual quality estimations of images with the vividness (red circles) and colorfulness (blue asterisk) change

Fig. 3 illustrates the interdependence between the quality judgments and the naturalness constraint. It is clear that there is a strong correlation between these notions. In fact, the correlation coefficient between the quality judgments and naturalness estimations computed over all of the scenes equals to 0.8196 for colorfulness, 0.7018 in the case of vividness test sets, and the overall correlation is equal to 0.7752.

With the increase of naturalness the quality also increases, which proves the preliminary assumption made, stating that in order for the image to be of good quality it should at least be perceived natural. The spread in the plot can be attributed to a lack of test images and a rough scale of spectral image attributes accepted in this

study. However, even such a small test set was enough to prove a connection between the attributes of spectral images, naturalness and overall image quality impression.

Even though we can see from the plot in Fig. 3 that there is a connection between the naturalness constraint and the quality judgments of the users, it is relatively difficult to predict what would be the effect of the naturalness change on image quality, and how fast does the quality decrease with the decrease in naturalness, which in turn can be varied through variation of any of the attributes of spectral images. Accordingly naturalness could serve as a universal image attribute that would allow modeling both image quality and the joint effect of attributes of spectral images on the overall perception of the image reproduction. In order to model the effect of naturalness on quality a preference distribution function of naturalness expressed in terms of JNDs has been constructed. According to [11] JND is a stimulus difference that yields a 75%:25% proportion in a forced-choiced paired comparison. In other words, JND is the smallest possible difference between two stimuli, noticeable by observers. The standard distinguishes two types of JND units, attribute and quality JNDs. In our study we have used the quality JNDs, which is a measure of the significance of perceived differences on overall image quality. Essentially all of the observers could detect the difference and identify which of the samples have had higher naturalness.

The preference data was obtained with the use of the rank order technique. The observers were presented small sets of stimuli (5 images at a time) and had to rank the images according to the quality of these. The result of such experiments is noisier than in a paired-comparison technique, however it significantly reduces sample handling. To construct the quality preference function value of each of the attributes of the highest rated image was identified for each scene and observer, and the fraction of times each position had been preferred was computed. To convert each fraction to a probability density, it was divided by the width of an attribute interval including the sample position. The resulting preference distribution function is presented in Fig. 4.

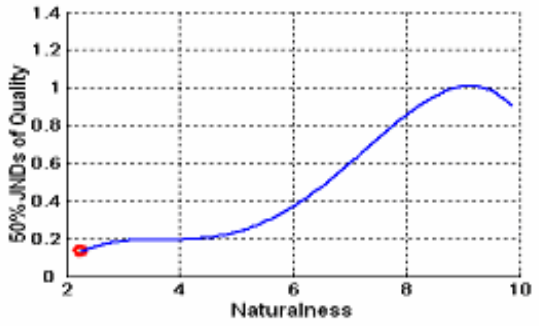


Fig. 4. Averaged preference distribution function of naturalness

A conclusion conforming to the preliminary assumption imposed upon the naturalness constraint can be drawn based on Fig. 4: with the increase of naturalness the quality of the images increases. However, at a certain point naturalness starts to decrease. Thus, high degree of naturalness is a necessary, but not a sufficient

condition for quality perception. A usefulness condition has to be satisfied as well, which in turn leads to a discriminability principle. Meaning that, even if the image gives an impression of being unnatural, it might be perceived as being of high quality, due to the fact that the information in the image is easily discriminable. Therefore, naturalness has a strong connection with the statistical characteristics of spectral images and image quality perception on the whole, being a necessary, but not a determinative factor.

4 Conclusions

In this paper, a preferential spectral image quality model has been presented. The model sets a relationship between the statistical characteristics of spectral images, overall quality, and perceived naturalness. The model described in this paper is an extension of a previously published spectral color appearance model. The original study has been extended to account for the naturalness constraint, i.e. the degree of correspondence between the image reproduced and the observers' perception of the reality.

Several conclusions can be drawn upon this study. One of the important inferences drawn from this work is that a strong connection between the statistical parameters of spectral images and the naturalness perception does exist. Particularly, not only there is a strong correlation between the colorfulness parameter, but it can also be said that there is a tendency for memory colors to be slightly more saturated compared with actual object colors, meaning that observers perceive slightly more colorful than original images to be more natural [13]. Moreover, considering the fact that observers generally discern more colorful images as being of higher quality, it can be stated that memory color reproduction influences the preferred color reproduction of the objects [4].

