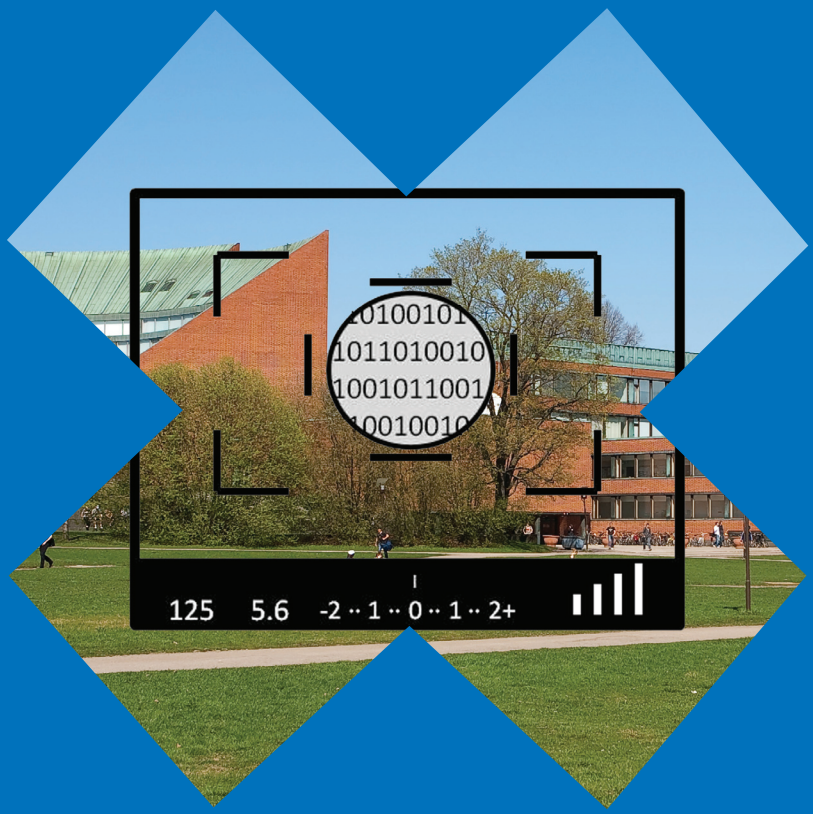
Department of Media Technology

# Reduced-reference methods for measuring quality attributes of natural images in imaging systems

Mikko Nuutinen




Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
**DOCTORAL DISSERTATIONS** 157/2012

# Reduced-reference methods for measuring quality attributes of natural images in imaging systems

**Mikko Nuutinen**

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall TU2 of the school on 7 December 2012 at 12.

**Aalto University**
**School of Science**
**Department of Media Technology**
**Visual Media Research Group**

**Supervising professor**
Pirkko Oittinen

**Preliminary examiners**
Professor Jon Y. Hardeberg, Gjøvik University College, Norway
Associate professor Zhou Wang, University of Waterloo, Canada

**Opponent**
Professor Janne Heikkilä, University of Oulu, Finland

Aalto University publication series
**DOCTORAL DISSERTATIONS** 157/2012



ISBN 978-952-60-4880-2 (printed)
ISBN 978-952-60-4881-9 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-4881-9

Unigrafia Oy
Helsinki 2012

Finland

441 697
Printed matter

Aalto University

# Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

**Author**
Mikko Nuutinen

**Name of the doctoral dissertation**
Reduced-reference methods for measuring quality attributes of natural images in imaging systems

**Publisher** School of Science

**Unit** Department of Media Technology

**Series** Aalto University publication series DOCTORAL DISSERTATIONS 157/2012

**Field of research** Media Technology

**Manuscript submitted** 17 April 2012 **Date of the defence** 7 December 2012

**Permission to publish granted (date)** 31 October 2012 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

**Abstract**

The main goal of the research field of image quality is to create a computational model capable of predicting the subjective visual quality of natural images and video. The model can be an alternative for expensive quality evaluations by human assessors. This dissertation aligns with this research field in the case of imaging systems, such as cameras, displays and printers.

The traditional approach to measuring the quality of imaging systems is based on test targets. These targets primarily facilitate description of the performance of a system in terms of how it reproduces and distorts simple test signals rather than measure the visual quality of natural images captured or shown by an imaging system.

This dissertation primarily contributes novel methods and algorithms for measuring the image quality attributes of natural images captured by cameras or printed by a printer. Both methods utilize reference image data in the reduced-reference mode. The method and algorithms developed for printers transform the printed natural test images into the form of the reference image by using a high-quality reference camera and multiple exposures. The methods and algorithms developed for camera images use a reference camera to capture scene information. The scene information is used to help measure the attributes of natural images. The main problem which needed to be solved concerns localization of areas in images from which different attributes can be measured. The challenge arises from a multidimensional distortion space in capture and display. The solution relies on low-level computational understanding of images.

The methods were evaluated with subjective data. Compared with the state-of-the-art computational or test target metrics, these methods were highly effective at predicting the quality attributes of natural images captured by different cameras or printed on different papers. According to the results, the proposed methods can replace test target methods and even small-scale subjective tests in some situations.

**Keywords** image quality, natural image, test-target, camera, printer, sharpness, noise, graininess, color, reduced-reference

**ISBN (printed)** 978-952-60-4880-2 **ISBN (pdf)** 978-952-60-4881-9

**ISSN-L** 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

**Location of publisher** Espoo **Location of printing** Helsinki **Year** 2012

**Pages** 184 **urn** http://urn.fi/URN:ISBN:978-952-60-4881-9

A! Aalto-yliopisto

**Tiivistelmä**

**Aalto-yliopisto, PL 11000, 00076 Aalto** www.aalto.fi

**Tekijä**
Mikko Nuutinen

**Väitöskirjan nimi**
Vähennetyn referenssin menetelmiä kuvasysteemien luonnollisen kuvan laatuattribuuttien mittaamiseen

**Julkaisija** Perustieteiden korkeakoulu

**Yksikkö** Mediatekniikan laitos

**Sarja** Aalto University publication series DOCTORAL DISSERTATIONS 157/2012

**Tutkimusala** Mediatekniikka

**Käsikirjoituksen pvm** 17.04.2012 **Väitöspäivä** 07.12.2012

**Julkaisuluvan myöntämispäivä** 31.10.2012 **Kieli** Englanti

☐ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

**Tiivistelmä**

Kuvanlaatututkimuksen päätavoite on aikaansaada laskennallinen malli, joka ennustaa luonnollisen kuvan tai videon visuaalista laatua. Laskennallinen malli korvaisi vaikeasti järjestettävät, aikaa vievät ja kalliit subjektiiviset kokeet. Tämä väitöstutkimus sijoittuu kuvanlaadun tutkimusalueelle ja keskittyy erityisesti kuvannussysteemien, kuten kameroiden ja tulostimien kuvanlaatumittauksen sovelluksiin.

Kuvannussysteemien laatua on perinteisesti määritetty testikenttämittauksilla. Testikenttämittaus karakterisoi kuinka kuvannussysteemi toistaa tai vääristää yksinkertaisia testisignaaleja. Testikenttämittaukset eivät kerro miten kuvannussysteemillä kuvattu, esitetty tai tulostettu luonnollinen kuva havaitaan visuaalisen laadun kriteereillä arvioituna.

Väitöstutkimuksessa kehitettiin algoritmeja ja menetelmiä sekä kameroille että tulostetulla kuvalle, jotka mittaa kuvanlaatuattribuutteja nimenomaan suoraan luonnollisista kuvista ja joilla voidaan korvata testikenttämittaukset. Kehitetyt menetelmät perustuu vähennetyn referenssin periaatteeseen. Tulostetun kuvan sovelluksissa tulostettu luonnollinen kuva digitoitiin käyttämällä korkealaatuista kameraa sekä useaa valotusta. Digitoinnin jälkeen alkuperäisen kuvan informaation avulla mitattiin tulostetuille luonnollisille kuville laatuattribuutteja. Kameran kuvan sovellukset käytti referenssikameraa, joka kaappasi näkymään liittyvää informaatiota. Näkymäinformaation avulla eri kameroilla kuvatuista luonnollisista kuvista mitattiin laatuattribuutteja.

Kehitettyjen menetelmien suorituskyky validioitiin subjektiivisella aineistolla. Verrattuna testikenttämittauksiin, kehitettyjen menetelmien suorituskyky oli korkea ennustettaessa kameralla kuvattujen tai tulostimella tulostettujen luonnollisten kuvien laatua. Tulosten perusteella ehdotetut menetelmät voivat korvata testikenttämenetelmien lisäksi myös pienen skaalan subjektiiviset testiasetelmat. Väitöstutkimus rajattiin käsittelemään ainostaan matalan tason laatuattribuutteja, kuten terävyys, kohina ja värintoisto.

**Avainsanat** kuvanlaatu, luonnollinen kuva, testikenttä, kamera, tulostin, terävyys, kohina, rakeisuus, väri, vähennetyn referenssin menetelmä

**ISBN (painettu)** 978-952-60-4880-2 **ISBN (pdf)** 978-952-60-4881-9

**ISSN-L** 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

**Julkaisupaikka** Espoo **Painopaikka** Helsinki **Vuosi** 2012

**Sivumäärä** 184 **urn** http://urn.fi/URN:ISBN:978-952-60-4881-9

# Preface

I wish to sincerely thank Professor Pirkko Oittinen for encouraging and supervising me throughout my postgraduate studies. I am also grateful to Emeritus Professor Hannu Saarelma for his feedback regarding this dissertation.

Next, I thank Timo Säämänen, Olli Orenius, Raisa Halonen and Tuomas Leisti for coauthoring the publications of this dissertation. I also want to thank all workers of the visual media research group.

Moreover, I am grateful to the pre-examiners, Professor Jon Y. Hardeberg and Associate Professor Zhou Wang, for their valuable comments. I would also like to thank the anonymous reviewers of the articles sent for peer review under the dissertation project. Further, I would like to thank Professor Janne Heikkilä for kindly agreeing to take the role of the opponent.

I am thankful for the funding of my work provided by the Academy of Finland funded VisiQ and VICOMM projects, TEKES funded DigiQ project and Nokia funded Camera I-V projects. Moreover, I am thankful for the travel grants from the Graphical Industry Research Foundation (GTTS).

Espoo, November 5, 2012

Mikko Nuutinen

# Contents

# List of Publications and Author's role

This dissertation consists of an overview and of the following publications, which are referred to in the text by their Roman numerals.

I. Nuutinen, M., Halonen, R., Leisti, T., Oittinen, P., "Reduced-reference Quality Metrics for Measuring the Image Quality of Digitally Printed Natural Image," Proc. of SPIE/IS&T Electronic Imaging 2010, Image Quality and System Performance VII. San Jose, California, USA, 17.-21.1.2010. [7529-18]. 12 p.

II. Nuutinen, M., Oittinen, P., "Recovering Spectral Data from Digital Prints with an RGB Camera Using Multi-exposure Method," Proc. of. IS&T's Fifth European Conference on Color in Graphics, Imaging and Vision, Joensuu, Finland, 14.-17.6.2010, pp. 120-125.

III. Nuutinen, M., Orenius, O., Säämänen, T., Oittinen, P., "A Reduced-reference Method for Characterizing Color Noise in Natural Images Captured by Digital Cameras," Proc. Of. IS&T/SID 18th Color and Imaging Conference, San Antonio, Texas, USA 8.-12.11.2010, pp. 80-85.

IV. Nuutinen, M., Orenius, O., Säämänen, T., Oittinen, P., "Reference Image Method for Measuring Quality of Photographs Produced by Digital Cameras," Proc. of SPIE/IS&T Electronic Imaging 2011, Image Quality and System Performance VIII. San Jose, California, USA, 23.-27.1.2011. [7876-22]. 14 p.

V. Nuutinen, M., Orenius, O., Säämänen, T., Oittinen, P., "Method for Measuring Sharpness in Natural Images Captured by Digital Cameras," EURASIP Journal on Image and Video Processing, Volume 2012:8.

VI. Nuutinen, M., Oittinen, P., "Measurement of Color Differences of Digital Cameras from Natural Images," Proc. of 7th International Symposium on Image and Signal Processing and Analysis, Dubrovnik, Croatia, 4.-6.9.2011. pp. 224-229.

The author of the dissertation was the main author of Publications (I-VI).

**Publications I and II:** The author developed and implemented the sample digitization process and the attribute measuring algorithms and methods for printed images. The printed samples for the printers were developed and printed by Raisa Halonen and others in the role of MSc students (Helsinki University of Technology).

**Publications III-VI:** The author developed and implemented the reference camera system and the attribute measuring algorithms and methods for digital cameras. The author was the main developer of the image clusters for the camera measurements.

**Publications I, III, IV and V:** The subjective tests for both the printed samples and photographed samples were developed and conducted in the Institute of Behavioural Science at the University of Helsinki (Evolving Media and Technology Research group). The author analyzed and applied the subjective raw data for the dissertation.

# Abbreviations

| | |
|---|---|
| ADC | Analog to Digital Conversion |
| APQC | Amplitude/Phase Quantization Code |
| CIELAB | CIE (International Commission on Illumination), Lightness, A (redness-greenness), B (yellowness-blueness) |
| CSF | Contrast Sensitivity Function |
| DAC | Digital to Analog Conversion |
| DCT | Discrete Cosine Transform |
| EPG | Electrophotography |
| FR | Full Reference |
| GGD | Generalized Gaussian Density |
| GIQM | Gradient Information-based Quality |
| HE | Horizontal Effect |
| HVS | Human Visual System |
| HR | Harris Response |
| ICC | International Color Consortium |
| IFC | Information Fidelity Criterion |
| IJ | Ink-Jet |
| JNB | Just-Noticeable Blur |
| JND | Just-Noticeable Difference |
| KLD | Kullback-Leibler Distance |
| LCC | Pearson Linear Correlation Coefficient |
| MOS | Mean Opinion Score |
| MS | Multi-Scale |
| MTF50 | Modulation Transfer Function 50% |
| NR | No Reference |
| NSS | Natural Scene Statistics |
| OR | Outlier Ratio |
| PSF | Point Spread Function |
| PQC | Phase Quantization Code |
| RGB | Red-Green-Blue |
| RMSE | Root Mean Square Error |

| | |
|---|---|
| ROCC | Spearman Rank-Order Correlation Coefficient |
| RR | Reduced Reference |
| SFR | Spatial Frequency Response |
| SIFT | Scale-Invariant Feature Transform |
| SNR | Signal-Noise Ratio |
| SSIM | Structural Similarity |
| SURF | Speeded-Up Robust Features |
| SVM | Support Vector Machine |
| SVR | Support Vector Regression |
| VIF | Visual Information Fidelity |

# Glossary

| | |
|---|---|
| Adaptive local feature | An image is divided into blocks with a predetermined initial order, and new positions are sought for the blocks by maximizing a feature-specific function. The feature vector includes the positions of the selected blocks. |
| Candidate block | A coarse block selected from the predetermined blocks. The blocks for actual measurements are selected from the group of candidate blocks. |
| Clarity | The easiness of distinguishing the content of the image. |
| Contrast | The magnitude of visual differences, global and local. |
| Computational method | An algorithmic method used to compute the quality or a feature from a digital image. |
| Correspondence block | A local block area in an image that equals the block area in the reference image. |
| Full-reference metric | A metric that requires a pixel-wise reference to compute the quality. |
| Global feature | A single feature value that is calculated from the all pixel values of a reference image. The feature vector includes the parameter set of image statistics or the no-reference metric value. |
| Graininess/noise | High- to mid-frequency unwanted random- or fixed-pattern intensity distortion on image. |
| High-level attribute | An abstract subjective criterion that cannot be directly related to an objective property of an image (e.g., naturalness). |
| Image content | The subject matter, objects and settings of an image. |

| | |
|---|---|
| Local feature | An image is divided into blocks, and the quality attribute-specific sensitivity values are calculated from the blocks. The feature vector includes the block-specific sensitivity values. |
| Low-level attribute | A concrete attribute that is associated with an objective property of an image, such as graininess. |
| Naturalness | The level of correspondence between a picture and the anticipated view or the real view in the memory of an assessor. |
| Natural image | A picture taken from a natural scene or from man-made objects. |
| No-reference metric | A metric that operates without a reference image or any information from one. |
| Overall quality | The overall quality rating of (or the overall value of) an image that depends on all of quality-related attributes of that image. |
| Reduced-reference metric | A metric that extracts features from the reference image and employs them as side information in the assessment. |
| Reference image | The original image that is considered to be of perfect quality or free of distortions (also called the anchor image in subjective tests). |
| Sharpness | The level of clarity of details and edges. |
| Test image | A digital or printed natural image used in subjective evaluations or objective measurements. |
| Test target | A target that contains one or several technical test fields. |

# 1. Introduction

## 1.1 Background

During the last few years, the number of consumer cameras has significantly increased. The maturity of digital camera technology and the integration of digital camera modules in mobile phones are the two main reasons for this growth. The number of captured photographs has increased at an even faster pace than the number of cameras, primarily because digital sensors have replaced expensive and tedious films as the image capture plane. Although more devices and methods of digitally sharing images have become available, printing these images continues to be a viable option.

Objective quality measurements used to characterize imaging systems are the focus of this dissertation. The term "imaging system" denotes a capture device, such as a digital camera, and a display device, such as a printer or a screen. Image communication systems (i.e., streaming over a network) are beyond of the scope of the associated problem areas. Figure 1 shows a generic imaging pipe where a camera captures a view and processes the information to represent the view. Afterwards, the captured information is transferred to the hard drive of a computer (or a network) and printed on paper or shown on a screen.

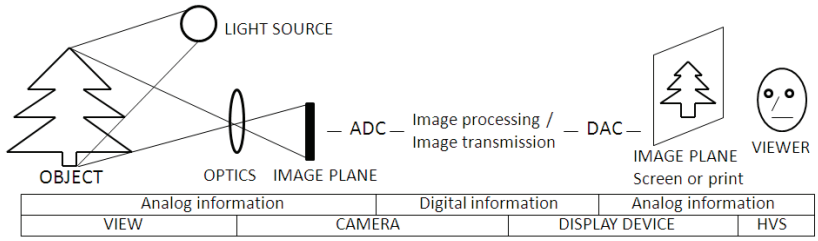



**Figure 1.** The information form in the imaging pipe is analog (view), digital (image processing and transmission) and analog (display).

The fundamental challenge to developing image quality metrics and methods for imaging systems arises from the analog-to-digital (ADC) and digital-to-analog (DAC) conversions that take place in the imaging pipe. The scene information captured by a camera is originally an analog signal

of an optical image. Before the captured information is digitally stored in a camera, an ADC is performed. Before captured information is printed on a paper or shown on a screen, DAC is required. That is, the form of the original or reference signal (optical or digital image) is different from that of the test signal (digital or optical image).

The research and development work on imaging systems in industrial settings is continuous and cyclical and requires robust, feasible analysis and validation methods. The development work on consumer and low-end cameras is particularly challenging because of the cost and size limitations of the devices. Cheap optics and small-sized sensors with small pixels result in noise and unsharp raw images. Advanced signal processing is a requirement for adequate image quality. In the future, new imaging applications will bring new challenges and requirements. For example, stereoscopic camera modules and different types of environment-sensitive camera network systems will become common. In addition, consumer cameras will implement new computational applications. Features such as high dynamic range (HDR) imaging, face recognition and panorama imaging are now basic features in many advanced consumer cameras. The development work on quality metrics should also be continuous and progressive. The new metrics and methods should be able to tackle the new requirements that new functionalities bring.

The three main quality-of-experience factors for imaging systems are usability, durability and image quality. Of these factors, image quality is probably the most important. Imaging systems are characterized using both subjective and objective methods. In this dissertation, the term “subjective method” denotes a test performed on test participants. The term “objective method” refers to an algorithm or a method based on the computational processes applied to the test images. The output of an objective method is a value related to the function of an imaging system or the quality of an image that is captured, shown or printed by an imaging system. In a subjective test, an observer rates test images based on the overall quality or quality attributes. Quality attributes include sharpness, graininess, naturalness, clarity, contrast and brightness. For example, in an image quality hierarchy, naturalness and clarity are high-level attributes, and sharpness and graininess are low-level attributes. The high-level attributes are more subjective, and personal preferences affect them more than the low-level attributes. The low-level attributes are more concrete and easier to evaluate. An image quality hierarchy describes the levels and relations among different quality attributes.

Subjective evaluations can function as the ground truth for image quality as long as the tests are planned well and executed carefully. However,

subjective measurements require a large number of assessors and are time-consuming in nature. In addition, subjective measurements cannot be used for applications that require real-time parametric control based on quality data.

The field of objective image quality research has tried to develop methods and models for predicting subjective quality and quality attributes as accurately as possible. Traditionally, test targets are employed in objective measurements of imaging systems. Test target data, however, do not correlate well with perceived image quality. The test target measurements primarily describe how imaging systems function rather than measure the perceived quality of images captured and processed by imaging systems.

The perceived quality of images relates to attributes such as naturalness and clarity. Both of these high-level quality attributes can only be assessed from natural images. Test target images cannot be used. In this dissertation, the term “natural image” refers to a picture taken from a natural scene or from man-made objects. The term “test target image” refers to a picture taken from known test fields, such as frequency-modulated bar patterns or tone- and color-modulated patches. Test targets allow us to determine how an imaging system distorts the known test signals when captured or rendered on display or in print.

## 1.2 The goal of the dissertation

The goal of this dissertation is to construct and evaluate methods and algorithms that measure quality attributes directly from natural images and thus replace test target images. If we reach this goal, the performance measurements of imaging systems can be streamlined because the tedious process of capturing test target images under strict laboratory conditions can be avoided. In addition, the same images can be used for subjective measurements. If both objective and subjective measurements could be performed from the same images, the relation between subjective and objective data and the constructs of the subjective data would be easier to establish. The ultimate goal is to replace subjective evaluations with objective computational models.

The study focuses primarily on the functionality of the metrics rather than specific imaging systems. With regard to the chronological order of the research, we first addressed the image quality attributes that utilized features from digital reference images. The application was printed images. We then shifted our focus to the determination of image quality attributes from the images captured by cameras. Specifically, we examined methods that utilize the information from the scenes to be acquired as reference

data. Because the development process follows this chronological order, the methods developed later for the camera measurements are more advanced, and the weight and novelty of the dissertation lies more in the camera measurements than in the printing measurements. However, the fundamental principles underlying the developed methods allow us to apply the methods with minor modifications to characterize any imaging system.

## 1.3 Dissertation contributions

The contributions of this dissertation are concerned with the methods, algorithms and metrics that allow us to use the reference approach to measure image quality attributes directly from the natural images captured by cameras or printed by printers. The proposed methods and metrics are verified using the data from subjective studies developed and accomplished by the research partner of University of Helsinki, as detailed in the List of Publication and Author's Roles on pages vi-vii. The six contributions of the dissertation are as follows:

The first contribution of the dissertation concerns the method used to digitize and transform printed test images into the format of a reference image. The original image can be used as a reference file for measurements of printed image quality. However, the form of the printed test image is analog, and it should be digitized before it can be compared with the reference image. Previous studies [17], [87] used a reflective scanner with a standard ICC profile to digitize printed color images. In this dissertation, a high-quality camera was characterized by colorimetric and spectral methods, which turned it into a color-accurate digitization device. Because the dynamic range of digital cameras (and reflective scanners) is the limiting factor of their imaging performance, we used multiple exposures to capture the sample images. Publications I and II presented the digitization methods.

The second contribution concerns the attribute metrics that utilize a reference image and compute the quality attributes for printed images. Reduced-reference (RR) type metrics were developed for the color contrast, sharpness and graininess attributes. The attributes were selected based on a subjective study. The state-of-the-art Full-reference (FR) and RR metrics compute the overall quality of an image instead of its attribute values. Although the previously proposed no-reference (NR) metrics compute the quality values of images distorted by a specific distortion, these metrics are also sensitive to other distortions. With the proposed method, we can compute attribute-specific values. In addition, as in the

printed images, the proposed metrics can be used for the images with a multi-dimensional distortion space. Publication I presented the proposed RR metrics for printed images.

The reference image is not available for camera quality measurements. The third contribution concerns a reference camera. Publication IV examined the usage of a reference camera. The principle is that a reference camera captures scene information in the form of a reference image. With the aid of the information, we can compute the quality attributes of natural images captured by cameras. The current practice computes the image quality attributes of imaging systems from test targets.

The fourth contribution relates to the methods employed to compute the sharpness, color noise and color difference of camera images. The contribution relates to the method and algorithms used to locate the optimal measurement positions from the captured scene. For the different attributes, the method uses different maximizing functions. The sharpness metric tries to find high energy areas, and the noise metric attempts to find low energy areas. The color metric searches for the most extensive color value set that can be captured from the scene. Publication III presented the noise metric, Publication V presented the sharpness metric and Publication VI presented the color difference metric.

The fifth contribution addresses the method and algorithms used to search the corresponding areas between the reference and test images. The method uses area descriptors to search the corresponding areas. We can use the proposed method for camera images because the principle of searching corresponding areas is local. The non-linear geometric differences between the images captured by different cameras (i.e., those without planar views and without constant shooting positions) create a set of problems. However, we can avoid these problems by using the proposed method. Publications V and VI described the method and algorithms in detail.

The sixth contribution is associated with the use of color information from natural images to compute image quality. We developed the RR type metrics to compute color contrast, color noise and color difference. The state-of-the-art algorithmic metrics do not utilize the color information of the test images. The proposed color contrast metric presented in Publication I is derived from the chromatic components of the CIELAB color space. The proposed color noise metric presented in Publication III utilized the components of the YCbCr color space to find the areas for computing chromatic noise in the test images. The proposed color difference metric presented in Publication VI expressed color in the

components of the CIELAB color space to compute the color value differences between the test images and the test scene.

## 1.4 Dissertation structure

This summary of the dissertation includes two parts based on Publications I-VI, which can be found in the appendixes.

The first part (Section 2) provides an overview of the previously proposed definitions of image quality, reviews the state-of-the-art test target and algorithmic methods and explains the different types of test images. In particular, Section 2 discusses the lack and the problems related to algorithmic methods, when applied to imaging systems.

The second part (Sections 3-5) covers the experimental part of the dissertation. Section 3 describes the developed methods, Section 4 shows and analyses the experimental results and Section 5 discusses the results and the contributions of the dissertation.

As for the abbreviations and definitions, the reader is referred to the respective sections at the beginning of the summary. Symbols are defined in the context of the equations in which they occur.

# 2. Survey of Image Quality Measurements

## 2.1 Definitions of image quality

Both Keelan and Engeldrum have used the term “excellence of image” as a definition of image quality. Keelan [50] stated, ”The quality of an image is defined to be an impression of its merit or excellence, as perceived by an observer neither associated with the act of photography, nor closely involved with the subject matter depicted”. Engeldrum’s [19] definition states, “Image quality is the integrated perception of the overall degree of excellence of an image”.

Janssen’s definition draws from the assumption that image quality is related to two perceptual attributes. Janssen [48] described “the quality of an image to be the degree to which the image is both useful and natural. The usefulness of an image to be the precision of the internal representation of the image, and the naturalness of an image to be the degree of correspondence between the internal representation of the image and knowledge of reality as stored in memory”.

This dissertation begins by pointing out that the image quality of an imaging system is determined by the quality attributes. In addition, we assumed that quality attributes can be classified into low- and high-level groups and that an image quality hierarchy can be formed. The high-level attributes are more subjective, and personal preferences affect the values of these attributes more than those of the low-level attributes. In addition, the meaning of image content is stronger with the high-level attributes than with the low-level attributes. High-level attributes cannot be directly related to an objective property of an image. The low-level attributes are concrete and general.

Image quality hierarchy describes the levels and links of different quality attributes in the construct of overall image quality. With the aid of an image quality hierarchy, we can also present an image measurement hierarchy. An image measurement hierarchy depicts the links and relations among different image quality metrics. The following sub-sections present

the subjective image quality attributes of different imaging systems and derive a general image quality hierarchy.

