The effect of scene content on image quality

Kyung Hoon Oh

Faculty of Media, Arts and Design

THE EFFECT OF SCENE CONTENT ON IMAGE QUALITY

KYUNG HOON OH

A thesis submitted in partial fulfilment of the
requirements of the University of Westminster
for the degree of Master of Philosophy.

The research programme was carried out in the Imaging
Technology Research Group at the University of Westminster.

August 2014

Abstract

Device-dependent metrics attempt to predict image quality from an ‘average signal’, usually embodied on test targets. Consequently, the metrics perform well on individual ‘average looking’ scenes and test targets, but provide lower correlation with subjective assessments when working with a variety of scenes with different than ‘average signal’ characteristics. This study considers the issues of scene dependency on image quality. This study aims to quantify the change in quality with scene contents, to research the problem of scene dependency in relation to device-dependent image quality metrics and to provide a solution to it.

A novel subjective scaling method was developed in order to derive individual attribute scales, using the results from the overall image quality assessments. This was an analytical top-down approach, which does not require separate scaling of individual attributes and does not assume that the attribute is not independent from other attributes. From the measurements, interval scales were created and the effective scene dependency factor was calculated, for each attribute.

Two device-dependent image quality metrics, the Effective Pictorial Information Capacity (EPIC) and the Perceived Information Capacity (PIC), were used to predict subjective image quality for a test set that varied in sharpness and noisiness. These metrics were found to be reliable predictors of image quality. However, they were not equally successful in predicting quality for different images with varying scene content.

Objective scene classification was thus considered and employed in order to deal with the problem of scene dependency in device-dependent metrics. It used objective scene descriptors, which correlated with subjective criteria on scene susceptibility. This process resulted in the development of a fully automatic classification of scenes into ‘standard’ and ‘non-standard’ groups, and the result allows the calculation of calibrated metric values for each group. The classification and metric calibration performance was quite encouraging, not only because it improved mean image quality predictions from all scenes, but also because it catered for non-standard scenes, which originally produced low correlations. The findings indicate that the proposed automatic scene classification method has great potential for tackling the problem of scene dependency, when modelling device-dependent image quality. In addition, possible further studies of objective scene classification are discussed.

Acknowledgements

My special thanks are due to all those who uncomplainingly and cheerfully supported my thesis with all round contributions. I am grateful to:

My supervisor, the director of the Imaging Technology Research Group (ITRG), Dr. Sophie Triantaphillidou, who has believed in my work and supplied patient yet powerful encouragement and input. Her direction and approval way from my MSc has been crucial. Also, I would like to thank her for the original idea for the project and for thus creating such a fascinating and challenging study.

My second supervisor, Professor. Ralph E Jacobson, whose generous guidance and invaluable instruction has ensured the completion of this project. He has given me the benefit of his experience and knowledge in compiling and collating this material.

My third supervisor, Dr. Alexandra Psarrou, whose inspiration and extension at imaging analysis area has guided me reflexively.

My colleagues in ITRG (Professor. Geoffrey G Attridge, Dr. Efthimia Bilissi, Mrs. Elizabeth Allen, Mr. Jae Young Park and Mr. John Smith) have provided me their advice and willingness to act like a discussion board.

Also, Professor. Hyun-Tae Kwak has been vital for his moral support since BSc.

I have received much help from my glorious English mentors, Mr. Ken Abhayaratna and Mrs. Angel Abhayaratna. I am indebted for their generosity and friendship.

This thesis is dedicated to my parents, brother and uncle, who have continued to provide for me with endless love.

Finally, I would like to thank my lovely wife, Mrs. Sungsil Oh, who gave me inspiration to carry on and motivated me to reach for my dream.

Contents

List of Figures

List of Tables

Abbreviations

ANSI: American National Standards Institute

CCD: Charge-coupled device

CCV: Colour Coherence Vector

CIE: Commission Internationale de l'Eclairage

CMOS: Complementary Metal Oxide Semiconductor

CMTA: Cascaded Modulation Transfer Acutance

CMY: Cyan, Magenta, Yellow

cpd: cycles per degree

CRI: Colour Reproduction Index

CRT: Cathode Ray Tube

CSF: Contrast sensitivity function

CV: Coefficient of variation

DCT: Discrete Cosine Transform

dpi: dots per inch

DQE: Detective Quantum Efficiency

DWT: Discrete Wavelet Transform

EPIC: Effective Pictorial Information Capacity

ESF: Edge Spread Function

GUI: Graphic User Interface

GLCM: Grey Level Co-occurrence Matrix

HVS: Human Visual System

iCAM: image Color Appearance Model

ICC: International Color Consortium

IEEE: Institute of Electrical and Electronics Engineers

IHIF: Integrated Hyperbolic Increment Function

ISO: International Standards Organisation

IQM: Image Quality Metric

ITRG: Imaging Technology Research Group

ITU-R: International Telecommunication Union -Radiocommunication Sector

JND: Just Noticeable Difference

JPEG: Joint Photographic Experts Group

LCD: Liquid Crystal Display

LoG: Laplacian of Gaussian

LSF: Line Spread Function

MTF: Modulation Transfer Function

MTFA: Modulation Transfer Function Area

NPS: Noise Power Spectrum

OECF: Opto-Electronic Conversion Function

PIC: Perceived Information Capacity

RGB: Red, Green, Blue

ROI: Region-of-Interest

s-CIELAB: spatial CIELAB

SFR: Spatial Frequency Response

SLR: Single Lens Reflex

SNR: Signal-to-Noise Ratio

SMTA: System Modulation Transfer Acutance

SQRI: Square-root Integral

SRS: Standard Reference Stimuli

sRGB: standard Red, Green, Blue colour space

TFT: Thin Film Transistor

TIFF: Tagged Image File Format

VDP: Visible Differences Predictor

Chapter 1

Introduction

Image quality evaluation is a key part in designing better imaging systems and deciding on the quality standards in image processing, e.g. the ratio of image compression. In general, the quality of an image is assessed using subjective (psychophysical) or objective (physical) scaling [1] (pp. 371-373).

The two scaling methods have their own benefits and drawbacks. Subjective scaling directly correlates to perceived image quality but is time-consuming, costly and complicated to implement. On the other hand, objective scaling is quick and less complex to implement but does not always provide reliable scales [2] (pp. 564). For example, the objective scales have been proven to successfully predict image quality in the laboratory, yet they have been poor predictors of quality when applied to digital systems in the field for some images, away from the laboratory.

Yendrikhovskij [3] (pp. 363) has noted that the current image technology does everything except one final and very crucial step, image quality appraisal. Many image quality models, mathematical formulas that are capable of predicting human perceptions of quality, have been proposed over the last fifty years [1] (pp. 371). However, as of yet, no definitive objective (physical) scaling method has been put forward [1] (pp. 390), [4] (pp. 3315).

Currently, the objective (physical) scaling methods can be classified into two groups [5] (pp. 12): i) the image-dependent (device-independent) models, and ii) the device-dependent (image-independent) models.

Image-dependent models attempt to predict the human perception of the image itself without knowledge of the imaging system's characteristics. These models are based on the difference between pairs of images, which involve a comparison of an image with either an ideal image or a non-ideal image.

If the reference image is the ideal image, then the compared image will have a lower quality scale. This is called the impairment approach [6] (pp. 251). An example is Daly's Visible Differences Predictor (VDP) [7] at image compression investigation. The ideal image is the original version, and the compared images are the compressed versions.

In the image-dependent models that use a comparison of an image with a non-ideal image, the quality result of the compared image changes as an increase or a decrease [8] (pp. 16). An example is image Color Appearance Model (iCAM) [9]. The non-ideal image is the original version, and the compared images are the image versions that vary in sharpness, resolution, noise or contrast. Johnson [8] (pp. 138) notes "what is interesting is this type of modelling can start to predict both magnitude and direction of the experimental sharpness scale."

These image-dependent models have an actual benefit in dealing with the problem of *scene dependency*. However, they are often unable to measure quality based on the imaging system variables.

The image-dependent modelling can be thought of as travelling throughout the left-hand side of the *Image Quality Circle,* as shown in Figure 1-1.

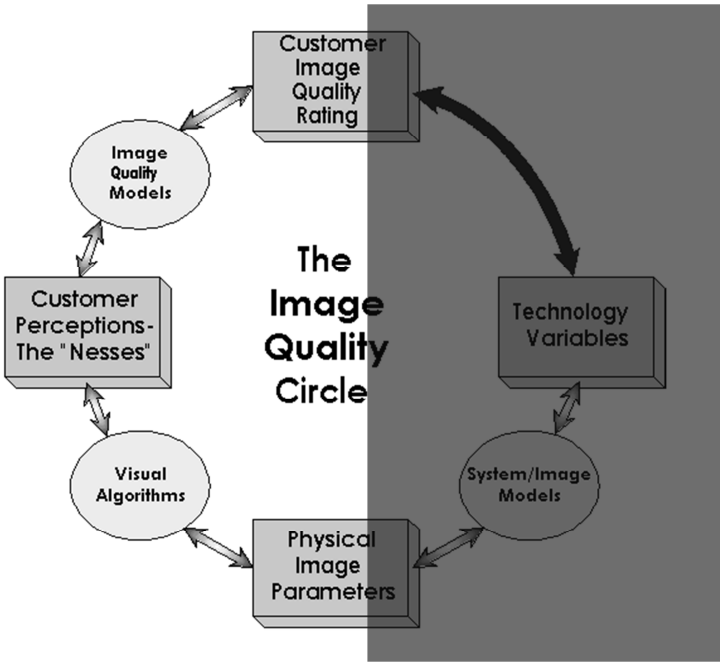


Figure 1-1.Image-dependent paths in the Image Quality Circle appear on the left-hand side (adapted from Fairchild [5] (pp. 16)) and (Image Quality Circle adapted from Engeldrum [6] (pp. 251)).

The second group models focus on an element of device-dependent image quality. This is applied in a straightforward fashion for the measurement of image quality of various system variables, such as the modulation transfer function (MTF), the noise power spectrum (NPS) and the gamma (γ).

These models are extremely powerful tools for measuring and predicting quality according to system variables, when knowledge of the images used is available. However, if knowledge of the image used is not available, these models have a well-known drawback, the scene dependency of image quality [10] (pp. 259), [11] (pp. 25). This is the reason that the models perform well on individual scenes and test targets, but provide lower correlation with subjective assessments when working with a variety of scenes.

The device-dependent image quality modelling falls into the right-hand side of the Image Quality Circle, illustrated in Figure 1-2.

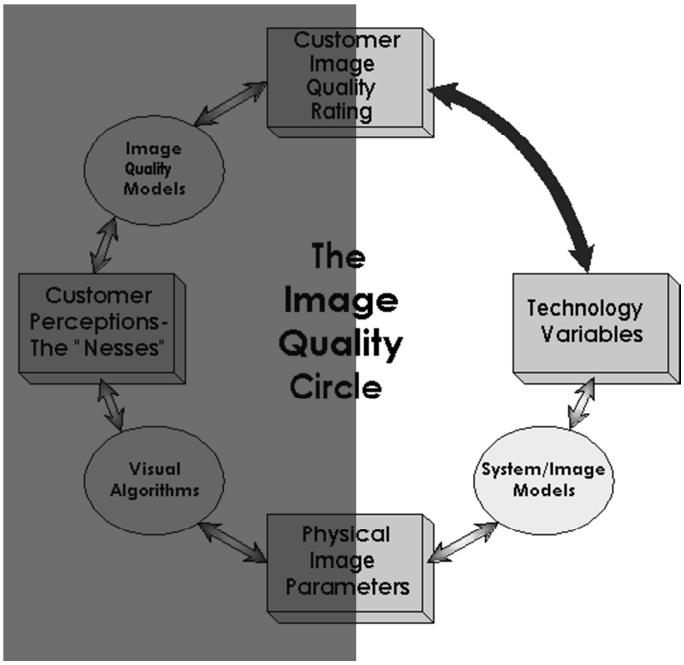


Figure 1-2. Device-dependent paths in the Image Quality Circle appear on the right-hand side (adapted from Fairchild [5] (pp. 13)) and (Image Quality Circle adapted from Engeldrum [6] (pp. 251)).

This study is concerned with device-dependent image quality models. It investigates the problem of scene dependency in device-dependent image quality models. In addition, it examines a solution with respect to scene dependency in device-dependent image quality models.

1.1 The impact of scenes on the perception of image quality

Image quality researches have confirmed that the scene content of the test images affects the observer's judgement [12] (pp. 9). For example, observers judged differently the sharpness of portraits and landscapes [13] (pp. 28). Portraits with low sharpness are usually preferred over over-sharpened portraits, while the opposite applies to landscapes.

Keelan [14] (pp. 131-135) has shown that the digital artefact of streaking is more evident in clear-sky image areas (i.e. relatively uniform, light areas) than in image areas of high-frequency signals and in extensive dark areas, which visually mask that streaking. Similarly, for a given print granularity, it has been shown that graininess (i.e. a subjective measure of photographic granularity) usually decreases with print density [15] (pp. 310), and hence dark areas in prints are less visually susceptible to the artefact. Also, scene characteristics, such as the spatial distribution of subjects [16] (pp. 663) and camera to subject distance [12] (pp. 13), have been shown to be important in the observer's preferences.

Scene dependency makes it difficult to design device-dependent image quality models, as well as to validate the models.

There are several ways of overcoming the problems caused by scene dependency [10] (pp. 262). One commonly employed method is to exclude subjective results obtained from non-standard scenes, when objective quality modelling. The validation of the models then uses the different set of standard scenes.

Another common technique is to employ a representative set (e.g. International Standards Organisation (ISO) 20462-3: 2005 set of test scenes [17] and 'Lena'), when objective quality modelling and validating. Nowadays many experimenters employ this technique [10] (pp. 262).

These, however, do not effectively represent the range and variety of different scenes that photographers, artists and consumers may wish to record and reproduce faithfully [10] (pp. 262-264).

1.2 Scene classification with respect to image quality

One possible way of overcoming the problem of scene dependency is a scene classification with respect to image quality. Keelan [14] (pp. 147) notes that "one way to characterize the variability associated with …scene susceptibility is to classify scenes … into small number of groups." The classification he proposed is as follows: 1) most

susceptible scenes 25%, 2) least susceptible scenes 25% and 3) intermediately susceptible scenes 50%.

In addition, Triantaphillidou *et al* [10] (pp. 269) proposed a method for test scene classification, which uses objective *scene descriptors* that correlate with subjective criteria on *scene susceptibility* to image quality attributes. Scene descriptors are derived to describe inherent scene properties that human observers refer to when they judge the quality of images.

Some studies consider scene classification either inspection of similarity [18, 19, 20] or clustering for image indexing [21, 22, 23, 24]. For example, Allen *et al* [18] (pp. 253-257) classified scenes into relatively identical groups at image quality, which was decreased by image compression. Their work was an effort to classify scenes into five groups by inspection. Teeselink *et al* [19] (pp. 553) also classified scenes by inspection. The study found that observers classify images into categories that reflect scene content. Mojsilovic and Rogowitz [20] (pp. 18-19) similarly classified scenes using inspection. The study demonstrated that image semantics play a large role in determining image similarity.

Yendrikhovskij [21] (pp. 406-408) sought to classify colour images by a k-means clustering algorithm. The effort was based on the colour coordinates, such as CIELAB in statistics of natural images. Szummer and Picard [22] (pp. 44) proposed a method for distinguishing between indoor and outdoor scenes, using colour histograms, autoregressive texture models and discrete cosine transform information. Vailaya *et al* [23] (pp. 1924-1930), [24] (pp. 421) described a method to classify vacation images into landscape/city, indoors/outdoors and sunset/mountain/forest, using colour histograms, colour coherence vectors (CCV), edge direction histograms and edge direction coherence vectors.

This study seeks to group scenes used in image quality investigation, based on scene descriptors that correlated with scene susceptibility in *sharpness* and *noisiness*. This will provide a fundamental basis for the selection of test-scenes and will allow meaningful grouping based on scene content in image quality measurement.

1.3 Aims and overview of the project

This research addresses the issue of scene dependency, which is a crucial factor in image quality. The main focus of this research is the impact of scene content on image quality.

This study of scene dependency in image quality has several goals. The first goal is to quantify the change of quality according to scene content. The second goal is to identify the problem of scene dependency in device-dependent image quality models. The final goal is to develop a solution for tackling the problem of scene dependency in device-dependent image quality models.

These aims will be addressed by considering the issues as set out in the following chapters:

Chapter 2 defines image quality and describes the differences between image fidelity and image quality. In addition, this chapter includes a detailed review of the image quality assessment, by subjective scaling and by objective scaling. In the subjective scaling section, the explanation of the nature of scene dependency and the perceived image quality that are effected by the blur and noise is included. The objective scaling section introduces image quality metrics and individual quality attribute assessments in terms of tone reproduction, colour reproduction, image resolution, sharpness and noise. This section also includes details on objective scene classification, based on scene descriptors to quantify various scene properties.

Experiments were conducted to characterise the imaging devices (i.e. a camera and a liquid crystal display (LCD)) used in the studies. The aim is to provide a means for producing accurate and reproducible results for image quality assessments. The tone and colour characterisation of the camera and the LCD devices are described in chapter 3 and chapter 4, respectively. Furthermore, the modulation transfer function (MTF) and noise characterisation of both devices are investigated in chapter 5. This chapter provides an account of the measurements of the soft display with the aid of a digital camera. The results of these measurements are presented in the chapter 5.

Chapter 6 deals with a series of subjective (psychophysical) experiments that investigate the effect of scene content on the perceived image quality. A novel approach is devised

for this investigation. The overall image quality is rated directly by the observer, and the result is then analysed to get the perceptual constraints that determine image quality. This chapter provides a detailed description of the experimental method used, followed by the results from the subjective experiments.

Chapter 7 describes an investigation of scene classification with respect to image quality. It involves 1) the investigation of various scene descriptors, derived to describe scene properties that influence image quality, 2) the correlation between scene descriptors and scene susceptibility parameters (from the results of chapter 6), and 3) the application of k-means clustering, using the selected scene descriptors for test-scene grouping. This chapter provides a detailed k-means clustering method with respect to image quality, followed by the results. The results from this chapter are employed to improve objective (physical) image quality scaling in chapter 8.

Chapter 8 sets out the details of an Effective Pictorial Information Capacity (EPIC) metric [74] and a Perceived Information Capacity (PIC) metric [71]. In addition, the various comparisons between EPIC and PIC results and perceived quality scales from chapter 6 are described. The comparisons are used to assess the success of objective device-dependent quality predictions. In addition, this chapter describes the method of improvement of the EPIC and PIC metrics, using the scene classification from chapter 7. Finally, the validity of the implementation is described.

Chapter 9 gives a discussion based on the results of the work as the system performance, subjective and objective image quality assessments. The conclusions and the recommendations for further work are presented in chapter 10.

Chapter 2

Image quality

This chapter describes the concept of image quality. A detailed account of image quality measurements via subjective (psychophysical) and objective (physical) evaluations is also presented. Some of these measurements have been applied in the experimental part of this research.

2.1 Image quality concept

According to ISO 12231: 2005 [25] (pp. 30), image quality is “an impression of the overall merit or excellence of an image, as perceived by an observer neither associated with the act of photography, nor closely involved with the subject matter depicted.” Although the definition is approved by ISO 12231: 2005, it cannot comprehensively explain the broader concept of image quality. This definition is slightly restricted. Keelan [26] (pp. 9) has pointed out “this narrow definition of image quality, which is based on third-party assessment, captures the artifactual, preferential and aesthetic attributes, but excludes personal attributes.” Jacobson [27] (pp. 7) has noted that “image quality has no single, unique definition yet.”

Furthermore, the concept of image quality is often confused with image fidelity. Therefore, it is necessary to distinguish between them. Silverstain and Farrell [28] (pp. 881) have stated that image fidelity is the ability to discriminate between two images, while image quality is the preference for one image over another. Klein [29] (pp. 75-76)

has distinguished between image fidelity and image quality, as the difference between the visibility of a factor and the degree to which that factor is bothersome. Berns [30] (pp. 107-108) has described the difference between image fidelity and image quality in the colour reproduction, as explaining that the fidelity of colour reproduction is concerned only with the least noticeable visible change, while quality of colour reproduction is far more than the least noticeable visible change.