The connection between the vividness parameters and naturalness is twofold. Part of the images used in the experiment exhibit a sharper drop in the naturalness judgments than the rest, in fact, in the second case the decrease in naturalness is such that the naturalness remains approximately close to the maximal value. A similar phenomenon has been found in the spectral color appearance model [3] concerning image quality judgments. Both of the phenomena can be attributed to a fact that images *inlab1* and *inlab5,* that exhibit an insignificant drop in both of the characteristics compared with the rest of the images, contain objects that attract the most of the observers' attention, in comparison with the objects situated at the background. Moreover, these objects lie in the red area of the spectrum, which assumes that observers are not susceptible to minor variations in these areas due to the properties of the human visual system. Thus, the drop in quality and in naturalness is less definitive.

The last conclusion is the connection between the naturalness constraint and the overall perceived image quality. Although with the increase in naturalness the quality of the images increases, naturalness of the image is a necessary, but not a sufficient condition for the high quality judgments. A usefulness condition has to be satisfied as well. Thus the peak of the quality judgments does not lie at the highest naturalness value, meaning that observers knowing that the image is unnatural would still

perceive the image as being of high quality. For the purpose of modeling the naturalness influence upon image quality a preference distribution function, describing the impact of the naturalness on the quality judgments in terms of JNDs has been constructed. The function can be used for spectral image quality prediction in terms of image naturalness.

In general, both the preferential spectral image quality and the spectral color appearance models can be attributed to a class of preferential image quality models and can serve as an efficient tool of image quality characterization and prediction.

References

1. Hardeberg, J. and Gerhardt, J.: Characterization of an Eight Colorant Inkjet System for Spectral Color Reproduction, in Procs. Second European Conf/ on Colour Graphics, Imaging and Vision, Aachen, Germany (2004) 263-267.
2. Rosen, M., Hattenberger, E. and Ohta, N.: Spectral Redundancy in a 6-ink Inkjet Printer, in Procs. of The Dig. Phot. Conference, Rochester, NY, USA (2003) 236-243.
3. Kalenova, D., Botchko, V., Jaaskelainen, T. and Parkkinen, J.: Spectral Color Appearance Modeling, in Proc. Dig. Phot. Conference, Rochester, NY, USA (2003) 381-385.
4. Fedorovskaya, E.A., de Ridder, H. and Blommaert, F.J.J.: Chroma Variations and Perceived Quality of Colour Images of Natural Scenes, J. Color res. and appl. 22 (1997) 96-110.
5. Buhr, J.D. and Franchino, H. D.: Color Image Reproduction of Scenes with Preferential Tone Mapping, U.S. Patent #5 447 (1995) 811.
6. de Ridder, H.: Saturation and Lightness Variation in Color Images of Natural Scenes, J. Imaging Sci. and Techn. 6(40) (1996) 487-493.
7. Janssen, R.: Computational Image Quality, (2001) 20-35.
8. de Ridder, H.: Naturalness and Image Quality: Saturation and Lightness Variation in Color Images of Natural Scenes, J. Imaging Sci. and Techn. 40 (1996) 487-498.
9. Botchko, V., Kälviäinen, H. and Parkkinen, J.: Highlight Reproduction Using Multispectral Texture Statistics, Proc. Third Int. Conf. on Multispectral Color Science, Joensuu, Finland (2001) 61-65.
10. Parraga, A., Brelstaff, G. and Troscianko, T.: Color and Luminance Information in Natural Scenes, J. of Opt. Soc. of America A 15 (1998) 3-5.
11. ISO/DIS 20462-1, Psychophysical Experimental Method to Estimate Image Quality – Part 1: Overview of Psychophysical Elements, International Organization for Standardization (2003).
12. ISO 3664, Graphic Technology and Photography: Viewing conditions, International Organization for Standardization (2000).
13. Newhall, S. M., Burnham, R. W. and Clark, J. R.: Comparison of Successive with Simultaneous Color Matching, J. of Opt. Soc. of America 47 (1957) 43-56.
14. Siple, P., and Springer, R. M.: Memory and Preference for the Color of Objects, Perception and Psychophysics 33 (1983) 363-370.`

Publication VII

KALENOVA, D., DOCHEV, D., BOCHKO, V., TOIVANEN, P., KAARNA, A.,
A Novel Technique of Spectral Image Quality Assessment Based on Structural Similarity Measure

A Novel Technique of Spectral Image Quality Assessment Based on Structural Similarity Measure