## 2.2 Subjective image quality attributes

The image quality of image processing and imaging systems has been studied and analyzed using attributes and viewpoints that have clear differences. The literature on image processing has often focused only on transmission and compression distortions, such as JPEG and JPEG2000 compression artifacts, white noise or packet loss distortions. The study of imaging systems appears to be a much more complex problem. For example, Keelan [50] classified the quality attributes of photographs (captured by a camera) into personal, aesthetic, preferential and artifactual groups. The low-level attributes of sharpness and noise are artifactual attributes. The low-level attributes of color balance, lightness and contrast are preferential attributes. Aspects such as lighting quality and image composition are aesthetic attributes. Features such as how an image preserves a cherished memory and conveys the essence of a subject are personal attributes. Keelan claimed that the classes describe how easy the attribute is to evaluate. Compared with the aesthetic and personal attributes, the artifactual and preferential attributes are straightforward to estimate.

Different imaging systems need to be characterized in terms of device-specific attributes and common attributes. Leisti et al. [59] and Pedersen et al. [88] have studied the quality attributes in the print context. Based on the results of a subjective interview test, Leisti et al. [59] claimed that the most important low-level attributes for prints are brightness of color, sharpness, graininess, brightness, color quality, gloss, contrast and lightness. The high-level attributes used to determine the meanings of the low-level attributes are realism, naturalness, clarity, depth and quality associations. Based on a literature survey, Pedersen et al. [88] identified a large set of attributes. The researchers compressed these attributes into six low-level attributes: color, lightness, contrast, sharpness, artifacts and physicality. The color attribute is related to hue, saturation and color rendition. The artifacts attribute includes noise, contouring and banding. The physicality attribute contains the physical parameters that affect image quality, such as paper properties and gloss.

Nyman et al. [83] and Radun et al. [92] studied the attributes of images captured by cameras. Nyman et al. [83] claimed that the important low-level attributes for the still image are sharpness, noise, lightness and color. The important attributes for video images were related to the same image

properties (sharpness, noise, lightness and color). In addition, the high-level attributes “good enough picture” and “distorted unnatural colors” were used for the video images. For the sound quality of the video images the attributes named as “good enough” and “noise” were used. Radun et al. [92] studied the image processing pipes of mobile camera phones. According to the results, the most important image quality dimensions were color shift, naturalness, darkness and sharpness. Radun et al. [92] claimed that the high-level attribute naturalness is a requirement for high-quality images, whereas quality can fail for other reasons in low-quality images. For example, a low-quality image can be dark and unsharp.

Murdoch et al. [77] studied the attributes of TV displays. They claimed that the important attributes for TV displays are brightness, contrast, color, sharpness and motion quality. Jumisko-Pyykkö et al. [49] studied the quality of mobile 3D video. According to the results, the quality of mobile 3D video experiences is related to classes such as visual depth, spatial and temporal properties, viewing experience, quality factors in general, audio quality, audiovisual quality and content.

## 2.3 Attribute relationships

The quality attributes interact with each other. If the application-specific attributes and links are known, the underlying causes explaining the subjective data are easier to find, and an objective overall quality model can be derived. For example, with the aid of the Bayesian networks derived by Eerola et al. [18], we can analyze the effects of quality attributes on the subjective overall quality of a printed image. The learnt networks connect the objective instrumental measurements of prints to the subjective opinion distribution of human observers. For example, according to the links of the network shown in Figure 2, the subjective high-level attribute naturalness affects the low-level attributes gloss, graininess and colorfulness.

I3A association [40] published an initial image measurement hierarchy for camera phones. In the hierarchy, the subjective low-level attributes are uniformity-ness, sharpness, hue-chroma (color) reproduction-ness and brightness (tone) reproduction-ness. The high-level attributes are genuineness, naturalness and usefulness. The image measurement hierarchy of I3A also shows different levels for the objective measurements. The lowest level of the objective measurements shows the technology variables of cameras. The image processing pipe and sensor design affect those parameters. For example, sharpening, denoising, demosaicing, automatic white balancing and auto-exposure are related to

image processing pipe design. Column noise, pixel design, spectral sensitivity, stack height and other issues are related to sensor design. The mid-level of the objective measurements shows the basic test target measurements or low-level technical measurements of digital cameras. The highest level shows the objective perceptual measurements that consider not only test target data but also the properties of the human visual system (HVS).

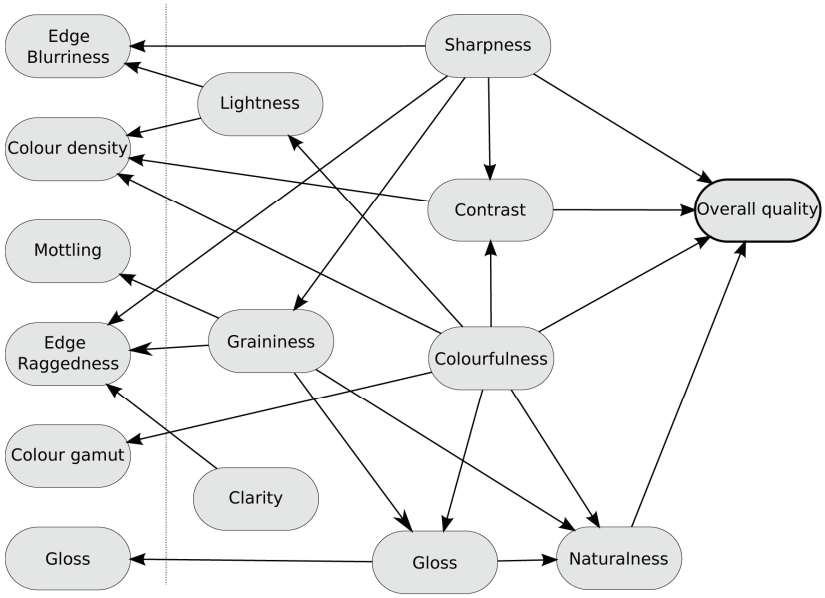



**Figure 2.** A Bayesian network structure found by Eerola et al. [18] for the quality attributes of a set of printed images (copied from Eerola et al. [18]).

Engeldrum [86] proposed the famous concept of Image Quality Circle (IQC) to help clarify the structures, links and elements of image quality rating and technology variables. The image measurement hierarchy of I3A and the IQC closely resemble each other. The objective level includes low-level technical properties (as technology variables), low-level technical measurements (as physical image parameters) and perceptual measurements (as customer perceptions). The subjective level includes low-level subjective attributes (as customer perception) and high-level subjective attributes (as customer image quality ratings).

## 2.4 The general measurement hierarchy for imaging systems

Figure 3 shows the image quality measurement hierarchy that served as the foundation of the metrics developed in this dissertation. The hierarchy is hypothetic and based on the above literature survey. The attributes of the hierarchy are based on their generality in the area of imaging systems.

Sharpness, contrast, noise, lightness and color are the low-level attributes. The color attribute is composed of attributes such as color balance, color accuracy and chroma. Color accuracy includes several metrics that characterize color difference.

Our hypothesis is that the overall image quality of a captured or printed image is a combination of the high-level attributes of clarity and naturalness. Naturalness is concerned with the correspondence between a picture and the anticipated view or the real view in the memory of an observer. Clarity determines whether the content of an image is easy or difficult to distinguish.

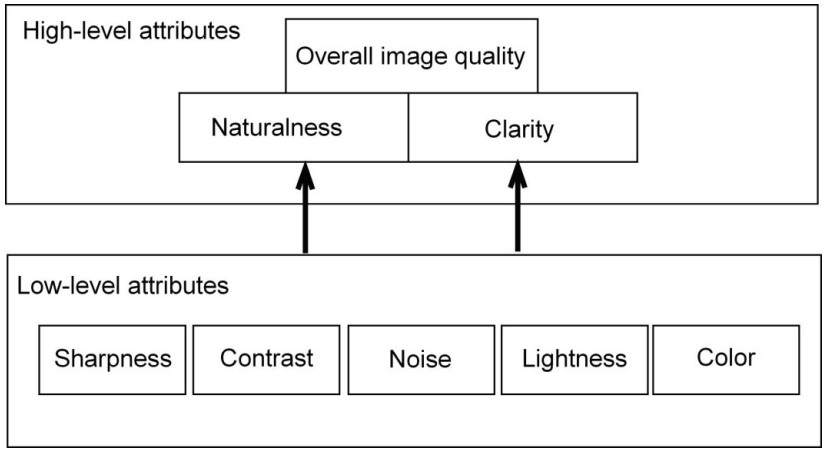



**Figure 3.** The image measurement hierarchy for imaging systems.

## 2.5 Test target metrics

There is a strong tradition of characterizing imaging systems by using test target metrics. In the case of printers, the measurements are performed by printing digital test target images and by using specific measurement devices to measure the reproduction from the printed paper. The measurement device can consist of, for example, a densitometer, a colorimeter or a spectrophotometer. Prior scholars have called the printer measurements “instrument measurements”. The data have been used to compute, for example, color reproduction and accuracy, print density, print gloss, sharpness, details and unevenness [90].

The test targets of digital cameras are measured by capturing the printed test targets under specific types and levels of illumination and by computing the metrics from the acquired signal. The ISO (International Organization for Standardization) has published camera measurement standards for sharpness [43], noise [45], lens optical distortion [47], ISO-

value [42], opto-electrical conversion function [44] and color [46] characterization and measurements.

The camera sharpness is measured by computing the frequency response of sharp edges (SFR, spatial frequency response). First, the gradient is calculated from the edge, and the Fourier-transformation gives the frequency response. It is known beforehand that the edges in the test target are sharp and that an accurate reproduction requires a camera that does not filter high frequencies or add energy in the edges. Publication V used MTF50 value as a reference for the proposed sharpness metric, which was computed directly from natural images. MTF50 is the spatial frequency at which MTF = 50% (i.e., at which the contrast has fallen to half of its value at low spatial frequencies).

The standard camera noise measurements are performed by determining the signal-to-noise-ratio (SNR) from the even patches. It is known beforehand that the patches are even in the test target and that the intensity variance calculated from the images captured by the camera exists because of noise. The annex of ISO 15739 standard [45] also introduces a visual noise (VN) test target metric. Publication III used the VN metric as a reference for the proposed color noise metric, which was computed directly from natural images. VN accounts for the properties of the HVS. The image data of each patch are converted to the spatial frequency domain by using a Fourier transform, which is applied to the color components of a uniform color space. The noise power spectra are weighted with the CSF, and the inverse Fourier transform is applied. The weighted sums of the three standard deviations for each axis in a uniform color space are calculated.

The camera color measurements are performed using specific color patch targets [85]. The metrics include the luminance, chroma and/or hue difference between the measured values from the test target patches and the computed values from the captured images. The measurement device can be, for example, a spectroradiometer or a colorimeter. Publication VI used the color difference values as ground-truth data for the proposed color difference metric, which was computed directly from natural images.

Initially, the camera test target metrics were designed for the manual cameras. Adaptive processing in modern cameras hinders the interpretation of measurement data. For example, signal sharpening does not add new details in an image, even though the spatial frequency response value that measures the detail reproduction increases as a result of the sharpening [84],[66]. Noise removal algorithms filter noise energy from the smooth areas of images, and the SNR value increases. The problem is that it is difficult to distinguish between the image structure

and noise energy, and a noise removal algorithm can also filter image details. The SNR value of a camera can be high, but image quality is low because the noise removal algorithm has filtered out the image details.

Instead of predicting the subjective quality, the data from the test target measurements can be used to characterize the devices. The characterization describes how an imaging system reproduces, distorts and manipulates those signals captured under ideal conditions. For example, how an edge, a smooth surface or a color patch is reproduced can be measured. The SFR can indicate whether the imaging system filters or amplifies frequencies. The SNR computes the smoothness of the even patches after capturing or printing the image. The color metrics depict how different combinations of color signals are interpreted. These factors do not determine how a complex view (natural image) is captured by a camera or how a subject would perceive the view if the image was shown on a display or printed on a paper.

Figure 3 shows the image measurement hierarchy derived from the literature review and used in this dissertation. Figure 4 expands the image measurement hierarchy in Figure 3 by including the system-level characterizations for which test target images are used. The whole measurement hierarchy outlines the distinct processes or levels for imaging system benchmarking studies. These studies need both subjective and objective data. A strong link exists between the objective test target metrics and the characterization values. However, the link between the test target and the subjective data is weak. The device characterization process measures the reproduction of the test targets captured or printed by an imaging system. The subjective quality measurement process requires natural pictures that are captured or printed by an imaging system. Test target pictures cannot be used. Only natural pictures can express the naturalness or clarity captured by a camera or printed by a printer. Both subjective and objective processes are tedious and require time resources. For the subjective tests, natural pictures should be taken, and for the objective study, the test target pictures should be taken in a controlled laboratory environment.

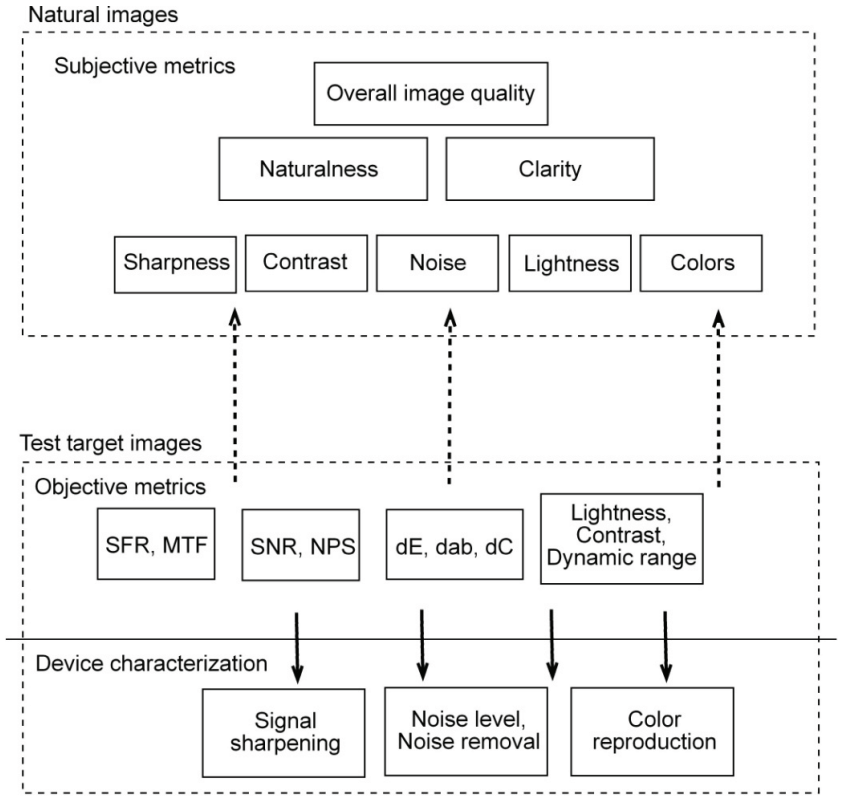



**Figure 4.** The whole image measurement hierarchy for benchmarking studies of imaging systems.

## 2.6 Algorithmic metrics

Previous scholars have mainly developed algorithmic metrics for image processing applications. Their objective is to estimate the overall quality directly from natural images.

Algorithmic metrics can be divided into three types: full-reference (FR), reduced-reference (RR) and no-reference (NR) metrics. The metrics are divided based on the availability or usage of a reference image. Figure 5 shows the basic requirements for the three types of metrics. An FR metric requires a pixel-wise reference image. That is, the original or reference image should be available, and the test image should be free from nonlinear geometrical local distortions. The corresponding pixel positions that relate the reference and test images should be known, or it should be possible to find them. The term “reference image” refers to an image whose visual quality or information capacity is high compared with that of the test images.

An RR metric requires some information from the original or reference image. An RR metric often computes a feature vector for the reference and test images. Unlike FR metrics, most RR metrics do not need pixel-wise reference images. The feature vector is often based on global statistics. For example, the feature vector can be composed of the parameters of given statistical distributions.

An NR metric does not need a reference or original image. The computed image quality metrics are based only on the information that is available from the test image. However, the performance of NR metrics is still limited. Traditionally, NR metrics are based on the assumption that a specific and known distortion type has distorted the image. NR metrics fail if the distortion space of the test image is multi-dimensional. NR metrics cannot handle test images with many concurrent distortion types, such as printed images or images captured by digital cameras.

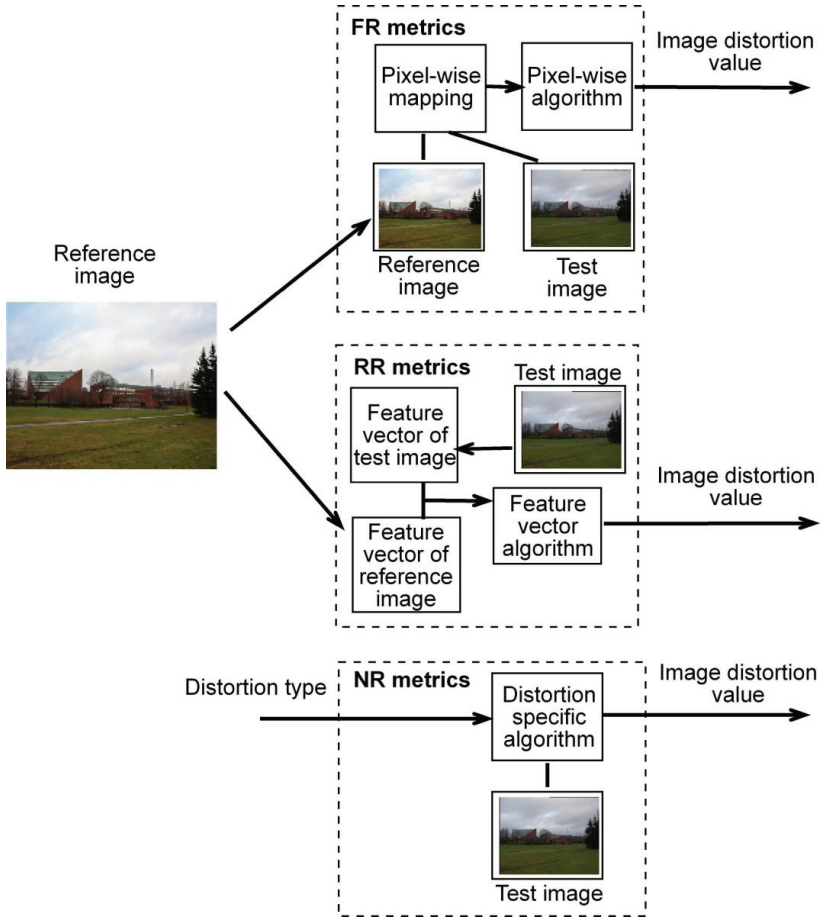



**Figure 5.** FR metrics require a pixel-wise reference image, RR metrics require a feature vector from the reference image and NR metrics require the distortion type as input data.

### 2.6.1 Full-reference metrics

Most FR metrics are based on HVS modeling [36], [8], the structural similarity principle or natural scene statistics (NSS) [99]. The HVS metrics are computationally complex. A HVS metric can include the components for preprocessing, channel decomposition, error normalization, error values pooling and fidelity value calculation [99]. The preprocessing component filters the images based on the point spread function (PSF) of the human eye. The channel decomposition component divides the signal into the scale and orientation information channels. The error normalization component filters the image by using, for example, CSF and the known masking functions. Finally, the error pooling component sums the pixel-specific error values between the reference and test images into one scalar number.

The recent FR metrics are mainly based on the structural similarity principle. The assumption is that the overall image quality relates strongly to the ease of image interpretation. These metrics are computationally simpler than the metrics based on the HVS. In addition, the performance of the metrics is comparable with that of the metrics based on the more complex approaches. For example, Sheikh et al. [101] used the LIVE image quality database to evaluate several FR algorithms whose codes were publicly available on the Internet or obtained from the authors. According to the results, the IFC, VIF, SSIM (MS) and JND metrics performed much better than the rest of the algorithms. The IFC and VIF are based on the NSS, the SSIM (MS) is based on image structure and the JND is based on the HVS.

The first metric that used the structural similarity between the reference and test images was the SSIM (Structural Similarity Index) metric [113]. The SSIM computes luminance, contrast and structural similarity values between the reference and test images. The structural similarity is measured using the cross-correlation value. Later, the SSIM metric was extended in several ways. For example, past scholars have used the complex wavelet domain [97], the edge-finding approach [10], the image gradient [11], different image scales, the SSIM (MS) [119], the added equalization dimension [5] and visual attention model weights [28] to enhance the performance of the original SSIM.

The SSIM metric has almost become the de facto standard for validating new image processing methods, such as image compression or super resolution. The idea is to show that a compression method does not change images or that a super resolution method can use low-resolution images to construct an image comparable with its high-resolution counterpart.

In addition to the SSIM metrics, past scholars have proposed many other structural FR metrics. For example, Zhang et al. [126] created the FSIM (feature similarity index metric), which includes two components. The components measure the significance and contrast of the images. Kim and Park [54] proposed the PQC (phase quantization code) metric, which is based on the phase difference. Later, Kim and Park [53] proposed the APQC (amplitude/phase quantization code) metric, which is an extended version of the PQC metric. In addition to the phase difference, the APQC metric calculates the amplitude difference. Kim et al. [51] proposed the GIQM (gradient information-based quality metric), which calculates the Harris response (HR) values from the gradient image. HR describes the structures of points in images. Shnayderman et al. [104] proposed a metric that compares the singular values of singular decomposition for the reference and test images. Narwaria and Lin [80] proposed a metric based on the singular vectors of singular decomposition. This metric assumes that the singular vectors have information related to the structural differences between the images. Ma et al. [68] proposed a metric that calculates the visual horizontal effect (HE) and the salience from the SSIM image. The HE quantifies the effect of image content, and the SSIM expresses the effect of the orientation of a stimulus. Han et al. [33] proposed a metric based on the U matrix diagonal values of LU factorization. This metric assumes that the U elements of the matrix relate to uniformity or homogeneity. Zhang et al. [127] proposed a metric that calculates the number of stable edge points for the reference and test images. This metric assumes that the number of edge points relates to the structural correctness.

### 2.6.2 Reduced-reference metrics

According to some estimations, the performance of FR metrics, in case of single distortion images, has reached a saturation point [60]; the predictions of state-of-the-art FR metrics are close enough to subjective evaluations. New extensions or modifications of the metrics will not significantly increase the performance. However, in the case of imaging systems, the applicability of FR metrics is limited because of the lack of pixel-wise reference images. Furthermore, the performance can be lower because of complex and multiple distortion sources [87]. Compared with the full-reference principle, the reduced-reference principle increases the number of use cases. For example, RR metrics can be used with video or image streaming applications where feature vectors are sent through the ancillary channel [114].

Several RR metrics are based on the NSS. Natural un-distorted images have certain statistical properties that hold across different contents. For

example, it is a known fact that the power spectrum of natural scenes can be modeled while assuming that the fractal law will hold (i.e., approximating the shape with dependency $1/f^a$, where f is frequency and a amplitude parameter) [26]. The amplitude parameter provides the shape of the power spectra. The NSS approach to RR metrics often models the marginal probability distributions of the coefficients of a transformation space. For instance, the Cheng and Cheng metric [14] fits the image gradient values to the generalized Laplace distribution model. Their metric calculates the Kullback-Leibler distance (KLD) and the variance difference between the distributions of the reference and test images. The Xue and Mou metric [122] and the Wang et al. metric [120] calculate the wavelet coefficients by using the steerable pyramid technology. The Xue and Mou metric [122] fits the wavelet coefficients to the Weibull distribution model:

$$p_{wbl}(w;\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{w}{\alpha}\right)^{\beta-1} e^{(-(w/\alpha)^{\beta})}, \quad w>0 \tag{1}$$

where $\alpha$ describes the scale, $\beta$ describes the shape of the distribution and $w$ is a coefficient. The parameters $\alpha$ and $\beta$ from different scales are the features of the reference image [122]. The Wang et al. metric [120] fits the wavelet coefficients to the Generalized Gaussian Density (GGD) model:

$$p_{GGD}(w;\alpha,\beta) = \left(\frac{\beta}{2\alpha\Gamma(1/\beta)} e^{(|w|/\alpha)^{\beta}}\right) \tag{2}$$

where $\Gamma(a)$ is the gamma function, $\alpha$ describes the scale and $\beta$ describes the shape of the distribution. The parameters $\alpha$ and $\beta$ and the error incurred while approximating the empirical coefficients of this distribution from the different subbands of the wavelet decomposition are the features of the reference image. The Li and Wang metric [61] fits the wavelet coefficients to the GGD after divisive normalization. Their metric computes the KLD, standard deviation, kurtosis and skewness between the reference and test images. Figure 6 shows a principle where the original or reference image statistics are modeled using a parametric model, and the parameters function as RR features (feature vector). Image quality is calculated by comparing the features of the original image and test image.

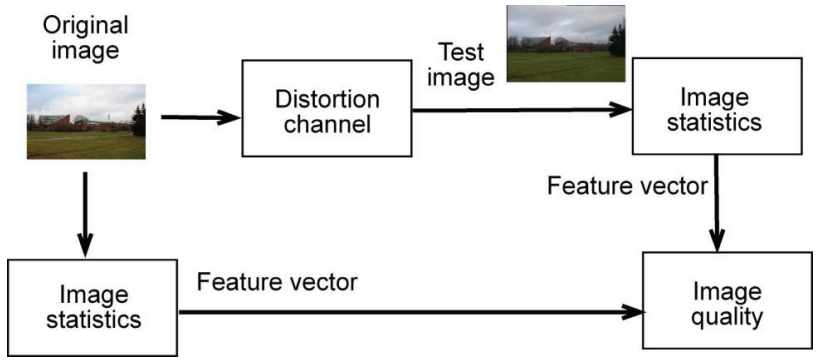
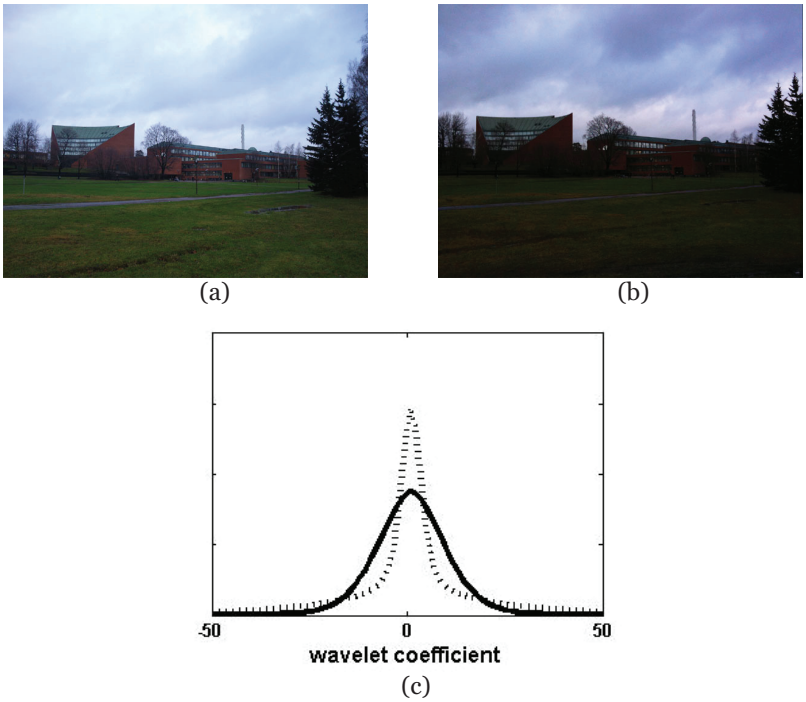



**Figure 6.** The main principle of many RR metrics is that a distortion modifies the known image statistics and that image quality is calculated using the feature vectors derived from the parameters of the statistical models.