2.2 Image quality assessment

Jacobson [27] (pp. 7) has noted that observers "are able to decide almost instantly whether a particular image is of good or poor quality, but for us to quantify how good an image is, and the scale of quality is far more difficult." Yendrikhovskij [3] (pp. 363) has mentioned that the "current image technology does everything except one final and very important step-image quality appraisal." Thus, quantifying our own preferences yields a fascinating scientific challenge.

Engeldrum [31] (pp. 312) has explained image quality evaluations in the form of an *Image Quality Circle* (Figure 2-1). It describes subjective (psychophysical) and objective (physical) image quality assessments. The Image Quality Circle involves four basic elements: 'customer image quality rating', 'customer perceptions', 'physical image parameters' and 'technology variables'.

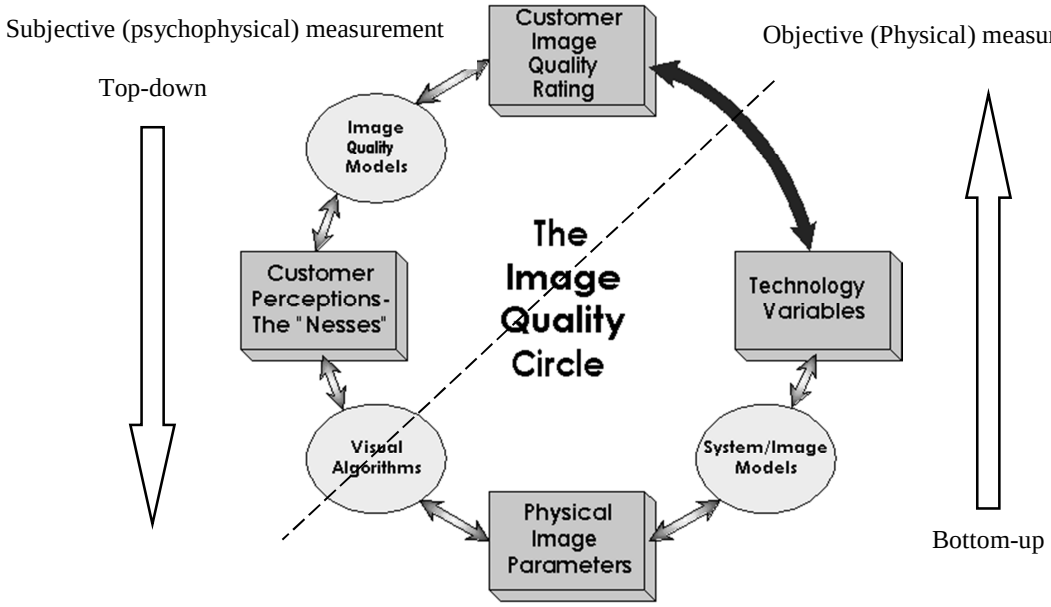


Figure 2-1. Image Quality Circle and approaches to quantifying image quality (adapted from Engeldrum [31] (pp. 312)).

The 'customer image quality rating' represents the customer's opinion of image quality. It consists of 'customer perceptions', such as graininess, sharpness and brightness. These perceptual attributes (Table 2-1) are related to the 'physical image parameters' which are evaluated by physical measurements (Table 2-2). The 'technology variables' cover an extensive number of variables. They are either fixed by the system being used or can be changed as part of the control of the system being optimised [31] (pp. 313).

Attributes	Perceptual description
Tone	The macroscopic contrast in an image.
Colour	The visual sensation in lightness, chrominance and saturation in an image.
Sharpness	The microscopic contrast in an image.
Resolution	The ability to depict spatial picture detail in an image.
Noise	The random and non-random spurious information in an image.

Table 2-1. Image attributes and their perceptual descriptions (adapted from Ford [32] (pp. 32) and Triantaphillidou [33] (pp. 38)).

Attributes	Physical image parameters
Tone	Gamma, Dynamic range, Tone reproduction curve (Characteristic curve), Histogram
Colour	Chromaticity values (*xyY*, $L^*u^*v^*$, $L^*a^*b^*$ etc.), Spectral power distribution, Pixel values, CIE colour differences
Sharpness	Acutance, Point spread function, Line spread function, Edge spread function, Modulation transfer function
Resolution	Resolving power, Image cell (cycles/mm, pixels/inch), dots per inch (dpi)
Noise	Granularity, Noise power spectrum (NPS), Autocorrelation function
Others	Information capacity, Detective quantum efficiency (DQE), Entropy, Power spectrum

Table 2-2. Physical measures relating to the objective evaluation of image quality (adapted from Triantaphillidou [33] (pp. 42) and Jacobson [34] (pp. 237)).

According to Yendrikhovskij [21] (pp. 396), subjective image quality can be studied from two perspectives: bottom-up and top-down (cf. Figure 2-1). The former is largely based on experimental research of the attributes of image quality. The latter is the process of analytical research of the perceptual constraints that determine image quality. Both approaches are valuable for image quality quantification.

2.2.1 Subjective evaluation

In the bottom-up perspective, the image quality rating is formed by the combination of individual visual attribute scales [31] (pp. 318), [35] (pp. 33), which are characteristics of an image that we see [36] (pp. 11-12). However, the individual attribute scaling has been a subject of discussion [33] (pp. 38), since image quality attributes are unlikely to be independent from other attributes.

There are many studies about the complicated relationships between attributes. For example, Biedermann and Frieser [37] (pp. 28) have studied the association between graininess and sharpness. Johnson and Fairchild [38] (pp. 28) have confirmed the relationship between sharpness and contrast. The individual attribute scaling is a relatively simplistic way of measuring image quality, because it does not take into consideration the complicated relationships between attributes.

In the top-down perspective, the overall image quality rating is directly judged by the observer. It produces “goodness” scales based on the overall image impression. The result can be then analysed to get the perceptual constraints that determine image quality. This is especially beneficial, because it does not require scaling of individual attributes and does not require the assumption that observers can see the quality attribute independently (cf. Chapter 6). Engeldrum [11] (pp. 22-23) has noted “it is common in image quality scaling studies that the quality judgement varies due to the variation of only one “ness.” The resulting scale from this sample set will not be one of image quality, although it may be labelled as such, but a scale of the single “ness” that varies in the sample set. Much care is needed in identifying the “nesses” in a sample set in order to avoid these pitfalls.”

2.2.1.1 Subjective scaling

The visual judgements produced by psychophysical scaling are quantified using one of the following scales: nominal, ordinal, interval or ratio [39] (pp. 678-680). From nominal to ratio scales, the information provided by the scale increases and the statistical operation complexity also increases. The ratio scales embody all the properties of the previous three scales.

- Nominal scales: These scales are used as indices for objects with numbers or names. These numbers or names have no meaning other than to identify the different items. An example of this type of scale would be the image categorises, such as architectural, natural scenes, portraits.
- Ordinal scales: These scales categorise variables associated with order. Items can be ranked in ascending or descending order, based on the magnitude of a certain visual attribute. However, there is no quantification information about the meaning of distances between items. The only mathematical operation that is valid for an ordinal scale is the use of the greater-than or less-than operators. An example of this type of scale would be the ordering from ‘best’ to ‘worst’ quality. The resulting scale would only reveal that one was greater than or less than others, but there would be no information as to how much greater or less.

- Interval scales: These scales have the property of distance to the ordinal scales. The mathematical operation that is valid for the interval scale is the use of the greater-than or less-than operators, as well as the differences between them. For example, in this type of scale if two images are separated by one unit, (e.g. just noticeable difference (JND) quality) away from one point on the scale, equal differences in scale values represent equal perceptual differences between the images.
- Ratio scales: These scales have all the properties of the previous three scales, with the addition of an origin. Thus, these scales hold the most mathematical power of all the scales. The mathematical operation that is valid for a ratio scale is the use of the greater-than or less-than operators, as well as the differences between them and an origin. However, it is argued that the ratio scales may not be useable in image quality, because the concept of the origin point provides philosophical and experimental difficulties. For example, what are the meanings of origin image quality in isolation assessment (assessment of single image)? [1] (pp. 380-381).

There are several scaling techniques to generate scales [40] (pp. 1115-1128). Most common techniques are as follows:

- Paired-comparison method: This is based on the Law of Comparative Judgement. During a test, observers are asked to indicate their preferred image, comparing displayed images (side-to-side) or prints. This method is a powerful technique for generating interval scales [40] (pp. 1117). It is particularly suited to assessing image quality when precise scalability is required. However, a large number of sample combinations can cause excessive observer stress, which can affect the accuracy and repeatability of results [41] (pp. v).
- Categorical method: The observer is presented with one stimuli image at a time, in identical viewing conditions. The observer is then asked to place the image into a specific category, e.g. International Telecommunication Union -Radiocommunication Sector (ITU-R). BT 500-12: 2009 [42] (pp. 18): 1. Bad, 2. Poor, 3. Fair, 4. Good, 5. Excellent, ISO 20462-2: 2005 [41] (pp. 2-3): 1. Favourable, 2. Acceptable, 3. Just acceptable, 4. Unacceptable, 5. Poor. This technique is useful when dealing with a

large number of samples. For this reason, this method was used in this study (cf. Chapter 6). Using this method, it is possible to derive ordinal and interval scales [40] (pp. 1119). The advantages of the category method include low stress and high stability [41] (pp. v). However, its scalability within a category is less precise, due to *range effect*. Cookingham [43] (pp. 90) has remarked that "observers seem to want to use most of the categories, perhaps so that they feel they are making some distinctions among samples, but usually they do not use the end categories much or at all, possibly in case a sample appears that is much higher or lower quality than seen previously."

- Rank ordering method: A series of stimuli images are presented, and the observer is asked to arrange the series in order of increasing or decreasing magnitudes. It is easily applied to the assessment of printed images. In practice, the number of printed images is limited to about 16 [44] (pp. 367). It is, however, a rather impractical method for assessing displayed images, since displays are not large enough to present all images at a reasonable size uniformly [33] (pp. 39).
- Magnitude estimation method: First, an image is presented to the observer, and the participant assigns a number to it. This image becomes a reference image throughout the test, i.e. an anchor image. Other images are then judged against the reference image. This method largely avoids being interpreted categories and complicated transformations [43] (pp. 91). However, results from this method are extremely difficult to replicate when the psychophysical experiments are conducted using non-expert observers [41] (pp. v). This is because the method is significantly increasing the amount of observer effort required [45] (pp. 15).

ISO 20462-2: 2005 [41] and ISO 20462-3: 2005 [17] describe two relatively new scaling methods.

- Triplet comparison method: It comprises of two steps, the "category step" and the "triplet comparison step." Firstly, the observer is asked to place the image into one of three categories, such as favourable, acceptable and unacceptable. The "category step" is to reduce the number of the samples to an appropriate number which is determined

by the purpose of the experiment. Secondly, the observer is asked again to place the sampled images into five categories, such as favourable, acceptable, just acceptable, unacceptable and poor. This step is conducted using the simultaneous scaling of three test stimuli at a time. The "triplet comparison step" derives a precise scaling [41] (pp. 2).

It satisfies the following requirements: it enables a large number of samples to be examined, provides precise scalability, provides low observer stress, suitable for ordinary (non-expert) observers, provides high repeatability of the results [41] (pp. 2). It has been proven to be effective at reducing the assessment time by about one-third, when compared with the paired-comparison method [41] (pp. 4).

- Quality ruler method: This is a technique for obtaining quality or attribute values for a test stimulus against a series of *univariate reference stimuli*, describing a series of test or reference stimuli that vary only in a single attribute of image quality. The observer is provided with a series of closely spaced stimuli of known separation (usually one to three JNDs), which vary in a single attribute of image quality and depict a single scene (often the same scene as that of the test stimulus). Other images are then judged against the ruler images [17] (pp. 5-9).

  One advantage of the method is that the reference stimuli are calibrated against a fixed standard numerical scale of quality, *scene-dependent ruler calibration* [17] (pp. 11). The standard reference stimuli (SRS) values are published on the I3A website (www.i3a.org, accessed on August, 2008).

More details of the experimental methods can be found in various texts [17, 40, 41, 42].

There are a number of factors that influence the results of psychophysical scaling: i) the choice of test samples, ii) the selection of observers, iii) the question including the instructions, iv) the viewing conditions and v) the duration of the experiment.

i) The choice of test samples: Engeldrum [11] (pp. 21) has identified four issues for consideration when selecting test samples.

1) What categories should the samples represent?

2) What range and distribution of “nesses” should the sample set contain?

3) What image size should be used?

4) What image content or image features should the samples contain?

When selecting test samples, the first consideration is the way in which images are selected - *sample of imagery*. Bartleson [46] (pp. 443-446) proposed five categories to describe samples of imagery: the random and independent sample, the stratified sample, the contrast sample, the purposeful sample and the incidental sample.

- The random and independent sample: Each image has an equal chance of being selected, and selection of one image has no influence on the selection of another. Although statistically interesting, it is difficult to implement in the real practice.
- The stratified sample: This is a more practical method as digital images with subclasses of portraits, landscapes and so on. It is a widely available method.
- The contrast sample: There is an interest in knowing the quality requirements or performance of a particular imaging device with respect to some class of imagery. This is common in a product development environment. For example, scenes with dominant edges may be selected in an investigation of the sharpness of an imaging device.
- The purposeful sample: This is either representative of a segment of the population or independently variable with regard to a chosen attribute. A purposeful sampling can be useful during product design. For example, scenes may be selected in an investigation of the acceptability of the levels of compression.
- The incidental sample: This represents a special collection of unique images (e.g. a standard image set [17]) that cannot be added to.

ISO 20462-1: 2005 [45] (pp. 6) suggests that for a sample of imagery, at least a minimum of 3 scenes[1] shall be used (and preferably at least 6 scenes should be used).

[1] Scene is the “content or subject matter of an image, or a starting image form which multiple stimuli may be produced through different experimental treatments” [45] (pp. 4).

A small number of scenes do not probably effectively represent the range and variety of different scenes. Extensive number of scenes can be preferable.

The second consideration is the range and distribution of "nesses." ISO 20462-1: 2005 [45] (pp. 6) suggests describing the nature of the variation among the test stimuli (other than scene content) in both subjective terms (image quality attributes) and objective terms (stimulus treatment or generation).

However, the desired range and distribution of "nesses" is often difficult. This is because there is no rigid rule. It is generally to have equal numbers of samples that uniformly span the range of "nesses" [11] (pp. 23). Engeldrum [11] (pp. 23-24) has noted that "it is more than worthwhile to expend the effort to select or generate a sample set that meets the scaling study requirements."

The third consideration is the image size, which is a well-known factor in image quality judgements [47] (pp. 116-119), [48] (pp. 72-75). In general, it is easier to keep the size of the image in a scaling study constant, thus eliminating or minimizing its influence on observer judgement [11] (pp. 24).

Also, the consideration should be taken of the image content or image elements. These may be selected on the basis of having different tonal ranges, dominant lines and edges, high-frequency information, etc. [1] (pp. 383).

ii) The selection of the observers: The choice of observers is often based on the target viewing population [33] (pp. 41). This has been a subject of discussion, since some argued that expert observers produce more accurate results than non-experts [14] (pp. 140).

According to ISO 20462-1: 2005 [45] (pp. 6), at least 10 observers should participate in the scaling, although 20 observers are preferred. However, an extensive number of observers should not be required, since no advantage was seen in having a large number of observations in the human observer studies [49] (pp. 2453), [50] (pp. 436). Using a large number of observers makes the experiment time-consuming.

Observers should also be checked for normal vision characteristics, such as colour vision and visual acutance [45] (pp. 6).

All selection criteria and the number of observers taking part should be reported [45] (pp. 6).

iii) The question including the instructions: The question needs to be handled carefully to achieve meaningful and useful results. The question needs to include a statement of what is to be evaluated by the observer, as well as the mechanics of the experimental procedure [45] (pp. 6).

iv) The viewing conditions: ISO 20462-1: 2005 [45] (pp. 7) and ISO 3664: 2000 [51] (pp. 13) set out recommended display viewing conditions. Table 2-3 presents the recommended viewing conditions for scale accuracy and repeatability of experiments.

	ISO 20462-1: 2005	**ISO 3664: 2000**
Display luminance level	> 60 cd/m^2	> 75 cd/m^2
Display White point	CIE illuminant D50	CIE illuminant D65
Surround luminous reflectance	< 64 lx	< 64 lx

Table 2-3. Recommended viewing conditions for display
(adapted from ISO 20462-1: 2005 [45] (pp. 7) and ISO 3664: 2000 [51] (pp. 13)).

v) The duration of the experiment: The experimental duration should not exceed a maximum of 60 minutes to avoid observer fatigue [45] (pp. 7).

Other issues and factors should be reported, including the complexity of the experiment, the adaptation of the viewing conditions and the surrounding environment [33] (pp. 41).

2.2.1.2 Scene dependency

As mentioned before (in the choice of test samples), the results of psychophysical scaling are highly dependent on the context of the image used. Triantaphillidou *et al* [10] (pp. 260-262) have described three different origins of scene dependency.

i) Scene dependency resulting from the observer's quality criteria (i.e. observer's preference): An example of the scene dependency from the observer's preferences is that observers judged the sharpness of portraits and landscapes differently [13] (pp. 28). Portraits with low sharpness are preferred over over-sharpened portraits, while the opposite applies to landscapes. The reason being that 'soft focus' renders the skin smoother and thus more pleasant to the viewer, whereas strong lines and edges are usually preferred when they are sharp. Scene characteristics, such as the spatial distribution of subjects [16] (pp. 663) and camera to subject distance [12] (pp. 13), are crucial scene dependence parameters in the observer's preferences.

ii) Scene dependency due to a visibility of an artefact: This is known as scene susceptibility [14] (pp. 131), as imaged scene quality is liable to be influenced or harmed by the visibility of an artefact. Variations in scene susceptibility occur when the same objective amount of an artefact, such as noise, streaking or banding is present in images, but it is more or less evident in different types of scenes or different areas with the same scene. Keelan [14] (pp. 131-135) shows that the digital artefact of streaking is more evident in clear-sky image areas (i.e. relatively uniform, light areas) than in image areas of high-frequency signals and in extensive dark areas, which visually mask that streaking. Similarly, for a given print granularity, it has been shown that graininess (i.e. a subjective measure of photographic granularity) usually decreases with print density [15] (pp. 310), and hence dark areas in prints are less visually susceptible to the artefact.

iii) Scene dependency due to digital processes or image processing algorithms: A classical example is image compression [18] (pp. 253-257), [32] (pp. 144-146). Applying the same objective amount of compression (i.e. compression ratio) in two different images, one with mostly high and the other with mostly low-frequency information, will discard different quantities of information, since both discrete cosine transform (DCT) [52] (pp. xx-xxiii) and discrete wavelet transform (DWT) [53] (pp. 22) compression schemes discard mostly high spatial frequencies.

2.2.1.3 Effect of blur and noise on image quality

Perceived image quality has been studied using manipulated images, which adjust various parameters. Nijenhuis [54] (pp. 61) measured the perceived image quality of a blurred image on a cathode ray tube (CRT) monitor at several blur levels. The blurred image was varied by Gaussian blur filtering, using different standard deviations. In this study, the image was an artificial picture consisting of an evenly lit square area of 0.1 m ×0.1 m with a luminance of 45 cd/m^2. The background was 0.28 m× 0.28 m with a luminance of 15.7 cd/m^2. The viewing distance was 4 m, corresponding with a field size 4° ×4°. Figure 2-2 shows the results for blurred images.