Diana Kalenova, Dobromir Dochev, Vladimir Bochko, Pekka Toivanen, Arto Kaarna; Department of Information Technology, Lappeenranta University of Technology; Lappeenranta, Finland; Digital Media Institute, Tampere University of Technology; Seinäjoki, Finland*

Abstract

In this work, a novel technique of objective spectral image quality evaluation is presented. The method is based on a Structural Similarity technique. The traditional approach, which deals primarily with gray-scale images, is extended to incorporate spectral data. The novel method has previously been tested against the conventional two-dimensional technique and proven to be more effective. The performance of the three-dimensional Structural Similarity Index presented in this paper is tested along with the previously proposed kernel similarity metrics and a subjective technique - Perceptual Image Distortion Map. The tests show that the proposed three-dimensional Structural Similarity Index performance is comparable to the rest of the measures in the task of spectral distortion evaluation.

Introduction

Digital imaging nowadays is undergoing dramatic changes. Appearance of capturing, recording and display systems that are capable of working with spectral data creates a whole set of problems that exist for conventional imaging, e.g. image quality assessment and adjustment [1,2]. By image quality, in this case, we mean the measure of the perceived difference from a reference image [3].

In this paper, a novel technique of spectral image quality estimation is proposed. The main objective for creation of such measures is primarily lossy compression applications. A quality measure should be established with a possibility of computing the distortion value dynamically as the information is discarded from the image. This kind of metric should also be able to account for the characteristics of the human visual system. Other areas of application include, among others, electronic museums, archiving and printing industry applications.

Several approaches to spectral image distortion measurement exist at the moment [4,5,6]. The choice of the method depends primarily on the end-user of the imaging chain. In case of applications that require high accuracy, the most appropriate method of image quality evaluation is subjective assessment. However, such methods require significant time and money consumption, which gives rise to the appearance of multiple objective measures. Most of these have emerged from gray-scale image metrics: mean-squared error, signal to noise ratio, percentage maximum absolute distortion etc. Nevertheless, none of these measures account for the characteristics of the human visual system [7]. One of the most popular solutions existing at the moment is the CIE recommended 1976 CIELAB and CIELUV color difference formulae [8]. However, these show significant discrepancies with the judgments obtained using the subjective technique. In an attempt to improve the perceptual uniformity of the measures several metrics have been developed [3]: CMC [9], BFD [10, 11], CIE94 [12] and CIEDE2000 [13]. A Blockwise Distortion Measure for Multispectral images (BDMM) has been suggested in [5]. The measure computes a quality estimate that corresponds to the human evaluation; however, it deals with the artifacts in the spatial direction and does not account for the specific spectral distortions.

The algorithm, described in this paper, is an extension of a Structural Similarity Index (SSIM) [7] that incorporates spectral data [14]. SSIM is based upon an assumption that human visual system is highly adapted to extracting structural information from the images. SSIM compares local patterns of pixel intensities, assuming that luminance and contrast are normalized [7]. As a result a gray-scale spectral distortion map is obtained, which shows the areas where the visible distortions are in the image, and how large the distortions are. The three-dimensional SSIM has already been tested against the two-dimensional conventional measure. The novel method has proven to be more efficient in the task of color and spectral image discrimination [14].

Color Similarity Measures

One of the most popular color similarity metrics so far has been the Euclidean distance [8] and measures based upon it. These have an advantage of simplicity in understanding and realization, however such metrics are not optimal. Euclidean distance calculates the difference between colors not taking into account the angle between color vectors, which produces a significant divergence for RGB image reproduction [15].

An alternative set of color similarity metrics was proposed in [15]. These consist of a set of kernel similarity measures that include polynomial, Gaussian radial basis (RBF), and sigmoid metrics. It was shown in [15] that the measures provide an excellent fit to the response of the human visual system in the task of image quality assessment.

Kernels, in general, can be assumed to be dot products of vectors in a certain feature space, meaning that if we have two vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in the input domain X, we can produce a mapping [16]:

$$\Phi : \mathrm{X} - \mathrm{F}$$

$$x \mapsto x := \Phi(x) \tag{1}$$

Polynomial kernel similarity measure can be presented as follows [12]:

$$S_{polynomial} = \left(x_i, x_j\right)^d \tag{2}$$

where d is a parameter of the sensitivity of the measure, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are input color vectors.

The Gaussian RBF kernel has the following form [16]:

$$S_{Gaussian} = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{3}$$

where $\sigma>0$, σ is the parameter of the sensitivity of the function.