The idea behind the NSS metric [14], [122], [120], [61] is that the orientation responses of natural images are highly kurtotic. The probability distribution of the responses of a natural image has a high peak and long tails. The marginal distributions change in different ways for different types of image distortions. Figure 7 shows the images captured by a high-quality camera and a low quality-camera as well as the marginal distributions of the wavelet coefficients (first scale vertical band) for the images. The distributions of the images differ, and the difference can be modeled. Compared with the distribution of the un-distorted image (high-quality image), the distribution of low-quality images (solid line) has a flat peak and short tails. For example, when $\beta < 2$ of the GGD (Equation 2), the tails are heavier than they are in the normal distribution, and when $\beta > 2$, the tails are lighter than normal. The tails may be heavy because of noise, and the tails may be light because of blurriness.

Some RR metrics utilize the contrast sensitivity function (CSF) before calculating the overall image quality. For example, the Li et al. metric [62] applies the directional filter bank (DFB) to wavelet decomposition. The metric filters the wavelet coefficients by using the CSF and calculates the threshold value. The threshold value determines whether a wavelet coefficient is visually discriminative. The metric compares the numbers of visually discriminative coefficients between the reference and test images. In the study of Li et al. [62], the free parameter for the visual threshold equation was determined by empirical tests. The Tao et al. metric [109] uses the contourlet domain to calculate the RR feature. The coefficients are filtered using the CSF, and the visual threshold is calculated for different subbands of the wavelet decomposition. The metric compares the number of visually discriminative coefficients between the reference and test images in different subbands. The Ming et al. metric [73] is the same as the Tao et al. metric [109], but the former also calculates the average H and S

values in the HSV color space. The Maalouf et al. metric [70] uses the Grouplet domain to calculate the RR features. The Grouplet transformation is based on the group matching method, which classifies the points in the same neighborhood into groups and defines the image geometry. The metric calculates the difference in value between the coefficients of the reference and test images after using the CSF. The Gao et al. metric [25] uses multiscale geometric analysis (MGA), the CSF and the Weber JND to perform calculations. The MGA includes many different transformations, such as wavelets, curvelets, bandlets, contourlets, wavelet-based contourlets and directed filter banks.




(a) (b) (c)

**Figure 7.** The image captured by the high-quality camera (a), the image captured by the low-quality camera (b) and the probability distributions of the wavelet coefficients for the high-quality (dashed line) and low-quality (solid line) images (c).

### 2.6.3 No-reference metrics

The literature has proposed many NR metrics, but finding quality measurements without having any knowledge of a reference is a difficult problem. Past scholars have often designed NR metrics to measure the distortion levels of images with a single distortion. The distortion is often JPEG or JPEG2000 image compression, blurriness or noise.

Image blurriness is often measured by calculating the width of edges that can be found from the test image. For example, Marziliano et al. [71] found the edges by using the Sobel-operator and calculated the edge widths as

pixels. Ferzli and Karam [24] proposed the just-noticeable blur (JNB) concept. If the edge width is higher than the JNB, the probability that an image is unsharp increases. Figure 8 provides an example where a sharpness metric finds edges and calculates the width of the edges from the edge profiles.

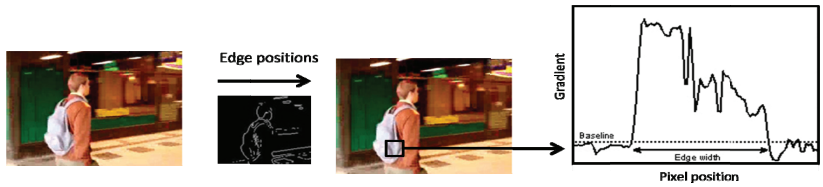



**Figure 8.** Image sharpness can be estimated by calculating the width of the image's edges.

Additionally, Liang et al. [63] and Caviedes and Gurbuz [7] proposed NR metrics based on edges. Liang et al. [63] calculated the histogram of vertical and horizontal gradient profiles. The shape of the histogram describes the sharpness of the image. Caviedes and Gurbuz [7] calculated sharpness values by using the kurtosis of the discrete cosine transform (DCT) values from an edge neighborhood. Scholars have also used singular values, eigenvalues and wavelet decomposition as sharpness metrics. Zhu and Milanfar [129] measured sharpness by using the singular values of a gradient image. Chen and Bovik [13] calculated sharpness by using the distributions of the gradient and wavelet-decomposition values. Wee and Paramesram [121] estimated sharpness by using the highest eigenvalues of a normalized image. The researchers expected that the dominating eigenvalues would relate to sharpness and that the less dominant eigenvalues would relate to noise.

Image noise is difficult to measure from natural images, and the literature shows only a few algorithms on this point. For example, Immerkaer [41] proposed a noise metric that utilizes the variance of the image filtered by the Laplace-operator. However, the method is also sensitive for the fine image structures. Tai and Yang [107] presented a noise metric that tries to compensate for the effect of edges on noise measurements. The metric tries to filter out edges before the noise calculations take place. The more general application area for the noise measurements than the image quality is the noise removing. Noise removing methods often need the threshold values for the algorithms before the image is filtered.

The JPEG and JPEG2000 compression methods incur at least two different types of image distortions. Both compression methods blur the image. In addition, the JPEG compression causes blockiness, and the JPEG2000 compression causes a ringing distortion. The blockiness appears as an artificial discontinuity between the adjacent blocks in an

image. The ringing distortion adds extra structural energy around the edges of an image. The NR metrics for the JPEG distortion often measure only the blockiness. The basic idea is that the block distortion forms a regular error signal on top of the original image signal [115], [112]. The error signal can be estimated, for example, from the peaks of the autocorrelation function or from the harmonic frequency in the frequency domain. Figure 9 shows an example where a method calculates the horizontal gradient of the image and estimates the level of blockiness from the autocorrelation function.

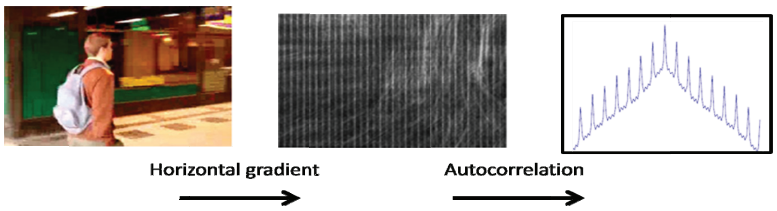



**Figure 9.** The block error signal can be estimated from the peaks of the autocorrelation function.

The JPEG metric proposed by Wang et al. [117] computes two components for the test image. The first measures the blockiness level, and the second measures activity. The JPEG2000 metric by Zhang and Le [125] measures the monotony of the pixels and the structure of the image content. The idea is that the second component compensates for the structural energy generated by a ringing distortion, whereas the other component calculates blurriness. The components of the Liu et al. [65] metric extract the regions that are likely to be impaired by a ringing distortion and quantify the visibility of the distortion. The visibility is measured by comparing the detected regions and the corresponding local background. The Sheikh et al. [100] metric for JPEG2000 is based on the joint distributions of the subbands of wavelet coefficients. The idea is to measure to what degree the JPEG2000 compression changes the distributions. The changes occur because of the signal quantization in the compression process.

The metrics were developed for the images with a single distortion source and for the applications with a known distortion type. The distortion space of imaging systems is multidimensional and includes a variety of distortion sources. The distortions of the images captured or printed by imaging systems can be dependent or independent of each other. Some distortion sources are known, but some are unknown. Because of the multidimensional distortion space of imaging systems, the performance of the NR metrics presented in the literature is low [15], [12]. For example, if

sharpness or blurriness is measured from the edge areas, the metric can interpret the noise energy of an imaging system or the block structure arising from the JPEG compression as an image structure. A noise metric can interpret the fine details of an image as noise energy.

For example, Ciancio et al. [15] and Chen and Bovik [12] tested NR sharpness metrics on the blurred image database (BID [2]). The BID includes subjective data and real images captured by digital cameras. According to the results of both studies, the performance of the NR sharpness metrics is not high for camera images. The Chen and Bovik [12] metric had the highest correlation at 0.586.

If the sharpness or blurriness of an image captured by a camera is measured, the intentionally unsharp background can also be a problem. The narrow depth of focus is a common method that photographers use to focus the viewer's attention on the subject of the picture. The principle of the metric proposed by Narvekar and Karam [78] tries to compensate for the unsharp background problem. The metric computes the sharpness only from the areas that are defined (according to a threshold value) to be sharp enough. However, the metric cannot handle the noise energy, and the performance of the metric is not high for camera applications. Based on the study presented in Publication V, the performance of the metric [78] equals the performance of the standard sharpness metric based on edge widths [71].

Recently, the research on NR metrics has focused more on applications without a priori knowledge about the specific distortion type. The approach relies on learning models, which input the feature values computed from the test image. The goodness of the methods depends on the features and the data used for learning purposes. The features are e.g. based on DCT, wavelets, curvelet transformations or Gabor filters.

For example, the Saad et al. metric [95] fits the DCT coefficients of the test image to the GGD model, and the features are derived from the parameters of the GGD model. The metric [95] was developed further by Saad et al. [94], who utilized statistical modeling and a different set of sample DCT statistics. Moorthy and Bovik [74] proposed a metric based on a framework with two phases. In the first phase, the probabilities for the different distortions are calculated. The probabilities are used as weighting factors for the second phase, when the quality values are calculated for the pre-determined distortion set. The GGD model parameters calculated from the wavelet coefficients function as features for both phases. The probabilities of the different distortions are modeled using the support vector machine (SVM). The image quality value is calculated using a support vector regression (SVR). Moorthy and Bovik [75] further

developed the metric [74] and replaced the parameters of the GGD model with 88 different features calculated from the wavelet coefficients. Figure 10 shows an example where the two-phase model is used to compute NR quality. The first phase inputs features $f_i$ and calculates the probabilities of the different learned distortions. The second phase applies the distortion-specific metrics for the test image and calculates the overall image quality by using the probabilities from the first phase as weighting factors.

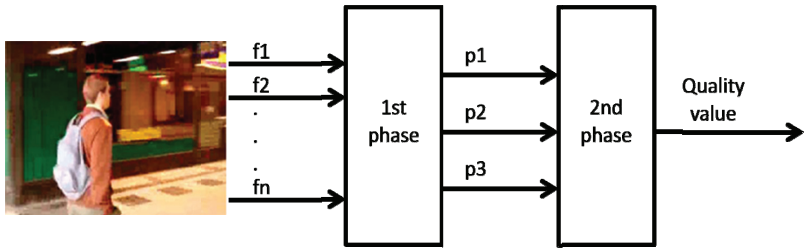



**Figure 10.** A two-phase model estimates the overall image quality: the first phase calculates the probabilities of different distortions, and the second phase calculates the overall image quality value by using the probabilities as weighting factors. $F_i$ denotes feature values and $p_k$ denotes the probabilities of the different distortions.

Shen et al. [103] proposed the HNR (hybrid no-reference) metric, which is based on the hybrid of the curvelet, wavelet and cosine transforms. The metric calculates the locations for the top coordinates of histograms. The metric assumes that the different distortions locate the coefficients in different clusters in the domain of the top coordinates. Li et al. [60] proposed a metric based on a neural network that inputs the following three features: phase congruency, image entropy and image gradient. Ye and Doermann [123] proposed a metric based on local texture analysis. It uses Gabor filters to capture statistics of image patches and a visual codebook to link local statistical properties and visual quality.

Although the abovementioned metrics have been used to measure different types of distortions from images, they do not solve the problem of a multidimensional distortion space. The metrics can be highly effective at predicting the effect of a single distortion, but the performance decreases if the image concurrently includes more than one distortion. In addition, the metrics have been used for specific distortion sets, such as the distortions that can be found in the popular LIVE image set. If the parameters of the model are learned, there should be knowledge of application-specific distortions. In addition, there should be knowledge of the features (metrics) that characterize the distortions. An image captured or printed by an imaging system has many different types and sources of distortions without any robust known features for characterizing these distortions.

### 2.6.4 Color information

Algorithmic metrics are usually applied only to the luminance or the intensity channel. This decision can be justified by the fact that the algorithmic metrics often determine the level of image deterioration in terms of the image structure. The HVS is more sensitive to changes in the luminance than to changes in the chrominance channels. However, some scholars have also suggested methods and metrics that utilize the components of the chromatic channels.

A simple FR type metric is the color error, which is expressed as the Euclidean distance ΔE between the reference and test images in the CIELAB space. The performance has been increased by utilizing the properties of the HVS. For example, the S-CIELAB metric [128] accounts for the sensitivity of the HVS to spatial frequencies before the color error values are calculated. The Hong and Luo [37] metric assigns a higher weight to the dominant colors and to the color with a greater difference when calculating the color error values.

If no reference image is available, the color metrics utilize the statistics of the images and different assumptions. For example, Yendrikhovskij [124] proposed a metric that computed the color naturalness of the image by using the mean and deviation values of the saturation component. Hasler and Süsstrunk [35] proposed a metric that computed the colorfulness of the image based on the mean and variation values of the chromatic components in the CIELAB space. The metric assumed that the perceived colorfulness of an image correlates with the mean and standard deviation in the chromatic plane.

## 2.7 Test and reference image digitization

In essence, test target measurements compare measured test signals with known reference signals. The main problem with the objective measurements of imaging systems using natural images is that these measurements are missing a reference signal (camera applications) or have different reference (digital) and test signal (analog) forms (printer applications). Table 1 lists the requirements for FR, RR and NR metrics when applied to camera and printer measurements. NR metrics can be used directly in the camera applications because the output of a camera is a digital image. FR and RR metrics always require a reference image. For the camera applications, reference images are missing. For the printer applications, a reference image is available because the original digital images can function as reference signals. The problem is that the printed

test images must be digitized before the test and reference signals can be compared.

**Table 1.** Requirements for the FR, RR and NR measurements when applied to camera and printer quality measurements

| | Camera | Printer |
|---|---|---|
| FR | Pixel-wise reference image | Test image digitalization |
| | | Pixel-wise image registration |
| RR | Reference image | Test image digitalization |
| NR | - | Test image digitalization |

A few studies have applied NR metrics to camera images [15], [12] or to printed images [30], [32]. In addition, prior scholars have proposed frameworks [17], [87] for applying FR metrics to printed images.

Halonen et al. [30], [32] applied NR metrics to measure the quality of printed images. In the study [30], the researchers used a scanner to digitalize the sample. In the studies presented in [32], Halonen et al. used the digital images that were digitized using the method developed in this dissertation. The digitization system was based on a high-quality camera and is described in detail in Section 3.5. By comparing the results of the study [30] using a scanner with those of the studies [32] using the camera system, we can conclude that the performance of the proposed camera system is high compared with that of a scanner.

Eerola et al. [17] and Pedersen et al. [87] proposed methods for applying FR metrics to printed images. Figure 11 shows the components of Eerola's method [17]. The method digitizes the printed image (hardcopy) by using a scanner, descreens the scanned image by using a Gaussian low-pass filter (GLPF) and registers the original and test images. The registration accuracy of the method is less than a pixel between the reference and test images. Eerola et al. claimed that the inaccuracy is compensated for by low-pass filtering before the quality computation. Pedersen et al. [87] assumed that because FR metrics often include low-pass filtering, they compensate for the pixel-wise inaccuracy of the registration.

In both studies [17] and [87], the researchers applied state-of-the-art FR metrics to measure the quality of printed images. In Eerola's study [17], the performance of some FR metrics was high with the test image set, whereas in Pedersen's study [87], the performance of FR metrics was low.

Pedersen's image set included fifteen images. The images were processed with two sRGB ICC profile versions (v2 and v4). In addition, the images were further processed to obtain eight different reproductions for each

original image. The images were printed, scanned and registered. The printed images were evaluated by 30 observers, and FR metrics were applied to the scanned and registered images. The Pearson correlation coefficients between the subjective evaluations and FR metrics were low for all of the implemented metrics.

Eerola's image set [17] included three images printed by ink-jet and electrophotography printers. The variation in images arose from the use of different grades of paper. The printed images were evaluated by 28 observers, and the scanned and registered images were measured by FR metrics. The state-of-the-art FR metrics accurately predicted the subjective evaluations. The highest correlation coefficients were over 0.96 for the image set printed by the ink-jet printer and over 0.86 for the images printed by the electrophotography printer.

Eerola's registration method [17] was based on global image transformation. First, SIFT (Scale-Invariant Feature Transform) computes the corresponding points between the reference and test images. Then RANSAC (random sample consensus) is used to find the best 2D-homography, and the test image is transformed into the reference image.

Pedersen's method is simpler, but it is more cumbersome to use than Eerola's method. Pedersen's method requires an image to be padded with a white border and equipped with four register marks before being printed. The method calculates a similarity transform based on the registration marks. Eerola's method is free of registration marks. It finds the corresponding points between the reference and test images and makes the global image transformation based on the points.

Both methods fail for those images that have been captured by different cameras. The methods cannot model the nonlinear geometrical transformations between the images. Eerola's and Pederesen's methods function only if the test images are captured from a planar surface (e.g., printed and scanned images). The methods fail to work for those images that have undergone geometrical distortions (e.g., camera images).

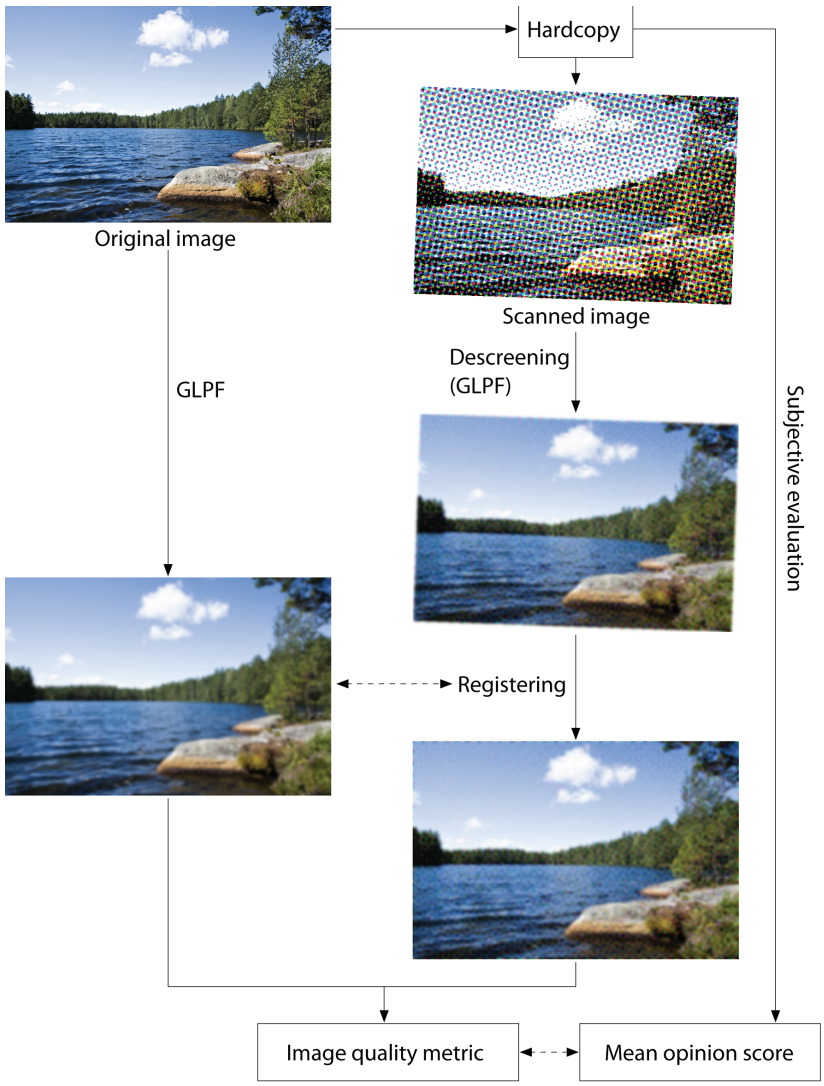



**Figure 11.** The structure of the Eerola et al. [17] framework and the data flow for computing full-reference image quality metrics for printed images. GLPF is Gaussian low-pass filter. (copied from Eerola et al. [17]).

## 2.8 Types of test images

For the purposes of image quality research, scholars have developed different test images. The publicly available test images help to develop algorithms and image processing methods that predict subjective image quality or, in the case of image processing, help to improve quality. The general requirement for the test images is that they should reflect typical problems from real-world applications. The problems should be visible from the displayed, printed or captured test images and should

differentiate the imaging systems or image processing algorithms of interest.

Table 2 groups the test image types into three categories: test image databases, digital test images and test image scenes. Each type serves different purposes.

Test image databases include sets of test images that have undergone some type of distortion and subjective data. Test image databases are used to measure the performance of image quality algorithms. The algorithm can be a full-reference or reduced-reference algorithm only if the database also includes the undistorted images. Without the undistorted images, only NR metrics are applicable.

Digital test images can be used to measure the performance of displays, printers and some image processing algorithms. These algorithms include image compression, image enhancement and tone-mapping methods. Test image scenes are used to measure the performance of camera systems.

**Table 2.** Test image types for measurements of visual image quality

| | Image database | Digital images | Image scenes |
|---|---|---|---|
| Measurement application | Image quality algorithms | Image processing algorithms, Displays and printers | Camera systems |
| References | LIVE [101], TID [89], IVC[58], A57 [9], MICT [38], BID [2] | Sony [106], Kodak [55], HDR images (Fairchild) [22] | Image clusters (I3A) [40] |

Each test image type is associated with a distinct measurement procedure. A test image databases have been used for measuring the performance of image quality algorithms. First, an objective image quality algorithm computes the objective data. Then, for example, the correlation coefficients between the objective data and subjective data are used to evaluate the performance of the algorithm.

The measurements of displays, printers or image processing algorithms cannot utilize a priori collected subjective data. The inputs are digital test images, and the differences between the inputs and outputs are measured. As a result, the subjective data of the output images need to be collected afterwards. With printers, the images are evaluated from prints. With displays and algorithms, the displayed images are evaluated. The objective data for displays and printers should be measured from the digitized

versions of the printed or displayed optical (analog) images. The objective data for processed images can be measured from the output image of an algorithm. Available test images are useful, because staging scenes and capturing processes can be avoided. In addition, if the same test images are used, the data from different measurement sets are more comparable.

The camera measurement applications cannot utilize existing digital test images. The performance of a camera system should be measured based on the images captured and processed by the camera system. The traditional method captures physical test target images (see Section 2.1). The predefined test image scenes make the image acquisition process faster and easier. In addition, the verified test image scenes ensure that the captured test images measure the critical aspect of camera performance. The subjective data are gathered by showing the captured images on a display or as prints. Objective data can be computed from the digital images captured and processed by the test cameras.

From the standpoint of this study, digital test images and test image scenes are relevant. Test image databases cannot be utilized for measuring the performance of imaging systems.

### 2.8.1 Digital test images

Some standard and general digital test image sets can be used to test image processing algorithms. The Sony sRGB standard image set [106] consists of two indoor images (portrait and party themes) and an outdoor image (picnic theme). The Kodak Lossless True Color Image Suit set [55] consists of twenty-two outdoor images and two indoor images.

Fairchild [22] presented a high-dynamic-range (HDR) test image set that includes HDR images with the colorimetric and color appearance data from a scene. Kuhna et al. [57] made HDR test images available. The test images were designed to test the performance of tone-mapping algorithms. The researchers chose images that were similar to the photographs consumers typically take.

Halonen et al. [31], [32] created test images for print quality evaluation purposes. The development work was started with three image contents (Figure 12). These images were chosen based on the presence of aspects important to image quality, such as memory colors (skin, sky and foliage), memory shapes (human face and cactus), different shades of natural green and areas with uniform colors and small details. Different types of surface materials ranging from shiny and smooth (e.g., porcelain and fabric) to detailed and textured (e.g., a zipper and rock) were also considered. The ultimate goal of Halonen et al. [31], [32] was to construct a single image for measuring all aspects of printed image quality. The development process included three test image versions. Figure 13a shows the first version of the

image, Figure 13b the second and Figure 13c the third. The first version of the image was highly engaging from a visual standpoint, the second version was highly colorful and the third version was the most natural [31].

In this dissertation, we used the image contents of Figure 12 to validate the proposed quality attribute metrics for printed images. The subjective measurement procedure and data are presented in Section 4.

(a) (b) (c)

**Figure 12.** Test images (i.e., man (a), cactus (b) and lake (c)) that include important aspects of image quality, such as memory colors, memory shapes, different shades of green and areas with uniform colors and small details.

(a) (b) (c)

**Figure 13.** Process of developing the test image to evaluate the print quality; first version (a), second version (b) and third version (c) [31].

### 2.8.2 Test image scenes

One step in moving from test target views to natural image scenes in camera characterization is to embed test targets or patches into a natural scene. Koivisto [56] followed this approach when he designed a test scene with hidden color patches to measure the color reproduction of digital cameras. Color measurements were performed by comparing the color values of the patches, which were measured by a color-calibrated camera and calculated from the images captured by test cameras. Figure 14a shows the scene and the locations of the hidden patches. Figure 14b shows the colors of the patches. The method proved to be promising, but the color accuracy of the calibrated camera was inadequate. The inaccuracy caused errors when calculating the reference color values of the hidden patches, and the performance of the method was only mediocre.


(a) (b)

Figure 14. **A test image scene with embedded color patches (a) and the colors of the patches (b) [56].**

The objective camera quality measurements of this dissertation were performed based on the images captured from natural test scenes. The test scenes were designed and selected to meet the benchmarking requirements of camera phones. The author of this dissertation was the main developer of the scenes.

The starting point of the scenes was the photospace approach described by I3A [40]. According to I3A, the photospace statistically describes the picture-taking frequency as a function of the subject illumination level L and the subject-to-camera distance D: PSD(L, D). The PSD is defined as a probability distribution: "the probability that an image is taken within a certain range of subject illumination and within a certain range of subject-camera distance" [40].

Segur [98] distinguished the photospaces of photographic utilization and photographic motivation. The photographic utilization space relates to a graph that describes where the camera users take photographs. The photographic motivation space relates to a graph that describes where the camera users would take photographs if possible. For example, compared with the range of a mobile phone camera, the operating range of a high-quality SLR camera is extensive; with the telephoto lens of SLRs, it is possible to photograph distant objects that could not be captured by camera phones. The test scenes used in this dissertation represent the photospace of photographic utilization.

In addition to I3A, Hultgren and Hertel [39] presented a photospace for camera phones. The study [39] drew its material from five cameras and 480 photos. According to the results, low illumination and short shooting distances dominate the probability distribution.

The photospace is also an interesting concept for video cameras. For example, Säämänen et al. [96] used it to define the concept of videospace by extending the photospace concept to three dimensions: scene lighting, subject-camera distance and object motion.

The photospace defined by I3A was divided into six parts, which are called clusters. Table 3 presents the definitions of the clusters [40]. A cluster defines the subject illuminance, subject-camera distance and scene descriptions for a typical scene captured by a mobile phone camera.

**Table 3.** Camera phone clusters defined by I3A [40]

| Cluster | Subject illuminance (Lux) | Subject-camera distance (m) | Typical scene description |
|---|---|---|---|
| 1 | < 50 | ≈ 1 | Close-up in dim-dark lighting conditions (indoor/outdoor) |
| 2 | 50-100 | ≈ 1 | Close-up in typical indoor lighting conditions (indoor/outdoor) |
| 3 | < 50 | > 4 | Small group in dim-dark lighting conditions (indoor/outdoor) |
| 4 | 50-100 | > 4 | Small group in typical indoor lighting conditions (indoor/outdoor) |
| 5 | > 3400 | 0.5-2 | Small group in cloudy bright to sunny lighting conditions (outdoor) |
| 6 | > 3400 | > 7 | Scenic landscape/large groups in cloudy bright to sunny lighting conditions (outdoor) |

We staged the scenes that are currently being used in ongoing projects and that were used in this dissertation with the following objectives in mind:

- be difficult to capture for typical camera phones,
- be able to differentiate camera phones,
- reveal camera-specific problems and
- represent views that typical camera phone users might capture with their cameras.