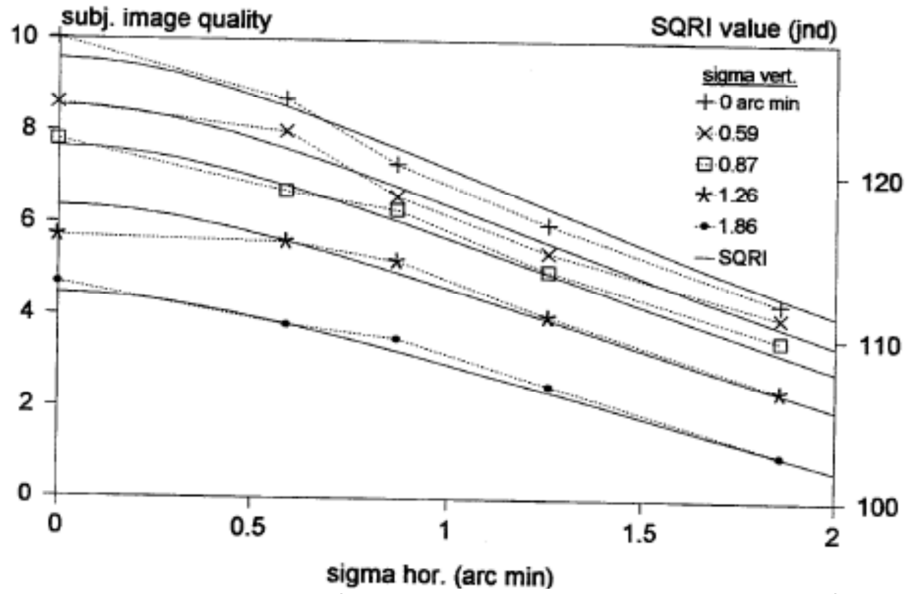


Figure 2-2. Subjective image quality of a blurred image at several combinations of horizontal and vertical directions. The broken line indicates the subjective image quality result and the solid line indicates the calculated SQRI (Square-root Integral) value
(produced by Nijenhuis [54] (pp. 61) and reproduced by Barten [55] (pp. 180)).

Kayargadde [56] (pp. 78- 96) measured the perceived noisiness, not the perceived image quality. In the study noise was added to a pure white field on a CRT monitor. The noise image was varied by Gaussian noise filtering, using different standard deviations. The images had a size of 0.17 m ×0.17 m and were viewed from a distance of 1.4 m corresponding with a field size of 7° × 7°. The luminance levels were 30 cd/m^2. Figure 2-3 shows the noisiness results.

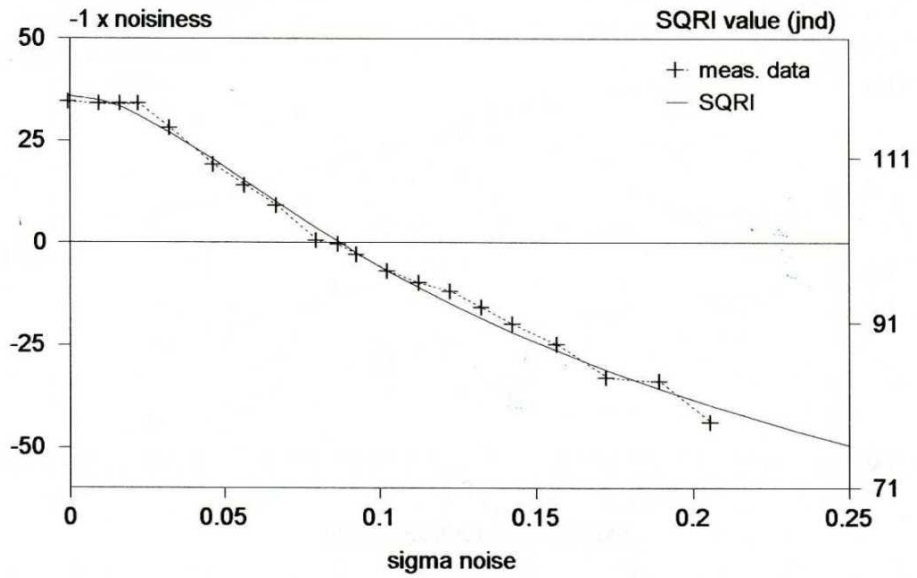


Figure 2-3. Scaled noisiness as a function of the sigma, measured at a white field on a CRT. The broken line indicates the subjective noisiness result and the solid line indicates the calculated SQRI value
(produced by Kayargadde [56] (pp. 87) and reproduced by Barten [55] (pp. 189)).

Kayargadde [56] (pp. 91-96) also investigated the effect of noise on two natural images (Figure 2-4). The noise image was generated by Gaussian noise filtering, using different standard deviations. The results were obtained under the following conditions: viewing distance 1.4 m, field size 9.8° × 9.8°, and image size 0.24 m ×0.24 m.

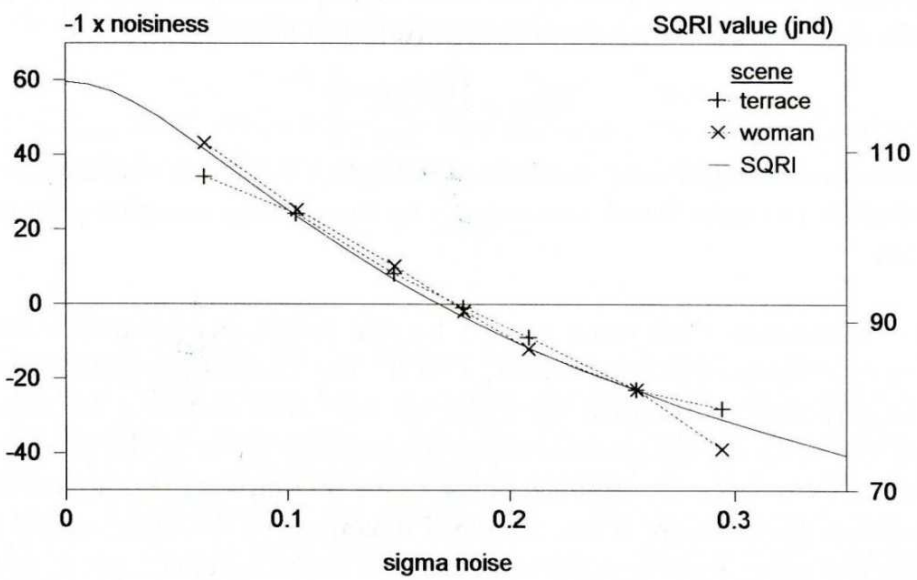


Figure 2-4. Scaled noisiness as a function of the sigma, measured for two natural scenes. The broken line indicates the subjective noisiness result and the solid line indicates the calculated SQRI value
(produced by Kayargadde [56] (pp. 92) and reproduced by Barten [55] (pp. 193)).

2.2.2 Objective evaluation

According to Komineck [57], “what the coding community desperately needs is an easy to compute error measure that accurately captures [the] subjective impression of human viewers.” Although image quality is the subjective impression of human viewers, it would require an objective measurement for quick and easy implementation.

2.2.2.1 Image quality metrics (IQMs)

Image quality metric (IQM) is the quantified measure of quality as numbers derived from physical measurements (cf. Table 2-2), which related to psychophysical attributes of the human eye [1] (pp. 371). A simplified diagram for IQM is shown in Figure 2-5.

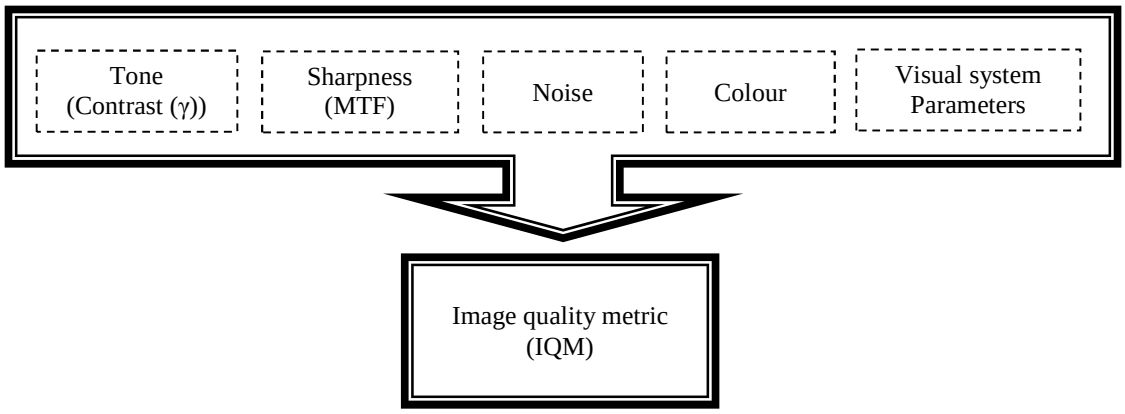


Figure 2-5. Simplified diagram of image quality metric
(adapted from Triantaphillidou and Jacobson [1] (pp. 372)).

A large number of different IQMs have been proposed in order to predict subjective image quality. However, individual metrics have a number of limitations, because of the multidimensional nature of image quality [58] (pp. 141). Nevertheless, these metrics have been used for image quality assessment since they work well when dealing with a single property, such as sharpness on an imaging system [1] (pp. 377-380). According to Farrell [59] (pp. 302), “when used properly, metrics can be powerful design evaluation tools.”

There have been various approaches for the universal physical measurement of image quality (Figure 2-6). Examples of IQMs are given in Table 2-4.

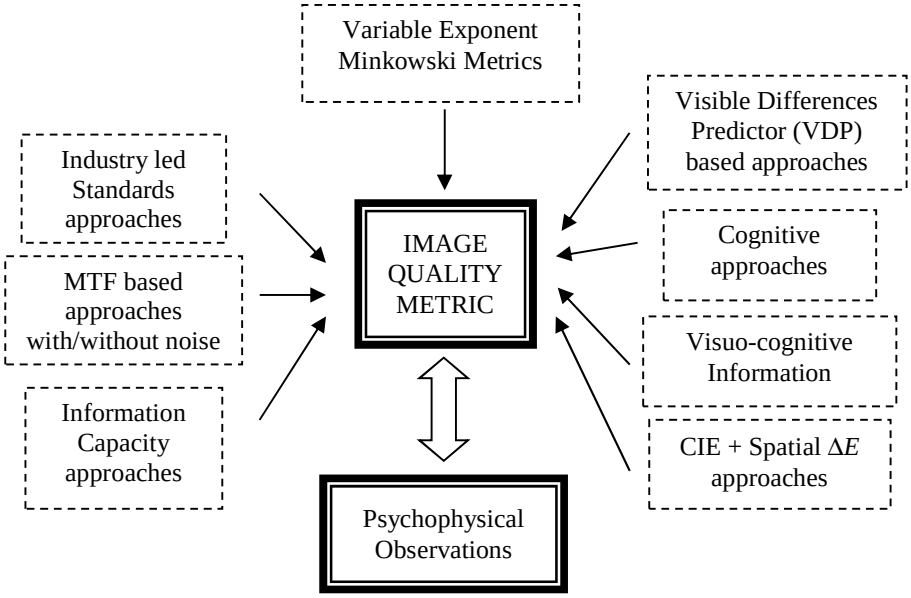


Figure 2-6. Various approaches to image quality metric (adapted from Jacobson [60] (pp. 55)).

Acronym	Full name
SMTA	System Modulation Transfer Acutance
CMTA	Cascaded Modulation Transfer Acutance
AMTA	A Modulation Transfer Acutance
MTFA	Modulation Transfer Function Area
SQRI	Square-root Integral
SNR	Signal-to-Noise Ratio
$SQRI_n$	Square-root Integral to include noise
PIC	Perceived Information Capacity
CRI	Colour Reproduction Index
s-CIELAB	Spatial CIELAB
iCAM	image Color Appearance Model
EPIC	Effective Pictorial Information Capacity

Table 2-4. Examples of image quality metrics (adapted from Jacobson [27] (pp. 7)).

- MTF based approaches with/ without noise: SMTA (system modulation transfer acutance) [61] (pp. 643-644) is one of the earliest proposals of a metric for sharpness. This sharpness metric was extended in to CMTA (cascaded modulation transfer acutance) [62] (pp. 1009) and AMTA (a modulation transfer acutance) [63] (pp. 125-128), which improved their correlation to the perceived sharpness.

  Another sharpness metric, which achieved wider acceptability as a standard in the USA for monochrome CRT displays [64] (pp. 17-18), is MTFA (modulation transfer function area) [65] (pp. 387). However, it has been shown to be a poor quality predictor when applied to digital systems [66] (pp. 348). [67] (pp. 69-98).

  This limitation of the MTFA metric led to the evolution of SQRI (square-root integral) of Barten [68] (pp. 2025) and its further modification to include noise, $SQRI_n$ (square-root integral to include noise) [69] (pp. 9-10), which was based on SNR (Signal-to-Noise Ratio) [70] (pp. 59).

  Another sharpness and noise metric based on SNR is known as PIC (perceived information capacity) [71] (pp. 7-10).

$$\mathrm{PIC} = \mathrm{k}_1 \left[\int_0^\infty \ln\left(1 + \frac{S(u)M_{eye}^2(u)}{N(u)M_{eye}^2(u)+N_{eye}(u)}\right)\frac{\mathrm{du}}{\mathrm{u}} \right]^{0.5} + \mathrm{k}_2 \qquad (2.1)$$

  where *S(u)* is the signal spectrum multiplied by the squared MTFs of each system component, $M_{eye}(u)$ is the MTF of the observer, $N(u)$ is the noise in the image, $N_{eye}(u)$ is a noise term for the observer, and k_1 and k_2 are constants.

- CIE + Spatial ΔE approaches: CRI (colour reproduction index) [72] (pp. 83-87) was the first colour IQM, which includes prediction of colour appearance under specified viewing conditions. Further metrics now exist, which allow the inclusion of the eye responses to the spatial frequencies of images, as well as their colour. The examples are Zhang and Wandell's s-CIELAB (spatial CIELAB) [73] and its modification, iCAM (image color appearance model) [9]. Although the s-CIELAB and iCAM are considered as image quality models, they might be extensions of fidelity measures. This is because the models are based on difference prediction.

- Information Capacity approach: EPIC (effective pictorial information capacity) [74] (pp. 7) is a sharpness and noisiness quality metric, based on information capacity theory [75]. The EPIC is derived from the *effective pixel dimension* in the image (*n*) and the number of *effective distinguishable levels* for each recording cell (*m*).

  The effective pixel dimension in the image (*n*) is calculated by:

$$n = \frac{A_{\mathrm{im}}}{\omega^2} \tag{2.2}$$

  where A_{im} is the area of the image, and ω is the effective pixel dimension.

  The number of effective distinguishable levels for each recording cell (dynamic range of a system) (*m*) is described by:

$$m \approx \frac{DS}{2\mathrm{k}\sigma} + 1 \tag{2.3}$$

  where *DS* is the difference between the maximum and minimum possible levels of the recording system, k is a constant and σ is the standard deviation in the imaging systems.

  The perceived capacity, *C*, is derived by:

$$C = n\log_2(m) \tag{2.4}$$

  The perceived capacity (*C*) is divided by the visual solid angle (Ω), which then yields the number of effective bits per visual steradian.

$$\mathrm{EPIC} = \frac{C}{\Omega} \tag{2.5}$$

$$\Omega = A_{\mathrm{im}}/r^2 \tag{2.6}$$

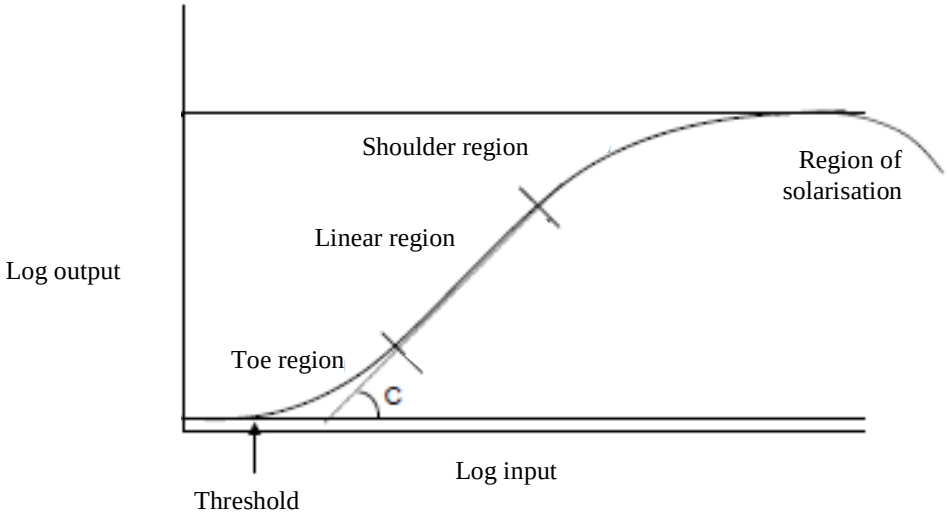
  where A_{im} is the area of the image, and *r* is the viewing distance.

2.2.2.2 Tone reproduction

Tone reproduction was studied firstly by Hurter and Driffield [76], and the theory of tone reproduction was established by Jones [77, 78, 79]. The terms were later formalised by Nelson [80]. It is concerned with the relationship between input-to-output intensities.

This relationship is described by the transfer function that involves the effective single parameter, gamma (γ).

The conventional photographic transfer function, known as the characteristic curve, is described by a function that relates the common logarithm of relative exposure and the reproduced density. The gamma (γ) is defined by the gradient of the straight line formed when plotting data on a logarithmic space (Figure 2-7) [81] (pp. 223-225).



$$\text{gamma}(\gamma) = \tan(c) \tag{2.7}$$

Figure 2-7. Typical characteristic curve and gamma
(adapted from Attridge [81] (pp. 224)).

The transfer functions for electronic still-picture cameras, known as the opto-electronic conversion function (OECF), are described by ISO 14524: 1999 [82]. The OECF is commonly described by the relationship between the original scene luminances (often expressed by luminance ratios, such as scene or print reflectance, or film transmittance) and the generated digital counts. The transfer function is often referred to a power function in linear-linear units. The equation is as follows [83] (pp. 384):

$$PV = o + gL^{\gamma_a} \tag{2.8}$$

where PV is the generated normalised pixel value, L is the input luminance ratio, o is the system offset, g is the system gain and γ_a represents the camera gamma.

The CRT transfer function is an approximate power function that describes the relationship between the input normalised pixel values and the generated normalised output luminances. The equation that commonly describes the CRT transfer function is as follows [84] (pp. 24):

$$L = o + gV^{\gamma_d} \tag{2.9}$$

where L is the displayed normalised luminance, V is the normalised input voltage, o and g are the offset (contrast) and gain (brightness) of the system. The γ_d is the gamma of the display system.

The relationship between the input pixel values and the output luminances in a LCD differs to that of the CRT [85] (pp. 609), [86] (pp. 2). The LCD native relationship is modelled as S-shape function [87] (pp. 191-193). Figure 2-8 shows the transfer functions of common CRT and LCD.

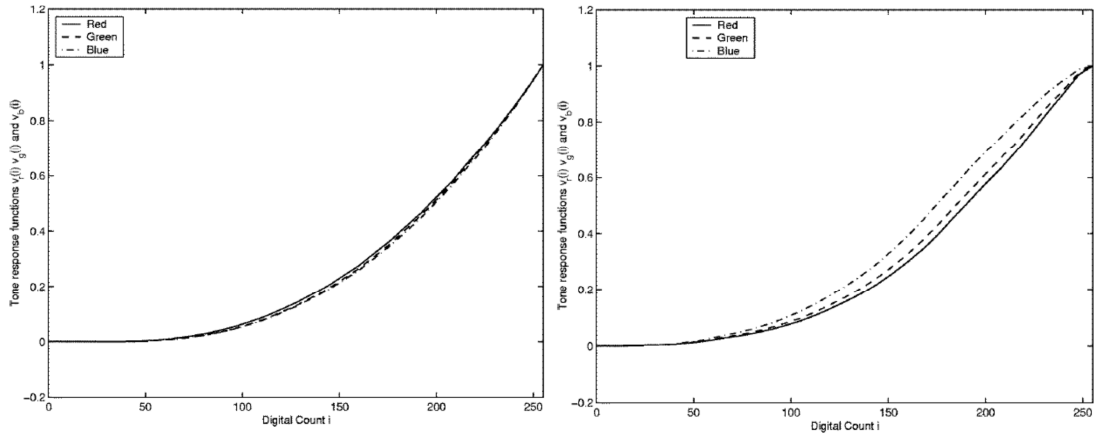


Figure 2-8. Typical transfer functions for CRT (left) and LCD (right)
(produced by Sharma [85] (pp. 612-614)).

Many LCD manufacturers build correction tables into the video card, so that the LCD's transfer functions mimic the CRT transfer functions [88] (pp. 15). Therefore, the above CRT equation (2.9) can be used to describe the transfer characteristics of such displays [83] (pp. 382).