And the sigmoid kernel based similarity can be presented as follows [16]:

$$S_{sigmoid} = \tanh(k * (x_i, x_j) + \vartheta) \quad (4)$$

where k and ϑ are variable sensitivity parameters.

In order to account for the characteristics of the human visual system the input data is multiplied by Spectral Luminous Efficiency Function for photopic vision [17] and illumination factor [18].

Structural Similarity Index

SSIM is based on an idea that the human visual system is highly adapted to extracting structural information from the images, which, in turn, can be defined as the attributes representing the structure of the objects in the scene, independent of the luminance and contrast [7]. SSIM is an objective measure of difference between a reference image (sometimes called original) and a modified image, and thus can be considered a quality metric of the second (processed) image.

Two-dimensional SSIM

Thus, given two spectral images, represented as vectors $\mathbf{x}_i$ and $\mathbf{x}_j$, as inputs, SSIM produces an output on a 0 to 1 scale, where 0 means that the images are "not similar at all" and 1 means "identical" [4]. The overall index is constituted of three parts: luminance, contrast and structure comparison, all three being relatively independent [7].

The overall measure is defined as follows [7]:

$$SSIM(x_i, x_j) = \left[l(x_i, x_j)\right]^{\alpha} \cdot \left[c(x_i, x_j)\right]^{\beta} \cdot \left[s(x_i, x_j)\right]^{\gamma} \quad (5)$$

where α, β, γ are non-negative parameters, used to adjust the importance of each of the components [6].

Luminance component $l(\mathbf{x}_i,\mathbf{x}_j)$ is estimated as [7]:

$$l(x_i, x_j) = \frac{2\mu_{x_i}\mu_{x_j} + C_1}{\mu_{x_i}^2 + \mu_{x_j}^2 + C_1} \quad (6)$$

where C_1 is a constant that is included to avoid instability when the sum of the squares of means is approximately zero and μ is the mean of the image [7].

Contrast component $c(\mathbf{x}_i,\mathbf{x}_j)$ is given as [7]:

$$c(x_i, x_j) = \frac{2\sigma_{x_i}\sigma_{x_j} + C_2}{\sigma_{x_i}^2 + \sigma_{x_j}^2 + C_2} \quad (7)$$

where C_2 is a constant and σ is the variance of the image [7].

The structure comparison component $s(\mathbf{x}_i,\mathbf{x}_j)$ is defined as follows [7]:

$$s(\mathbf{x_i}, \mathbf{x_j}) = \frac{\sigma_{x_i x_j} + C_3}{\sigma_{x_i}\sigma_{x_j} + C_3} \quad (8)$$

where C_3 is a small constant given to avoid instability and σ_{xixj} is the covariance of $\mathbf{x}_i$ and $\mathbf{x}_j$.

Constants C_1, C_2 and C_3 can be computed as [7]:

$$C_1=(K_1L)^2;\ C_2=(K_2L)^2;\ C_3=C_2/2 \quad (9)$$

where L is the dynamic range of pixel values and $K_1<<1$, $K_2<<1$ are two scalar constants.

SSIM can be applied in a pointwise manner, but it is better to use the Gaussian weighting function $\mathbf{w} = \{w_i | i = 1,2,\ldots,N\}$, normalized to unit sum ($\sum w_i = 1$), as the windowing approach. The local statistics are then computed using the weights $\boldsymbol{w}$ [7]**.** And the overall SSIM image quality measure is computed by averaging all of the local windows in the image. This is done due to a number of reasons. For one thing image statistics are on the most part highly spatially non-stationary, the same can be assumed of the distortions introduced into the image. Moreover, localized measures provide more information about the quality degradation [7].

Three-dimensional SSIM

In the case of the three-dimensional SSIM measure the weighting function should be different, thus it is computed as follows [19]:

$$h(x, y, z) = \sqrt{2\pi}\sigma A e^{-2\pi^2\sigma^2\left(x^2+y^2+z^2\right)} \quad (10)$$

Experiments

Experiments were performed on spectral images of natural scenes from [20]. Two images – inlab2 and inlab5 were selected. Each image has the following dimensions: 256x256 in the spatial dimension and 31 components in the spectral dimension. Images were captured by a CCD (charge coupled device) camera in a 400-700 nm wavelength range at 10 nm intervals.

First, both of the images were compressed using PCA (principal component analysis) down to two principal components. The color reproductions of original images and reconstructed after compression are given in Fig.1.