Table 4 shows the descriptions and two images per cluster as examples.

The images in Cluster 1 simulate a bar or restaurant image. They are close-up photos in dark lighting conditions. The illuminance is 2 lux, and the images are mainly exposed by camera flash. The short shooting distance sets the requirements for flash and signal gain tuning.

Clusters 2 and 3 simulate a living room environment. Cluster 2 is a close-up photo in typical indoor lighting conditions, and Cluster 3 is a photo of a small group in dim lighting conditions. The illuminance levels are 100 lux and 10 lux, respectively. In particular, Cluster 3 sets the requirements for flash power because the illuminance is low and the shooting distance is long (4 m).

Cluster 5 simulates a tourist image, and Cluster 6 simulates a landscape image. Cluster 5 is a photo of a small group in cloudy to sunny lighting conditions. Cluster 6 is a typical landscape photo. Cluster 4 is a studio image that device manufacturers use to make signal-processing adjustments or other measurements. Clusters 1, 2, 3, 5 and 6 are views that mobile phone users might be expected to capture with their cameras.

In this dissertation, the image scenes of Table 4 were used to validate the proposed quality attribute metrics for camera images. The subjective measurement procedure and gathered data are presented in Section 4.

**Table 4.** The illuminances, shooting distances and scene descriptions for the camera quality measurement clusters

| Cluster | Subject illuminance (lux) | Subject-camera distance (m) | Scene description | Images |
|---|---|---|---|---|
| 1 | 2 | 0.50 | Close-up in dark lighting conditions |  |
| 2 | 100 | 1.50 | Close-up in typical indoor lighting conditions |  |
| 3 | 10 | 4.0 | Small group in dim lighting conditions |  |
| 4 | 1000 | 1.50 | Studio image |  |
| 5 | > 3400 | 3.0 | Small group in cloudy bright to sunny lighting conditions |  |
| 6 | > 3400 | > 50 | Landscape image in cloudy bright to sunny lighting conditions |  |

# 3. Reference Image Quality Measurement Methods for Imaging Systems

This dissertation is grounded on the argument that the objective quality measurements of imaging systems can be simpler and the accuracy with which subjective quality is predicted can be higher if two conditions are met: 1) a reference image is available; and 2) instead of test targets, natural images are used. The next few sections summarize the methods proposed in Publications I-VI to enable the use of natural images.

Figure 15 shows the proposed image quality measurement framework. The illustration shows two applications of the framework: 1) camera measurements and 2) printing and display measurements. The inputs of camera measurements are the images captured by reference and test cameras, whereas the inputs of printing/display measurements are digital images.

The two main components of the framework are “characterization of image” and “computation of image quality”. In addition, the application of printing/display measurements uses a component called “digitization of print/displayed image”. The boxes with gray backgrounds describe the devices under study. In the camera measurement test, the cameras output digital test images for the “computation of image quality” component. With the printing/display measurement, the printed or displayed test images are digitized and fed to the “computation of image quality” component. The output of the framework is the value of an image quality attribute.

The inputs of the “characterization of image” component are reference images and the specification of a quality attribute. A reference image is used to compute a feature vector that can help compute the quality attributes of natural images.

The reference image in camera measurements is an image captured by a reference camera from the same test scenes used for the test images. The reference image for print/display measurements is a high-quality digital test image. The quality of the reference images should be high enough to compute robust features for the image quality metrics. For example, in this

dissertation, we took the reference images by using a high-quality professional SLR camera with a high-quality lens.

In this dissertation, we focused on the reduced-reference approach. With camera measurements full-reference approach can be problematic, because of non-linear geometric differences (perspective and lens distortions) between reference and test images (i.e., those without planar views and without constant shooting positions). For example pixel-wise reference images are missing. Section 3.3 further explains the problem. With printing measurements, the full-reference approach is feasible. Eerola et al. [17] and Pedersen et al. [87] have proposed methods for applying FR metrics to printed images. In this dissertation, we developed novel methods for applying RR methods for printed images. We wanted to study if it is possible to measure quality attributes without computational complex transformations between reference and test images and pixel-wise comparisons. Our reduced-reference approaches are simpler than the full-reference approaches proposed earlier.

Section 3.1 describes the function of the "characterization of image" component. Section 3.2 describes the attribute metrics that were developed and integrated into the "computation of image quality" component. Section 3.4 describes the "digitization of print/displayed image" component.

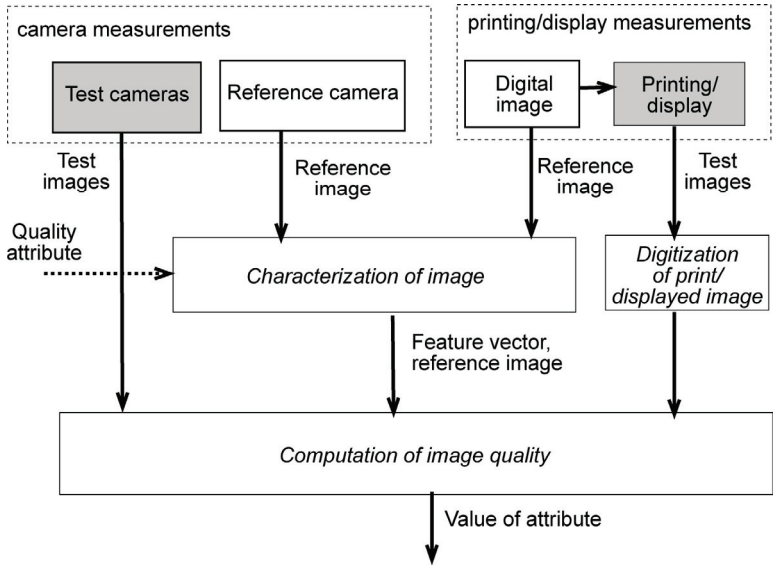



**Figure 15.** Image quality measurement framework for imaging systems (cameras and printing/display): the camera measurement application inputs reference image and test images captured by reference and test cameras, the printing/display measurement application inputs digital images and the output is the value of an image quality attribute.

## 3.1 Characterization of image

The "characterization of image" component extracts reduced reference information from the reference images. This dissertation applied three different characterization principles to the reference images. The principles are called the global feature, local feature and adaptive local feature.

Table 5 summarizes the metrics used in the dissertation. We classified the metrics according to the characterization principle. The metrics that were tested for printed images use the global feature or the local feature principles, whereas the metrics tested for cameras use the global feature or the adaptive local feature principles.

Figure 16a illustrates the global feature principle. In this principle, a single global feature is computed from all of the pixel values of a reference image. The global feature can be a parameter set of image statistics or the value of a no-reference metric. Figures 16b and 16c show the local feature and adaptive local feature principles. Both principles take advantage of a reference image's local values. Under the local feature principle, an image is divided into blocks, and quality attribute-specific sensitivity values are calculated for the blocks. The feature vector is fed to the "computation of image quality" component.

The adaptive local feature principle seeks new positions for the predetermined initial blocks by maximizing a quality attribute-specific sensitivity function. The process consists of two phases. First, the candidate blocks are found. Second, the blocks to be measured are selected from the group of candidate blocks. A feature vector includes the positions of the selected blocks. The feature vector and reference image are fed to the "computation of image quality" component.

**Table 5.** The metrics proposed in Publications I-VI grouped according to the characterization principles

| Characterization principles | Attribute | Metrics | Publication | Section |
|---|---|---|---|---|
| Global feature | Overall image quality | $D_1$, $D_2$, $D_3$ | IV | 3.2.1 |
| | Color contrast | CC | I | |
| Local feature | Graininess | G | I | 3.2.2 |
| | Sharpness | $S_1$ | I | |
| Adaptive local feature | Color noise | N | III | 3.2.3 |
| | Sharpness | $S_2$ | V | |
| | Color difference | $\Delta L$, $\Delta C$, $\Delta H$ | VI | |

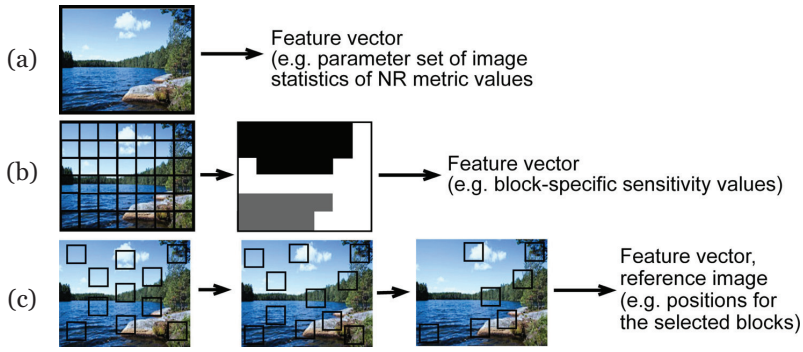



**Figure 16.** The RR features from reference images were calculated using three different principles: global feature (a), local feature (b) and adaptive local feature (c).

## 3.2 Computation of image quality

The next several sub-sections present the metrics and methods used to compute the quality metrics in Publications I-VI. The metrics are classified according to the principle of image characterization (Table 5).

### 3.2.1 Global feature characterization

The metrics defined in this sub-section are based on the principle of global feature characterization. The global feature principle computes the feature vector from all of the reference image's pixels (see Figure 16a).

**Overall image quality metrics**

The RR metrics from the literature cannot be applied to digital cameras because of the lack of reference images. Publication IV proposed to capture the reference images by using a high-quality reference camera. Three state-of-the-art RR metrics were implemented in the "computation of image quality" component. Two of the implemented metrics [120], [61] are based on the NSS approach. The third metric [20] is based on the simple NR metrics and their baseline values calculated from the reference image.

The first implemented metric was the Wang et al. RR metric [120]. It decomposes images into three scales and four orientations by using the steerable pyramid technology. The wavelet coefficients from the subbands of the reference and test images are fitted using the GGD model (see Equation (2)). The parameters $\alpha$ and $\beta$ of the GGD model from the different subbands of the wavelet decomposition are the features of the reference image. Wang et al. compute image quality $D_1$ by Equation (3):

$$D_1 = \log\left(1 + \left(\frac{1}{C_0}\right)\sum_{i=1}^{12}\left|kld^i(p^i \,\|\, q^i)\right|\right) \tag{3}$$

where $p^i$ and $q^i$ are the probability functions of the i-th subband in the reference and test images estimated by the parameters $\alpha$ and $\beta$ of the GGD model, respectively; $kld^i$ is the estimate of the Kullback-Leibler distance (KLD) between $p^i$ and $q^i$; and $C_o$ is a constant used to control the scale of the metric.

The second implemented metric was the Li and Wang RR metric [61]. The metric computes features by using divisive normalization (DN). Li and Wang claimed that DN accurately models the local behavior of the HVS. The metric performs the DN transformation for the wavelet coefficients. The wavelet decomposition in our implementation included three scales and four orientations. For the wavelet coefficient $w_c$, the new normalized value $\widetilde{w}_c$ is calculated by Equation (4):

$$\widetilde{w}_c = w_c / z \tag{4}$$

$$z = (Y^T C_U^{-1} Y / N)^{0.5} \tag{5}$$

where the covariance matrix $C_U = E[UU^T]$ is estimated from all of the subbands before the local $z$ is calculated. $N$ is the length of vector $Y$. Vector $Y$ includes thirteen wavelet coefficients: nine coefficients are from the neighborhood of the wavelet coefficients $w_c$, including the coefficient $w_c$; one coefficient is from the parent band; and three coefficients are from the other orientation bands. The normalized wavelet coefficients from the subbands of the reference and test images are fitted using the GGD model. The image quality $D_2$ is computed using Equation (6):

$$D_2 = \sum_{i=1}^{12} \log\left(1 + \left(\frac{1}{C_0}\right) kld^i (p^i \parallel q^i)^{\delta_1} (d_\sigma^k)^{\delta_2} (d_\kappa^k)^{\delta_3} (d_s^k)^{\delta_4}\right) \tag{6}$$

where $C_o$ is a positive constant, $p^i$ and $q^i$ are the probability functions of the i-th subband in the reference and test images estimated by the parameters $\alpha$ and $\beta$ of the GGD model, $kld^i$ is the estimate of the Kullback-Leibler distance (KLD) between $p^i$ and $q^i$ and

$$d_\sigma = |\sigma_R - \sigma_T| \tag{7}$$

$$d_\kappa = |\kappa_R - \kappa_T| \tag{8}$$

$$d_s = |s_R - s_T| \tag{9}$$

where $\sigma_R$, $\kappa_R$, $s_R$ and $\sigma_T$, $\kappa_T$, $s_T$ are the standard deviation, kurtosis and skewness values computed from the reference and test image, respectively. The metric $D_2$ has five parameters, $\delta_1$, $\delta_2$, $\delta_3$, $\delta_4$ and $C_0$, which should be learned from the data.

The third metric used in this dissertation was proposed by Engelke et al. [20]. It is based on five features $f_i$. Feature $f_1$ measures blockiness and uses

the NR metric [117]. Feature $f_2$ measures blurriness and uses the NR-type edge width metric [71]. Features $f_3$ and $f_4$ measure the ringing distortion and use the NR-type image activity metrics [27]. Feature $f_5$ measures the intensity masking and lost block distortion by using image histograms. The image quality $D_3$ is calculated by Equation (10):

$$D_3 = \left[ \sum_{i=1}^{5} k_i^p \left| f_{i,R} - f_{i,T} \right|^p \right] \tag{10}$$

where $k_i$ is the weighting factor, $f_{i,R}$ is the value of feature i in the reference image and $f_{i,T}$ is the value of feature i in the test image. The metric $D_3$ has one parameter, $p$, which should be learned from the data.

We implemented the Wang et al. metric [120] into the proposed image quality measurement framework because it is a well-known metric in the RR research field, and its algorithm code was available from [118]. The Li and Wang metric [61] was selected because it is the state-of-the-art metric. This metric determines the non-linear weighting of the wavelet coefficients before the KLD. The metric was implemented by adding the non-linear weighting component in the Wang et al. metric. The Engelke et al. metric [20] was selected because it uses the five distinct NR-type image features instead of the global statistic approach. These features were either easy to implement, or their algorithm codes were available [116].

**Color contrast metric**

Publication I proposed a metric for measuring the color contrast of images. The proposed metric assumes that natural images contain one or more objects with specific hues. The object is perceived as more colorful and brighter if the contrast of its texture is high. If the contrast of the texture is low, the object is perceived to be pale or dim. The idea behind the metric is to redirect the axes of the color space to the direction of an image's dominant color and to calculate the standard deviation of the color points.

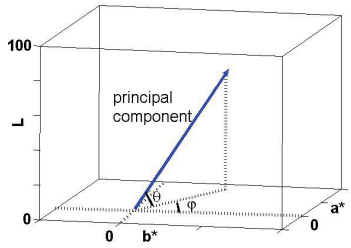
The proposed metric computes the first principal component for the reference image data in the CIELAB space, and two color clusters are defined. The first principal component shows the direction of the maximum deviation of the color data in the color space, and the hue of the dominant color can be described. The first color cluster includes the neighborhood points of the principal component. The second color cluster includes all of the other points. The neighborhood points of the first principal component should fulfill the following conditions:

$\theta - \theta_{th} < \theta_{point} < \theta + \theta_{th}$ and $\varphi - \varphi_{th} < \varphi < \varphi - \varphi_{th}$,

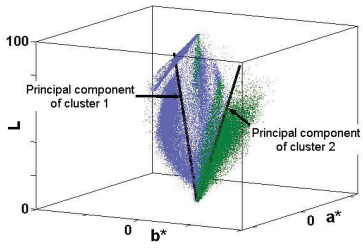
where $\theta$ and $\varphi$ are the angles of the principal component and $\theta_{th}$ and $\varphi_{th}$ are the threshold angles (Figure 17b). The values of the threshold angles were chosen based on the empirical study. Next, the principal component is also calculated for color cluster 2. The data of the two principal components define the feature vector.

The coordinates of the calculated principal components describe the directions of an image's two main color hues. For example, for the image of the lake (Figure 17a), the principal component of the first cluster describes the color contrast of the blue sea and sky, and the principal component of the second cluster describes the color contrast of the green foliage (Figure 17c).

(a)




(b)




(c)

**Figure 17.** Test image "lake" (a), the points of the first cluster are the neighborhood points of the first principal component (b), the color contrast metric is based on the standard deviations calculated for the directions of two principal components (c).

The "computation of image quality" component of the image quality measurement framework (Figure 15) transforms the pixel values of a test image into the new space defined by the feature vector. The axes of the new space are composed of the principal components calculated in the "characterization of image" component. The color contrast metric CC is calculated by Equation (11):

$$CC = \sqrt{\sigma_{p1}^2 + \sigma_{p2}^2} \tag{11}$$

where $\sigma_{p1}$ and $\sigma_{p2}$ are the standard deviations along the first and second dimensions of the new space, respectively.

### 3.2.2 Local feature characterization

To compute the metrics defined in this sub-section, we divide the images into adjacent blocks, and the block-specific sensitivity values are calculated (see Figure 16b). The value of the feature vector is derived from the sensitivity values.

Publication I proposed the graininess and sharpness metrics by using the principle of local feature characterization. The block-specific energy values $g$ are computed from the reference image by Equation (12) for the graininess metric, and the parameter $\beta$ values are computed by Equation (13) for the sharpness metric:

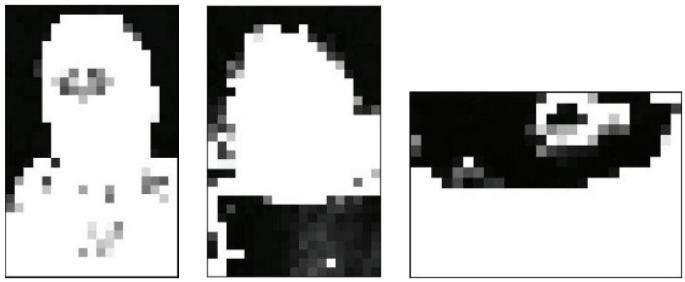
$$g = \frac{1}{M}\sum \left| w_i^p \right| \quad (12),$$

$$p(w;\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|w_i|/\alpha)^\beta} \quad (13).$$

Equation (12) is power function and $w_i$ is the wavelet coefficient in the block. $M$ is the number of wavelet coefficients in the block. Equation (13) is the GGD model. $\Gamma$ is the gamma function and $\alpha$ and $\beta$ are the parameters of the GGD.

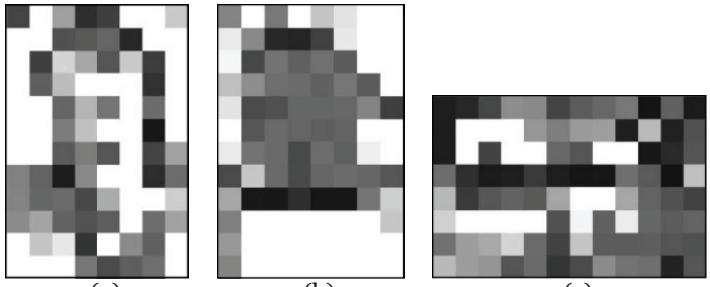
The energy values $g$ are calculated for the wavelet coefficients of the first scale and the $\beta$ values of the second scale. We used the first scale for $g$ because we assumed that the perceived graininess is high-frequency energy. We computed the parameter $\beta$ values for the second scale to compensate for the graininess energy of the first scale. The assumption is that the perceived sharpness relates to the reproduction of mid-frequency energy. Because the metric uses the second scale and handles only the mid-frequency energy, it is non-sensitive to high-frequency graininess.

Figure 18 shows the $g$ values, and Figure 19 shows the $\beta$ values for the following test images: man (a), cactus (b) and lake (c). If the intensity of the block is low (dark blocks), the block is well suited for the metrics. A low $g$ value refers to a low energy block that is appropriate for the graininess metric, and a low $\beta$ value refers to a high-energy block that is appropriate for the sharpness metric.



(a) (b) (c)

**Figure 18.** Graininess is calculated from the low-energy pixel blocks (dark regions): test images of man (a), cactus (b) and lake (c) shown in Figure 12. The low-energy block indicates a smooth area that is appropriate for graininess measurements.



(a) (b) (c)

**Figure 19.** Sharpness is calculated from the high-energy blocks (dark regions): test images of man (a), cactus (b) and lake (c) shown in Figure 12. The high-energy block indicates a texture area that is appropriate for sharpness measurements.

Figure 20a shows the $g$ values and Figure 20b the $\beta$ values of the blocks for the reference image and a set of test images sorted in ascending order. The graininess and sharpness of the test images are calculated from the n blocks, whose values are small. The feature vector includes the value of n calculated from the reference image. The n for the graininess metric is the point where the value of $g$ for the reference image starts to increase. The n for the sharpness metric is the point where $\beta = 0.5$. The value of $\beta$ is based on an empirical study (Publication I).

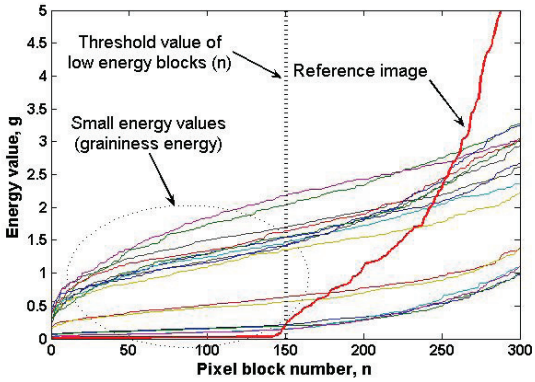
The "computation of image quality" component of the image quality measurement framework (Fig 15) calculates the graininess value G by using Equation (14) and the sharpness value $S_1$ by using Equation (15):

$$G = \frac{\sum_{i=1}^{n} g_{hi} + g_{d1i} + g_{vi} + g_{d2i}}{4} \tag{14}$$

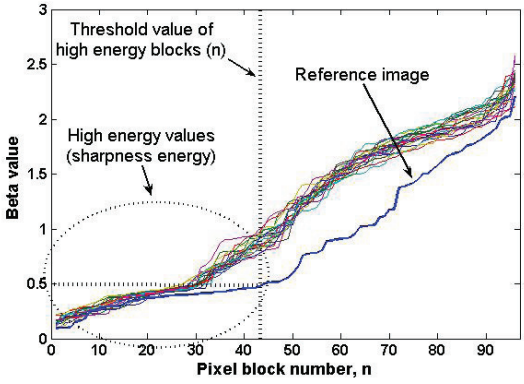
$$S_1 = \frac{\sum_{i=1}^{n} \beta_{hi} + \beta_{d1i} + \beta_{vi} + \beta_{d2i}}{4} \tag{15}$$

where $g_{hi}$, $g_{d1i}$, $g_{vi}$ and $g_{d2i}$ are the $g$ values in the horizontal, first diagonal, vertical and second diagonal directions from block i of the group of n

lowest $g$-valued blocks. $\beta_{hi}$, $\beta_{d1i}$, $\beta_{vi}$ and $\beta_{d2i}$ are the $\beta$ values in the horizontal, first diagonal, vertical and second diagonal directions from block i of the group of n lowest $\beta$-valued blocks.




(a)




(b)

**Figure 20.** The number of low energy blocks for the graininess metric is the point where the value of $g$ for the reference image starts to increase (a); the number of high-energy blocks for the sharpness metric is the point of the empirically determined threshold ($\beta$ = 0.5) (b).

### 3.2.3 Adaptive local feature characterization

The metrics defined in this sub-section seek new positions (candidate blocks) for the predetermined blocks by maximizing the sensitivity function (see Figure 16c). The blocks to be measured are selected from the candidate blocks.

#### Color noise metric

Publication III presented a color noise metric by using the adaptive local feature principle. The blocks are selected based on three features: the chromatic energy, achromatic energy and brightness of the block. The chromatic energy of the blocks should be low. The blocks can have achromatic structural energy, but this structure should be composed

primarily of random textures rather than edges. Random textures in a scene can be beneficial to noise measurements for two reasons. The first reason is that achromatic texture-like surfaces in scenes are sensitive to color noise in digital camera images. The second and more important reason is that texture-like surfaces present challenges for noise reduction algorithms in cameras. If the structure is edge-like, then a noise reduction method can easily filter the noise away from the neighboring smooth area of the edges. If the structure is a random texture, then it is difficult to separate the noise energy from the energy of the image structure by using computational methods.

In addition, the intensity of the selected blocks should not be too low or high. If a block is too bright, then it becomes saturated for images produced by low-end cameras. If the block is too dark, then a low-end camera may not detect its structural energy, and the camera image-processing software may apply strong noise reduction to the block.

We applied the method in the YCbCr space. With an opponent color space, we can separate achromatic information from chromatic information. The method operates on the principle that the blocks are initially located in the reference image (Figure 21a). The method searches for new locations for the blocks across a limited neighborhood in the Cb and Cr channels by maximizing Equation (16):

$$(CP_y, CP_x) = \arg\max_{CP_y, CP_x} (CO_E) \quad when \quad dist(IP, CP) < T \tag{16}$$

$$CO_E = \sum_{a,b} P^2_{\phi,d}(I_1, I_2) \tag{17}$$

where the co-occurrence energy feature $CO_E$ is calculated within the block. $P_{\phi d}(I_1,I_2)$ describes the probability that two pixels with intensity levels $I_1$ and $I_2$ appear in the window separated by a distance $d$ in direction $\phi$. The more homogeneous the block is, the higher the value of $CO_E$ on a scale from 0 to 1. The homogeneity metric $CO_E$ is calculated as an average of its values at 0, 45 and 90 degrees. The blocks in the new locations are called candidate blocks. The aim is to find the homogeneous areas in the chromatic Cb and Cr channels. The function *dist()* sets a distance constraint between the initial and candidate points (*IP* and *CP*). The *IP* are the center coordinates for the predetermined blocks (Figure 21a), and the *CP* are the center coordinates for the candidate blocks (Figure 21b).

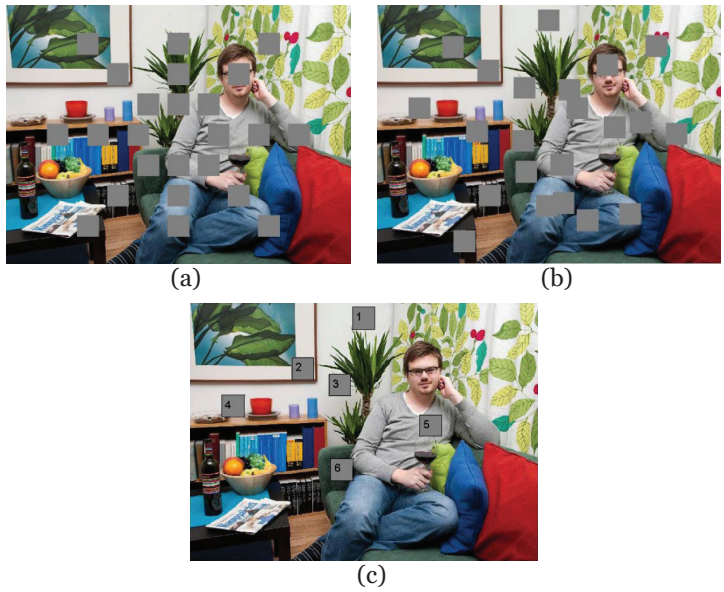



**Figure 21.** The reference image with the blocks in a predetermined symmetric order (a), the reference image with the blocks when the homogeneous metric was maximized for the Cb channel (b) and the reference image with the most homogeneous blocks (n = 6) for the Cb channel (c).