In an imaging chain, it is possible to involve image encoding, e.g. sRGB and Adobe RGB. The sRGB transfer function is not a pure power function. At exceptionally low relative luminances, the transfer function is linear. At relative luminances larger than 0.003130, the encoding gamma is equal to the exponent 1/2.4, the offset is -0.055 and the gain is 1.055 [89] (pp. 12). The Adobe RGB 1998 transfer function is a pure power function, which has a gamma of 2.199 [90] (pp. 9-16). The encoding transfer functions are described in Appendix A. Figure 2-9 illustrates the sRGB and Adobe RGB 1998 transfer functions.

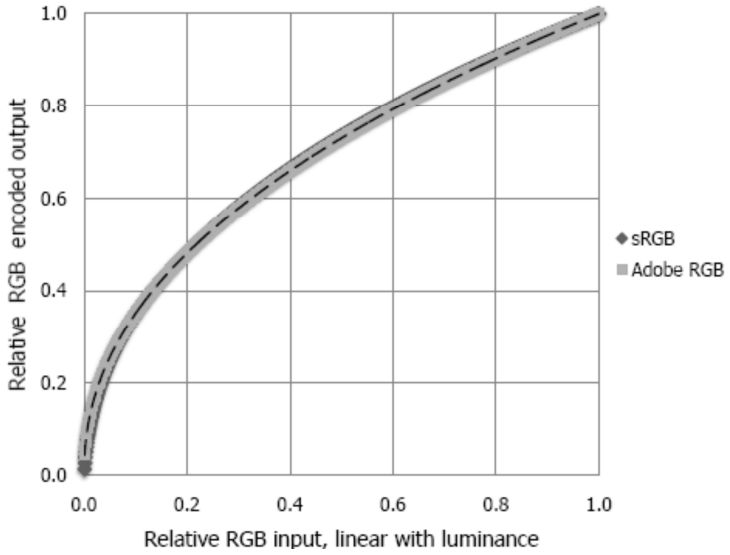


Figure 2-9. sRGB (solid line) and Adobe RGB 1998 (dotted line) transfer functions (produced by Triantaphillidou [83] (pp. 387)).

Jones [78] (pp. 235-239) introduced a descriptor of the overall tone reproduction. The overall gamma ($\gamma_{overall}$) can be calculated by multiplying each component gamma value of the imaging chain. Following equation sets out the way of calculating the overall gamma in imaging systems:

$$\gamma_{overall} = \gamma_{system_1} \times \gamma_{system_2} \times \ldots \times \gamma_{system_n} \tag{2.10}$$

Optimal overall gamma (γ_o) is achieved by a gamma correction (γ_c) with the overall gamma ($\gamma_{overall}$). The optimal overall gamma can be expressed by [91] (pp. 205):

$$\gamma_o = \gamma_{overall} \times \gamma_c \tag{2.11}$$

The optimal overall gamma can only be considered when the viewing conditions are known. Typically, the optimal overall gamma values vary between 1.0 and 1.5. Some of the reported optimal overall gamma values are listed as follows [83] (pp. 379):

- Reflection prints: close to 1.1 to compensate for flare in bright viewing conditions.
- Viewing monitor/ television: approximately 1.25 in dim environments.
- Transparencies projected motion pictures: 1.5 in dark condition (typically a gamma of 1.6 to compensate also for flare).
- Displayed digital images: between 1.1 and 1.15 in office settings.
- Displayed sRGB images on CRT: 1.125 for dim illumination conditions and flare.

2.2.2.3 Colour reproduction

The colour of an object depends on three components, the light source, the chemical and physical properties of the object and the human visual system (HVS) (Figure 2-10) [92] (pp. 54). The first component is the light source which provides the electromagnetic energy. This energy is modulated by the physical and chemical properties of the object. The modulated energy is perceived by the HVS.

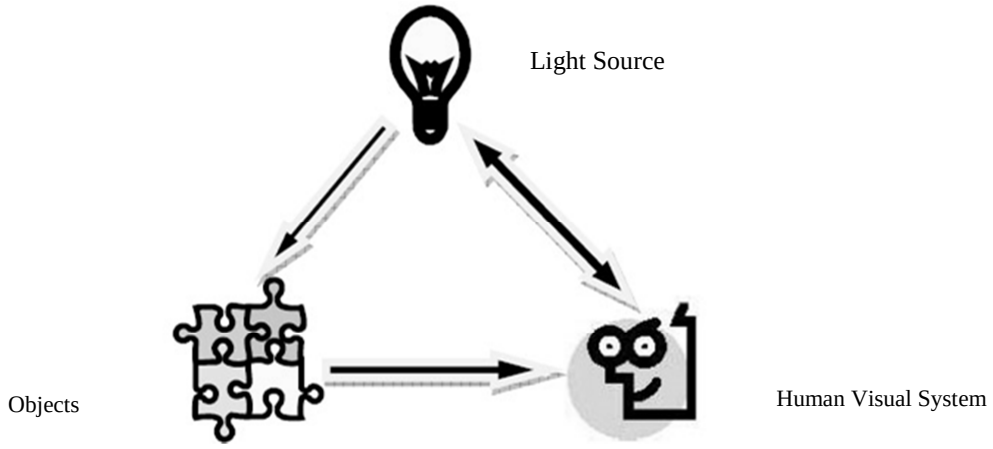


Figure 2-10. The triangle of colour. Colour exists due to the interaction of light sources, objects, and the human visual system
(adapted from Fairchild [92] (pp. 55)).

The description of colour in images is traditionally based on spectral colour definition or *colorimetry* [93] (pp. 205-207). Colorimetry is based on rules of matching colour using additive colour mixtures, i.e. it is possible to match all the colours of the spectrum. When this is done the result is represented by three curves, referred to as the *colour matching functions* [94, 95].

In 1931, the Commission Internationale de l'Eclairage (CIE) [96] first assigned colour matching functions, which represent the CIE 1931 2° *standard colorimetric observer* for the *XYZ* primaries.

$$X = \mathrm{k} \int_{\lambda} R(\lambda) I(\lambda) \bar{x}(\lambda) \mathrm{d}\lambda \tag{2.12}$$

$$Y = \mathrm{k} \int_{\lambda} R(\lambda) I(\lambda) \bar{y}(\lambda) \mathrm{d}\lambda \tag{2.13}$$

$$Z = \mathrm{k} \int_{\lambda} R(\lambda) I(\lambda) \bar{z}(\lambda) \mathrm{d}\lambda \tag{2.14}$$

where k is a normalizing constant which is defined differently for relative and absolute colorimetry, $R(\lambda)$ is the spectral data of the coloured object, $I(\lambda)$ is the spectral power distribution of the irradiating illuminant $\bar{x}, \bar{y}$ and $\bar{z}$ are the colour matching functions of the CIE standard colorimetric 2° observer, and λ is the range of wavelength of the visible electromagnetic spectrum (CIE publication 15.2 [97] (pp. 23) recommends that for practical purposes 5nm intervals over the range 380-780 nm be used).

In 1964, the CIE suggested the primaries X_{10}, Y_{10} and Z_{10}, basing them on the colour matching functions of the standard colorimetric 10° observer. The comparison between the colour matching functions for the 2° and the 10° observers is illustrated in Figure 2-11 [98] (pp. 99).

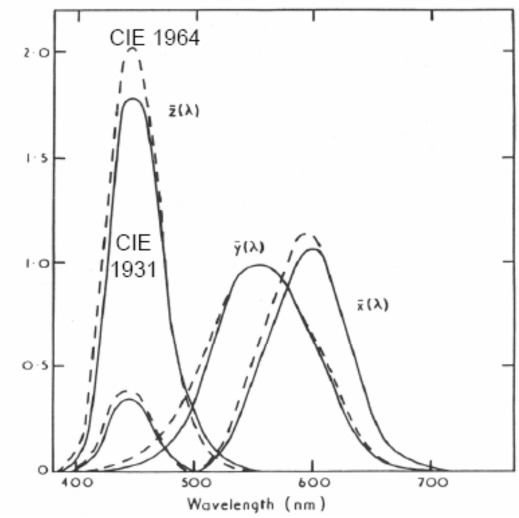


Figure 2-11. Colour matching functions related to the standard observer CIE 1931 (visual field of 2°, continuous line) and to the standard observer CIE 1964 (visual field of 10°, dotted line) (produced by Hunt [98] (pp. 99)).

The *x*, *y* and *z* chromaticity coordinates are calculated to determine the stimulus chromaticity in the CIE *XYZ* system. The sum of three chromaticity coordinates is one. Chromaticity coordinates are plotted in CIE *xy* chromaticity diagram. The diagram shows only two-dimensional stimuli information. The *Y* tristimulus value is usually reported providing full colour information (*x*,*y*,*Y*).

$$x = X/(X + Y + Z) \tag{2.15}$$

$$y = Y/(X + Y + Z) \tag{2.16}$$

$$z = Z/(X + Y + Z) \tag{2.17}$$

The CIE *xy* chromaticity diagram does not provide perceptual uniformity. A lot of efforts were invested into producing a perceptually uniform diagram. This is achieved by the CIE *u'v'* chromaticity diagram (the CIE 1976 uniform chromaticity diagram). The CIE *u'*,*v'* coordinates are derived from either *X*,*Y*,*Z* or *x*,*y*,*z* by the following equation:

$$u' = 4X/(X + 15Y + 3Z) = 4x/(-2x + 12y + 3) \tag{2.18}$$

$$v' = 9Y/(X + 15Y + 3Z) = 9y/(-2x + 12y + 3) \tag{2.19}$$

where the third chromaticity coordinate, *w'* is equal to 1-*u'*-*v'*.

In 1976, the CIE 1976 *L*u*v** (CIELUV) and CIE 1976 *L*a*b** (CIELAB) colour spaces were recommended to represent colour in three-dimensions that approximately correlated with the perceived lightness, chroma and hue of a stimulus [99] (pp. 29-33).

- Lightness is the brightness of an area judged relative to the brightness of a similarly illuminated reference white [100] (pp. 86 & 90).

$$\text{Lightness} = \text{Brightness}/\text{Brightness}(\text{White}) \tag{2.20}$$

- Chroma is the colourfulness of an area judged as a proportion of the brightness of a similarly illuminated reference white [100] (pp. 87 & 90).

$$\text{Chroma} = \text{Colourfulness}/\text{Brightness}(\text{White}) \tag{2.21}$$

- Hue is the attribute of a visual sensation according to which an area appears to be similar to one, or to proportions of two of the perceived colours, red, yellow, green and blue [100] (pp. 85).

The CIELAB colour space uses two colour coordinates *a**(approximate redness-greenness) and *b**(approximate yellowness-blueness), and a lightness coordinate *L** (Figure 2-12). These coordinates are derived from the *XYZ* tristimulus values of the stimulus and the X_n, Y_n, Z_n of a specified reference white.

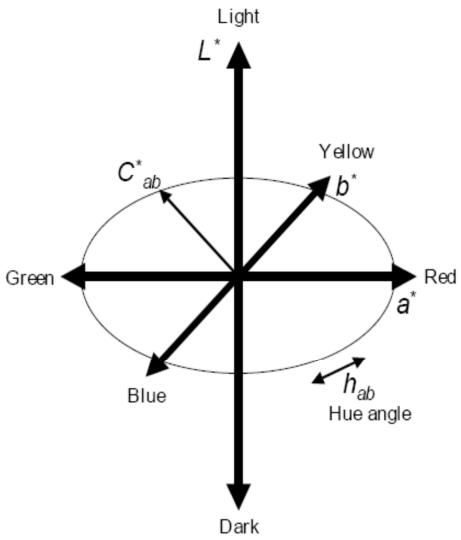


Figure 2-12. Cylindrical representation of CIELAB colour space
(produced by Fairchild [92] (pp. 80)).

$$L^* = 116\, f\left(\frac{Y}{Y_\mathrm{n}}\right) - 16 \tag{2.22}$$

$$a^* = 500\left[f\left(\frac{X}{X_\mathrm{n}}\right) - f\left(\frac{Y}{Y_\mathrm{n}}\right)\right] \tag{2.23}$$

$$b^* = 200\left[f\left(\frac{Y}{Y_\mathrm{n}}\right) - f\left(\frac{Z}{Z_\mathrm{n}}\right)\right] \tag{2.24}$$

$$f(x) = \begin{cases} x^{\frac{1}{3}} & x > 0.008856 \\ 7.787(x) + \frac{16}{116} & x \le 0.008856 \end{cases} \tag{2.25}$$

In the CIELAB space, the chroma, C^*_ab, and hue angle by h_ab are calculated the following equation:

$$C^*_\mathrm{ab} = \sqrt{a^{*2} + b^{*2}} \tag{2.26}$$

$$h_\mathrm{ab} = \tan^{-1}\left(\frac{a^*}{b^*}\right) \tag{2.27}$$

In the CIELAB space, colour differences (ΔE^*_ab) are measured as the Euclidean distances between the coordinates of the two stimuli (Figure 2-13). The colour differences can be broken down into the components of lightness, chroma and hue.

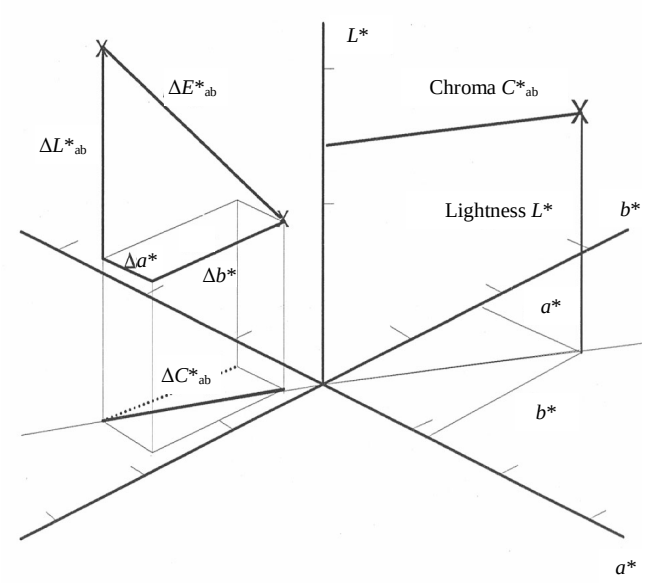


Figure 2-13. Geometrical illustration of the colour difference (ΔE^*_ab) in CIELAB colour space (adapted from Triantaphillidou [101] (pp. 32)).

$$\Delta E^*_{ab} = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \quad (2.28)$$

$$\Delta E^*_{ab} = \sqrt{\Delta L^{*2} + \Delta C^{*2}_{ab} + \Delta H^{*2}_{ab}} \quad (2.29)$$

Various modifications (e.g. CMC(l:c), CIE94, CIE DE2000) have been developed to improve the uniformity of colour difference measurement, based on the CIELAB colour space. The CMC(l:c) [102] (pp. 130) is a modification widely used in the field of textiles. The modified CIE94 (ΔE^*_{94}) [103] (pp. 8-11) utilises specified reference conditions, and it is used in various industries. It includes weighting functions, K_L, K_C and K_H for better representation of colour difference. The CIE DE2000 (ΔE_{00}) [104], [105] (pp. 348-350) was developed as an improvement to the CIE94 colour difference formula. It includes corrections for variation in colour difference perception dependent on lightness, chroma, hue and chroma-hue interaction. The CIE DE2000 has been retained using the same conditions as the CIE94 model, presented in Table 2-5.

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{K_L S_L}\right)^2 + \left(\frac{\Delta C'}{K_C S_C}\right)^2 + \left(\frac{\Delta H'}{K_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{K_C S_C}\right)\left(\frac{\Delta H'}{K_H S_H}\right)} \quad (2.30)$$

where $S_L = 1+0.015(L'_m-50)^2/[20+(L'_m-50)^2]^{1/2}$ with $L'_m = (L'_1 + L'_0)/2$

$S_C = 1+ 0.045(C'_m)$ with $C'_m = (C'_1 + C'_0)/2$

$S_H = 1+ 0.015(C'_m)\ (T)$

with $T = 1-0.17\cos(h'_m-30°)+0.24\cos(2h'_m)+ 0.32\cos(3h'_m+6°)- 0.20\cos(4h'_m-63°)$
with $h'_m = (h'_1 + h'_0)/2$

$R_T = -\sin(2\Delta\theta)R_c$ with $\Delta\theta = 30\exp[-((h'_m-275°)/25°)^2]$

$R_c = 2[C'^7_m/(C'^7_m+25^7)]^{1/2}$

$L' = L^*$, $a' = a^*(1+G)$, $b' = b^*$

$G = 0.5[1- ((C^{*7}_{ab,m}/(C^{*7}_{ab,m}+25^7))^{1/2}]$ with $C^*_{ab,m} = (C^*_{ab,1} + C^*_{ab,0})/2$

K_H, K_C and K_L are parametric factors, which may be chosen other than 1 if experimental conditions deviate from reference condition.

	Reference conditions
Illumination	CIE illuminant D65 stimulator
Illuminance	1000 lx
Observer	normal colour vision
Background field	uniform, neutral grey L^*=50
Viewing mode	object (such as colour patch)
Sample size	greater than 4° subtended visual angle
Sample separation	minimum sample separation achieved by placing the sample pair in direct edge contact
Sample colour-difference magnitude	0 to 5 CIELAB units
Sample structure	no visually apparent pattern or non-uniformity

Table 2-5. CIE 94 and CIE DE2000 colour difference's reference conditions (produced by CIE 116: 1995 [103] (pp. 6), CIE 142: 2001 [104] (pp. 2) and reproduced by Bilissi [106] (pp. 44-45)).

2.2.2.4 Resolution

Image quality also depends on structural image properties, such as resolution and sharpness. Resolution describes the finest detail that may be recorded by a system [110] (pp. 4).

In a silver based photographic system, the resolution is determined by measuring the resolving power of the total system, and it is defined as the number of lines per distance that can be resolved by an observer, i.e. the lowest value of the test pattern where the individual black and white lines can no longer be distinguished [33] (pp. 53). The resolving power is strongly dependent on a line target and quality at each stage of the complete system, i.e. lens, photographic, microscopic and visual system [111] (pp. 81).

In digital imaging systems, the resolution is described as pixel resolution and spatial resolution. The pixel resolution is expressed by the number of elements per unit area. The spatial resolution is expressed by the number of line pairs per unit distance or cycles per unit distance [33] (pp. 53-54).

The resolution can be measured by MTF (cf. Section 2.2.2.5). Generally, the corresponding point to a 10% MTF can be measured as the *limiting resolution* (Figure 2-14) [112] (pp. 34).

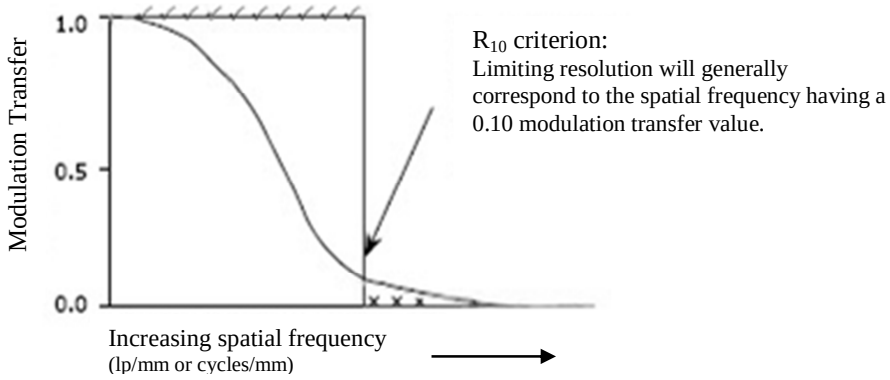


Figure 2-14. Limiting resolution in modulation transfer function (MTF) (adapted from Burns [112] (pp. 34)).

2.2.2.5 Sharpness

Sharpness is the perception of micro-image contrast [113] (pp. 443). It is commonly measured using modulation transfer function (MTF) and acutance. The MTF measurement has been a more successful measuring tool than the acutance, since the MTF is a function describing the behaviour of the system at different spatial frequencies and falls to some threshold value [33] (pp. 55). The MTF measurement has been used as a detailed image quality prediction measure for more than fifty years [114] (pp. 7).