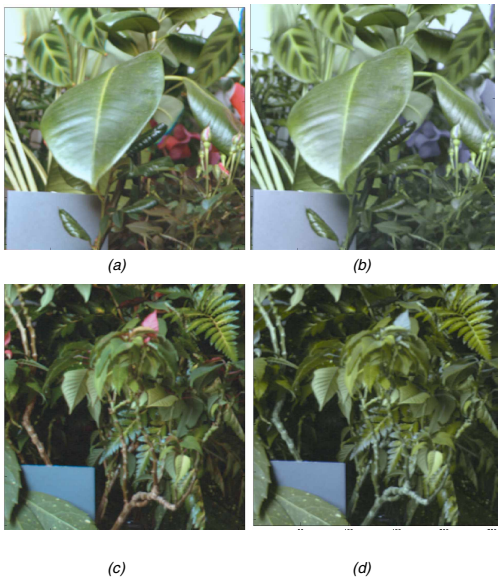


Figure 1*. Color reproduction of spectral images inlab2 (a) original, (b) reconstruction after compression (PCA 2); inlab5 (c) original, (d) reconstruction after compression (PCA 2)*

Both of the images were multiplied afterwards by Spectral Luminous Efficiency function for photopic vision [17] and illumination factor [18]. The areas of color difference are clearly visible in the images, and concentrate primarily in red and brown regions.

Then difference maps were computed using the three-dimensional SSIM (Eq. 5, 10) proposed in this paper, and three previously proposed kernel similarity metrics [15]: polynomial kernel (Eq. 2), Gaussian radial basis kernel (Eq. 3) and

sigmoidal kernel (Eq. 4). For the latter three, similarity between images was computed on a pixelwise basis. The resulting maps are shown in Fig. 2. The level of the intensity in the maps corresponds to the similarity scale: from black "not similar at all" to white "identical".

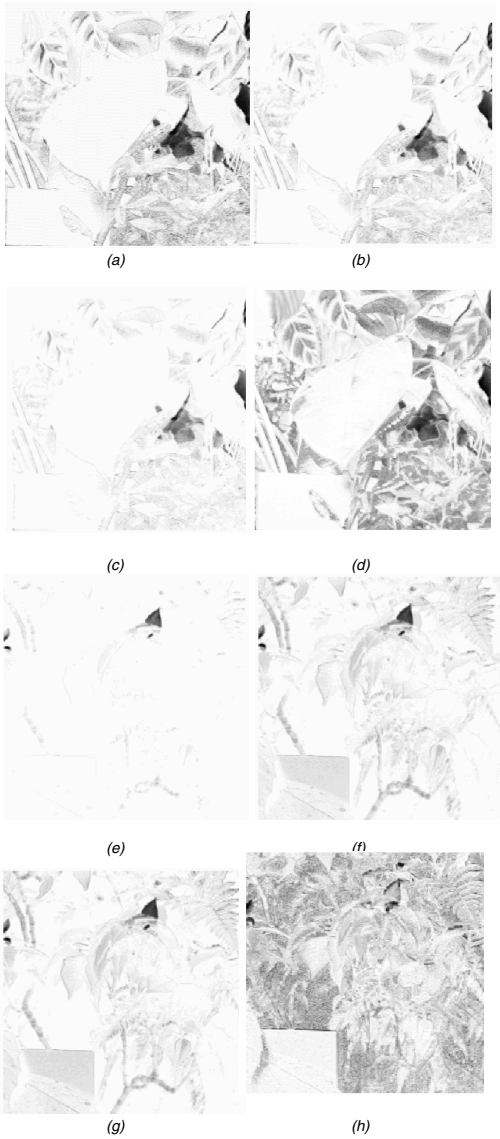


***Figure 2**. Difference maps. Inlab2 (a,b,c,d); inlab5 (e,f,g,h). (a,e) polynomial kernel; (b,f) Gaussian radial basis function; (c,g) sigmoidal kernel and (d,h) three-dimensional-SSIM*

Looking at Fig. 2, it can be stated that the difference maps produced using the kernel measures present a similar to a certain extent result, while the output of the three-dimensional SSIM gives a slightly different result specifically far more regions in the image are shown to be different.

Experimental Results

The accuracy of kernel similarity measures and the extended SSIM was tested using Perceptual Image Distortion Map (PIDM) [21]. PIDM is an empirical measure of the distribution of errors in the images [21]. PIDM can be obtained either on a pixelwise basis, or locally, with different marker sizes and shapes.