Next, the "characterization of image" component selects n blocks from the group of candidate blocks with the highest $CO_E$ values for the Cb and Cr components (Figure 21c shows the candidate blocks for the Cb component when n = 6). Equations (18) and (19) depict the texture and brightness, respectively, of the Y component from the selected candidate blocks:

$$CO_{idm}(\phi,d) = \sum_{a,b;a\neq b} \frac{P^{\lambda}_{\phi,d}(I_1,I_2)}{\left|I_1 - I_2\right|^{\kappa}} \tag{18}$$

$$B = \frac{1}{M}\sum Y_i \tag{19}$$

where $M$ is the number of pixels in block $i$. The co-occurrence feature $CO_{IDM}$ is calculated as an average of its values at 0, 45 and 90 degrees. The aim is to find both the smooth and textured achromatic areas. Feature $CO_{IDM}$ obtains a higher value on the scale ranging from 0 to 1 if the block pixel intensity values are close to each other. That is, if the $CO_{IDM}$ value is small, the intensity structure in the block is more texture-like than smooth and vice versa. The two blocks with the lowest and highest $CO_{IDM}$ values from the n blocks with $B$ values between the $B_{min}$ and $B_{max}$ levels are then chosen as the selected blocks.

The pixel locations of the selected blocks in the reference image are fed to the "computation of image quality" component. In the next step, the blocks that correspond to the selected blocks should be found in the test images.

These blocks are called correspondence blocks. The search for the correspondence area is described in detail in Section 3.4.

The "computation of quality metrics" component calculates the color noise values N for a test image by using Equation (20):

$$N = noise_s + noise_t \tag{20}$$

$$noise_s = w_{std,y} + k_1 w_{std,cb,h} + k_2 w_{std,cr,h} \tag{21}$$

$$noise_t = k_1 w_{std,cb,l} + k_2 w_{std,cr,l} \tag{22}$$

where $noise_s$ is the smooth area component, $noise_t$ is the texture area component and $k_i$ represent the weighting factors. $w_{std,cb,l}$ and $w_{std,cr,l}$ are the standard deviations of the wavelet coefficients for the Cb and Cr blocks with the lowest $CO_{IDM}$ values (texture blocks), and $w_{std,y}$, $w_{std,cb,h}$ and $w_{std,cr,h}$ are the standard deviations of the wavelet coefficients for the block with the highest $CO_{IDM}$ value (smooth blocks). The wavelet coefficients for the Cb and Cr components are calculated from the second scale of the wavelet decomposition, and the wavelet coefficients for the Y component are calculated from the first scale.

Figure 22 summarizes the scheme of the proposed metric. First, the candidate blocks that maximize the homogeneity metric are sought for the Cb and Cr channels. Next, the smooth and textured blocks are selected from the candidate blocks based on the texture and brightness metrics applied to the Y channel. The correspondence blocks of the reference camera are searched from the test camera images and the texture and smooth area noise components are calculated.

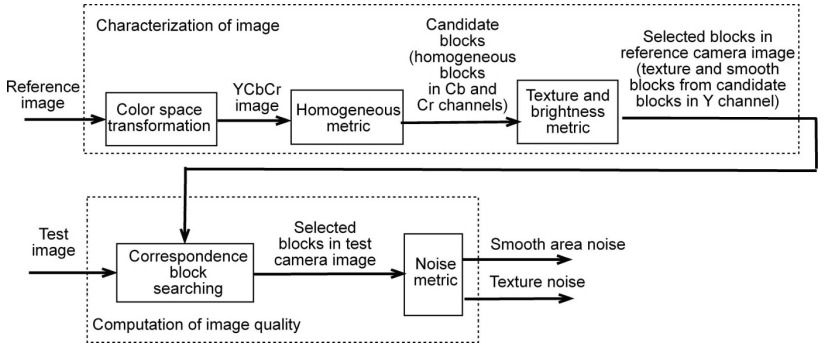



**Figure 22.** Block diagram for the proposed color noise metric. The dashed boxes show the components of the image quality measurement framework.

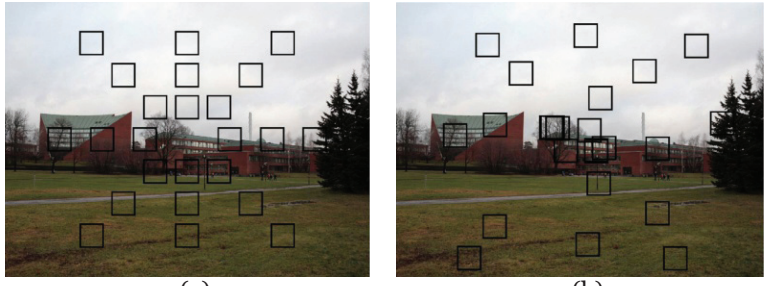
**Sharpness metric**

Publication V presented the sharpness metric using the adaptive local feature principle. First, the candidate blocks are located by Equation (23):

$$(CP_y, CP_x) = \arg\max_{CP_y, CP_x}(STD) \quad when \quad dist(IP, CP) < T \tag{23},$$

$$STD = \frac{1}{M} \sum_{CP_y - \sqrt{M}/2 < j}^{j < CP_y + \sqrt{M}/2} \sum_{CP_x - \sqrt{M}/2 < k}^{k < CP_x + \sqrt{M}/2} w_{j,k}^2 \tag{24}.$$

Equation (23) maximizes the standard deviation (*STD)* of the wavelet coefficients $w$ within the block. $M$ is the number of pixels in block. Figure 23a shows the *IP* points for the local characterizations, and Figure 23b shows the locations of the *CP* points. The "characterization of image" component selects the m highest valued candidate blocks by using Equation (24). Figure 24 shows the selected blocks for five image contents when m = 5. The effects of block size and the number of m were studied in Publication V. The performance was highest when the block size was 100 x 100 pixels and the number of m ranged from 5 to 8.



(a) (b)

**Figure 23.** The reference image with the blocks in a predetermined symmetric order (a); the reference image with blocks when the sharpness function for the blocks is maximized (b).

**Figure 24.** Five reference images with the selected blocks (m = 5).

The locations of the selected blocks are fed to the "computation of image quality" component. The correspondence areas between the reference and test images are searched. The sharpness value $S_2$ for a test image is computed by Equation (25):

$$S_2 = \sum_{i=1}^{m} \frac{1}{(\sqrt{M}-b)^2} \sum_{j>b/2}^{j<(\sqrt{M}-b/2)} \sum_{k>b/2}^{k<(\sqrt{M}-b/2)} w_{i,j,k}^2 \tag{25}$$

where the ($j$, $k$) are the pixel coordinates in a correspondence block, $M$ is the size of the correspondence block, $b$ is a parameter for the reduced measurement area and $w_{i,j,k}$ is the wavelet coefficient. The parameter $b$ is used to compensate for the fact that the edge areas of the correspondence blocks can include structures from outside of the original candidate block's area. If the candidate block's size is $M$ pixels, the measurement area in the test camera image is $(M^{0,5}\text{-}b)^2$ pixels.

Figure 25 summarizes the scheme of the proposed sharpness metric. First, the candidate blocks that maximize the standard deviation of the wavelet coefficients are sought for the Y component of the reference image. Next, the m highest energy blocks are chosen for the group of selected blocks. The sharpness values are calculated from the correspondence blocks of the test camera images.

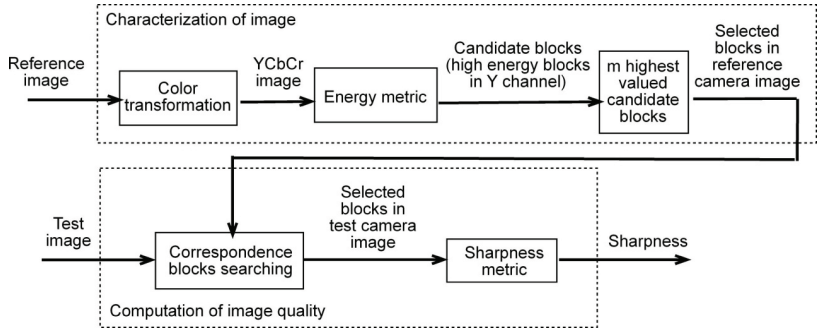



**Figure 25.** Block diagram for the proposed sharpness metric. The dotted boxes show the components of the image quality measurement framework.

**Color difference metric**

Publication VI presented color difference metrics by using the adaptive local feature principle. The color difference values of cameras should be measured using a color sample set that is as extensive as possible. The method uses a two-phase process to select the blocks from a scene captured by a color-calibrated reference camera for the sample set. First, the method seeks the candidate blocks by maximizing the average chroma and minimizing the standard deviation of the hue angle when (initially located) blocks are moved within a limited neighborhood by using Equation (26):

$$(CP_i, CP_j) = \arg\max_{CP_i, CP_j} \left[ k \cdot C_{ave,ij} + \frac{(1-k)}{H_{std,ij}} \right], \quad when \quad dist(IP, CP) < T \tag{26}$$

where $C_{ave}$ is mean chroma and $H_{std}$ is the standard deviation of the hue angle in the block. The constant $k$ is a weighting factor. Figure 26a shows an example of initial blocks, and Figure 26c shows an example of candidate blocks.

Second, m measure blocks are selected from the candidate blocks. The measure blocks are selected from the candidate blocks by using an iterative search. First, a histogram of m equal-width bins is formed for the average hue angle values of the candidate blocks. The next iteration rounds decrease the bin width until m bins have at least one data point each. The term “data point” refers to an average chroma and hue value pair of a candidate block. From the bins containing more than one data point, the point with the highest chroma value is selected.

Figure 27 shows an example in which the chroma values of the candidate blocks are shown as a function of the hue angle values. The selected blocks (m = 18) are from different bins, and their chroma values are the highest ones in the bin. Figure 26e shows the selected blocks on a scene.

Figures 26b and 26d show how the candidate block selection spreads the color values on the ab-plane of the CIELAB color space because the chroma values of the blocks are maximized by Equation (26). Figures 26d and 26f show how the iterative search decreases the number of data points while maintaining an extensive color value set.

Next, the coordinate values of the selected blocks are fed to the “computation of image quality” component, the corresponding blocks are located (as described in Section 3.3) and the color difference values are calculated by comparing the color values of the reference and test images.

The results for the luminance ΔL, hue ΔH, and chroma ΔC difference metrics were presented in Publication VI.

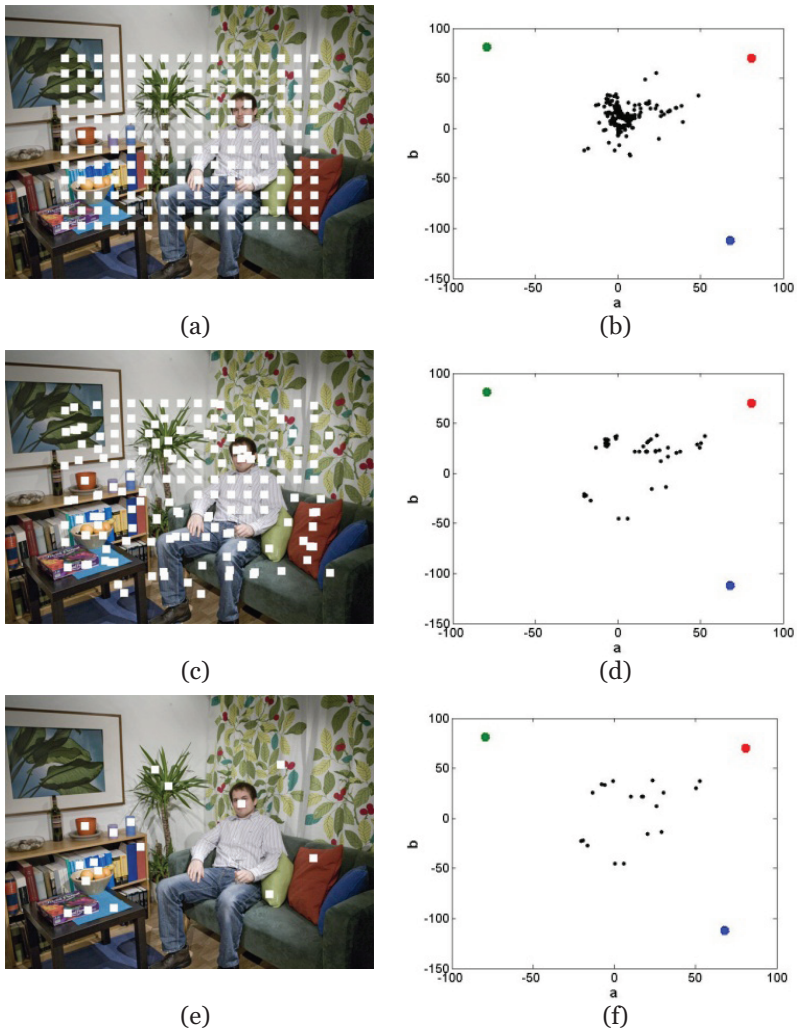



**Figure 26.** The pixel blocks are initially located on a regular grid arrangement (a). The initial candidate blocks are searched for within a limited neighborhood (c). *m* measuring blocks are selected from within the initial candidate areas (e). The color values of the initial blocks, initial candidate blocks and selected measuring blocks are shown on ab-planes (b), (d) and (f), respectively.

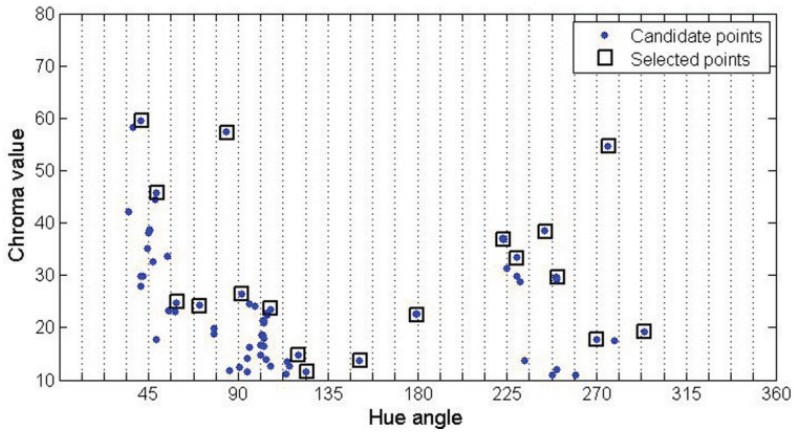



**Figure 27.** Chroma values of the candidate blocks are shown as a function of the hue angle values. The candidate block with the highest chroma value in the hue angle bin is selected for the sample set of the color measurement.

Figure 28 summarizes the scheme of the proposed color metric. First, the candidate blocks that maximize the average chroma and that minimize the standard deviation of the hue are sought using the CIELAB color space. Next, the most extensive color value set as possible (selected blocks) is searched using a method based on the hue value histogram. The color difference values are calculated by comparing the color values of the measuring blocks between the reference and test camera images.

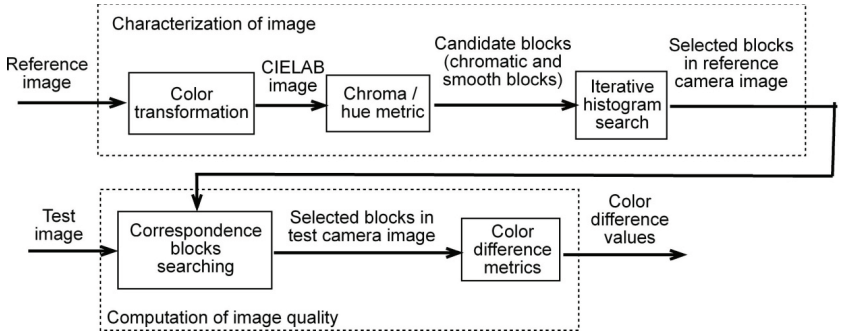



**Figure 28.** Block diagram for the proposed color difference metric. The dotted boxes show the components of the image quality measurement framework.

## 3.3 Search for correspondence blocks

If the adaptive local feature principle (described in Section 3.2.3) is used to characterize the reference image, the corresponding blocks should be found in the test images. In the schemes for the proposed methods (Figures 22, 25 and 28) the component of the correspondence blocks searching shows the point of which the searching process is done. Figure

29 shows an example where three example measuring areas were found in the reference image (left) and the corresponding areas were located in a test image (right).

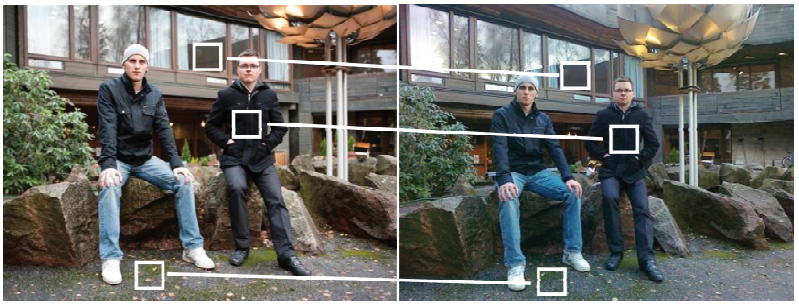


**Figure 29.** The measuring areas from the reference image (left) and the corresponding areas from the test image (right).

Locating the corresponding areas from the camera images is not a straightforward process. Figure 30 shows an example where the pixel regions were cropped from the images produced by different cameras. The images were captured such that they are as similar as possible. The image regions were cropped using the same pixel coordinates for each image. Clearly, searching for correspondence blocks while using only pixel-coordinate values does not work.

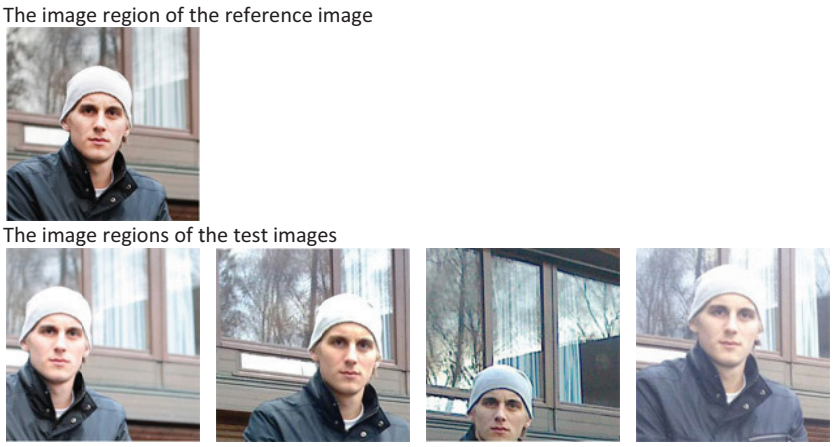



**Figure 30.** The image regions of the reference and the four test images are cropped using the same pixel coordinates. It is apparent that searching the correspondence blocks by using only the pixel-coordinate values generates inaccurate results.

Publications III, V and VI proposed a method to search for correspondence blocks by using local area descriptors (e.g., SIFT [67] (III, V) or SURF [29] (VI)). Figure 31 shows an example where correspondence points (n = 8) are used for searching the center point of measuring block in

a test camera image. In the first step, the correspondence points are located from an image pair of the reference and test cameras by matching the area descriptors. Next, the vectors starting from the $n$ nearest correspondence points are directed to the center point of the measuring block on the reference image. The center point in the test camera image is estimated by calculating the angles and lengths of the vectors between the center point and the correspondence points in the reference image. The center point of a measuring block in the test camera image is the average of the vector endpoints from the correspondence points.

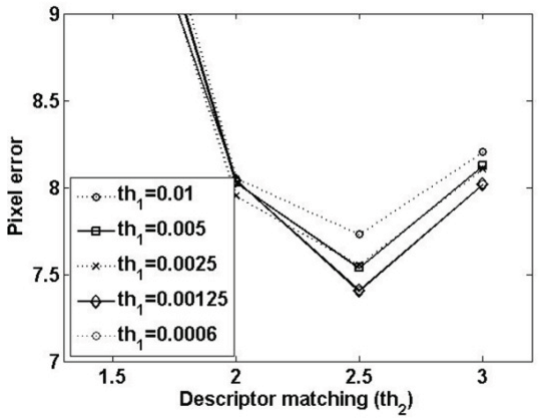
(a) (b)

**Figure 31.** The center points of the correspondence areas from the reference image (a) and test image (b) are approximated by calculating the mean point of the vector heads from the correspondence feature points. The lines denote the vectors whose lengths and directions are calculated from the reference image. The points denote the correspondence-feature points. The crosses denote the correspondence-area centers.

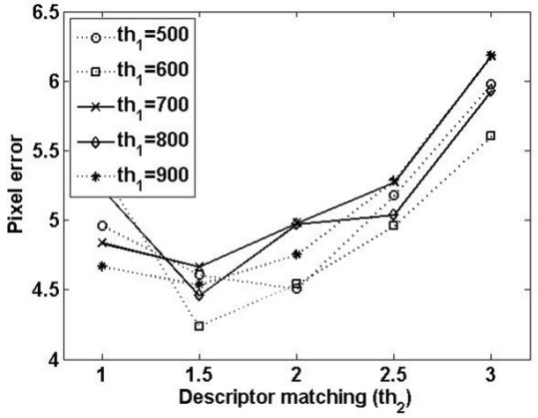
Publication VI presented the performance comparison between SIFT and SURF when applied in the camera image measurement framework. The performance metric was the pixel distance between the ground truth and the predicted center point of the measuring block. The ground truth data were collected manually by utilizing the "cpselect" function in MATLAB. The prediction error was studied as a function of the threshold values of th1 (interest point detection) and th2 (descriptor matching). The remaining parameter values of the implemented SIFT and SURF codes were set at their default values.

For the SIFT, the threshold of the interest point is related to the contrast values of key points. For the SURF, the threshold of an interest point is related to the value of the Hessian's determinant. For the SIFT, a higher value yields fewer points, whereas for the SURF, a higher value gives more points. The threshold value of the descriptor matching relates to the distance between descriptors D1 and D2. The nearest-neighbor distance-ratio matching strategy was used for both methods. A descriptor D1 is matched to a descriptor D2 only if the distance dist(D1, D2) between them is smaller than the distance of D1 to all of the other descriptors divided by the threshold.

Figures 32a and 32b show the average pixel errors for the SIFT and SURF, respectively, as functions of th1 and th2. The pixel error is the average value for the scene in Figure 26, which was captured by eight different cameras. Six cameras were mobile phone cameras, one was a consumer compact camera and one was an SLR camera. The reference camera was a Canon 5D with a Canon EF 24-70/2.8 L USM lens. Before the pixel error calculations, the images were all scaled to a size of 1200 × 900 pixels.




(a)




(b)

**Figure 32.** Average pixel error values between the predicted and ground truth as a function of the interest point and descriptor-matching threshold values for SIFT (a) and SURF (b).

The lowest average pixel error for the SIFT was 7.4 pixels (th1 = 0.00125 and th2 = 2.5), with a standard deviation of 1.5 pixels between the cameras. The lowest average pixel error for the SURF was 4.2 pixels (th1 = 600 and th2 = 1.5), with a standard deviation of 1.0 pixels. The sizes of the measuring areas in the studies described in Section 4 ranged from 25 × 25 pixels to 125 x 125 pixels. The image size was 1200 × 900 or 1600 x 1200

pixels. According to these dimensions, the pixel errors of 4.2 or 7.4 pixels were small, and the locations of the measuring blocks on the test images were sufficiently accurate.

### 3.4 Digitization of print

Before the quality of printed (/displayed) images can be measured using an algorithmic metric, the test samples should be digitized. Eerola [17] and Pedersen [87] used an ICC-profiled scanner to digitize their samples. The digitization process proposed in Publications I and II uses a high-quality camera system instead of a scanner to digitize a sample. Digital camera RAW-imaging enables fully manual settings. In addition, the lighting environment can be adjusted and characterized, which is impossible or difficult to do with a scanner.

Figure 33 shows the structure of the "digitization of print" component. The component captures n exposures, compensates for the photometric distortion of the camera, makes color space transformation from camera RGB to the CIELAB color space by using the colorimetric (Publication I) or spectral-based (Publication II) camera characterization and forms the high dynamic Lab image.

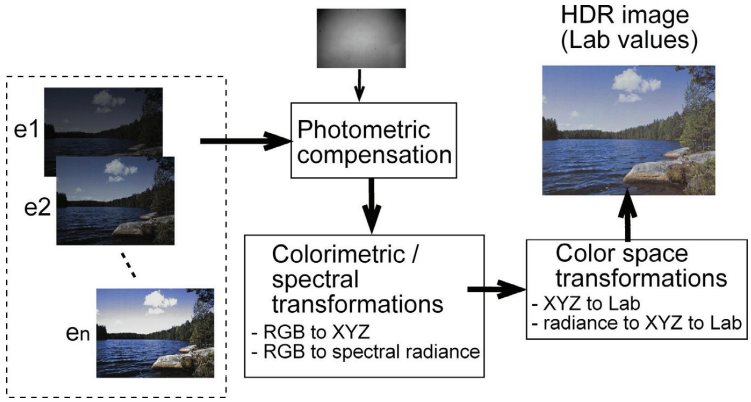



**Figure 33.** Sample digitization component captures n exposures, compensates for the photometric distortion of the camera, makes color space transformation from camera RGB to the CIELAB color space by using colorimetric or spectral-based camera characterizations and forms the high dynamic Lab image.

The study reported in Publication I used the colorimetric camera characterization. The transformation from multi-exposure RGB image to XYZ image was based on the 3x5 transformation matrix **W**:

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \end{bmatrix} \cdot \begin{bmatrix} r_i \\ g_i \\ b_i \\ r_i g_i b_i \\ 1 \end{bmatrix} \tag{27}$$

where [$x$ $y$ $z$] represents absolute XYZ values, [$r$ $g$ $b$] represents the corresponding camera linear raw response and $p_{ij}$ are the fitting parameters for the transformation matrix.

Publication II proposed a multi-exposure method for spectrally characterizing the camera. The transformation matrixes were calculated based on the method proposed by Shen and Xin [102]. They assumed that the training samples $\mathbf{u}_i$ that are closer to a testing sample $\mathbf{u}$ are usually more reliable and should thus contribute more to the estimation of the transformation matrix $\mathbf{W}_{\mathbf{shen}}$. The researchers calculated the weights $\alpha_i$ for $\mathbf{u}_i$ as:

$$\alpha_i = (2\pi)^{-3/2} \left|\Sigma_{UU}\right|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{u}_i - \mathbf{u})^T \Sigma_{UU}^{-1} (\mathbf{u}_i - \mathbf{u})\right] \tag{28}$$

where $\mathbf{\Sigma}_{UU}$ is the covariance matrix of $\mathbf{u}_i$. By incorporating the weighting, the mean square error between the measured and the predicted spectra can be formulated as:

$$J = \frac{1}{L}\sum_{i=1}^{L} \left\| \mathbf{W}_{shen}\alpha_i \mathbf{u}_i - \alpha_i \mathbf{r}_i \right\|^2 \tag{29}$$

and the transformation matrix $\mathbf{W}_{\mathbf{shen}}$ can be estimated.

To evaluate the performance of the multi-exposure method, we compared it to the traditional single-exposure method. The 180 colour patches of the Gretag Macbeth DC test target were used as training samples, and the 24 colour patches of the Gretag Macbeth CC test target were used as testing samples. The digital test target images were printed on six paper grades (p1-p6). We photographed the samples by using a Canon EOS 5D camera. The multi-exposure method produced an RGB image by selecting the intensity values from the exposures $e_k$ and $e_l$ ($1 \le k,l \le n$). The single-exposure method used only exposure $e_l$ to produce the RGB image. Exposure $e_l$ was the optimal value of the lighting environment computed by the camera processing.

We measured the ground truth spectral data of the printed samples by using the Photo Research PR-670 spectroradiometer. The measured values of the DC test target were used to calculate the weighting factors $\alpha_i$ for Equation (29) when the CC test target patch-specific transformation matrixes were estimated.