The theory of MTF is strictly valid in linear, isotropic and spatially invariant systems [2] [115] (pp. 124), [116] (pp. 187). However, digital imaging systems are non-linear, since they are non-isotropic and non-stationary systems [117] (pp. 231). Compensation for the non-linearity present in a digital system is necessary for the accurate evaluation of the MTF. The non-linear compensation is able to be achieved by correcting the system for transfer compensation or restricting the test target to a very low contrast [118] (pp. 232-233).

[2] The impulse response of digital imaging systems is not rotationally symmetrical due to the rectangular shape of the pixels. A stationary imaging device yields the same response to a point source at any position within its field of view [33] (pp. 87).

There are various MTF measurement methods, including a periodic (sinusoidal) signal method and non-periodic methods [113] (pp. 446). Different methods generally yield different results [117] (pp. 235), [119] (pp. 811).

The periodic (sinusoidal) signal method is the original method. The MTF commences with the definition of a sinusoidal input:

$$I(x) = a + b\cos(2\pi\omega x) \tag{2.31}$$

where a is the average signal level (offset), b is the amplitude and ω is the spatial frequency.

The original input modulation, $M_{in}(\omega)$, is calculated by:

$$M_{in}(\omega) = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{b}{a} \tag{2.32}$$

where I_{max} represents the maximum intensity, and I_{min} represents the minimum intensity of the sinusoid.

The reduced output sinusoidal signal is expressed by:

$$I'(x) = a + b'\cos(2\pi\omega x + \varepsilon) \tag{2.33}$$

where a is the average signal level (offset), b' is the reduced amplitude, ω is the spatial frequency and ε is the phase difference.

The reduced output modulation, $M_{out}(\omega)$, is calculated by:

$$M_{out}(\omega) = \frac{I'_{max} - I'_{min}}{I'_{max} + I'_{min}} = \frac{b'}{a} \tag{2.34}$$

where I'_{max} represents the maximum intensity, and I'_{min} represents the minimum intensity of output sinusoid.

The modulation transfer factor, $M(\omega)$, is the ratio of output modulation, $M_{out}(\omega)$, to input modulation, $M_{in}(\omega)$ (Figure 2-15). The MTF is plotted as the modulation transfer factor against spatial frequency (cf. Figure 2-14).

$$M(\omega) = \frac{M_{out}(\omega)}{M_{in}(\omega)} \tag{2.35}$$

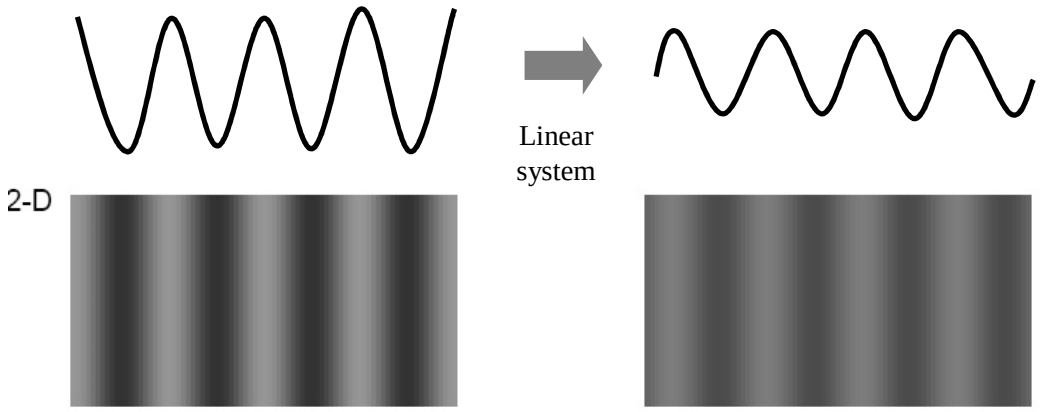


Figure 2-15. Imaging a sinusoidal exposure
(adapted from Jenkin [114] (pp. 5)).

The non-periodic MTF methods[3] include the edge gradient technique and the slanted edge technique.

In the edge gradient technique, the MTF is derived from the modulus of the Fourier transform of the line spread function (LSF) (Figure 2-16). The LSF is obtained by differentiation of the edge spread function (ESF) using a reasonably produced edge.

$$M(\omega) = \left| \int_{-\infty}^{+\infty} L(x) e^{-2\pi i \omega x} dx \right| \tag{2.36}$$

where $L(x)$ is the LSF, x is the distance and ω is the spatial frequency.

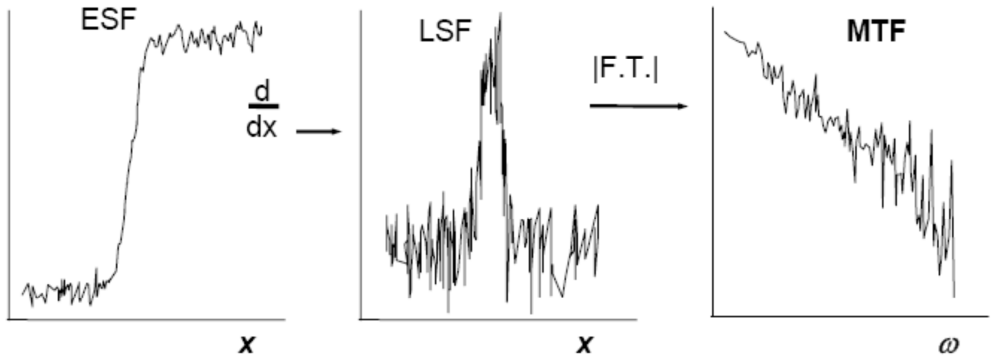


Figure 2-16. Derivation of modulation transfer function (MTF) from the edge spread function (ESF) via the line spread function (LSF)
(produced by Axford [123] (pp. 407)).

[3] Noise method [120, 121] and dead leaf method [122] are also non-periodic MTF methods.

The basic relationships between MTF and spread functions are illustrated in Figure 2-17 [114] (pp. 16).

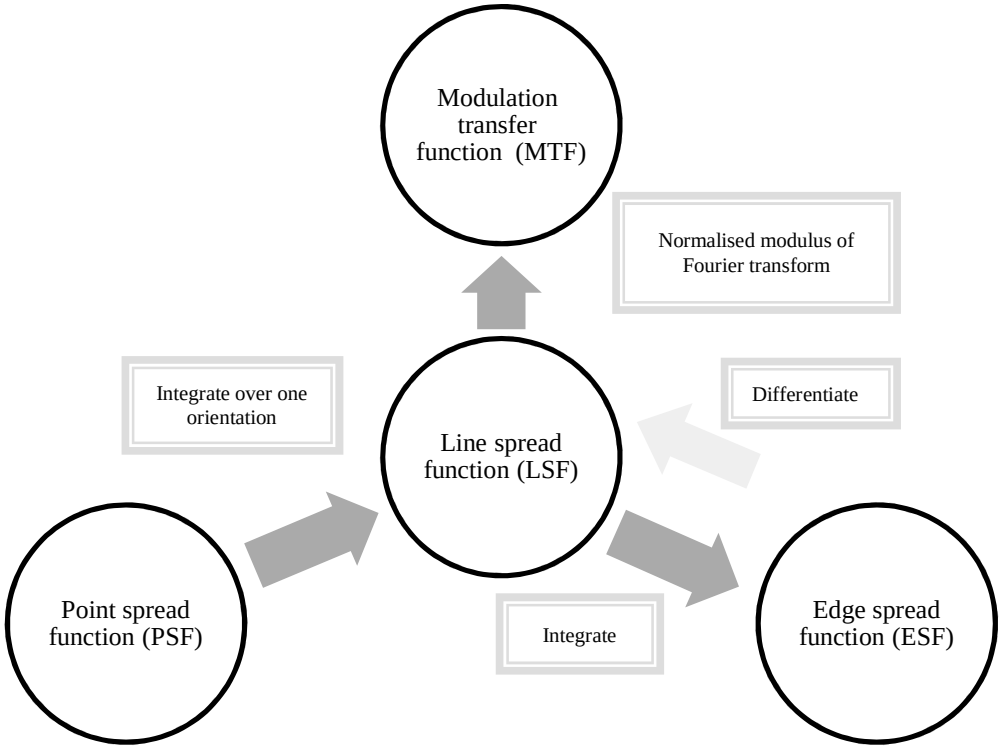


Figure 2-17. Relationships between spread functions and modulation transfer function (MTF) (adapted from Jenkin [114] (pp. 16)).

The basic edge gradient technique was sensitive to phase differences between the experimental target and sampling comb and also the under sampling [114] (pp. 25). To overcome these problems, Reichenbach *et al* [124] (pp. 172-174) originally developed an extended edge technique based on a sloping target. ISO 12233: 2000 [110] has revised this slanted edge technique.

This technique is based on a sloping edge target, i.e. a small amount of rotation (5-10°) is intentionally introduced into the edge target [110] (pp. 5-9). The target is imaged in the usual manner. Even illumination of the target is provided using two light sources [110] (pp. 10-11).

Each row value in the imaged edge represents the response of the system to an edge that is slightly shifted with relation to the previous row. The image code values are linearised by inverting the OECF of the camera. After that, a one-dimensional super-sampled line spread function is formed using the derivatives of the image data. Using the first line as reference points, the data points from all the other lines are placed into one of four "bins" between these reference points, according to the distance from the edge for that particular line. This creates a single super-sampled "composite" line spread function, having four times as many points along the line as the original image data. A hamming window is then applied to reduce the effects of noise, and the normalised modulus of the Fourier transform is calculated to produce the spatial frequency response (SFR).

The SFR is generated from uniformly super-sampled slanted edge profiles where the input edges are of sufficient optical quality. The SFR of the system, SFR(ω), is a combination of the frequency content of the edge target, $M_{edge}(\omega)$, and the MTF of the system, $M(\omega)$ [117] (pp. 233):

$$\mathrm{SFR}(\omega)=M_{edge}(\omega) \times M(\omega) \tag{2.37}$$

According to Burns [125] (pp. 135), the input edge used needs to be of a sufficient optical quality, for the SFR to be taken as an estimate of the MTF of the system. Otherwise, the output modulation can be divided by the input target modulation, frequency by frequency, to yield the system MTF.

A diagram of a super-sampled edge construction is shown in Figure 2-18. The SFR algorithm is shown in the flow-chart form in Figure 2-19.

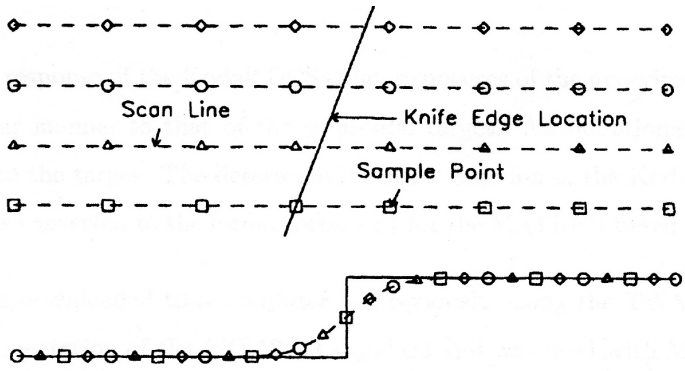


Figure 2-18. Schematic of CCD/CMOS detector array and super-sampled edge construction (produced by Reichenbach [124] (pp. 172)).

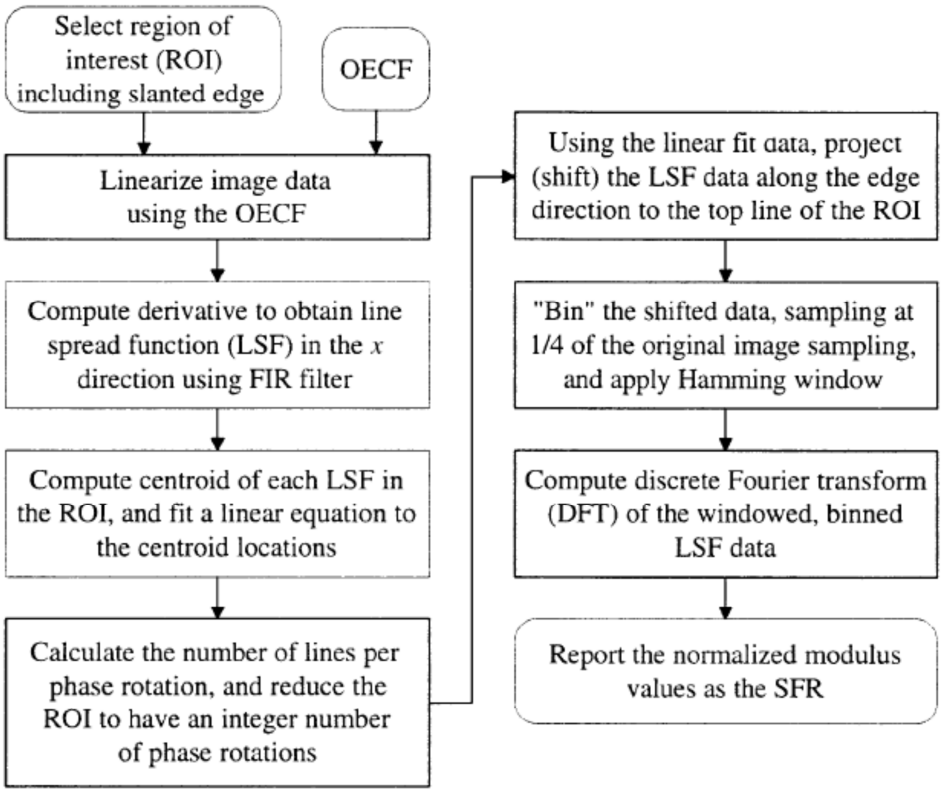


Figure 2-19. Flow-chart of spatial frequency response (SFR) measurement algorithm (produced by ISO 12233: 2000 [110] (pp. 13)).

2.2.2.6 Noise

According to ISO 15739: 2003 [126] (pp. 2), noise is defined as "the unwanted variations in the response of an imaging system." Theoretically, it varies randomly in the image plane and is independent from the input signal. Therefore, the noise is quantified as a random process [127] (pp. 413).

The causes of image noise in digital imaging systems are random partitioning of exposing light, photo-electronic noise, electronic noise, quantization noise and Poisson exposure noise. If an image is converted from a light intensity or exposure distribution to an electrical form, there will be photo-electronic noise. Electronic noise will occur in any system with electronic components. Poisson exposure noise means there is a randomness of photons in a nominally uniform exposure distribution. The photo-electronic noise and the electronic noise are the main cause of image noise at the output of a charge-coupled device (CCD) imaging array [113] (pp. 433-434), [127] (pp. 413).

Noise can be described in density units (denoted *D*) where the fluctuations might be luminance or digital values [113] (pp. 438).

In the luminance fluctuations, the noise is defined as:

$$\sigma_A{}^2 = \lim_{x \to \infty} \frac{1}{2x} \int_{-x}^{x} \Delta D^2(x) \mathrm{d}x \tag{2.38}$$

where $\Delta D(x)$ is the deviation of $D(x)$ from mean density.

In the digital value fluctuations, the noise is defined as:

$$\sigma_A{}^2 = \frac{1}{N-1} \sum_{i=1}^{N} \Delta D_i^2 \tag{2.39}$$

where ΔD_i is the i_{th} measured value of the deviation. N values are recorded, e.g. N is 1000.

Instead of evaluating the mean square density deviation of a noise trace, the noise power spectrum (NPS) or Wiener spectrum is directly calculated [113] (pp. 438):

$$N(\omega) = \frac{1}{x} \left| \int_0^x \Delta D(x) e^{-2\pi \mathrm{i} \omega x} \mathrm{d}x \right|^2 \tag{2.40}$$

where $N(\omega)$ is the noise power spectrum, x is the range of integration and $\Delta D(x)$ is the measured fluctuation at position x.

The noise levels usually correspond to an image signal, thus it should describe the signal-to-noise ratio for the imaging system [127] (pp. 420-421).

ISO 15739: 2003 [126] (pp. v) suggests the specified operating conditions be reported along with the measurement results, since the noise performance of an image sensor may vary significantly, e.g. exposure time, operating temperature and relative humidity, illumination, white balance and ISO speed.

2.2.2.7 Models of the human visual system (HVS)

The contrast sensitivity function (CSF) is widely used as the physical model of the HVS in the determination of image quality. It measures the sensitivity to gratings of different spatial frequencies as the reciprocal of the modulation threshold [128] (pp. 627-628).

Barten [129] (pp. 63-64) has provided a physical model for determining the contrast sensitivity of the human eye. This has been tested against a wide range of CSF measurements from various labs. It takes into account variations in luminance (Figure 2-20) and viewing angle (Figure 2-21). It is described by [130] (pp. 39):

$$\frac{1}{M_t(u)} = \frac{M_{opt}(u)}{k}\left[\frac{2}{T}\left(\frac{1}{X_0^2}+\frac{1}{X_e^2}+\frac{u^2}{N_e^2}\right)\left(\frac{1}{\eta\rho E}+\frac{\Phi_0}{1-\exp[-(u/u_0)^2]}\right)\right]^{-\frac{1}{2}} \qquad (2.41)$$

where $M_{opt}(u)$ is the optical MTF of the eye defined as a Gaussian distribution, k is the signal-to-noise ratio (k=3.0), u is the spatial frequency in cycles per degree (cpd), T is the integration time (T=0.1s), X_0 is the image size (degrees), X_e is the maximum angular integration size (X_e=12°), N_e is the estimated maximum number of cycles over which the eye integrates (N_e=15 cycles), η is the quantum efficiency (η=0.03), ρ is a photon conversion factor of the light units, E is the retinal illuminance (trolands), Φ_0 is the spectral density of the neural noise ($\Phi_0 = 3.0 \times 10^{-8}$ s deg^2), and u_0 is the frequency above which lateral inhibition ceases (u_0= 7 cycles per degree).

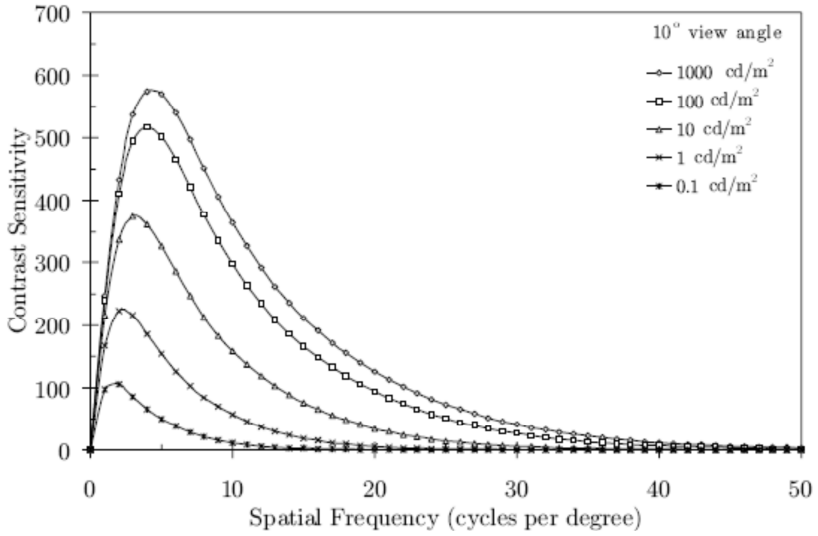


Figure 2-20. Contrast sensitivity functions (CSFs) at different levels of luminance (produced by Barten [68] (pp. 2026) and reproduced by Jacobson [27] (pp. 8)).