Five subjects were presented two sets of images, consisting of an original and a compressed image (Fig. 1). The users were asked to mark the regions that appeared different with a rectangular digital marker of size 4 by 4 pixels with different levels of gray-level intensity. Where black means "not similar at all" and white "identical". The subjects were instructed to mark the whole image area. Fig. 3 presents the mean of all subject maps [11,12].

***Figure 3**. Perceptual Image Distortion Map for images: (a) inlab2; (b) inlab5*

Fig. 3 clearly indicates that PIDM presents a practically excellent fit to the difference map calculated through the use of Gaussian RBF. Nevertheless, certain errors exist, which can be attributed to the fact that the marker size and shape caused several inaccuracies in stamping identical regions several times. Variance across the subjective judgments was equal to 0.0103, which is quite low and in turn means that subjects of the PIDM experiment were consistent in their estimations.

However, the results obtained using the [11] SSIM show significant difference with the results obtained using the PIDM technique.

Comparison of PIDM, SSIM and kernel metrics is given in Table 1, where each of the cells in first three columns present the mean deviation of the error maps, obtained through the use of polynomial, Gaussian RBF and sigmoidal kernels and the extended SSIM for each of the images, from the values of PIDM. Last column presents the value of the deviation of SSIM [7] error image from the PIDM.

Table 1. Comparison of SSIM, kernel metrics and PIDM

	Polynomial	Gaussian RBF	Sigmoidal	SSIM
Inlab 2	0.0499	0.0395	0.0551	0.1636
Inlab 5	0.0395	0.0291	0.0581	0.0986

PIDM presents a full map of empirical distortion data, which can be used in the task of evaluation of the accuracy of the metrics presented. Thus, looking at Table 1 it can be concluded that the most accurate evaluation of the human response in the quality estimation task is obtained through the

use of Gaussian RBF kernel, while the worst one with SSIM, although the deviation between these is not large.

Taking into consideration all of the above it can be stated that the kernel similarity measures and SSIM are quality evaluation techniques that accurately predict the response of a human visual system in a distortion evaluation task. The values of these vary in the range from 0 to 1, representing the difference values from "not similar at all" to "identical". From the point of view of the probability theory it can be stated that these measures presents a probability of the subject identifying a certain pixel as similar, which allows avoiding time and money consuming procedure of expert survey, and gives the possibility of computing the distortion values dynamically as the information is discarded from the image, as for example in a lossy compression task.

Conclusions

In this paper a novel technique of spectral image quality evaluation using Structural Similarity Measure was proposed. The algorithm, described in this paper, is an extension of a Structural Similarity Index (SSIM) [7] that incorporates spectral data [14]. SSIM is based upon an assumption that human visual system is highly adapted to extracting structural information from the images. The algorithm, given in this paper, computes a localized difference between the original and the distorted images. A gray-scale image distortion map is obtained as a result, where the intensity of each of the pixels corresponds to the value of the similarity between them, which, in turn, shows the areas where the visible distortions are in the image, and how large the distortions are. The overall index is constituted of three parts: luminance, contrast and structure comparison, all three being relatively independent [7]. As a windowing approach a three-dimensional Gaussian windowing function was used. The overall measure is obtained via averaging.

SSIM [7] was tested against several images of natural scenes [20] (with spectral distortions introduced into the images) along with several previously proposed kernel similarity measures [15] and a Perceptual Image Distortion Map [21], where PIDM is an empirical measure of the distribution of errors in the images [21]. The choice of the kernel similarity measures for comparison is not incidental - these have previously proven to be effective in the task of spectral distortion evaluation [15]. It was shown in [15] that these mimic closely the response of the human visual system in a task of quality evaluation and can be considered among the best approaches to evaluation of spectral distortions introduced into the spectral images. The kernel similarity measures chosen are polynomial, Gaussian radial basis and sigmoidal kernels.

The three-dimensional SSIM has already been tested against the two-dimensional conventional measure. The novel method has proven to be more efficient in the task of color and spectral image discrimination [14].

Comparing the results obtained in this work, it can be stated that SSIM performs slightly worse than the rest of the measures whilst Gaussian RBF gives a practically excellent fit to the human observer evaluation. Thus it can be concluded that performance of the SSIM is comparable to the rest of the measures in the task of spectral distortion evaluation.

References

[1] J. Hardeberg, and J. Gerhardt, Characterization of an Eight Colorant Inkjet System for Spectral Color Reproduction, Proc. CGIV, pg. 263. (2004).

[2] M. Rosen, E. Hattenberger, and N. Ohta, Spectral Redundancy in a 6-ink Inkjet Printer, Proc. PICS, pg. 236. (2003).