Figure 34a shows the CIEDE2000 mean color error values of different paper grades using single-exposure and multi-exposure methods for the testing samples. Figure 34b shows the maximum color error values of the different paper grades. Figure 35 shows the CIEDE2000 color error values for paper p3 ordered in ascending order by the measured luminance of the patches. According to the results, the multi-exposure method improves mostly the reconstruction performance of the dark patches. This result was expected. The multi-exposure method detects the lower luminance levels more linearly than the single-exposure method does.

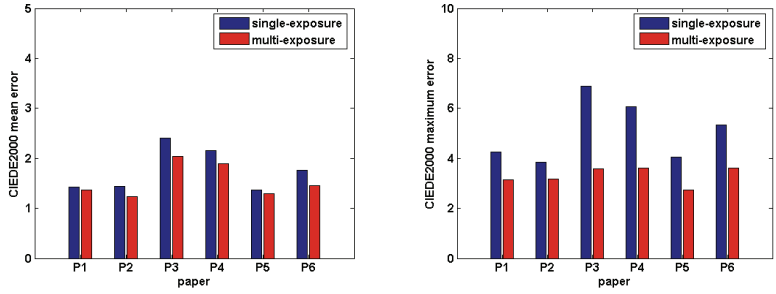



**Figure 34.** CIEDE2000 mean (a) and maximum (b) colour error values of different paper grades using single-exposure and multi-exposure methods for the testing samples.

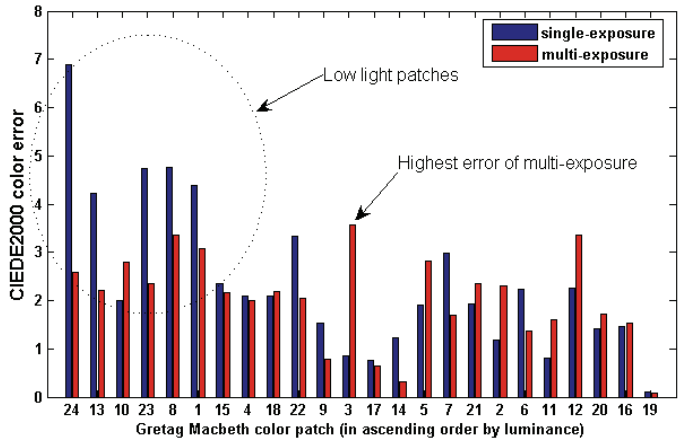



**Figure 35.** CIEDE2000 colour error values for paper p3 ordered in ascending order by measured luminance of patches.

# 4. Experiments

The experimental results in Publications I-VI fall into two groups. The overall quality of the images was measured in Publications I and IV and the quality attributes were measured in Publications III, V and VI. Publication I presented the simple quality model of printed images, which is composed of the color contrast, graininess and sharpness attributes (CC, G and $S_1$). Publication IV presented the results when three RR metrics of overall quality ($D_1$, $D_2$ and $D_3$) were implemented into the image quality measurement framework (Figure 15). Publications III, V and VI determined the quality attributes (N, $\Delta L$, $\Delta C$, $\Delta H$ and $S_2$) of images captured by digital cameras.

The experimental results for the overall quality measurements and for the quality attributes are reviewed in Sections 4.2 and 4.3, respectively. Section 4.1.1 and Section 4.1.2 present and analyze the subjective data for the printed samples and camera samples, respectively. Table 6 lists the metrics and shows the grouping.

**Table 6.** The experimental results of the dissertation are presented in two sections: Section 4.2 presents the results for the overall quality metrics, and Section 4.3 presents the results for the attribute metrics. Section 4.1 presents the subjective data

| Attribute | Section | Publication | Metrics | Application |
|---|---|---|---|---|
| Overall quality | 4.2 | I | CC, Equation (11)<br>G, Equation (14)<br>$S_1$, Equation (15) | Printed image |
| | | IV | $D_1$, Equation (3)<br>$D_2$, Equation (6)<br>$D_3$, Equation (10) | Camera |
| Noise | 4.3 | III | N, Equation (20) | Camera |
| Sharpness | | V | $S_2$, Equation (25) | Camera |
| Color difference | | VI | $\Delta L$, $\Delta C$, $\Delta H$ | Camera |

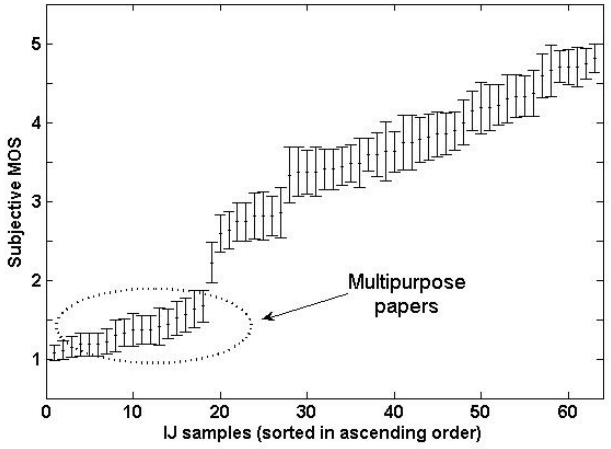
## 4.1 Subjective experimental data

### 4.1.1 Subjective quality data for printed images

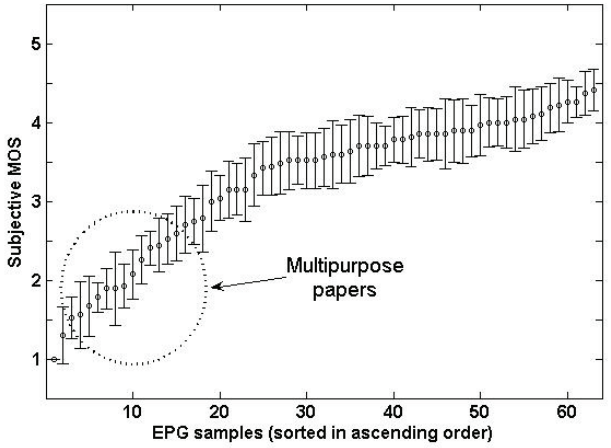
In this dissertation, the image contents of Figure 12 were used to test the proposed quality attribute metrics in the context of digitally printed images. The images include small details, uniform areas and colors. The variation in printed images arose from the use of different grades of paper (fifteen electrophotographic (EPG), six multipurpose and fifteen ink-jet (IJ) papers). The size of the printed images was 10 cm x 15 cm. The EPG papers were printed using an electrophotographic printer, the IJ papers were printed by an ink-jet printer and the multipurpose papers were printed with both methods. For more details on the printing process, see [30].

The subjective tests consisted of a quality evaluation task and self-report using the IBQ (Interpretation Based Quality) method [91]. The observers (n = 27) were university students and naïve as regards to image quality. The image samples were presented on a table covered with a gray tablecloth. The illuminance level was 2200 lux, and the color temperature was 5000 K. The quality evaluation task provided information about the experienced quality of the images (mean opinion score values, MOS). The observers were asked to select the best sample (IQ = 5) and the worst sample (IQ = 1). Afterwards, the observers rated the samples on a scale from 1 to 5. The purpose of the self-report was to obtain information about the relevant subjective attributes that influenced the visual quality evaluation. In practice, the observers were asked to provide the reasons behind their evaluations. For more details on the subjective tests, see [32].

We used the self-report data to select the quality attributes for the printed image quality measurements of this dissertation. The MOS data was used to test the performance of the proposed metrics. Figure 36 shows the MOS values for the IJ and EPG samples sorted in ascending order. The error bars added in the figures show the 95 % confidence intervals. The MOS data of the multipurpose papers fall within the dashed circles. The quality of the multipurpose papers was lower than the quality of the dedicated IJ or EPG papers. In addition, the quality differences between the multipurpose and dedicated IJ papers (Fig. 36a) were higher than the quality differences between the multipurpose and dedicated EPG papers (Fig. 36b). Two or three compact clusters can be distinguished in the subjective MOS data of the IJ samples. In contrast, the data of the EPG samples are more coherent.




(a)




(b)

**Figure 36.** The subjective MOS values for the IJ (a) and EPG (b) samples sorted in ascending order.

Table 7 presents the ten image quality attributes most frequently used to describe the printed samples. Sharp, unsharp and grainy were the three most used attributes for the IJ samples, regardless of the image content. These attributes were also the most frequently used for the EPG samples, but there were some differences between the image contents. With the test image of the cactus (of the EPG samples), the attribute “sharp” was the fourth most used, and the attribute “unsharp” was not in the group of the ten most used attributes. In contrast, with the test images of the man and the lake, “sharp” was the most frequently used attribute.

**Table 7.** The ten subjective attributes most frequently used to describe the printed samples; the data represent the number of times an attribute was mentioned.

| Man | | | | Cactus | | | | Lake | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| EPG | | IJ | | EPG | | IJ | | EPG | | IJ | |
| Sharp | 97 | Unsharp | 109 | Grainy | 90 | Grainy | 114 | Sharp | 97 | Sharp | 121 |
| Grainy | 85 | Sharp | 107 | Faded | 83 | Sharp | 110 | Unsharp | 81 | Unsharp | 116 |
| Unsharp | 67 | Grainy | 98 | Good colors | 83 | Unsharp | 94 | Good colors | 71 | Grainy | 92 |
| Good colors | 62 | Faded colors | 95 | Sharp | 62 | Faded colors | 83 | Grainy | 64 | Faded colors | 88 |
| Faded colors | 62 | Faded | 82 | Faded colors | 50 | Clear | 76 | Streaking | 63 | Unclear | 83 |
| Clear | 50 | Unclear | 81 | Dark | 47 | Deep colors | 75 | Faded colors | 63 | Good colors | 70 |
| White dots | 46 | Good colors | 65 | White dots | 44 | Unclear | 74 | Unclear | 40 | Deep colors | 61 |
| Streaking | 38 | Deep colors | 63 | Uniform print | 41 | Good colors | 73 | Not uniform print | 39 | Faded | 56 |
| Unclear | 35 | Clear | 54 | Clear | 41 | Faded | 63 | White dots | 39 | Clear | 48 |
| Faded | 34 | White dots | 46 | Gray | 40 | Contrast good | 39 | Deep colors | 39 | Matt | 35 |

With the IJ samples, the fourth most used attribute was “faded colors”, regardless of image content. With the EPG samples, “faded colors” was the fifth or sixth most used attribute. With the EPG samples, “good colors” was the third or fourth most frequently used attribute. With the IJ samples, “good colors” was in the group of the ten most used attributes. The third most frequently used color attribute was “deep colors”. In the case of the IJ samples, it was used relatively often.

The attribute “good colors” relates to the overall color quality. The other color attributes characterize more specific color properties. We assume that the attributes “deep color” and “faded color” comprise a bipolar dimension. Deep color is the positive pole and faded color is the negative pole of this dimension. In addition, we assume that the attribute “faded” has an effect on this dimension.

The subjective frequencies of the attributes suggest that the dimensions of sharpness, graininess and color contrast can be used to evaluate the quality of printed images. The dimension of sharpness is composed of the attributes “sharp” and “unsharp”. The dimension of graininess is composed of the attributes “grainy” and “white dots”. The dimension of color contrast is composed of the attributes “deep colors”, “faded colors” and “faded”.

The context of this space is the natural image, which is printed at the size of 10 cm x 15 cm. The proposed metrics for the dimensions were presented in Section 3.3.

### 4.1.2 Subjective quality data for camera images

We validated the objective metrics for camera applications proposed in this dissertation by using two Data sets. The test images for Data set I were captured in autumn (left-side images in Table 4), and those for Data set II were captured in winter (right-side images in Table 4). The most notable differences between the images of the data sets can be found in the outdoor Clusters 5 and 6. The differences between Clusters 1, 2 and 3 relate only to the persons in the images and their clothes. Cluster 4 is identical for the

two sets. Data set I was used to evaluate the performance of the proposed metrics described in Sections 4.2.2, 4.3.1, 4.3.2 and 4.3.3. Data set II was used for the study described in Section 4.3.1.

The images of both data sets were captured by test cameras and a high-quality reference camera. The quality levels of the test cameras ranged from low to moderate; the cameras consisted of low-, moderate- and high-quality mobile phone cameras and moderate-quality compact cameras. The pixel counts of the cameras ranged from 3 to 12 Mpix. The reference camera was a Canon EOS 5D with a Canon EF 24-80 mm lens. The performance (e.g., signal-to-noise ratio and detail reproduction) of the reference camera was considerably higher than that of the cameras to be tested. This difference was a required and sufficient condition.

The images were scaled to a 1600 x 1200 pixel size for the subjective tests. In addition, we added black borders around the images to match the image file resolution with the display resolution (1920 x 1200). The test setup included two Eizo ColorEdge CG241W displays and a small display. The test image was shown on one display, and the reference image (Data set I) or several reference images (Data set II) were shown on the other. The user interface of the observer was on the small display.

The observers first rated the overall quality of a test image before evaluating the values of the quality attributes. Test images representing one cluster at a time were shown. The order of the images and clusters were randomized for the observers. The viewing distance was approximately 80 cm, and the ambient illuminance was 20 lux. The displays were calibrated based on the sRGB standard.

Data set I included seventy-eight test images (13 cameras x 6 clusters), and Data set II had eighty-four test images (14 cameras x 6 clusters). All observers were naïve with respect to image quality (n = 25 for Data set I, and n = 30 for Data set II). With Data set I, the subjective reference image was shown on one display during the test, and the test images were shown on the other display. The image-quality value of the subjective reference image was set to be 90 on a scale ranging from 0 to 100. The subjective reference image functioned as a high-quality anchor image. The quality value of 90 out of 100 left some latitude for the observers in the case of high-quality test images. The chosen quality value of the anchor image is not critical.

With Data set II, before a single test image of a given content was shown, all of the test images of the clusters in question were shown to the observer as a slide show. This process was repeated before each test image was evaluated. In addition, the observers were instructed to give the lowest rating and highest rating in every cluster. The differences between the

subjective data in Data sets I and II are analyzed in greater detail in Publication V.

Figure 37 shows the MOS scores and the 95% confidence intervals of Data set I sorted in ascending order. We can observe that there are clear differences in the scales between the clusters. The scales for Clusters 1 and 3 are wider than the scales for the other clusters. This difference is attributable to the illuminance levels. The illuminance levels of Clusters 1 and 3 are low, and the quality differences between the images are clear. The impact of the flash power can be seen by comparing the values between Clusters 1 and 3. The shooting distance was longer for Cluster 3 than for Cluster 1. The mid-level MOS values are lower for Cluster 3 than for Cluster 1 because low-power LED flash cameras have an exposure power that is too low for longer distances. Cluster 4 shows that the image quality of a modern low-end camera saturates if the structure of the view is simple and the illuminance level is sufficiently high. The observers had difficulty seeing the differences between the images. The images for Cluster 4 were sharp, the noise level was low and the colors were balanced because of the easy content.

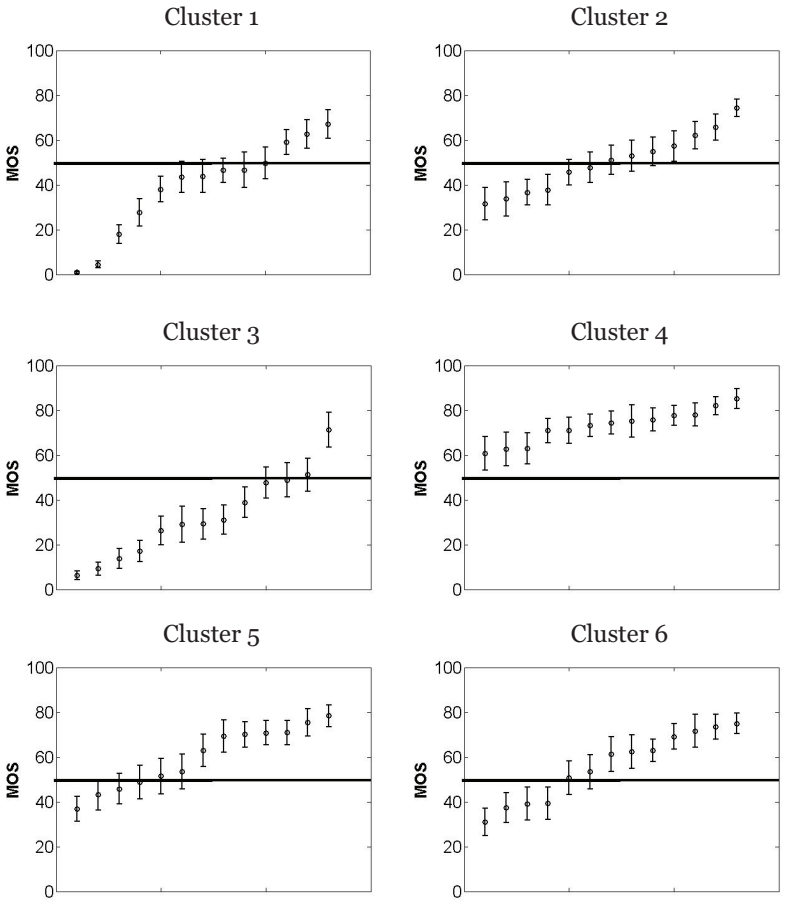



**Figure 37.** Subjective MOS on the vertical axis with 95% confidence intervals sorted in ascending order (on the horizontal axis) for clusters 1-6 with 13 cameras (Data set I).

## 4.2 Performance of objective methods for computing overall image quality

### 4.2.1 Camera images

Publication IV examined the overall quality of camera images from the standpoint of predicting subjective quality. The aim was to determine the benefits of a reference camera for camera image quality measurements. The six test image views (clusters) were photographed by different digital cameras (Table 4). Section 2.5.2 described the views, and Section 4.1.2 outlined the experimental procedure for gathering subjective data. More details about the subjective tests can be found in Publication IV. We implemented three state-of-the-art RR metrics for the image quality measurement framework depicted in Figure 15. The metrics $D_1$, $D_2$ and $D_3$ are defined by Equations (3), (6) and (10), respectively.

Before the analyses, we fitted the values of the metrics by using the three-parameter logistic function:

$$D_{pred} = p_1 - \frac{p_1}{1 + \exp(-p_2(x_i - p_3))} \quad (30)$$

where $p_i$ are the fitting parameters of the model, $D_{pred}$ is the predicted image quality and $x_i$ is the metric value for image i. The fitting parameters were obtained by calculating the minimum least-squares, non-linear regression using the fminsearch function in MATLAB.

We measured the performance as the Pearson linear correlation (LCC) and Spearman rank-ordered correlation (ROCC) between the computational metrics and the subjective data (MOS). Additionally, we used the outlier-ratio (OR) metrics. Table 8 shows the results. Boldface indicates the best performer. Based on the LCC and ROCC values, the performances of the $D_1$ [120] and $D_2$ [61] metrics were slightly better than the performance of the $D_3$ [20] metric. The image cluster-specific performance analyses can be found in Publication IV.

**Table 8.** Performance of the RR metrics when applied to the proposed camera measurement framework. Boldface indicates the best performer.

| Metric | LCC | ROCC | OR ($>2*\sigma_{mos}$) | OR ($>1*\sigma_{mos}$) |
|---|---|---|---|---|
| $D_1$, (Wang et al. [120]) | 0.8030 | **0.7982** | 0.0641 | **0.1539** |
| $D_2$, (Li & Wang [61]) | **0.8159** | 0.7916 | **0.0513** | 0.2179 |
| $D_3$, (Engelke et al. [20]) | 0.7753 | 0.7671 | **0.0513** | 0.3077 |

#### 4.2.2 Printed images

The aim of Publication I was to determine the applicability of a reference image to printed image quality measurements. In addition, the aim was to find the quality space of printed images. Section 4.1.1 described the experimental procedure for gathering the subjective data. The test images are shown in Figure 12. Based on the subjective interview data (Table 7), we developed metrics for the sharpness, graininess and color contrast attributes. Section 3.2 described the metrics in detail. The printed images were digitized using the method described in Section 3.4. Equation (27) was used for the camera characterization. Because the MOS values for the image contents were always scaled to the interval 1-5, we normalized the attribute metric values to a common scale [0-1]. For the sake of simplicity, the overall value of image quality IQ is predicted by a linear model:

$$IQ = \frac{k_1(1 - CC) + k_2 G + k_3 S_1}{3} \quad (31)$$

where $k_i$ are the weighting factors and CC, G and $S_1$ are the color contrast, graininess and sharpness values calculated by Equations (11), (14) and (15), respectively.

We estimated the weighting factors $k_i$ for the IJ and EPG samples by using the fmincon function in MATLAB to maximize the LCC between the MOS and IQ. For the IJ samples, the maximum LCC was 0.989 when $k_1$ = 1.90, $k_2$ = 1.07 and $k_3$ = 0.01. For the EPG samples, the maximum LCC was 0.906 when $k_1$ = 0.28, $k_2$ = 0.73 and $k_3$ = 1.95. The weighting factors show that the color contrast highly influenced the overall image quality of the IJ samples, whereas the sharpness metric highly influenced the overall image quality of the EPG samples.

However, if the weighting factors are set as constant values ($k_i$ = 1), the LCC is 0.980 for the IJ samples and 0.883 for the EPG samples. Because of the generalization capability, we executed the performance analysis below by using constant weighting factor values ($k_i$ = 1). By using the constants weighting factors we can prove that the IQ metric described by Equation (31) has a generalization capability over different sample sets.

We measured the performance of the proposed method with the LCC, ROCC and root-mean-square-error (RMSE) metrics (Table 9). The reference metric was the $D_1$ proposed by Wang et al. [120] and defined in Equation (3). Before the performance comparison, we fitted the data (of proposed metric and the $D_1$) by using the five-parameter logistic function [101]. The values of the LCC, ROCC and RMSE show that the proposed IQ metric was more effective than the $D_1$ at predicting printed image quality.

Figure 38 shows the subjective MOS values as a function of the predicted MOS (proposed metric and the $D_1$). Figures 38a and 38b show the results for the EPG samples, and Figures 38c and 38d show the results for the IJ samples. The performance of the proposed IQ was especially high compared with that of the $D_1$ for the IJ samples. The reason for the difference is the color contrast term included in the model of the proposed IQ defined in Equation (31). According to the studied attribute weighting factors $k_i$, the color contrast highly influenced the overall image quality of the IJ samples. The $D_1$ metric used only the luminance information of the images and could not find differences between the IJ images as clear as those found by the proposed IQ.

**Table 9.** LCC, ROCC and RMSE values for the EPG and IJ samples. Boldface indicates the best performer.

| Content | Metric | EPG | | | IJ | | |
|---|---|---|---|---|---|---|---|
| | | LCC | ROCC | RMSE | LCC | ROCC | RMSE |
| All | Proposed IQ | **0.884** | **0.727** | **0.353** | **0.981** | **0.962** | **0.196** |
| | $D_1$ (Wang et al. [120]) | 0.797 | 0.631 | 0.419 | 0.898 | 0.830 | 0.408 |
| Man | Proposed IQ | **0.875** | **0.748** | **0.351** | **0.977** | **0.973** | **0.208** |
| | $D_1$ (Wang et al. [120]) | 0.859 | 0.680 | 0.416 | 0.836 | 0.762 | 0.591 |
| Cactus | Proposed IQ | **0.878** | **0.772** | **0.426** | **0.988** | **0.969** | **0.163** |
| | $D_1$ (Wang et al. [120]) | 0.772 | 0.617 | 0.475 | 0.926 | 0.871 | 0.326 |
| Lake | Proposed IQ | **0.932** | 0.636 | **0.283** | **0.977** | **0.945** | **0.218** |
| | $D_1$ (Wang et al. [120]) | 0.866 | **0.721** | 0.366 | 0.954 | 0.932 | 0.307 |

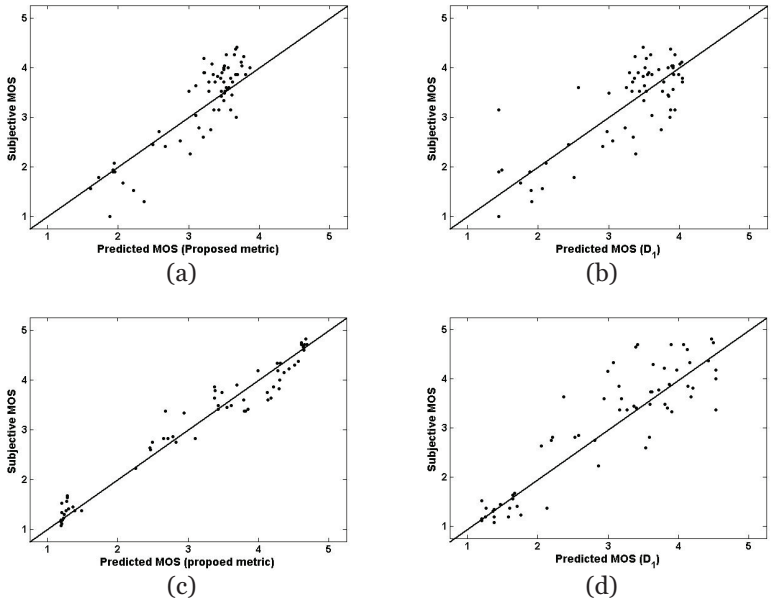
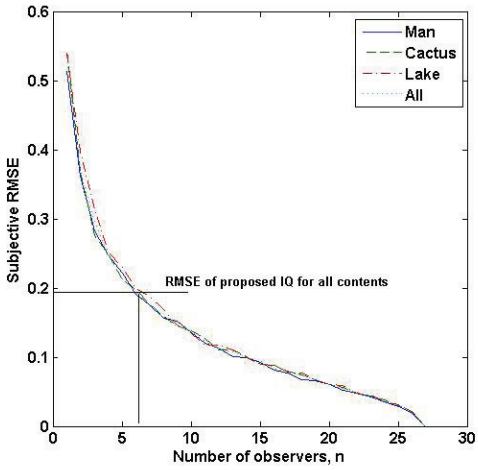
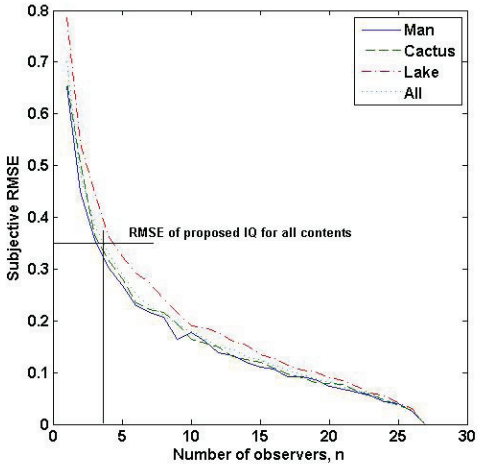



**Figure 38.** Subjective MOS as a function of the predicted MOS: the proposed metric for EPG (a) and for IJ (c) and the Wang et al. [120] metric $D_1$ for EPG (b) and for IJ (d).

Figure 39 shows the RMSE values for the subjective data as a function of the number of observers. The subjective RMSE values were calculated by comparing the average of n observer values with the mean values for all of the observers. For example if n = 3, the mean value of 3 observers was compared with the mean of all observers. We randomly selected different observer combinations from the group of all 27 observers, and the subjective RMSE was the average value for all combinations. The RMSE values of the proposed IQ for all of the contents were added in the figures. The comparison between the subjective RMSE and the RMSE for objective IQ indicates that the accuracy of the proposed IQ metric is higher than the accuracy of five randomly selected observers for the IJ samples and three randomly selected observers for the EPG samples.