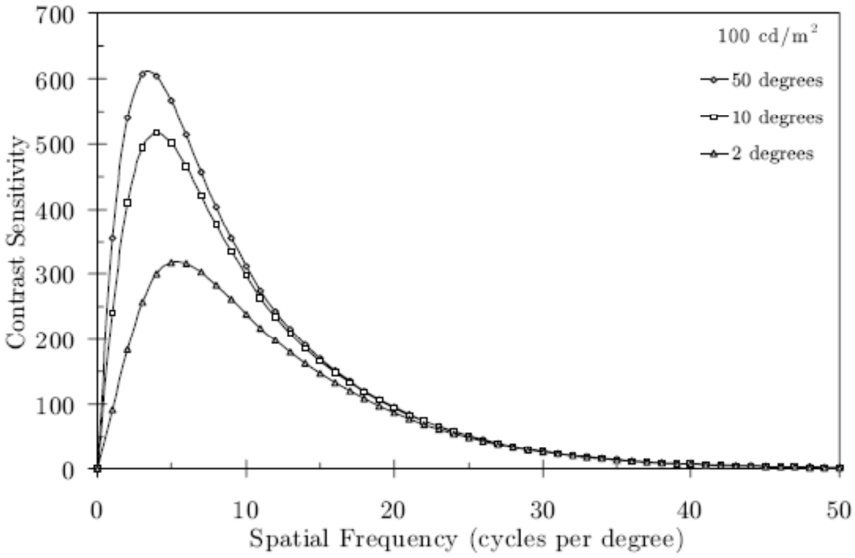


Figure 2-21. Contrast sensitivity functions (CSFs) at different angles of view (produced by Triantaphillidou [33] (pp. 65) using the Barten's CSF model [69] (pp. 8)).

Farrell [59] (pp. 299) was concerned with the limitations of single-channel metrics that are based on the CSF. This CSF is a relatively simple function employed in image quality investigation. The study was considered as the contribution to the independent visual multi-channels functions employed in image quality investigation [59] (pp. 300).

2.2.2.8 Objective scene classification

As mentioned before (in the section 2.2.1.2), the subjective quality depends upon the scene content of the test images. The nature of scene dependency causes problems in modelling and predicting image quality.

The scene classification with respect to image quality might be one possible way of overcoming the problem of scene dependency in image quality modelling [14] (pp. 147), [131] (pp. 413).

There are different approaches for the scene classification, such as inspection and clustering. Objective scene classification is an approach to identify groups of scene having or producing correlated responses across the test stimuli [10] (pp. 269).

The stages of objective scene classification are shown in Figure 2-22. These stages are interrelated and depend on the classification results [132] (pp. 7).

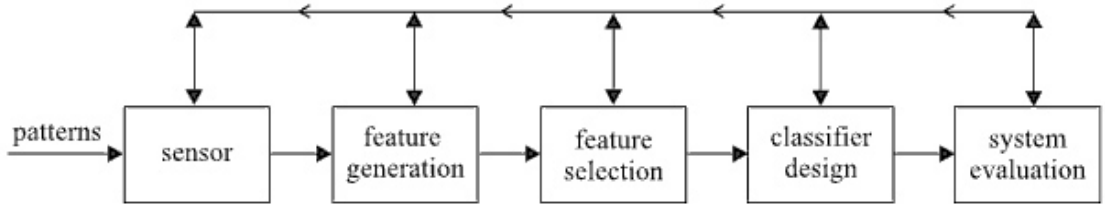


Figure 2-22. Basic stages involved in objective scene classification
(produced by Theodoridis and Koutroumbas [132] (pp. 7)).

For the classification, a number of features are generated, and then the "best"[4] of them is adopted [132] (pp. 6).

i) Feature generation: The goal of this stage is to quantify scene features that will be used to classify the image. As the image does not directly provide mathematical information in its natural state, the features of the image are encoded, so that the relevant information in the image is represented as data.

[4] This is important for successful classification and serves different purposes. This is an unresolved challenge within the imaging science community [10] (pp. 269).

There are a large number of feature generation techniques to quantify scene information, as a numerical output. They include second-order statistical measures and measurement from edge detection.

Second-order statistical measurements are normally indicators of texture in the imaged scene [133] (pp. 272), [134] (pp. 666-670). The second-order statistical measurements are referred to as the grey level co-occurrence matrix (GLCM), approximation of the joint probability distribution of pairs of pixels (Figure 2-23) [135] (pp. 562-563).

$$P(i,j) = \frac{N(i,j)}{M} \tag{2.42}$$

where M is the total number of pixels in the image, $N(i, j)$ denotes the number of occurrences between two pixel values i and j: the second pixel (b) is specified at distance r and angle θ from the first pixel (a).

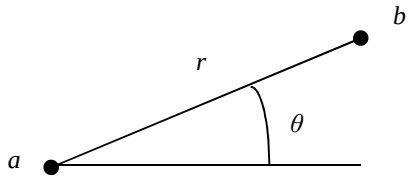


Figure 2-23. Relationship of a pair of pixels for grey level co-occurrence matrix (GLCM) (adapted from Pratt [135] (pp. 562)).

Figure 2-24 presents the example of GLCM using the 0° angle and 1 pixel distance.

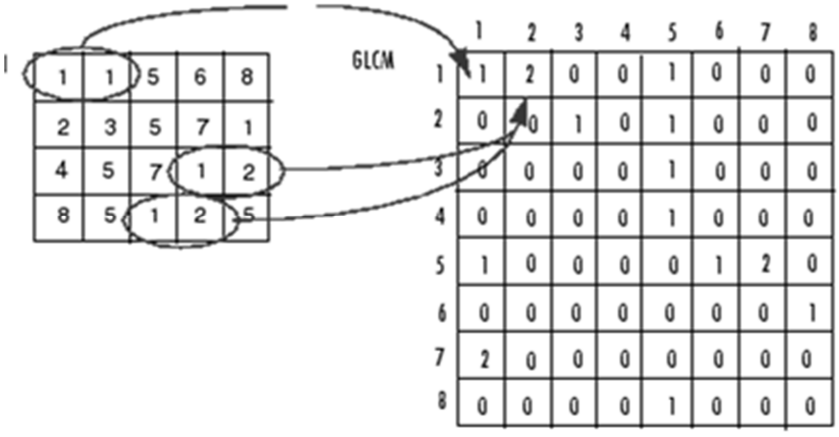


Figure 2-24. Example of grey level co-occurrence matrix (GLCM), employed the 0° angle and 1 pixel distance in a pair of pixels (produced by MathWorks [136]).

Common second-order statistical measurements are listed below:

- Contrast (or inertia): relates to the scene texture variation.

$$\text{Contrast} = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} |i-j|^2 P(i,j) \tag{2.43}$$

- Homogeneity: relates to the scene texture variation. Contrast (inertia) and homogeneity are strongly, but inversely, correlated in terms of equivalent distribution in the pixel pairs population.

$$\text{Homogenity} = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} \frac{P(i,j)}{1+|i-j|} \tag{2.44}$$

- Correlation (or linearity): relates to the scene texture variation with linearity.

$$\text{Correlation} = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} \frac{(i-m_x)\big(j-m_y\big)P^2(i,j)}{\sigma_x\sigma_y} \tag{2.45}$$

where

$$m_x = \sum_{i=0}^{k-1} i \sum_{j=0}^{k-1} P(i,j), \qquad m_y = \sum_{j=0}^{k-1} j \sum_{i=0}^{k-1} P(i,j)$$

$$\sigma_x^2 = \sum_{i=0}^{k-1} (1-m_x)^2 \sum_{j=0}^{k-1} P(i,j), \quad \sigma_y^2 = \sum_{j=0}^{k-1} (1-m_y)^2 \sum_{i=0}^{k-1} P(i,j)$$

- Energy: relates to the disorders in scene texture. The highest energy values occur when the grey level distribution has a constant or periodic form.

$$\text{Energy} = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} P(i,j)^2 \tag{2.46}$$

where $P(i,j)$ is the joint probability distribution of pairs of pixels in the matrix.

Edges in an image are fundamentally important, because they often provide an indication of the physical extent of objects within the image [137] (pp. 491). The edge in an image is defined as a set of connected pixels that lie on the boundary between two regions [138] (pp. 572).

The edge can be detected by the first- and second-order derivatives of the grey level edge profile, as an edge detector [139] (pp. 124). Figure 2-25 illustrates the first- and second-order derivatives.

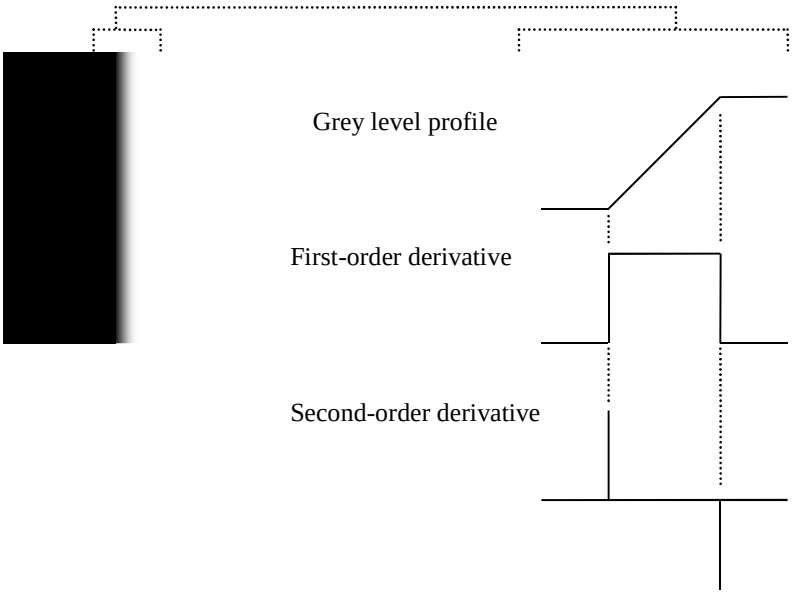


Figure 2-25. Two regions separated by a vertical edge and detail near the edge, showing a grey level profile, first- and second-order derivatives of the profile (produced by Gonzalez and Woods [138] (pp. 574)).

The first-order derivatives are based on various approximations of the 2-D gradient, such as Sobel, Prewitt and Robert [138] (pp. 578). Examples of Sobel and Prewitt are presented in Figure 2-26.

Sobel

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Prewitt

-1	0	+1
-1	0	+1
-1	0	+1

Gx

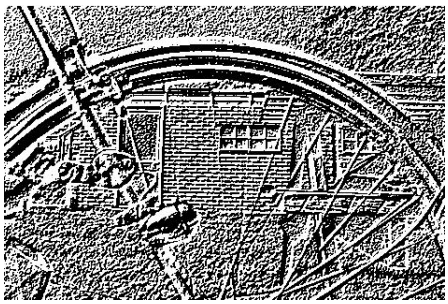
+1	+1	+1
0	0	0
-1	-1	-1

Gy

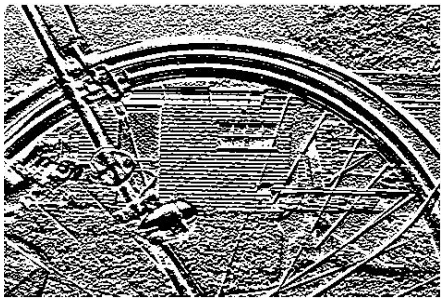
Figure 2-26. Examples of approximations of the 2-D gradient

Figure 2-27 illustrates the response of two components of the gradient, G_x and G_y, as well as the original image and the sum edge image.

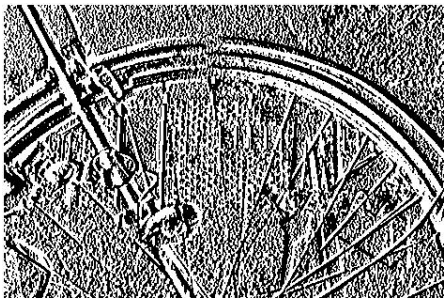
Original image



Sum edge image



Sobel edge image in the x-direction, G_x



Sobel edge image in the y-direction, G_y

Figure 2-27. Response image of two components of the gradient, G_x (bottom-left) and G_y (bottom-right), as well as its original “Bike” image (top-left) and the sum edge image (top-right).

The gradient of an image f(*x,y*) at location (*x,y*) is defined in the vector [138] (pp. 577-580).

$$\nabla \mathrm{f} = \begin{bmatrix} G_x \\ G_y \end{bmatrix} \tag{2.47}$$

The magnitude of the edge was computed by omitting the square-root operation in the vector.

$$G = \sqrt{G_x^2 + G_y^2} \tag{2.48}$$

The direction of the edge was also computed in the gradient vector.

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_x}{G_y}\right) \tag{2.49}$$

ii) Feature selection: This could be complicated, since sometimes the salient features are not easy to select [140] (pp. 11), [141] (pp. 141). Thus, prior knowledge plays a major role in the feature selection.

The feature selection includes statistical hypothesis testing [142] (pp. 166). This helps discard easily recognizable "bad" choices. The aim of the test is to identify which of the following hypothesis is correct [143] (pp. 140):

H_1 (alternative hypothesis): The values of the feature differ significantly
H_0 (null hypothesis): The values of the feature do not differ significantly

The decision is reached on the basis of experimental evidence, supporting the rejection or not of the null hypothesis. This is accomplished by exploiting statistical information, referred to as the significance level. 10%, 5% and 1% significance levels are used to describe results as 'almost significant', 'significant' and 'highly significant.'

The statistical hypothesis testing is a clear indication of the strength of the relation between variables. If the two variables are not approximately normally distributed, a rank correlation coefficient is used. An example of such a rank correlation coefficient is the Spearman's correlation coefficient (r_s). The correlation coefficients range between -1.0 (indicating perfect anti-correlation) and 1.0 (indicating perfect

correlation), with 0 denoting no correlation at all [144] (pp. 219-223), [145] (pp. 80-81).

iii) Classifier design: Once the feature has been selected, the classification method is considered. The classification is typically the final step for identifying the object, which is a step that is automatically carried out without using human visual perception. There are a number of suggested classification methods, which include supervised/ unsupervised classification [146] (pp. 414).

Clustering is an unsupervised classification method, where there is no training data requirement [147] (pp. 397). In addition, clustering includes k-means clustering. The main advantages of k-means clustering are simplicity and speed, which allows it to be run on large data sets [148] (pp. 526-528), [149].

The k-means clustering consists of several steps [21] (pp. 406-408). The first step is to define a fixed number of clusters, k. The choice of k is exceedingly influential in clustering; an inappropriate choice of k may yield poor results while the correct choice of k is often ambiguous. Possible methods for choosing k include empirical and numerical methods [150] (pp. 750). The empirical method is usually preferred [149]. Once k is chosen, then modifications of the distances between all points in the n^{th} cluster (n varying from 1 to k) and the centre of the cluster are applied. The main idea for their modifications is that the average distances between all points in each cluster and the central point are minimal. During these modifications, new cluster centres are allocated using Euclidean distances. The modification stops when the average distances from all points in the n^{th} cluster and the new central point are minimised.

iv) System evaluation: Once the results of the clustering algorithm have been obtained, the classification accuracy is assessed by Theodoridis and Koutroumbas [132] (pp. 6) and Duda *et al* [140] (pp. 15).

2.3 Summary

This chapter concentrates on the theory of image quality and its measurements.

The broader concept of image quality has no unique single definition yet. Currently, there is a narrow definition of image quality based on third-party, i.e. artefactual, preferential and aesthetic attributes. A clearer and broader definition of image quality might be required.

In general, the image quality has been measured using subjective (psychophysical) or objective (physical) scaling.

Subjective image quality measures are studied from bottom-up and top-down. The bottom-up perspective is common in image quality scaling studies. The image quality rating is formed by the combination of individual visual attribute scales, which are characteristics of an image that we sense (could see). However, individual attribute scaling has been a subject of discussion. This is because individual attribute scaling is a relatively simplistic way of measuring image quality, as it ignores the complicated relationships between attributes. A top-down perspective tends to avoid these pitfalls.

Objective image quality measures are studied for speed and are less complex to implement. However, as of yet, no definitive objective scaling method has been put forward. For this ultimate aim, further objective measurement requires a full understanding of image formation by the imaging system all of which involve the end-user, and of the scene content of the test images used for the evaluation.

Chapter 3

Camera characterisation

This chapter presents the digital image acquisition device (camera) characterisation and its calibration. Initial examination of the fundamental behaviour of the system is described with respect to spatial uniformity, tone reproduction and colour reproduction.

The characterisation can be defined as the description of the collective qualities of a system. The calibration is the process of maintaining the device with a fixed, known characteristic [151] (pp. 272-275). The characterisation and calibration form a pair and are interrelated [152] (pp. 388). Figure 3-1 shows the general characterisation and calibration process of input devices.

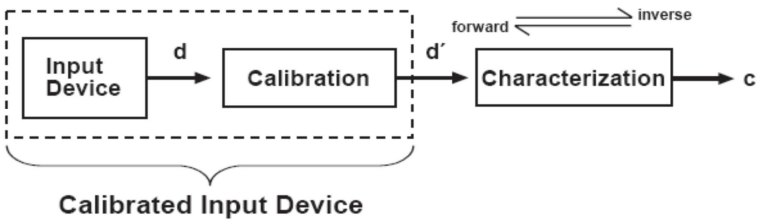


Figure 3-1. Characterisation and calibration for input devices (produced by Bala [151] (pp. 273)).

3.1 Digital camera description

The digital image acquisition device used in this work for image capture was a Canon EOS-1Ds full frame digital single lens reflex (SLR) camera [153]. This camera uses a complementary metal–oxide–semiconductor (CMOS) with Bayer RGB primary colour filter, which has 50% green cells arranged in a checkerboard and alternating cells of red and blue (2×2 matrix of R+2G+B pixels form one colour pixel). Figure 3-2 presents the colour filter array in the camera.

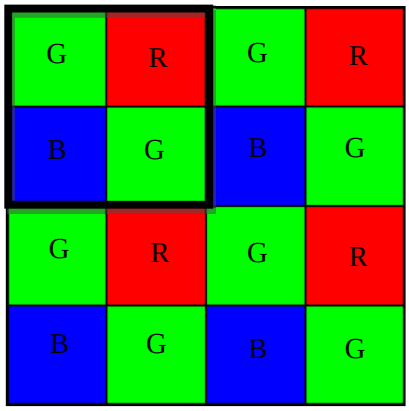


Figure 3-2. Arrangement of RGB sensors in a Bayer colour filter (adapted from Ray [154] (pp. 121)).

The camera had a 35mm SLR body, enabling the use of a range of lenses and optical accessories. A Canon EF 28-135mm *f* 3.5-5.6 IS USM zoom lens was used in conjunction with this body.

The camera software consists of two parts: the firmware and the driver. The firmware controls the fundamental operations of the camera body and captures the image information. The driver is a programme for a computer, which enables the import of the image to the computer. Some of the specifications of the camera and lens are in Table 3-1.

Component	Specifications
Camera body	Imaging sensor size: 35.8 x 23.8 mm Pixel resolution: 4064 x 2704 Colour filter system: RBG primary colour filter array Low-pass filter: Located in front of the CMOS sensor, non-removable Recording image format: JPEG, RAW (12-bit) Colour encoding: sRGB and Adobe RGB. White balance settings: auto, daylight, shade, overcast, tungsten light, fluorescent light, flash, custom, colour temperature setting, personal white balance (Total 10 settings) ISO speed range: 100-1250 (in 1/3-stop increments)
Lens	Focal length: 28-135 mm Maximum aperture: 1:3.5-1:5.6 Diagonal angle of view: 75° - 18°
Driving software	EOS-1Ds Firmware File Viewer Utility 1.2

Table 3-1. Technical specifications for the image capture system (adapted from Canon [153, 155]).

3.2 Conditions for capturing

A number of test charts were photographed during the camera characterisation. In order to capture the image, the camera was mounted on a tripod with the optical axis of the lens being orthogonal to the plane of the target. The camera exposure was operated in manual mode, and the lens was focused manually with a 50 mm focal length and an aperture of *f* 11.

Two gas-filled tungsten lamps (3200 K), with 312 mired value [156] (pp. 21), were placed on each side at 45° angle and 1m away from the test targets, so that all areas of the target were uniformly illuminated (Figure 3-3). Even illumination of the scene was assessed using a Kodak R-27 18% grey card. Nine measurements were taken on different points of the grey card using a SEKONIC L-308s light meter and a Minolta CL-200 colorimeter. The even illumination presented in the light meter, and the uniformity of illumination ranged from 1939 (-3.7%) to 2092 (3.9%) lx in the colorimeter. EN 61966-9: 2004 [157] (pp. 7) notes that the uniformity of illumination shall be less than 5 %.