[3] Lindsay W. MacDonald, M. Ronier Luo eds, Color Image Science: Exploiting Digital Media, (Wiley, UK, 2002) pg. 357.

[4] M. Hild, On the Effectiveness of Color Similarity Measures in Background-Frame Differencing Application", Proc. CGIV, pg. 306. (2002).

[5] A. Kaarna, and J. Parkkinen, Blockwise Distortion Measure for Lossy Compression of Multispectral Images, Proc. EUSIPCO, pg. 2197. (2000).

[6] M. J. Ryan, and J. F. Arnold, A Suitable Distortion Measure for the Lossy Compression of Hyperspectral Data, Procs. IGARSS, pg. 2056. (1998).

[7] Z.Wang, A.C.Bovik, H.R.Sheikh and E.P.Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Trans. Image Processing, 13 (2004).

[8] CIE Colorimetry, 2nd edition, (CIE, Vienna, 1986).

[9] F. J. J. Clarke, R. McDonald and B. Rigg, "Modification to the JPC79 Colour-difference Formula", J. of the Soc. of Dyers and Colorists, 100, (1984).

[10] M. R. Luo and B. Riggs, "BFD (l:c) color-difference formula, Part I – development of the formula", J. of the Soc. of Dyers and Coloroists,103, (1987).

[11] M. R. Luo and B. Riggs, "BFD (l:c) color-difference formula, Part II – performance of the formula", J. of the Soc. of Dyers and Coloroists,103, (1987).

[12] Industrial Color-Difference Evaluation, (CIE, Vienna, 1995).

[13] G. Sharma, W. Wu and E. N. Dalal, "The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations", Color Research and Application, 30, 1, (2005).

[14] D. Dochev, V. Bochko, D. Kalenova, P. Toivanen and A. Kaarna, 3D Similarity Index For Evaluating Quality Of Lossy Compressed Spectral Images, Proc. CGIV, (2006), to be published.

[15] D. Kalenova, P. Toivanen and V. Botchko, Spectral Image Distortion Map, Proc. ICPR, pg. 668. (2004).

[16] Berhard Scholkopf, Alexander Smola, Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond, (MIT Press, Cambridge, MA, 2002) pg. 235.

[17] Spectral Luminous Efficiency Function for Photopic Vision, Publication CIE 86-1990.

[18] CIE Standard Illuminant D65 Relative Spectral Power Distribution, CIE 15.2-1986.

[19] R. Gonzalez and R. Woods, Digital Image Processing (Addison-Wesley, New York, NY,1992) pg. 194.

[20] A. Parraga, G. Brelstaff and T. Troscianko, Color and Luminance Information in Natural Scenes, J. of Optical Soc. of America A, 15, 3 (1998).

[21] X. Zhang, E. Setiawan and B. A. Wandell, Image Distortion Maps, Proc. CIC, pg. 120. (1997).

Author Biography

Diana Kalenova received her MSc in Information Technology from Lappeenranta University of Technology (2003,) and is now working on her PhD at the same University. Her work has focused on spectral image analysis, color appearance models, statistical characteristics of images and image quality issues. She is a member of Pattern Recognition Society of Finland and IAPR.

.

ACTA UNIVERSITATIS LAPPEENRANTAENSIS

318. UOTILA, TUOMO. The use of future-oriented knowledge in regional innovation processes: research on knowledge generation, transfer and conversion. 2008. Diss.

319. LAPPALAINEN, TOMMI. Validation of plant dynamic model by online and laboratory measurements – a tool to predict online COD loads out of production of mechanical printing papers. 2008. Diss.

320. KOSONEN, ANTTI. Power line communication in motor cables of variable-speed electric drives – analysis and implementation. 2008. Diss.

321. HANNUKAINEN, PETRI. Non-linear journal bearing model for analysis of superharmonic vibrations of rotor systems. 2008. Diss.

322. SAASTAMOINEN, KALLE. Many valued algebraic structures as measures of comparison. 2008. Diss.

323. PEUHU, LEENA. Liiketoimintastrategisten vaatimusten syntyminen ja niiden toteutumisen arviointi keskisuurten yritysten toiminnanohjausjärjestelmähankkeissa: Tapaustutkimus kolmesta teollisuusyrityksestä ja aineistolähtöinen teoria. 2008. Diss.

324. BANZUZI, KUKKA. Trigger and data link system for CMS resistive plate chambers at the LHC accelerator. 2008. Diss.

325. HIETANEN, HERKKO. The pursuit of efficient copyright licensing – How some rights reserved attempts to solve the problems of all rights reserved. 2008. Diss.