(a)




(b)

**Figure 39.** Subjective RMSE as a function of the number of observers for the IJ samples (a) and EPG samples (b).

To determine which differences between the proposed metric and $D_1$ are statistically significant, we performed the variance test. The test is the same as the one used in previous studies [17], [101]. The assumption is that the residuals (the difference between the MOS and the predicted MOS values) are normally distributed. We tested the normality by using a Kurtosis-based criterion, according to which the residuals are Gaussian if a kurtosis is between 2 and 4 [101]. The F-test was used to test whether the variances of the residuals are identical, i.e. whether the two sample sets come from the same distribution. The null hypothesis is that the residuals of both metrics are expressions from the same distribution and are statistically indistinguishable with 95% confidence.

According to the variance test, there is significant difference between the proposed metric and the metric $D_1$ with respect to the IJ samples. The difference is not significant for the EPG samples with 95% confidence. According to the Kurtosis-based criterion, the assumption of Gaussian residuals is not met for the EPG samples assessed by the $D_1$ metric.

We also compared the overall computational complexity of the proposed metric and $D_1$. Table 10 lists the time (in seconds) taken to compute the metrics for an image of resolution 1500 x 1000 on a 3.0 GHz dual-core PC with 4 GB of RAM. The total computation time is divided between the reference and test images. It is clear that $D_1$ outperforms the proposed metric in terms of computational complexity. In this study we have used unoptimized MATLAB implementations and computational complexity can therefore be reduced if needed. The computational complexity of the proposed IQ is the sum of $S_1$, G and DC. Table 11 lists the times taken to compute the $S_1$, G and DC metrics, respectively.

**Table 10.** Computational complexity analysis of the proposed metric and $D_1$

| Metric | Reference image time (s) | Test image time (s) | Total time (s) |
|---|---|---|---|
| Proposed IQ* | 92.3 | 56.6 | 148.9 |
| $D_1$ | **13.7** | **5.2** | **18.8** |

*sum of the $S_1$, G and DC metrics

**Table 11.** Computational complexity analysis of the $S_1$, G and DC metrics.

| Metric | Reference image time (s) | Test image time (s) | Total time (s) | Percentage of time (%) |
|---|---|---|---|---|
| $S_1$ | 49.1 | 20.3 | 69.4 | 46.6 |
| G | 21.4 | 21.2 | 42.6 | 28.6 |
| DC | 21.7 | 15.1 | 36.8 | 24.7 |
| | | | 148.9 | 100 |

## Data reliability

The subjective MOS scores and the prediction of overall quality of the printed samples correlated strongly with one another. The LCC was 0.98 for the IJ samples and 0.88 for the EPG samples. The proposed application-specific metric was compared with the state-of-the-art generic RR metric, which was implemented into the proposed image quality measurement framework. The performance of the proposed metric was higher than that of the RR metric. For the IJ samples, the performance difference between the proposed metric and the state-of-the-art metric was also statistically significant.

From the perspective of data reliability, the number of observers (n = 28) and the number of samples (n = 21) were high enough. However, the low

number of different image contents (n = 3) can decrease the data reliability. We stress that three contents cannot validate the general image quality metrics, even if the selection of the contents were based on aspects strongly related to image quality. However, with the contents, we can and have proven that the proposed metrics and image quality measurement framework for printed images are promising.

## 4.3 Performance of objective methods for computing image quality attributes

### 4.3.1 Sharpness

Publication V examined the sharpness attribute of digital cameras. The sharpness metric $S_2$ was defined in Equation (25). The method was tested using two data sets (Data set I and Data set II). Section 4.1.2 described the experimental procedure for gathering subjective data. More details for the subjective tests can be found in Publication V.

The proposed sharpness metric $S_2$ was compared with the state-of-the-art NR and RR metrics as well as the test-target metrics. The NR sharpness metrics were from Marziliano et al. [71], Ferzli and Karam [24] and Narvekar and Karam [78]. The RR metric was the $D_1$ metric proposed by Wang et al. [120]. In addition, we captured the Mica test-target [110] images under laboratory conditions and calculated the MTF50 test target values. Before the analyses, the values of the metrics were fitted using the three-parameter logistic function defined in Equation (30).

Table 12 shows the LCC values, and Table 13 presents the coefficients of determination $R^2$ for the metrics. The results suggest that the performance of the $S_2$ is higher than the performance of the reference metrics. When we fitted the data over all of the clusters, the LCC and $R^2$ of the proposed metric $S_2$ were the highest. The cluster-specific performance of the $S_2$ was the highest except for Clusters 4 and 5 in Data set I and Cluster 5 in Data set II. In these cases, the performance of the $D_1$ metric or the test-target MTF50 was the best.

Figure 40 shows the subjective sharpness values as a function of the proposed metric and $D_1$. Figure 40 shows that Data set I (Figure 40a and 40b) was easier for both metrics than Data set II (Figure 40c and 40d). Compared with the performance of the application-specific sharpness metric proposed here, the performance of the generic image quality metric $D_1$ was particularly low for Data set II.

According to the $R^2$ values, $S_2$ can explain 72% and 69% of the total variation in the subjective sharpness values of Data sets I and II, respectively. We cannot account for the remaining 28% and 31% of the

total variation in subjective sharpness values in Data sets I and II, respectively.

**Table 12.** The Pearson linear correlation coefficients LCC of the Marziliano et al. [71], Ferzli and Karam [24], Narvekar and Karam [78] and Wang et al. ($D_1$) [120] metrics and of the proposed metric $S_2$ (M = 100x100, m = 5) with subjective sharpness

| | Data set I | | | | | |
|---|---|---|---|---|---|---|
| | Marziliano | Ferzli | Narvekar | $D_1$ | MTF50 | $S_2$ |
| Cluster 2 | 0.676 | 0.500 | 0.488 | 0.618 | 0.629 | **0.840** |
| Cluster 3 | 0.489 | 0.432 | 0.494 | 0.886 | 0.547 | **0.898** |
| Cluster 4 | 0.676 | 0.254 | 0.498 | **0.782** | 0.545 | 0.748 |
| Cluster 5 | 0.683 | 0.105 | 0.296 | 0.361 | **0.788** | 0.684 |
| Cluster 6 | 0.845 | 0.471 | 0.425 | 0.775 | 0.554 | **0.920** |
| Over all | 0.733 | 0.477 | 0.563 | 0.786 | 0.650 | **0.848** |
| | Data set II | | | | | |
| | Marziliano | Ferzli | Narvekar | D1 | MTF50 | S2 |
| Cluster 2 | 0.443 | 0.415 | 0.240 | 0.410 | 0.600 | **0.770** |
| Cluster 3 | 0.142 | -0.045 | -0.057 | 0.664 | 0.595 | **0.880** |
| Cluster 4 | 0.774 | 0.574 | 0.725 | 0.051 | 0.752 | **0.805** |
| Cluster 5 | 0.793 | 0.716 | 0.723 | **0.815** | 0.800 | 0.761 |
| Cluster 6 | 0.820 | 0.886 | 0.655 | 0.681 | 0.696 | **0.931** |
| Over all | 0.589 | 0.511 | 0.404 | 0.538 | 0.669 | **0.828** |

**Table 13.** The coefficients of determination [%] of the Marziliano et al. [71], Ferzli and Karam [24], Narvekar and Karam [78] and Wang et al. ($D_1$) [120] metrics and of the proposed metric $S_2$ (M = 100x100, m = 5) with subjective sharpness

| | Data set I | | | | | |
|---|---|---|---|---|---|---|
| | Marziliano | Ferzli | Narvekar | $D_1$ | MTF50 | $S_2$ |
| Cluster 2 | 45.698 | 25.000 | 23.814 | 38.192 | 39.564 | **70.560** |
| Cluster 3 | 23.912 | 18.662 | 24.404 | 78.500 | 29.921 | **80.640** |
| Cluster 4 | 45.698 | 6.452 | 24.800 | **61.152** | 29.703 | 55.950 |
| Cluster 5 | 46.649 | 1.103 | 8.762 | 13.032 | **62.094** | 46.786 |
| Cluster 6 | 71.403 | 22.184 | 18.063 | 60.063 | 30.692 | **84.640** |
| Over all | 53.729 | 22.753 | 31.697 | 61.780 | 42.250 | **71.910** |
| | Data set II | | | | | |
| | Marziliano | Ferzli | Narvekar | $D_1$ | MTF50 | $S_2$ |
| Cluster 2 | 19.625 | 17.223 | 5.760 | 16.810 | 36.000 | **59.290** |
| Cluster 3 | 2.016 | 0.203 | 0.325 | 44.090 | 35.403 | **77.440** |
| Cluster 4 | 59.908 | 32.948 | 52.563 | 0.260 | 56.550 | **64.803** |
| Cluster 5 | 62.885 | 51.266 | 52.273 | **66.423** | 64.000 | 57.912 |
| Cluster 6 | 67.240 | 78.500 | 42.903 | 46.376 | 48.442 | **86.676** |
| Over all | 34.692 | 26.112 | 16.322 | 28.944 | 44.756 | **68.558** |

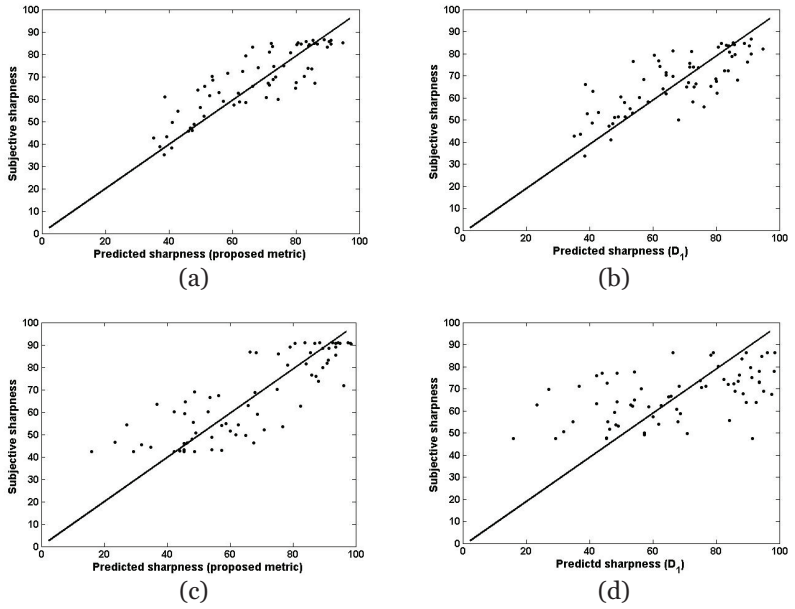



**Figure 40.** Subjective sharpness as a function of predicted sharpness: the proposed metric $S_2$ for Data set I (a) and for Data set II (c) and the Wang et al. [120] metric $D_1$ for Data set I (b) and for Data set II (d).

We tested the independence of the image contents by using a cross-validation method. Both Data sets I and II were divided into five groups, each of which represented one image cluster. We estimated the fitting parameters of the logistic function by using data from the other four groups. These four groups functioned as the training data. The fifth group was used as the testing data. Thus, the testing was performed five times. All of the groups (clusters) functioned once as testing data. Table 14 shows the mean LCC, ROCC and RMSE values of $S_2$ and $D_1$ for the training and testing groups. Table 14 clearly demonstrates that the performance of the proposed metric was also high for the testing data. The mean performance was clearly higher than the performance of the generic image quality metric $D_1$.

**Table 14.** The mean validation performance values for the Wang et al. [120] RR metric $D_1$ and the proposed metric $S_2$ (M = 100x100, m = 5)

| | Metric | ROCC | | LCC | | RMSE | |
|---|---|---|---|---|---|---|---|
| | | Training | Testing | Training | Testing | Training | Testing |
| Data set | $S_2$ | **0.838** | **0.781** | **0.843** | **0.811** | **8.812** | **10.241** |
| I | $D_1$ | 0.766 | 0.551 | 0.774 | 0.411 | 10.153 | 10.902 |
| Data set | $S_2$ | **0.821** | **0.785** | **0.836** | **0.807** | **12.318** | **13.460** |
| II | $D_1$ | 0.577 | 0.635 | 0.630 | 0.576 | 16.407 | 19.195 |

Figure 41 shows the RMSE values for the subjective data as a function of the number of observers. Different observer combinations were randomly selected from the group containing all 25 observers for Data set I and the group containing all 30 observers for Data set II. The subjective RMSE is the average value of the combinations. The RMSE values of the proposed sharpness metric $S_2$ for the validation data (see Table 14) were added in Figure 41. The figure shows that the accuracy of the proposed metric is higher than the accuracy of two randomly selected observers.

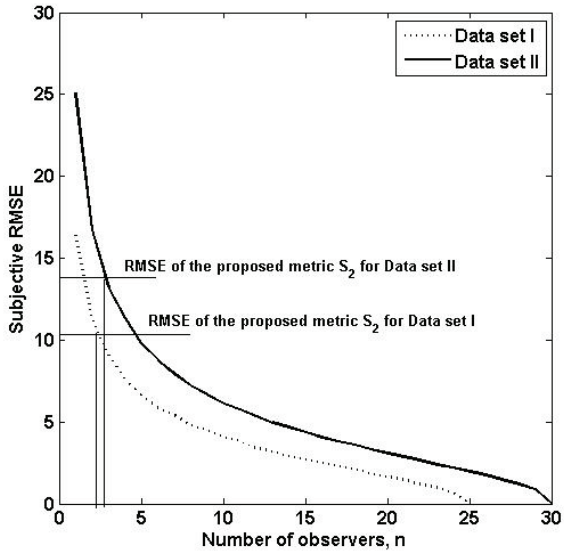



**Figure 41.** Subjective RMSE as a function of the number of observers for Data sets I and II.

To ascertain which differences between the proposed sharpness metric and the $D_1$ metric (Eq. (3)) proposed by Wang et al. [120] are statistically significant, we executed the variance test. According to the variance test, there is a significant difference between the proposed metric and $D_1$ with respect to the Data set II samples. The difference is not significant for the Data set I samples with 95% confidence.

We also compared the overall computational complexity between the proposed $S_2$ metric and the $D_1$. Table 15 lists the time (in seconds) taken to compute the metrics for an image of resolution 1600 x 1200 on a 3.0 GHz dual-core PC with 4 GB of RAM. The total computational time is divided between the reference and test images. It is clear that $D_1$ outperforms the proposed metric in terms of computational complexity. In this study we used unoptimized MATLAB implementations and computational complexity can therefore be reduced if needed. The computational complexity of the proposed $S_2$ is the sum of the blocks shown in Figure 25.

Table 16 lists the times taken to compute the blocks of “Energy metric”, “m highest candidate blocks”, Correspondence block search” and “Sharpness metric” .

**Table 15.** Complexity analysis of the proposed $S_2$ and $D_1$ metrics

| Metric | Reference image time (s) | Test image time (s) | Total time (s) |
|---|---|---|---|
| Sharpness, $S_2$ | 31.7 | 83.9 | 115.6 |
| $D_1$ | **14.9** | **8.0** | **23.0** |

**Table 16.** Complexity analysis of the blocks of the proposed $S_2$ metric

| Image type | Block (see Figure 25) | Time (s) | Percentage of time (%) |
|---|---|---|---|
| Reference image | Energy metric | 31.4 | 27.1 |
| | m highest candidate blocks | 0.3 | 0.3 |
| Test image | Correspondence block search | 83.9 | 72.6 |
| | Sharpness metric | 0.0 | 0.0 |
| | | 115.6 | 100 |

### 4.3.2 Color noise

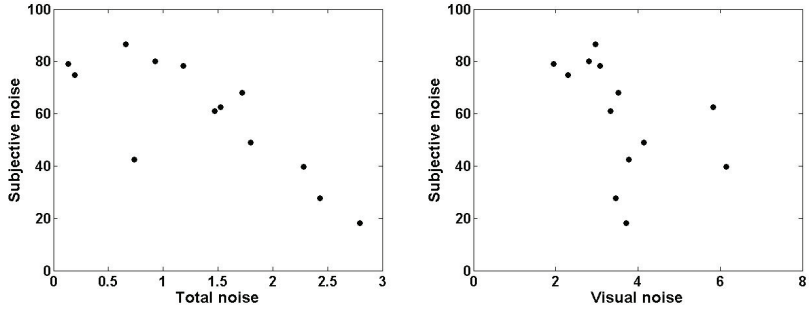
Publication III studied the color noise attribute of digital cameras. Equation (20) calculates the total noise metric N, and Equations (21) and (22) estimate the texture and smooth area noise components, respectively. Cluster 2 in Table 4 was used to validate the method. University students were used as the observers (n = 25). More details of the subjective tests can be found in Publication III. The prediction accuracy of N was compared with the visual noise test target metric [45] and the NR noise metric proposed by Immerkaer [41]. The visual noise was measured using the Gretag Macbeth test target under the lighting conditions of Cluster 2.

Table 17 shows the LCC and ROCC values for the metrics. The performance of the texture noise component was only moderate, but the performances of the smooth area component and total noise component were rather high compared with those of the reference noise metrics.

Figure 42 shows the subjective noise as a function of the proposed total noise metric and the visual noise test target metric. Compared with the visual noise test target metric, the proposed total noise metric accurately predicts the subjective noise of the samples. One clear outlier is evident. The proposed total noise predicts low noise value for that sample.

**Table 17.** The LCC and ROCC values of the proposed texture, smooth and total noise metrics and visual noise test target and Immerkaer's NR noise metrics

| Metric | LCC | ROCC |
|---|---|---|
| Proposed texture noise metric, $noise_t$ | -0.635 | -0.676 |
| Proposed smooth noise metric, $noise_s$ | **-0.837** | -0.775 |
| Proposed total noise metric, N | -0.800 | **-0.786** |
| Visual noise test target metric [45] | -0.471 | -0.709 |
| Immerkaer's NR noise metric [41] | -0.144 | -0.115 |




(a) (b)

**Figure 42.** Subjective noise as a function of predicted noise: the proposed total noise metric N (a) and the test target visual noise metric (b).

### 4.3.3 Color difference metrics

Publication VI presented the study on the color difference metrics of digital cameras. Figure 26 shows how the selected blocks were located. The color difference values were measured from the images captured by eight cameras using the scenes 1 and 2 shown in Figure 43.

The ground-truth test-target data were measured using the Gretag Macbeth CC test target under the lighting conditions of scenes 1 and 2. The color values of the test target were measured by the spectroradiometer, and the target was photographed by the cameras to be tested. Three basic test-target color difference values were calculated: luminance, chroma and hue differences ($\Delta L$, $\Delta C$ and $\Delta H$, respectively).

The LCC and ROCC between the proposed and (ground-truth) test target color difference values are shown in Table 18. Some of the correlations are significant for the luminance and chroma metrics, but the correlations of the hue metric are weak. Based on the t-test, the correlation value of 0.62 is significant ($df = 6$, $p = 0.05$).

One reason for the low correlation of the hue metric may be that the hue difference is highly dependent on the sample set. Because the proposed method selects the color samples based on the scene, the selected samples

were different from the Gretag Macbeth CC color patches. As more global values, the luminance and chroma are less dependent on the sample set.

When comparing the correlations calculated for the different scenes, we found it notable that Scene 2 differed from the test target more than Scene 1 did. The reason for this difference may be that Scene 1 was easier for color calibration of the reference camera because of its higher illuminance level.

(a) (b)

**Figure 43.** The proposed method was applied to scenes 1 (a) and 2 (b).

**Table 18.** The linear correlation (LCC) and rank-ordered correlation (ROCC) coefficients between the proposed method and the test target method

| Color metric | Scene 1 | | Scene 2 | |
|---|---|---|---|---|
| | LCC | ROCC | LCC | ROCC |
| ΔL | 0.73 | 0.81 | 0.61 | 0.52 |
| ΔC | 0.80 | 0.55 | 0.64 | 0.60 |
| ΔH | 0.14 | 0.17 | -0.08 | 0.31 |

### 4.3.4 Data reliability

The evaluation of the proposed sharpness metric for camera images according to the block diagram of Figure 25 constituted the most extensive part of the dissertation. According to the results, the performance of the sharpness metric is high compared with those of the state-of-the-art metrics that are applicable to camera images. In Data set II, the differences were also statistically significant.

The proposed color noise metric was evaluated using a single cluster. According to the results, the metric is promising. Only one image cluster was used in the test, and it decreased the reliability of the results. The color difference method was evaluated using two views. The evaluation was based only on the objective data. The assumption was that the test target data represent the ground truth.

The correlation coefficients of the proposed camera image methods were lower than the correlation coefficients of the corresponding methods

proposed for printed images. The main reason for this difference is that the camera image application is more difficult to measure than the printed image application. The reference images of the printed image application are simpler to reproduce and use than those of the camera image application. The form of an original reference image for the printed image application is digital, and test samples that should be digitized are planar (papers). In the case of cameras, the reference images (scenes) are captured by projecting 3D scenes onto a 2D plane (image sensor). In addition, the different optical properties of test cameras create perspective differences between test images.

# 5. Discussion

The proliferation of digital cameras and the growth of digital image and video applications have caused the issue of image quality to continue to be a lively research theme. Within this field, computing the quality of imaging systems with analog-digital and digital-analog conversions is a sub-field that has received less attention than the computation of quality distortions arising from, for instance, image compression or transmission.

The main goal of this dissertation was to create methods and algorithms that characterize imaging systems in a user-relevant manner by directly measuring relevant and objective image quality attributes from natural images. The traditional approach to measuring the quality of imaging systems uses test targets. However, the overall quality of the images captured, printed or displayed is perceived through high-level attributes. These attributes can only be evaluated from natural images. High-level attributes such as naturalness and clarity are not relevant concepts in the case of test target images.

In principle, reference-free (NR) algorithms would be ideal for computing the quality of imaging systems. However, the performance of NR metrics is low for imaging systems because even the state-of-the-art NR metrics can handle only images with one type of distortion [75], [94]. An image captured by a camera or printed by a printer can suffer from several types of distortions. For example, a NR metric for a blurriness distortion can interpret the noise energy in a blurry image as the image structure and assess the image as sharp.

If quality is computed from a natural image captured or printed by an imaging system rather than an all-digital image communication system, the type (or lack) of the reference image gives rise to the main problem. For the printer applications, a digital reference image is available, but printed images (i.e., test images) are in analog (optical) form. Test images should be digitized before the measurements are taken. For example, a previous study [17] proposed a framework for digitizing test images by using a scanner. Compared with this method [17], the digitization device in this dissertation was a high-quality camera, and the method was based on multiple exposures. A device of the proposed method could be a fully-

manual scanner, but with a camera e.g. lighting environment is easier to control. Because the study [17] applied state-of-the-art FR algorithms to the framework, precise registration between the reference and test images was a requirement. The proposed reduced-reference (RR) metrics for printed images are registration-free in principle. These metrics compute a feature vector for the reference image and utilize it if the attribute values are computed from test images digitized from prints.

For camera applications, the reference images are missing. The proposed methods use a high-quality reference camera to capture reference images. The performance of the reference camera should be high enough. Noise energy and image structure should be differentiable under all image-capturing conditions. In addition, an accurate color calibration should be possible. The areas for measurements are searched from the reference images, and the corresponding areas are located from the images captured by cameras. The search method is based on the correspondence points computed by the area descriptors in the neighborhood of the reference image's measurement areas. Only the corresponding points in the neighborhood of the measurement areas are used. A traditional image registration approach is based on a global principle. The image homography is computed by covering all of the images' pixel areas. Because the proposed method is based on a local principle, it compensates for the geometrical distortions and perspective differences between the images captured by different cameras. The local principle of the method eliminates the need for camera-specific optical calibrations.

Traditional full-reference (FR) metrics compute error maps between reference and test images. Overall quality is often computed as an average value of the map. The quality attributes computed by the proposed RR metrics are more suited for research and development work on imaging systems with many types of distortions than metrics that express quality with a single number. The attributes were selected based on a literature review and a subjective study of printed images. We developed sharpness, graininess and color contrast metrics for printing applications and sharpness, color noise and color difference metrics for camera applications. In particular, the color contrast metric played an important role in predicting the overall quality of the printed samples in this dissertation. The state-of-the-art computational image quality metrics for natural images use only luminance information and bypass color information.

In the dissertation, different methods were presented for camera and printer applications. Because of the order and progress of the research, the metrics for camera applications use a more advanced method to select the

local blocks to be measured. By applying the so-called corresponding blocks approach also in the printed image methods, the performance of the computational metrics would inevitably increase, because only the relevant regions would be used for measurements. The same most optimal blocks approach can also be applied to display measurements. The test images shown on a screen could be digitized by a high-quality camera, and the quality attributes can be computed accordingly.

According to the results, the performance of the proposed methods was promising in most cases, and the goal of the dissertation was achieved. To support this point, we compared the proposed methods and the data from the test target measurements, the state-of-the-art algorithmic metrics and the subjective tests. By comparing the accuracy of the proposed methods with that of the test target-based methods, we showed that considerable advances were made in comparison with the traditional approach to measuring the image quality attributes of imaging systems. The correlation coefficients between the subjective data and the objective metrics indicated that the performance of the proposed sharpness and noise metrics exceeded the performance of the MTF50 and visual noise test target metrics (see Tables 12 and 17). In addition, the proposed color metrics correlate with the color metric values measured from the physical color patches (Table 18).

By comparing the accuracy of the proposed metrics with that of the general-purpose algorithmic metrics from the field of image processing, we revealed that the proposed methods reached higher performance levels in the context of imaging systems than the state-of-the-art metrics (see Tables 9, 14 and 17). For example, according to the coefficients of determination, the proposed sharpness metric for camera applications explained more of the subjective data than the state-of-the-art metrics. This finding supports the notion that imaging system-specific algorithms need to be developed.

One of the long-term goals of the research area for this dissertation is to make time-consuming routine subjective evaluations by test participants obsolete. By comparing the prediction uncertainty of the proposed methods with the variations of the subjective data, we found that the proposed methods are sufficiently accurate and can replace small-scale subjective tests (See Figures 39 and 41). The RMSE values for the subjective and objective data showed that the proposed objective methods can predict a mean opinion score more accurately than a random observer for the images used in this dissertation.

The proposed methods measured the sharpness, color contrast, graininess, color noise and color difference attributes. In addition, it is

possible to compute the attributes of lightness and luminance contrast by using the same methods. Luminance contrast can be measured e.g. from local edge areas that can be found in a reference image. In addition, the values computed by the proposed sharpness metrics correlate strongly with subjective contrast. Lightness can be measured by, for example, comparing the global or local mean lightness values of reference and test images.

The proposed methods can be modified to measure video quality as well. A reference video can be captured by a high-quality video camera, and the areas suitable for computing quality attributes can be located from this video by using the proposed methods. In the future, most consumer cameras will include features such as HDR and stereoscopic 3D-imaging. The proposed methods, which are based on the reference images captured by a high-quality camera, can also be modified to assess HDR and stereoscopic images. For example, HDR images can be measured by using a reference camera calibrated for lightness, and stereoscopic images can be measured by using, for example, two calibrated reference cameras.

However, NR methods and high-level attributes and factors are the next main research topics in the research field of image quality. The use cases of the proposed reduced-reference methods for imaging systems are limited mostly to the field of research and development. The applications of image processing pipe tuning and imaging system benchmarking can utilize methods that require some information from a reference image. However, a wide range of real-life consumer applications, such as image retrieval or organizing systems, can apply only reference-free metrics. In addition, the applications of assisted photography would benefit from robust NR metrics. The goal can be to assist users to take better photographs by providing feedback such as framing instructions, which are based on object detection and quality algorithms.

The high-level attributes of naturalness and clarity require more advanced methods than the low-level attributes studied here. Computational image understanding and better models for human viewing of images are vital. For example, in a previous study [81], we showed that the performance of a reference-free method is high if the metrics are computed from the face areas in camera-captured images.

In addition, the theoretical constructs and relations of the quality attributes should be studied. The relations between low- and high-level attributes must be better understood before comprehensive methods can be constructed. Moreover, the difference between the quality evaluations of low- and high-quality images should be studied because it is known that the quality perceptions of low- and high-quality images are based on different attributes and aspects [82]. For example, the high-quality image

is sharp and noiseless, and the quality evaluation is probably based primarily on the reproduction of memory colors, texture and/or perspective rather than global sharpness or noise.