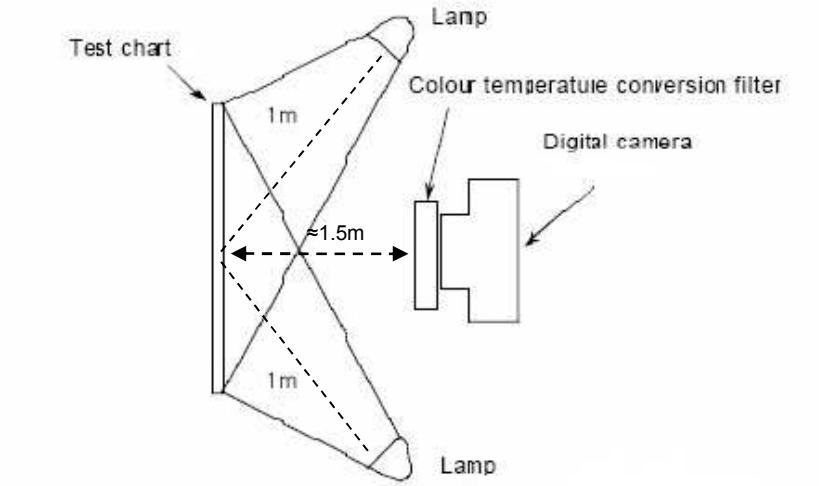


Figure 3-3. Equipment arrangement for the camera measurements (adapted from EN 61966-9: 2004 [157] (pp. 20)).

Kodak conversion filters were then placed in front of the lens, to obtain a fixed white point D65 (X_n=0.313, Y_n=0.329). The D65 is a commonly used standard illuminant and represents average daylight [156] (pp. 25).

The white balance setting of the camera was set to manual, to stop the automatic function of the camera compensating. EN 61966-9: 2004 [157] (pp. 25) notes that the white balance control should be considered as part of the camera characterisation. A previous study showed the importance of white balance control. Different colour errors were identified depending to the white balance settings, and the largest colour errors occurred where the auto white balance setting was used [158] (pp. 219).

3.3 Spatial uniformity

The uniformity of the Canon EOS-1Ds was evaluated using a Kodak R-27 18% grey card. Firstly, a number of measurements were taken to ensure the uniformity of the target, because a non-uniform test target is likely to produce step intensities.

The target was divided into 25 (5×5) squares (Figure 3-4). The CIE *XYZ* values in each square on the test target were recorded using a Color-Eye 7000A spectrophotometer,

which has reported a good inter-instrument agreement and repeatability, i.e. 0.8 average ΔE^*_{ab} at inter-instrument agreement and 0.1 ΔE^*_{ab} maximum at repeatability for a white tile [159]. The CIE colour difference (ΔE^*_{ab} and ΔE_{00}) between the middle area and across the target was then calculated. The reason for using two colour difference equations is that ΔE^*_{ab} is well established and ΔE_{00} is a relatively recent modification which models approximately uniform colour space. Thus, ΔE^*_{ab} is useful for comparing these results with previous studies and ΔE_{00} provides more meaningful insight into perceptual information.

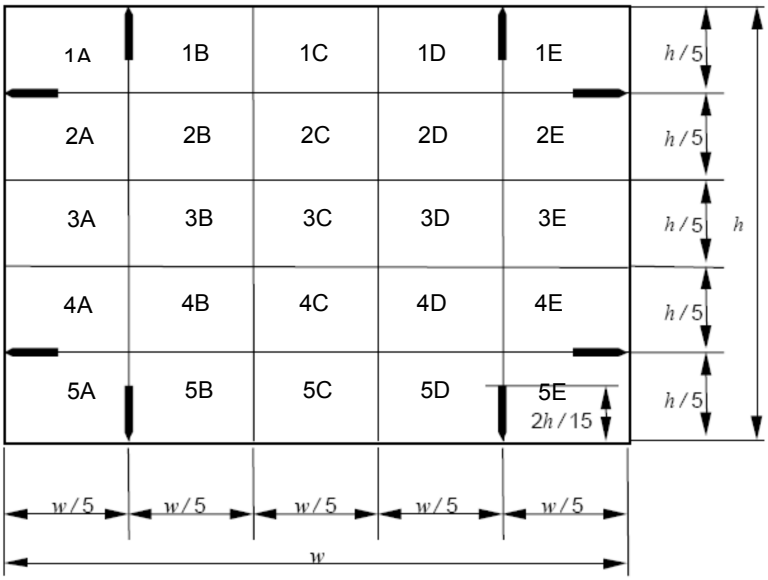


Figure 3-4. Twenty-five patches on a Kodak R-27 18% grey test target for the camera spatial uniformity
(adapted from EN 61966-9: 2004 [157] (pp. 11)).

The target was found to be almost uniform in all areas; the average colour differences produced 0.49 ΔE^*_{ab} and 0.38 ΔE_{00} (Figure 3-5). The colour differences in all areas were below perceptibility thresholds in uniform areas (≈ 1.0 ΔE) and complex images, 3.00 ΔE^*_{ab} [160] and 4.12 ΔE^*_{ab} [161] (pp. 66-67).

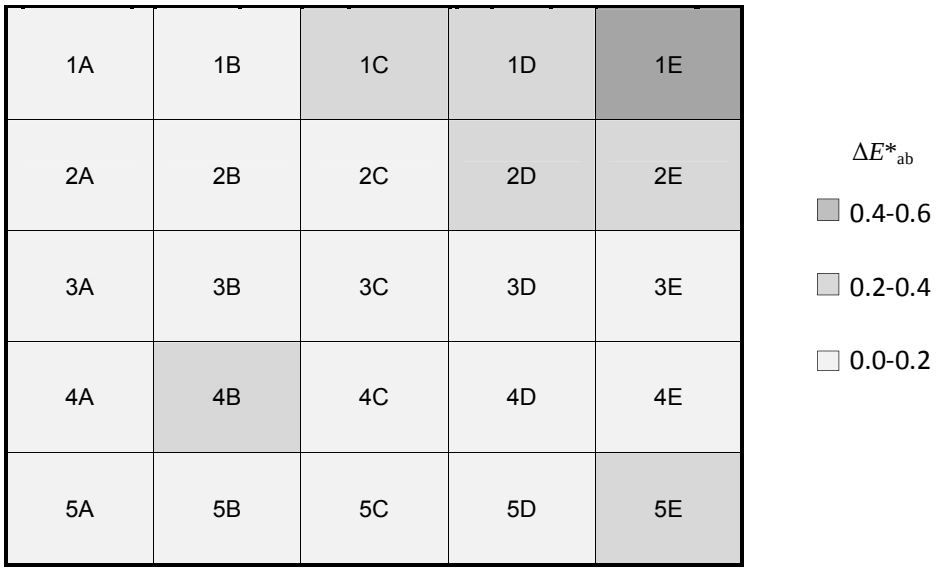


Figure 3-5. Spatial uniformity on the test target, expressed in $\Delta E^*_{ab.}$

After confirmation of the test target uniformity, the target was imaged twice using sRGB settings and 12-bit capture. Once the first image was captured, the target was rotated 180° before the second image was taken. Then, the imaged target was downloaded and saved as 8 bits per channel uncompressed data, Tagged Image File Format (TIFF) file. This was because the natural scenes used in the image quality assessment later were also downloaded and displayed as 8-bit TIFF uncompressed images (cf. Section 6.1 and 6.3).

The downloaded image was divided into 25 (5×5) squares (cf. Figure 3-4). The pixel values for the red, green and blue channels were recorded in each square. The results were then converted to tristimulus values *X*, *Y* and *Z*, via the sRGB encoding transformation (cf. Appendix A). The tristimulus values for each area were compared to these of the middle area in order to calculate the colour difference between them.

The non-uniformity of the camera is shown in Figure 3-6. The actual fluctuations of the 24 patches ranged from 0.15 to 5.38 in ΔE^*_{ab} and from 0.12 to 2.82 in ΔE_{00}. The top-right area (1E) was shown to be the most non-uniform system area. This result was shown to be between perceptibility and acceptability thresholds in complex images, i.e. the definitions for perceptibility threshold being set at 3.00 and acceptability threshold at

6.00 in ΔE^*_{ab} [160]. Where an acceptability threshold has been determined it is typically twice as large as that for perceptibility [162] (pp. 69-70), [163] (pp. 199).

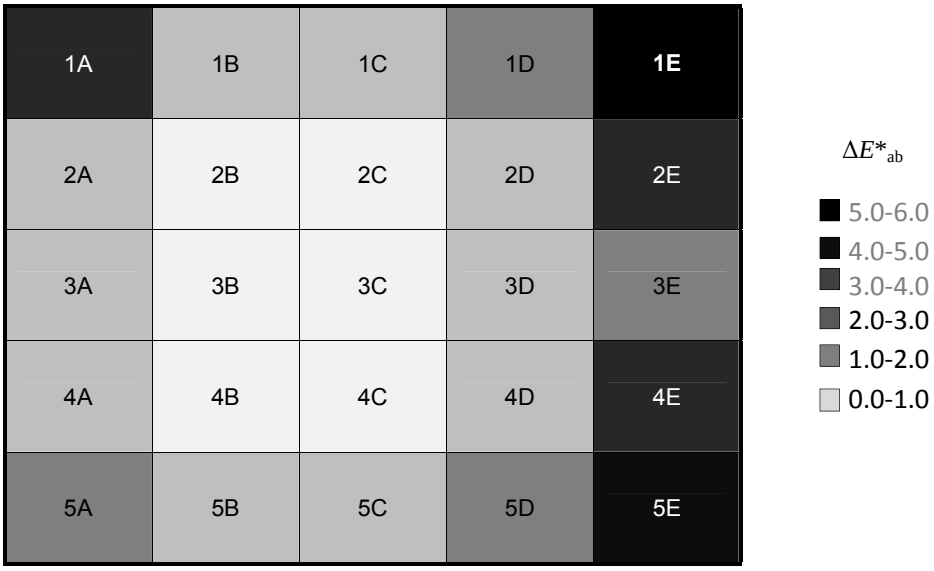


Figure 3-6. Spatial uniformity on the camera, expressed in ΔE^*_{ab}.

These results guided the rest of the investigations, determining that all images would be captured using the central area of the captured frame (cf. Section 3.4 & 3.5, section 5.1 & 5.2 and section 6.1).

3.4 Tone reproduction

A Kodak Q-60R2 reflection test target was used for the tone reproduction characterisation of the camera. The Kodak Q-60R2 includes a 24-step grey scale with Density $_{min}$ and Density $_{max}$ of 0.11 and 2.16, respectively. The target is currently an American National Standards Institute (ANSI) and ISO standard (Figure 3-7) [164].

The tone reproduction of the camera was evaluated for different colour settings. This was to identify the optimal colour setting and the requirement of gamma correction for optimal overall gamma.

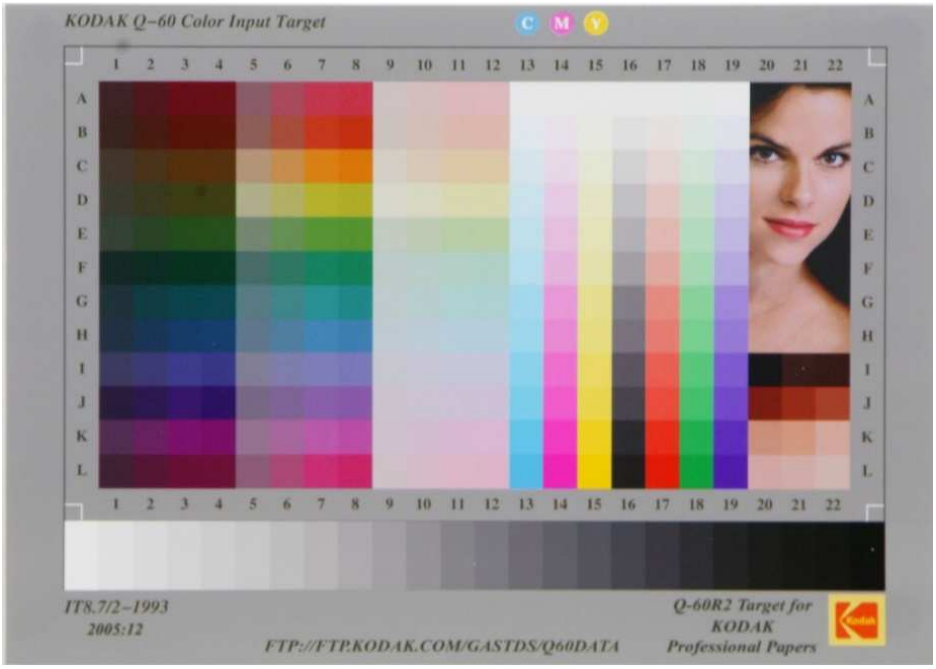


Figure 3-7. A Kodak Q-60R2 test target
(reproduced from Kodak [164]).

The camera is capable of capturing images at colour setting of between 1 and 5 (Table 3-2).

Setting number	Description
1	Sets a natural-looking hue and chroma. Effective for bringing out the subject's natural colour tone
2	Sets a hue and chroma suitable for portraits. Effective for rendering good skin tones.
3	Sets a hue and chroma similar to high-chroma slide film. Effective for making the colours clear.
4	An image corresponding to Adobe RGB colour space is created. It is useful for profile conversion to Adobe RGB or fine-tuning of chroma, because the colour reproduction range is much wider than the standard sRGB setting. Adobe RGB must be selected for profile conversion, because the ICC (International Color Consortium) profile is not attached to the image shot by the camera. Also, chroma adjustment is required as chroma is lower under sRGB circumstance.
5	Sets an image low-chroma. Effective for making colour tone moderate.
The colour settings space for 1,2,3 and 5 is sRGB	

Table 3-2. Colour settings and their descriptions of the camera
(adapted from Canon [153] (pp. 48)).

First of all, a calibrated Macbeth TR 924 reflection densitometer was employed to determine the densities of the original grey scale. Mean densities for each patch were obtained from three measurements.

The target was then photographed using a variety of colour settings as listed in Table 3-2, i.e. four sRGB and one Adobe RGB. All captured images were downloaded and saved (cf. Section 3.3). Then mean pixel values for each captured grey scale step were measured using Scion Image software [165]. This measurement was taken using 100×100 pixels from the central part of the patches.

The measured camera OECF is shown in Figure 3-8. The tone reproduction had little variation with respect to the colour settings used. These minimal differences indicate that the camera's tone reproduction is affected only slightly by the sRGB/ Abobe RGB colour encoding. This result is unsurprising since both the transfer functions have similar trends (cf. Figure 2-9).

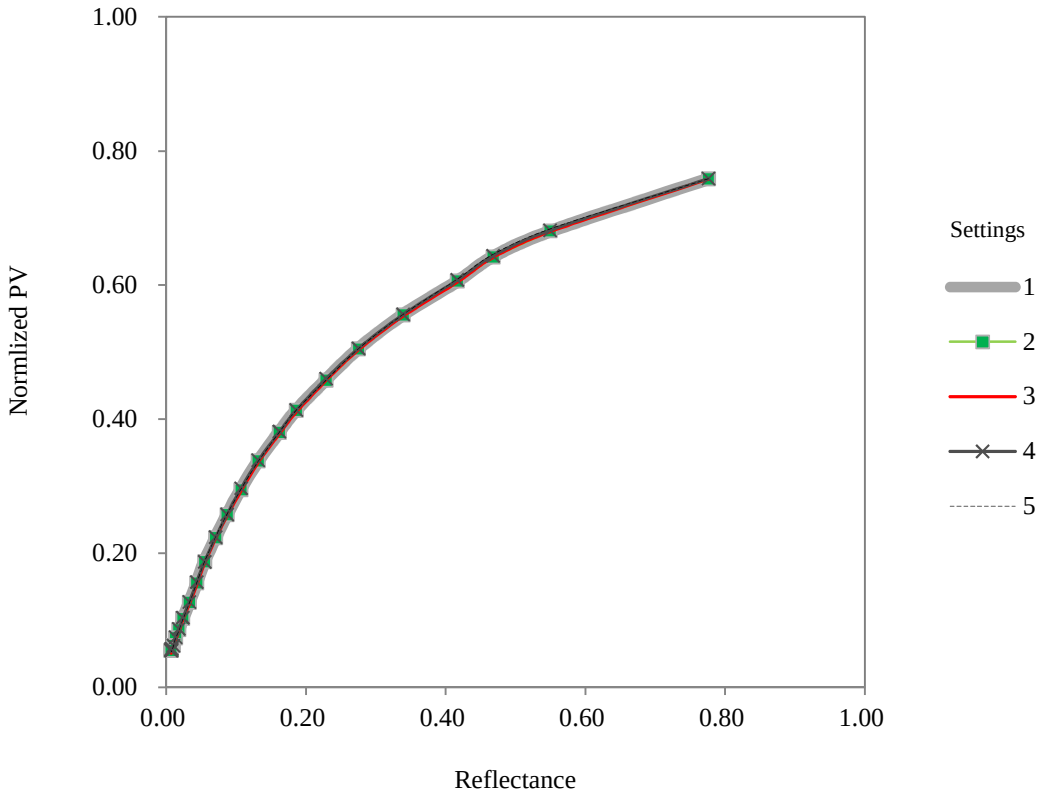


Figure 3-8. OECF curves of the camera at five different colour settings, plotted in linear-linear units.

In Figure 3-9, the logarithm of normalised pixel values against the density values for each patch are plotted, because the normal practice is for the photographic characteristic curve to be plotted in log-log units [83] (pp. 377).

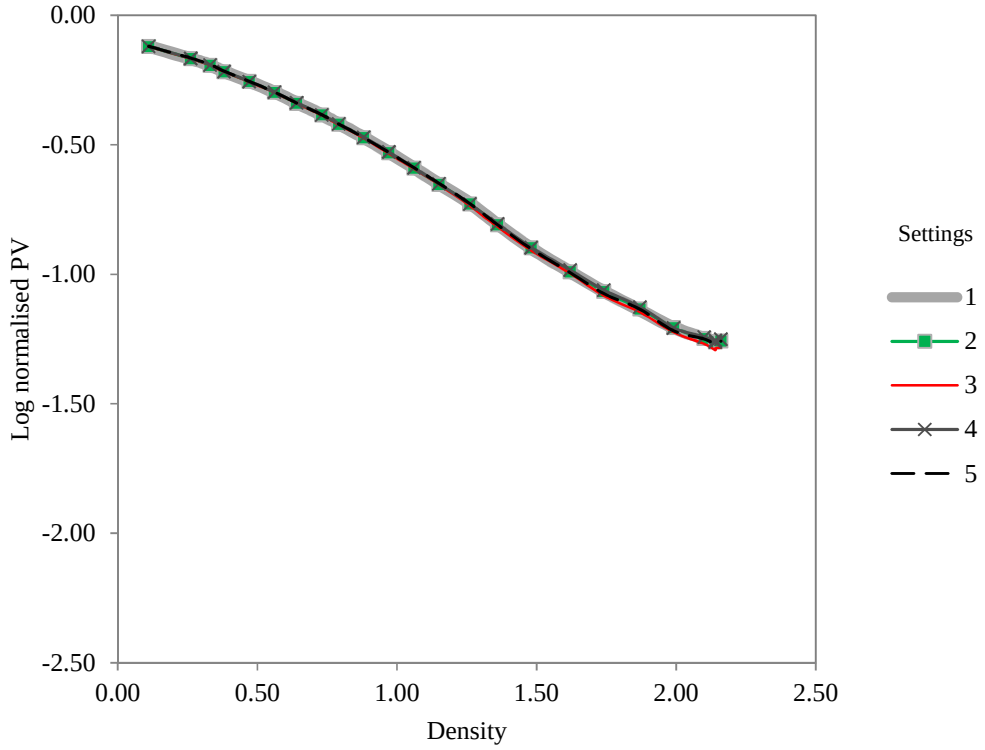


Figure 3-9. OECF curves of the camera at five different colour settings, plotted in log-log units.

The gamma, offset and correlation coefficient derived with linear regression for each setting are shown in Table 3-3.

Settings	Gamma (γ)	Offset (o)	Correlation Coefficient (r)
1	0.603	0.021	0.997
2	0.602	0.021	0.997
3	0.615	0.027	0.997
4	0.601	0.021	0.997
5	0.607	0.024	0.997

Table 3-3. Gamma, offset and correlation coefficient of the camera at five different colour settings (1,2,3& 5: sRGB, 4: Adobe RGB).