326. SINTONEN, SANNA. Older consumers adopting information and communication technology: Evaluating opportunities for health care applications. 2008. Diss.

327. KUPARINEN, TONI. Reconstruction and analysis of surface variation using photometric stereo. 2008. Diss.

328. SEPPÄNEN, RISTO. Trust in inter-organizational relationships. 2008. Diss.

329. VISKARI, KIRSI. Drivers and barriers of collaboration in the value chain of paperboard-packed consumer goods. 2008. Diss.

330. KOLEHMAINEN, EERO. Process intensification: From optimised flow patterns to microprocess technology. 2008. Diss.

331. KUOSA, MARKKU. Modeling reaction kinetics and mass transfer in ozonation in water solutions. 2008. Diss.

332. KYRKI, ANNA. Offshore sourcing in software development: Case studies of Finnish-Russian cooperation. 2008. Diss.

333. JAFARI, AREZOU. CFD simulation of complex phenomena containing suspensions and flow through porous media. 2008. Diss.

334. KOIVUNIEMI, JOUNI. Managing the front end of innovation in a networked company environment – Combining strategy, processes and systems of innovation. 2008. Diss.

335. KOSONEN, MIIA. Knowledge sharing in virtual communities. 2008. Diss.

336. NIEMI, PETRI. Improving the effectiveness of supply chain development work – an expert role perspective. 2008. Diss.

337. LEPISTÖ-JOHANSSON, PIIA. Making sense of women managers´ identities through the constructions of managerial career and gender. 2009. Diss.

338. HYRKÄS, ELINA. Osaamisen johtaminen Suomen kunnissa. 2009. Diss.

339. LAIHANEN, ANNA-LEENA. Ajopuusta asiantuntijaksi – luottamushenkilöarvioinnin merkitys kunnan johtamisessa ja päätöksenteossa. 2009. Diss.

340. KUKKURAINEN, PAAVO. Fuzzy subgroups, algebraic and topological points of view and complex analysis. 2009. Diss.

341. SÄRKIMÄKI, VILLE. Radio frequency measurement method for detecting bearing currents in induction motors. 2009. Diss.

342. SARANEN, JUHA. Enhancing the efficiency of freight transport by using simulation. 2009. Diss.

343. SALEEM, KASHIF. Essays on pricing of risk and international linkage of Russian stock market. 2009. Diss.

344. HUANG, JIEHUA. Managerial careers in the IT industry: Women in China and in Finland. 2009. Diss.

345. LAMPELA, HANNELE. Inter-organizational learning within and by innovation networks. 2009. Diss.

346. LUORANEN, MIKA. Methods for assessing the sustainability of integrated municipal waste management and energy supply systems. 2009. Diss.

347. KORKEALAAKSO, PASI. Real-time simulation of mobile and industrial machines using the multibody simulation approach. 2009. Diss.

348. UKKO, JUHANI. Managing through measurement: A framework for successful operative level performance measurement. 2009. Diss.

349. JUUTILAINEN, MATTI. Towards open access networks – prototyping with the Lappeenranta model. 2009. Diss.

350. LINTUKANGAS, KATRINA. Supplier relationship management capability in the firm´s global integration. 2009. Diss.

351. TAMPER, JUHA. Water circulations for effective bleaching of high-brightness mechanical pulps. 2009. Diss.

352. JAATINEN, AHTI. Performance improvement of centrifugal compressor stage with pinched geometry or vaned diffuser. 2009. Diss.

353. KOHONEN, JARNO. Advanced chemometric methods: applicability on industrial data. 2009. Diss.

354. DZHANKHOTOV, VALENTIN. Hybrid LC filter for power electronic drivers: theory and implementation. 2009. Diss.

355. ANI, ELISABETA-CRISTINA. Minimization of the experimental workload for the prediction of pollutants propagation in rivers. Mathematical modelling and knowledge re-use. 2009. Diss.

356. RÖYTTÄ, PEKKA. Study of a vapor-compression air-conditioning system for jetliners. 2009. Diss.

357. KÄRKI, TIMO. Factors affecting the usability of aspen (Populus tremula) wood dried at different temperature levels. 2009. Diss.

358. ALKKIOMÄKI, OLLI. Sensor fusion of proprioception, force and vision in estimation and robot control. 2009. Diss.

359. MATIKAINEN, MARKO. Development of beam and plate finite elements based on the absolute nodal coordinate formulation. 2009. Diss.