# 6. Conclusions

Image quality is a key performance indicator of digital cameras, displays and printers. As a research area, a major challenge arises from the shift in interest from characterizing the performance of devices to computationally predicting the quality of images and video as they are perceived and experienced by users. This dissertation addresses this challenge by focusing on quality attributes, particularly low-level attributes. These attributes are components in the constructs of high-level quality attributes and in judgments of overall quality as well.

The goal of the dissertation was to create algorithms, verify their performance and apply them to methods of determining quality attributes directly from natural images. This approach contrasts with the standard practice of measuring imaging system performance based on test targets. The study lies at the intersection of the imaging and signal-processing traditions of image quality research.

The key quality computation issues that differentiate imaging systems from image communication systems are the analog-to-digital and digital-to-analog conversions, which are part of the imaging pipelines. A variety of nonlinear distortions and artifacts are generated in the conversions. This has the implication that it is feasible to try to apply other principles of quality computation than the full-reference principle successfully employed in the signal-processing research tradition. This study focuses on the reduced-reference principle and uses reference images for image attribute metrics.

The contributions of the dissertation concern both the analog-to-digital conversions and reference image metrics for imaging systems. As the methods proved to predict subjective attribute scores more accurately than test target methods, we can conclude that the proposed methods facilitate computation of the quality attributes of natural images. In terms of the correlation coefficients between the objective data and subjective data, the performance of the objective methods also exceeded that of the state-of-the-art algorithms. This suggests that the methods of this dissertation are applicable to research on the constructs of high-level attributes. Also they can be employed to replace small scale subjective tests.

In the course of the research carried out in this dissertation several possibilities for further research have been identified. For instance, the proposed methods can be modified to measure video, HDR and stereoscopic imaging quality. Proposed methods can be extended to measure high-level attributes such as naturalness and clarity for printed or camera images. This requires, however, more advanced methods than the ones we used with the low-level attributes in this dissertation. Computational image understanding and more advanced models for human viewing or even human behaviour should be developed and implemented in the proposed image quality measurement framework.

A restriction of the proposed methods is the need for reference images. The measurements of printed images require a digital original image, and the measurements of camera images require reference images captured by a high-quality camera. However, progress in reference-free methods requires new and more advanced understanding of images from both objective and subjective standpoints.

# References


1. Artmann, U., Wueller, D., “Interaction of image noise, spatial resolution, and low contrast fine detail preservation in digital image processing,” *Proc. of SPIE* 7250, [72500I] (2009).
2. BID – Blurred Image DataBase, http://www.lps.ufrj.br/profs/eduardo/ ImageDatabase.htm.
3. Brémond, R., Tarel, J.-P., Dumont, E., Hautiére, N., “Vison models for image quality assessment: one is not enough,” *Journal of Electronic Imaging*, 19(4) , 043004 (2010).
4. Cao, F., Guichard, F., Hornung, H., “Measuring texture sharpen of a digital camera,” *Proc. of SPIE* 7250 [72500H] (2009).
5. Capodiferro, L., Mangiatordi, F., Claudio, E. D. D., Jacovitti, G., “Equalized Structural Similarity Index for Image Quality Assesment” *Proc. of ISPA* 2011, 420-424 (2011).
6. Carnec, M., Le Callet, P., Barba, D., “Objective quality assessment of color images based on a generic perceptual reduced reference,” *Signal Processing: Image Communication*, 23(4), 239-256 (2008).
7. Caviedes, J., Gurbuz, S., “No-Reference Sharpness Metric Based on Local Edge Kurtosis,” *Proc. of ICIP* 2002, 53-56 (2002).
8. Chandler, D. M., Hemami, S. S., “VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images,” *IEEE Transaction on Image Processing*, 16(9), 2284-2298 (2007).
9. Chandler, D. M., Hemami, S. S., “VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images,” http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html.
10. Chen, G.-H., Yang, C.-L., Po, L.-M., Xe, S.-L., “Edge-Based Structural Similarity for Image Quality Assessment,” *Proc. of Int. Conf. Acoustic, Speech, and Signal Processing*, 14-19 (2006).
11. Chen, G.-H., Yang, C.-L., Po, L.-M., Xe, S.-L., “Gradient-Based Structural Similarity for Image Quality Assessment,” *Proc. of Int. Conf. Image Processing*, 2929-2932 (2006).
12. Chen, M. J., Bovik, A. C., “No-Reference image blur assessment using multiscale gradient,” *EURASIP Journal on Image and Video Processing* 2011, 3 (2011).
13. Chen, M.-J., Bovik, A. C., “No-Reference Image Blur Assessment using Multiscale Gradient,” *Proc. of QoMEX* 2009, 70-74 (2009).
14. Cheng, G., Cheng, L., “Reduced reference image quality assessment based on dual derivative priors,” *IEEE Electronic Letters*, 45(18), 937-939 (2009).
15. Ciancio, A., Costa, A. L. N., Silva, E. A. B., Said, A., Samadani, R., Obrador, P., “No-Reference Blur Assessment of Digital Pictures Based on Multifeature Classifiers,” *IEEE Transaction on Image Processing*, 20(1), 64-75 (2011).

16. CPIQ Phase 1 v.1.10. “Camera Phone Image Quality – Phase 1 – Fundamentals and review of considered test methods.” International Imaging Industry Association (I3A), 50 (2007).
17. Eerola, T., Lensu, L., Kälviäinen, H., Kämäräinen, J.-K., Leisti, T., Nyman, G., Halonen, R., Oittinen, P., ”Full Reference Printed Image Quality: Measurement Framework and Statistical Evaluation,” *Journal of Imaging Science and Technology*, 54(1), 010201 (2010).
18. Eerola, T., Lensu, L., Kämäräinen, J.-K., Leisti, T., Ritala, R., Nyman, G., Kälviäinen, H., “Bayesian network model of overall print quality: construction and structural optimisation,” *Pattern recognition letters,* 32(11), 1558-1566 (2011).
19. Engeldrum, P. G., “A Theory of Image Quality: The Image Quality Circle,” *Journal of Imaging Science and Technology,* 48(5), 446-456 (2004).
20. Engelke, U., Kusuma, M., Zepernick, H.-J., Caldera, M., “Reduced-reference metric design for objective perceptual quality assessment in wireless imaging,” *Signal Processing: Image Communication,* 24(7), 525-547 (2009).
21. Engelke, U., Zepernick, H.-J., “Framework for optimal region of interest-based quality assessment in wireless imaging,” *Journal of Electronic Imaging*, 19(1), 011005 (2010).
22. Fairchild, M., “The HDR Photographic Survey,” *Proc. of CIC 2007*, 233-238 (2007).
23. Ferzli, R., Karam, L. J., "JNB Sharpness Metric Software," http://ivulab.asu.edu/Quality/JNBM.
24. Ferzli, R., Karam, L. J., “A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur (JNB),” *IEEE Transaction on Image Processing*, 18(4), 717-728 (2009).
25. Gao, X., Lu, W., Tao, D., Li, X., “Image Quality Assessment Based on Multiscale Geometric Analysis,” *IEEE Transaction on Image Processing*, 18(7), 1409-1423 (2009).
26. Geisler, W. S., “Visual Perception and the Statistical Properties of Natural Scenes,” *Annual Review of Psychology*, 59, 167-192 (2007).
27. Gunawan, I. P., Ghanbari, M., “Efficient Reduced-Reference Video Quality Meter,” *IEEE Transaction on Broadcasting*, 54(3), 669-679 (2008).
28. Guo, A., Zhao, D., Liu, S., Fan, X., Gao, W. “Visual Attention Based Image Quality Assessment,” *Proc. of ICIP 2011,* 3358-3361 (2011).
29. H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, “Speeded-up robust features (SURF),” *Computer Vision and Image Understanding (CVIU)*, 110(3), 346-359 (2008).
30. Halonen, R., Leisti, T., Oittinen, P., “The influence of image content and paper grade on image quality attributes computed from printed natural images,” *Proc. of IS&T NIP* 24, 459-462 (2008).
31. Halonen, R., Nuutinen, M., Asikainen, R., Oittinen, P., ”Development and measurement of the goodness of test images for visual print quality evaluation,” *Proc. of SPIE* 7529, [752909] (2010).
32. Halonen, R., Nuutinen, M., Oittinen, P., Leisti, T., Nyman, G., Mettänen, M., Ritala, R., Eerola, T., Lensu, L., Kämäräinen, J.-K., Kälviäinen, H., “Fusion of Digital and Visual Print Quality – Final Report of DigiQ Project,” *Graphic Arts in Finland*, 39(3), 1-108 (2010).

33. Han, H.-S., Kim, D.-O., Park, R. H., "Structural Information-based Image Quality Assessment using LU Factorization," *IEEE Transaction on Consumer Electronics*, 55(1), 165-171 (2009).
34. Han, H.-S., Park, R. H., "New Image Quality Metric using the Harris Response," *IEEE Signal Processing Letters*, 16(7), 616-619 (2009).
35. Hasler, D., Süsstrunk, S., "Measuring colourfulness in natural images," *Proc. of SPIE* 5007, 87-95 (2003).
36. Hekstra, A. P. , Beerends, J. G., Ledermann, D., de Caluwe, F. E., Kohler, S., Koenen, R. H., Rihs, S., Ehrsam, M., Schlauss, D., "PVQM – A perceptual video quality measure," *Signal Processing: Image Communication,* 17(10), 781-798 (2002).
37. Hong, G., Luo, R., "Perceptually based colour difference for complex images," *Proc. of SPIE* 4421, 618-621 (2002).
38. Horita, Y., "MICT Image quality Evaluation Database," http://mict.eng.u-toyama.ac.jp/mictdb.html.
39. Hultgren, B. O., Hertel, D. W., "Megapixel mythology and photospace: Estimating photospace for camera phones from large image sets," *Proc. of SPIE* 6808, [680818] (2008).
40. I3A, "CPIQ Initiative Phase 1 White Paper: Fundamentals and review of considered test methods," 74 (2007).
41. Immerkaer, J., "Fast Noise Variance Estimation," *Computer Vision and Image Understanding*, 64(2), 300-302 (1996).
42. ISO 12232:2006(E). Photography – Digital still cameras – Determination of exposure index, ISO speed ratings, standard output sensitivity, and recommended exposure index. ISO 2006: 17.
43. ISO 12233:2000(E). Photography – Electronic still-picture cameras – Resolution measurements. ISO 2000: 32.
44. ISO 14524:1999(E). Photography – Electronic still-picture cameras – Methods for measuring opto-electronic conversion functions (OECFs). ISO 1999: 21.
45. ISO 15739:2003(E). Photography – Electronic still-picture cameras – Noise measurements. ISO 2003: 26.
46. ISO 17321-1:2006(E). Graphic technology and photography – Colour characterization of digital still cameras (DSCs) – Part 1: Stimuli, metrology and test procedures. ISO 2006: 25.
47. ISO 9039:1994(E). Optics and optical instruments – Quality evaluation of optical systems – Determination of distortion. ISO 1994: 14.
48. Janssen, R., "Computational Image Quality," SPIE – The International Society for Optical Engineering, Bellingham, Washington, 145 (2001).
49. Jumisko-Pykkö, S., Strohmeier, D., Utriainen, T., Kunze, K., "Descriptive Quality of Experience for Mobile 3D Video," *Proc. of NordiCHI 2010*, 266-275 (2010).
50. Keelan, B. W., "Handbook of Image Quality," Markel Dekker Inc., New York, 516 (2002).
51. Kim, D. O., Han, H.-S., Park, R.-H., "Gradient Information-Based Image Quality Metric," *IEEE Transaction on Consumer Electronics*, 56(2), 930-936 (2010).
52. Kim, D. O., Park, R.-H., "Evaluation of image quality using dual-tree complex wavelet transform and compressive sensing," *IEEE Electronics Letter*s, 46(7), 494-495 (2010).
53. Kim, D. O., Park, R.-H., "Image Quality Assessment Using the Amplitude/Phase Quantization Code," *IEEE Transaction on Consumer Electronics*, 56(4), 2756-2762 (2010).

54. Kim, D. O., Park, R.-H., “Image Quality Measure Using the Phase Quantization Code,” *IEEE Transaction on Consumer Electronics*, 56(2), 937-945 (2010).
55. Kodak, “Lossless True Color Image Suit,” http://r0k.us/graphics/kodak.
56. Koivisto, L., “Measuring the color reproduction of digital camera with hidden color targets” MSc. Thesis, TKK/Faculty of Electronics, 2009.
57. Kuhna, M., Nuutinen, M., Oittinen, P., “Method for Evaluating Tone Mapping Operators for Natural High Dynamic Range Images,” *Proc. of SPIE* 7876, [78760O] (2011).
58. Le Callet, P., Autrusseau, F., “Subjective quality assessment IRCCyn/IVC database,” http://www.irccyn.ec-nantes.fr/ivcdb/.
59. Leisti, T., Radun, J., Virtanen, T., Halonen, R., Nyman, G., “Subjective Experience of Image Quality: Attributes, Definitions and Decision Making of Subjective Image Quality,” *Proc. of SPIE* 7242, [72420D] (2009).
60. Li, C., Bovik, A. C., Wu, X., “Blind Image Quality Assessment Using a General Regression Neural Network,” *IEEE Transaction on Neural Networks, 22(5), 793-799* (2011).
61. Li, Q., Wang, Z., “Reduced-Reference Image Quality Assessment Using Divisive Normalization-Based Image Representation,” *IEEE Journal of Selected Topics in Signal Processing,* 3(2), 202-211 (2009).
62. Li, X., Tao, D., Gao, X., Lu, W., ”A natural image quality evaluation metric,” *Signal Processing,* 89(4), 548-555 (2009).
63. Liang, L., Chen, J., Ma, S., Zhao, D., Gao, W., “A no-reference perceptual blur metric using histogram of gradient profile sharpness,” *Proc. of ICIP* 2009, 4369-4372 (2009).
64. Liu, H., Heynderickx, I., “Visual Attention in Objective Image Quality Assessment: Based on Eye-Tracking Data,” *IEEE Transactions on Circuits and Systems for Video Technology*, 21(7), 971-982 (2011).
65. Liu, H., Klomp, N., Heynderickx, I., “A No-Reference Metric for Perceived Ringing Artifacts in Images,” *IEEE Transaction on Circuits and Systems for Video Technology,* 20(4), 529-539 (2010).
66. Loebich, C. M., Wueller, D., Klingen, B., Jaeger, A., “Digital Camera Resolution Measurement Using Sinisoidal Siemens Stars,” *Proc. of SPIE* 6502, [65020N] (2007).
67. Lowe, D. G., “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, 60(3), 91-110 (2004).
68. Ma, L., Li, S., Ngan, K. N., “Visual Horizontal Effect for Image Quality Assessment,” *IEEE Signal Processing Letters*, 17(7), 627-630 (2010).
69. Ma, L., Li, S., Zhang, F., Ngan, K. N., “Reduced-Reference Image Quality Assessment Using Reorganized DCT-Based Image Representation,” *IEEE Transaction on Multimedia*, 13(4), 824-829 (2011).
70. Maalouf, A., Larabi, M.-C., Fernandez-Maloigne, C., “A Groupled-Based Reduced Reference Image Quality Assessment,” *Proc. of IEEE QoMEX 2009*, 59-63 (2009).
71. Marziliano, P., Dufaux, F., Winkler, S., Ebrahimi, T., “Perceptual Blur and Ringing Metrics: Application to JPEG2000,” *Signal Processing: Image Communication,* 19(2), 163-172 (2004).

72. Meur, O. L., Ninassi, A., Callet, P. L., Barba, D., ”Overt visual attention for free-viewing and quality assessment tasks Impact of the regions of interest on a video quality metric,” *Signal Processing: Image Communication*, 25(7), 547-558 (2010).
73. Ming, Y., Yingchun, G., Dongming, Z., “A Method for Reduced-reference Color Image Quality Assessment,” *Proc. of IEEE CISP* 2009, 1-5 (2009).
74. Moorthy, A. K., Bovik, A. C., “A Two-Step Framework for Constructing Blind Image Quality Indices,” *IEEE Signal Processing Letters*, 17(5), 513-516 (2010).
75. Moorthy, A. K., Bovik, A. C., “Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality,” *IEEE Transaction on Image Processing*, 20(12), 3350-3364 (2011).
76. Moorthy, A. K., Bovik, A. C., “Visual Importance Pooling for Image Quality Assessment,” *IEEE Journal of Selected Topics in Signal Processing*, 3(2), 193-201 (2009).
77. Murdoch, M. J., van Etten, M., Heynderickx, I., ”Effect of Environmental Factors on Perception of Image Attributes,” *Proc. of ICIS 2010*, 153-157 (2010).
78. Narvekar, N. D., Karam, L. J., “A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD),” *IEEE Transaction on Image Processing,* 20(9), 2678-2683 (2011).
79. Narvekar, N. D., Karam, L. J., “CPBD Sharpness Metric Software,” http://ivulab.asu.edu/Quality/ CPBD.
80. Narwaria, M., Lin, W., “Objective Image Quality Assessment Based on Support Vector Regression,” *IEEE Transaction on Neural Networks*, 21(3), 515-516 (2010).
81. Nuutinen, M., Orenius, O., Säämänen, T., Oittinen, P., ”Potential of face area data for predicting sharpness of natural images” *Proc. of SPIE* 7876, [7876-36] (2011).
82. Nyman, G., Häkkinen, J., Koivisto, E.M., Leisti, T., Lindroos, P., Orenius, O., Virtanen, T., Vuori, T., “Evaluation of the visual performance of image processing pipes: information value of subjective image attributes,” *Proc. of SPIE* 7529, [752905] (2010).
83. Nyman, G., Leisti, T., Lindroos, P., Radun, J., Suomi, S., Virtanen, T., Olives, J.-L., Oja, J., Vuori, T., ”Measuring multivariate subjective image quality for still and video cameras and image processing system components,” *Proc. of SPIE* 6808, [68080N] (2008).
84. Okano, Y., “MTF Analysis and its Measurements for Digital Still Camera,” *Proc. of IS&T 50th Annual Conference*, 383-387 (1997).
85. Orava, J., Jääskeläinen, T., Parkkinen, J., “Color errors of digital cameras,” *Color research and applications*, 29(3), 217-221 (2004).
86. Engeldrum, P., “A Theory of Image Quality: The Image Quality Circle,” *Journal of Imaging Science and Technology*, 48(5), 446-456 (2003).
87. Pedersen, M., Amirshasi, S. A., “Framework for the Evaluation of Color Prints Using Image Quality Metrics,” *Proc. of CGIV 2010*, 75-81 (2010).
88. Pedersen, M., Bonnier, N., Hardeberg, J. Y., Albregtsen, F., “Attributes of image quality for color prints,” *Journal of Electronic Imaging*, 19(1), [011016] (2010).
89. Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Astola, J., Carli, M., Battisti, F., ”TID2008 – A Database for Evaluating of Full-Reference Visual Quality Assessment Metrics,” *Advances of Modern Radioelectronics*, 10, 30-45 (2009).

90. Puukko, P., Niemi, K., “Instrumental measurement of print quality,” In: “Print Media – Principles, Processes and Quality” Edited by P. Oittinen, H. Saarelma, Papermaking Science and Technology, 317-334 (2009).
91. Radun, J., Leisti, T., Häkkinen, J., Ojanen, H., Olives, J.-L., Vuori, T., Nyman, G., ”Content and quality: Interpretation-based estimation of image quality,” *ACM Transaction on Applied Perception (TAP)*, 4(4), 21 (2008).
92. Radun, J., Leisti, T., Virtanen, T., Häkkinen, J., Vuori, T., Nyman, G., ”Evaluating the Multivariate Visual quality Performance of Image-Processing Components,” *ACM Transaction on Applied Perception,* 7(3), 16 (2010).
93. Redi, J. A., Gastaldo, P., Heynderickx, I., Zunino, R., “Color Distribution Information for the Reduced-Reference Assessment of Perceived Image Quality,” *IEEE Transaction on Circuits and Systems for Video Technology*, 20(12), 1757-1769 (2010).
94. Saad, M. A., Bovik, A. C., Carrier, C., “DCT Statistics Model-Based Blind Image Quality Assessment,” *Proc. of ICIP 2011*, 3154-3157.
95. Saad, M. A., Bovik, A. C., Charrier, C., “A DCT Statistics-Based Blind Image Quality Index,” *IEEE Signal Processing Letters*, 17(6), 583-586 (2010).
96. Säämänen, T., Virtanen, T., Nyman, G., “Videospace – Classification of video through shooting context information,” *Proc. of SPIE* 7529, [752906] (2010).
97. Sampat, M. P., Wang, Z., Gupta, S., Bovik, A. C., Markey, M. K., “Complex Wavelet Structural Similarity: A New Image Similarity Index,” *IEEE Transaction on Image Processing*, 18(11), 2385-2401 (2009).
98. Segur, R. K., “Using Photographic Space to Improve the Evaluation of Consumer Cameras,” *Proc. of IS&T’s PICS Conference,* (2000).
99. Sheikh, H. R., Bovik, A. C., “Image Information and Visual Quality,” *IEEE Transaction on Image Processing*, 15(2), 430-444 (2006).
100. Sheikh, H. R., Bovik, A. C., Cormack, L., “No-Reference Quality Assessment Using Natural Scene Statistics: JPEG2000,” *IEEE Transaction on Image Processing,* 14(11), 1918-1927 (2005).
101. Sheikh, H. R., Sabir, M. F., Bovik, A. C., “A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms,” *IEEE Transaction on Image Processing,* 15(11), 3441-3452 (2006).
102. Shen, H.-L., Xin, J.H., “Colorimetric and Spectral Characterization of a Color Scanner Using Local Statistic,” *Journal of Imaging Science and Technology*, 48(4), 342-346 (2004).
103. Shen, J., Li, Q., Erlebacher, G., “Hybrid No-reference Natural Image Quality Assessment of noisy, Blurry, JPEG2000, and JPEG Images,” *IEEE Transaction on Image Processing*, 20(8), 2089-2098 (2011).
104. Shnayderman, A., Gusev, A, Eskicioglu, A. M., “An SVD-based Grayscale Image Quality Measure for Local and Global Assessment,” *IEEE Transaction on Image Processing*, 15(2), 422-429 (2006).
105. Siddiqui, H., Bouman, C. A., “Hierarchical Color Correction for Camera Cell Phone Images,” *IEEE Transaction on Image Processing*, 17(11), 2138-2155 (2008).
106. Sony, “sRGB standard images 1999,” http://www.colour.org/tc8-04/test_images/Sony.

107. Tai, S.-C., Yang, S.-H., “A Fast Method for Image Noise Estimation Using Laplacian Operator and Adaptive Edge Detection.” *Proc. of 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP)* 2008, 1077-1081 (2008).
108. Tao, D., Li, X., “Reduced-Reference IQA in Contourlet Domain,” *IEEE Transaction on Systems, Man, and Cybernetics – Part B: Cybernetics,* 39(6), 1623-1627 (2009).
109. Tao, D., Li, X., Lu, W., Gao, X., “Reduced-reference IQA in Contourlet Domain,” *IEEE Transaction on Systems, Man, and Cybernetics – Part B: Cybernetics*, 39(6), 1623-1627 (2009).
110. Tervonen, A., Nivala, I., Ryytty, P., Saari, H., Ojanen, H., Viinikanoja, J., ”Integrated measurement system for miniature camera modules,” *Proc. of SPIE* 6196, [61960L] (2006).
111. Tong, Y., Konik, H., Cheikh, F. A., Tremeau, A., “Full Reference Image Quality Assessment Based on Saliency Map analysis,” *Journal of Imaging Science and Technology*, 54(3), 030503 (2010).
112. Uzair, M., Fayek, D., “An Efficient No-Reference Blockiness Metric for Intra-Code Video Frames,” *Proc. of WPMC* 2011, 1-5 (2011).
113. Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Transaction on Image Processing*, 13(4), 600-612 (2004).
114. Wang, Z., Bovik, A. C., “Reduced- and No-Reference Image Quality Assessment,” *IEEE Signal Processing Magazine*, 28(6), 29-40 (2011).
115. Wang, Z., Bovik, A. C., Evan, B. L., “Blind measurement of blocking artifacts in images,” *Proc. of IEEE ICIP* 2000, 981-984 (2000).
116. Wang, Z., Sheikh, H. R., Bovik, A. C., “No-reference perceptual quality assessment of JPEG compressed images,” *Proc. of IEEE ICIP* 2002, 477-480 (2002), Matlab code: http://anchovy.ece.utexas.edu/~zwang/research/nr_jpeg_quality/index.html
117. Wang, Z., Sheikh, H. R., Bovik, A. C., “No-reference perceptual quality assessment of JPEG compressed images,” *Proc. of IEEE ICIP* 2002, 477-480 (2002).
118. Wang, Z., Simoncelli, E. P., “Reduced-Reference Image Quality Assessment,” http://www.cns.nyu.edu/~zwang/files/research/rriqa/index.html.
119. Wang, Z., Simoncelli, E. P., Bovik, A. C., “Multi-Scale Strutural Similarity for Image Quality Assessment,” *Proc. of Asilomar Conf. Signal, Systems, and Computers*, 1398-1402 (2003).
120. Wang, Z., Wu, G., Sheikh, H. R., Simoncelli, E. P., Yang, E.-H., Bovik, A. C., “Quality-Aware Images,” *IEEE Transaction on Image Processing*, 15(6), 1680-1689 (2006).
121. Wee, C.-Y., Paramesram, R., “Image Sharpness Measure Using Eigenvalues,” *Proc. of ICSP 2008*, 840-843 (2008).
122. Xue, W., Mou, X., “Reduced reference image quality assessment based on weibull statistics,” *Proc. of IEEE QoMEX* 2010, 1-6 (2010).
123. Ye, P., Doermann, D., “No-reference Image Quality Assessment Based on Visual Codebook,” *Proc. of ICIP 2011*, 3150-3153.
124. Yendrikhovskij, S. N., Blommaert, J. J., de Ridder, H., “Color reproduction and the naturalness constraint,” *Color Research & Application*, 24(1), 52-67 (1998).

125. Zhang, J., Le, T. M., “A New No-Reference Quality Metric for JPEG2000 Images,” *IEEE Transaction on Consumer Electronics,* 56(2), 743-750 (2010).
126. Zhang, L., Zhang, L ., Mou, M., Zhang, D., “FSIM: A Feature Similarity Index for Image Quality Assessment,” *IEEE Transaction on Image Processing,* 20(8), 2378-2386 (2011).
127. Zhang, M., Mou, X., Zhang, L., “Non-Shift Edge Based Ratio (NSER): An Image Quality Assessment Metric Based on Early Vision Features,” *IEEE Signal Processing Letters,* 18(5), 315-318 (2011).
128. Zhang, X., Wandell, B.A., “A Spatial Extension of CIELAB for Digital Color Image Reproduction,” *Proc. of the SID Symposium* 1996, 731-734 (1996).
129. Zhu, X., Milanfar, P., “A No-Reference Sharpness Metric Sensitive to Blut and Noise,” *Proc. of QoMEX* 2009, 64-69 (2009).

# Errata

**Publication I**

In Subsection 3.3, the image sharpness metric S should be

$$S = \frac{\sum_{i=i}^{n} \beta_{hi} + \beta_{d1i} + \beta_{vi} + \beta_{d2i}}{4}$$

**Publication V**

In Subsection 4.1, ‘Dataset 14 cameras’ should be ‘Dataset II 14 cameras’.

Aalto-DD 157/2012

ISBN 978-952-60-4880-2
ISBN 978-952-60-4881-9 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934
ISSN 1799-4942 (pdf)

**Aalto University**
**School of Science**
**Department of Media Technology**
**www.aalto.fi**

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

**DOCTORAL
DISSERTATIONS**