The 0.6 gamma was approximately equivalent to the gamma of photographic film negatives, when conventionally developed [81] (pp. 246). The results indicate that the gamma settings of the camera are intentionally set by the manufacture, to provide an optimal overall gamma of between 1.0 and 1.5 (cf. Section 2.2.2.2) in default display gamma 1.8 (Mac) and 2.2 (PC), i.e. if the display has a gamma 1.8 or 2.2, the camera gamma will make the optimal overall gamma equal to 1.1 or 1.3.

The additivity of the camera was also investigated by comparing between the combined RGB response and the neutral response. The combined RGB response was calculated by weighting each colour channel equally. The transfer functions for the combined RGB response and the neutral response are presented in Figure 3-10. The result showed that the combined RGB response closely matched the neutral response.

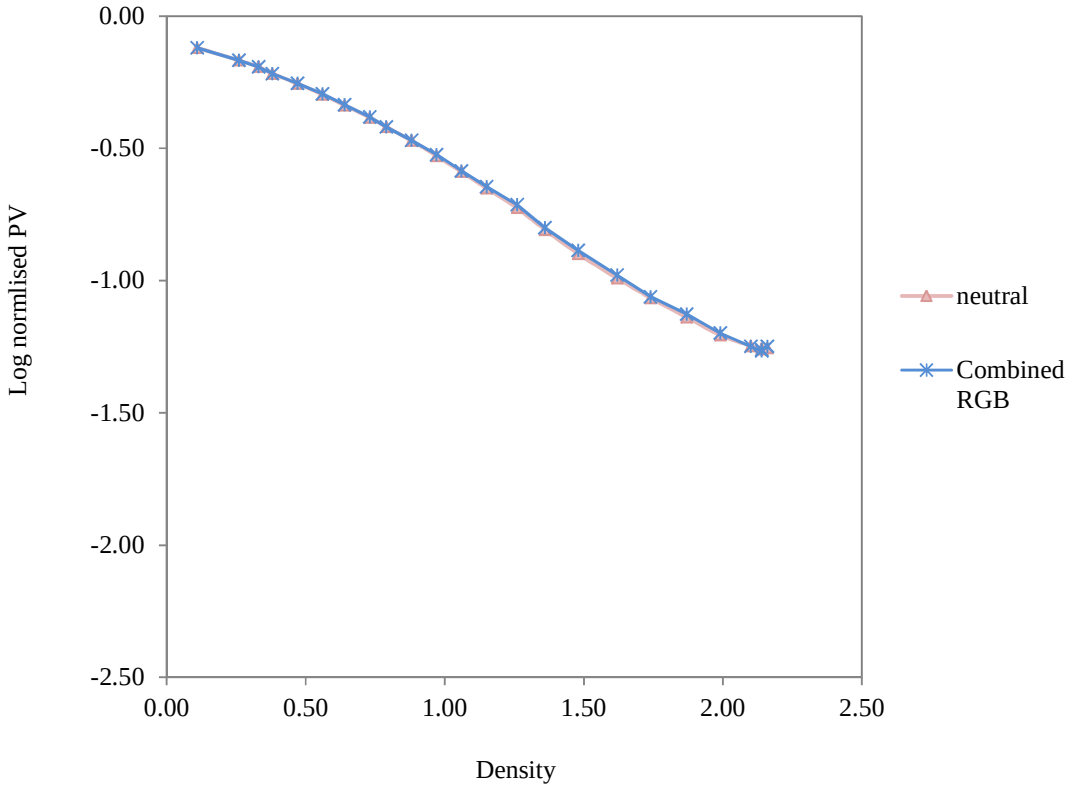


Figure 3-10. The camera additivity at the colour setting 1, comparison between the combined RGB response and the neutral response.

3.5 Colour reproduction

The Kodak Q-60R2 reflection test target was again used for the colour reproduction characterisation of the camera. The Kodak Q-60R2 consists of 264 colour patches, which include 12 colour samples of skin tones in columns 20-22, a series of single (CMY), double (RGB) and triple (K) dye colours in columns 13-19, and 12 samples each of 12 hues in columns 1-12. The columns 4, 8 and 12 represent the maximum chroma in three lightness levels, and a 24-step grey scale (cf. Figure 3-7) [164].

The camera's colour reproduction was assessed for colour accuracy and where necessary colour compensation was carried out to ensure accurate colour reproduction.

The device colour reproduction is classified into three groups: using (a) a custom profile, (b) a general profile and (c) a process profile [166] (pp. 25-28). The custom profile is designed to find the relationship between device-dependent and device-independent data using a suitable number of test colours. Typical custom profile method includes a three-dimensional lookup table model [167], a polynomial regression model [168, 169, 170, 171] and a neural network model [170, 171]. The general profile is supplied by the vendor. The process profile is an easy to implement method according to standard conditions, e.g. sRGB [89] and Adobe RGB [90].

3.5.1 Colour reproduction using sRGB/ Adobe RGB

First, mean CIE *XYZ* values for each patch were obtained from three different measurements, using the Color-Eye 7000A spectrophotometer [159]. The colorimetric data for all colour samples were obtained using the CIE 2° colorimetric observer and the D65 illuminant.

Then, the target was recorded at five different colour settings, and the captured images were transferred to a computer. The mean pixel values for the red, green and blue channels were then measured by averaging 100×100 pixels from the central part of each patch, using Scion Image software [165]. The results were then converted to tristimulus values *X*, *Y* and *Z*, via the two encoding transformations: sRGB and Adobe RGB encoding (cf. Appendix A).

Finally, the CIE *XYZ* colorimetric data between the measurement (original) and the calculation were assessed using two colour difference equations, CIELAB (ΔE^*_{ab}) and CIE DE 2000 (ΔE_{00}).

Figure 3-11 illustrates the process of colour reproduction characterisation for the camera.

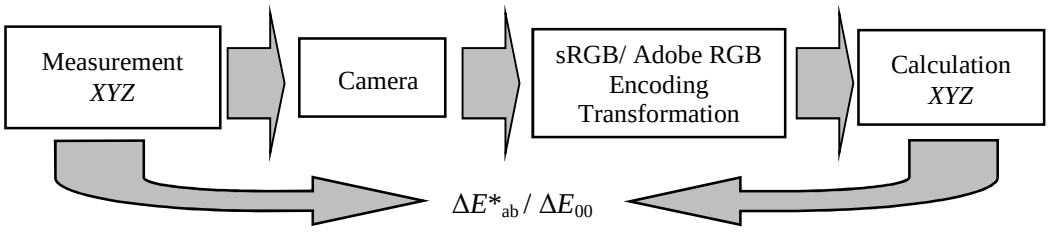


Figure 3-11. Process for the camera colour reproduction characterisation, using process profiles.

The results of the camera's colour reproduction characterisation are presented in Table 3-4. The results are the mean, median[1] and maximum colour differences between the measurement *XYZ* and the calculation *XYZ* at all patches.

Setting	ΔE^*_{ab}			ΔE_{00}		
	Mean	Median	Max	Mean	Median	Max
1	16.85	17.11	49.39	8.31	8.08	25.34
2	17.04	17.34	50.27	8.22	7.98	25.60
3	16.27	16.53	40.03	7.89	7.42	24.81
4	17.18	17.36	51.46	8.22	7.95	25.41
5	18.16	17.79	53.59	9.10	8.46	26.18

Table 3-4. Descriptive statistics (mean, median, maximum) for colour differences (ΔE^*_{ab} and ΔE_{00}) between measured values and calculated values at five different colour settings.

The accuracy of the camera's colour reproduction was found to be rather low. The mean colour differences were more than 16.00 in ΔE^*_{ab} and 7.00 in ΔE_{00}. Orava and Jaaskelainen [158] (pp. 219) investigated colour errors in digital cameras and found means of 13.1 in ΔE^*_{ab} and 8.1 in ΔE_{00}. Possible reasons for this phenomenon include 1) the sRGB and the Adobe RGB encoding have colour reproduction errors due to the

[1] Median is often used when data are skewed, meaning that the distribution is not a normal distribution [172] (pp. 38-39).

narrow colour gamut [173] (pp. 86-98), and 2) the camera's colour reproduction has been intentionally set this way by the manufacture based on the observer's preference.

Also, the result demonstrated that the colour reproduction of the camera had little variation in relation to the colour settings. The smallest mean colour difference was achieved with the colour setting 3, i.e. 16.27 in ΔE^*_{ab} and 7.89 in ΔE_{00}. The manufacture's claim of the setting being colour clear appeared to be correct (cf. Table 3-2).

3.5.2 Colour characterisation using polynomial regression model

Since the accuracy of the camera's colour reproduction using sRGB and Adobe RGB was found to be unsatisfactory, the custom profile was used to ensure accurate colour reproduction. The custom profile was applied to a polynomial regression model, which assumes that the correlation between scene (CIE *XYZ*) and image (RGB) can be approximated by a set of simultaneous equations. The process of the polynomial regression model is illustrated in Figure 3-12.

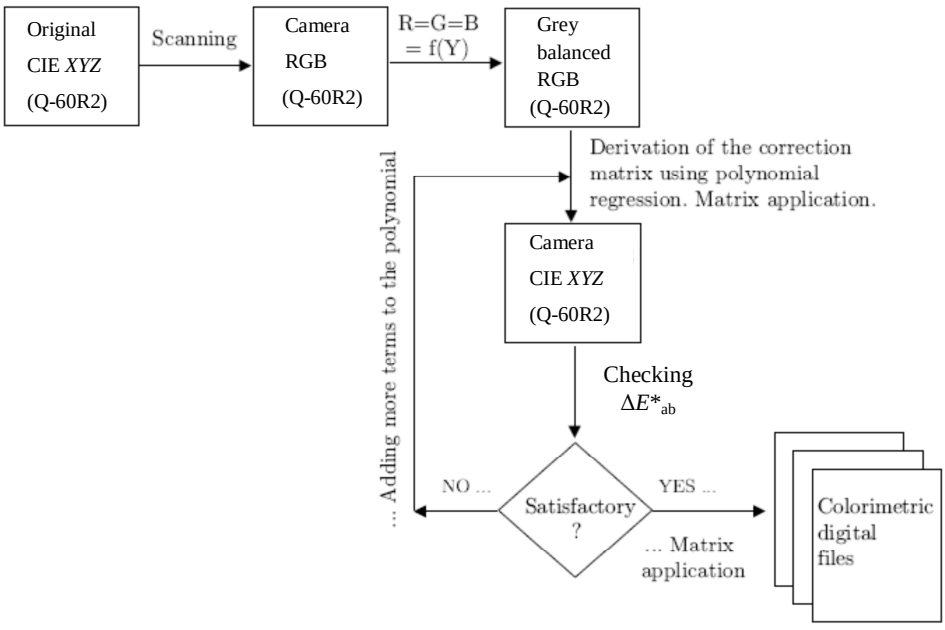


Figure 3-12. Process for the camera colour reproduction characterisation, using a custom profile, a polynomial regression model
(adapted from Triantaphillidou [33] (pp. 81)).

The polynomial regression model consists of two main steps: 1) the grey balance of red, green and blue signal (optional but preferable) and 2) the derivation of colour matrixes [174] (pp. 282).

1) Grey balance red, green, blue signal: Grey balancing process was achieved by setting R=G=B=f(Y) for the neutral patches of the test target, where f(Y) was a function of luminance *Y*. MATLAB code reproduced and applied for the purpose [33] (pp. 83):

$$\text{RedLUT}=\text{interp1}(\text{RedL}, \text{Y255}, 0{:}255);$$
$$\text{GreenLUT}=\text{interp1}(\text{GreenL}, \text{Y255}, 0{:}255);$$
$$\text{BlueLUT}=\text{interp1}(\text{BlueL}, \text{Y255}, 0{:}255);$$

where the RedLUT, GreenLUT and BlueLUT are red, green and blue look-up-tables, and the function Yi=interp1(Xo,Yo,Xi) interpolates to find Yi, the values of the underlying function Yo at the points in the vector Xi. The vector Xo specifies the points at which the data Yo is given. The RedL, GreenL, BlueL are vectors that contain the measured red, green and blue pixel values of the neutral patches in the images, rescaled from 0 to 255.

$$\text{RedL} = [R_1, R_2, R_3, \ldots R_{20}, R_{21}, R_{22}]$$
$$\text{GreenL} = [G_1, G_2, G_3, \ldots G_{20}, G_{21}, G_{22}]$$
$$\text{BlueL} = [B_1, B_2, B_3, \ldots B_{20}, B_{21}, B_{22}]$$

The Y255 is the vector containing the original reflectance measurements in the test target, rescaled from 0 to 255.

$$\text{Y255} = [L_1, L_2, L_3, \ldots L_{20}, L_{21}, L_{22}]$$

The look-up-tables were implemented to obtain the grey balanced image.

$$\text{Imageout}(:, :, 1) = \text{RedLUT}(\text{Image}(:, :, 1)+1);$$
$$\text{Imageout}(:, :, 2) = \text{GreenLUT}(\text{Image}(:, :, 2)+1);$$
$$\text{Imageout}(:, :, 3) = \text{BlueLUT}(\text{Image}(:, :, 3)+1);$$

where the Imageout is the grey balanced image, Image is the captured image and the parameters (:, :,1), (:, :,2) and (:, :,3) are referring to red, green and blue channels.

2) Derivation of colour matrixes: A number of colour correction matrixes with polynomial terms (*m*) were derived, using the following equation (Table 3-5).

$$M = (R^T R)^{-1} R^T H \quad (3.1)$$

where R is the matrix of independent variables of size $3 \times m$, H is the vector of dependent variables and R^T denotes the transposition of R and R^{-1} its inverse.

m	**Polynomial Terms**
3	*R,G,B*
4	*R,G,B,RGB*
6	*R,G,B,* R^2,B^2,G^2
7	*R,G,B,* R^2,B^2,G^2,*RGB*
9	*R,G,B,* R^2,B^2,G^2, *RG, GB,BR*
10	*R,G,B,* R^2,B^2,G^2, *RG, GB,BR,RGB*
12	*R,G,B,* R^2,B^2,G^2, *RG, GB,BR,* R^3, G^3,B^3
13	*R,G,B,* R^2,B^2,G^2, *RG, GB,BR,* R^3, G^3,B^3,*RGB*

Table 3-5. Polynomial terms (*m*) used in the matrix derivation.

The transform matrix M was then calculated into a CIE *XYZ* value. The best M was assessed using the colour differences between the original and calculated colours.

Table 3-6 describes the colour difference values, ΔE^*_{ab} and ΔE_{00}, between the original *XYZ* and the calculated *XYZ*. The results show that the colour difference values decrease when the number of polynomial terms increases[2]. This phenomenon has been reported by many researchers [168] (pp. 167-169), [169] (pp. 80-81).

[2] The black point addition is a possibility to produce a better fit in the regression [169] (pp. 83).

Polynomial terms (*m*)	ΔE^*_{ab}			ΔE_{00}		
	Mean	**Median**	**Max**	**Mean**	**Median**	**Max**
3	9.93	9.85	20.60	6.85	6.66	14.51
4	8.14	7.72	16.71	5.10	4.64	9.21
6	6.33	5.60	12.23	4.84	4.39	9.76
7	5.47	5.41	12.36	4.20	4.24	7.10
9	5.37	4.95	10.20	3.87	3.64	7.69
10	4.49	4.60	8.80	3.29	3.04	6.14
12	3.88	3.74	6.67	2.83	2.68	5.66
13	3.79	3.92	6.92	2.72	2.27	5.37

Table 3-6. Descriptive statistics (mean, median, maximum) for colour differences (ΔE^*_{ab} and ΔE_{00}) between measured values and calculated values at different polynomial terms (*m*).

The acceptability threshold for colour difference in complex images was reached at Matrix 7, but the perceptibility colour difference was unattained up to Matrix 13, i.e. the perceptibility threshold is 3.00 and the acceptability threshold is 6.00 in ΔE^*_{ab} (Figure 3-13).

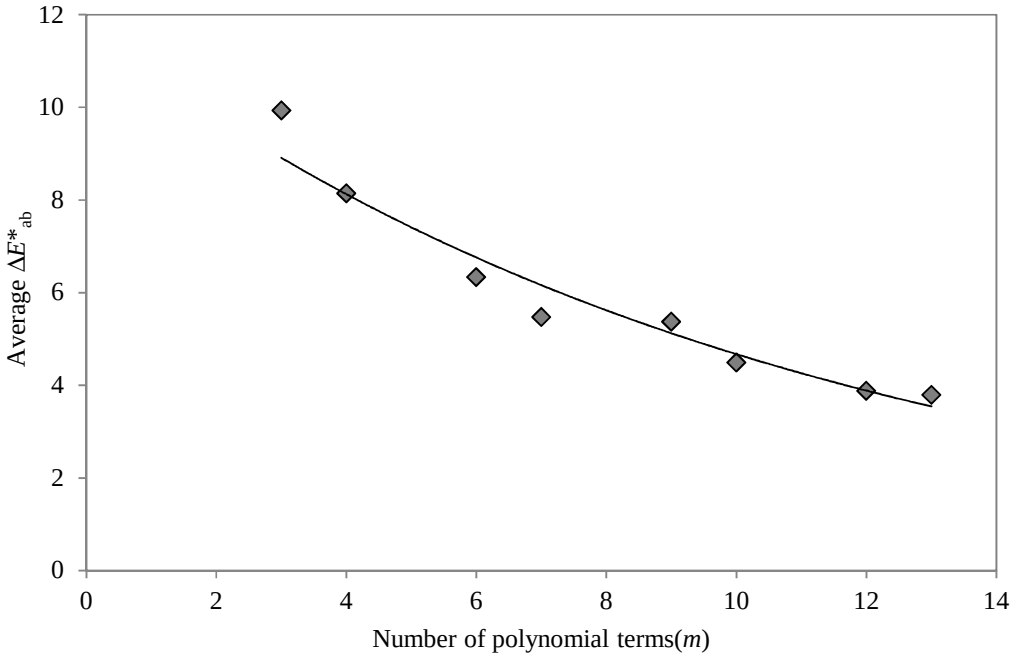


Figure 3-13. Effect of the number of terms (*m*) in the polynomial regression.

Overall, the polynomial regression model resulted in visually equivalent colour reproduction with high polynomial terms. However, the process is complicated and time-consuming. In addition, the task was only carried out in a specified standard condition. In ISO 17321-1: 2006 [175] (pp. 4), it has been noted that “when target-based characterisation is used, the resultant characterisation data is only applicable for similar geometric and spectral illumination characteristics.” Thus, the model is extremely limited in relation to real-life photography (cf. Chapter 6). Further extensive investigations are required for accurate colour reproduction in the real scene images.

3.6 Summary

A Canon EOS-1Ds SLR camera was characterised in terms of spatial uniformity, tone reproduction and colour reproduction. The following summary lists the main findings regarding the camera's characteristics.

1. The spatial variations of the camera ranged from 0.15 to 5.38 in ΔE^*_{ab} and from 0.12 to 2.82 in ΔE_{00}. The highest colour difference was identified as being in the top-right corner of the capturing frame. The result was greater than the perceptibility colour difference in complex scenes, 3.00 in ΔE^*_{ab}. Therefore, the rest of the investigations were carried out using the central area of the image (cf. Section 5.1 & 5.2 and section 6.1).
2. The camera's tone reproduction had little variation in relation to the colour settings used. In terms of tone reproduction, it was possible to use any setting desired. For this project sRGB was used for image capture.
3. With regard to the camera's tone reproduction, the gamma and offset of the linear regression was approximately equal to 0.60 and 0.02 with a 0.997 correlation coefficient. The gamma settings of the camera were intentionally set by the producer for the optimal overall gamma of between 1.0 and 1.5.
4. The camera's colour reproduction was investigated using two methods: 1) process profiles of all of the camera's colour settings and 2) polynomial regression model, which is a type of custom profile. The polynomial regression model provided visually equivalent colour reproduction with high polynomial terms. However, the process was found to be time-consuming and complicated to implement. The method had limitations for a natural scene. On the other hand, camera colour reproduction at sRGB/ Adobe RGB encoding was quick and less complex to implement. Thus, the camera setting with the smallest colour difference - setting 3, was applied in order to optimise the colour image for the rest of studies (cf. Section 5.1 & 5.2 and section 6.1). Also, the sRGB was designed to match the display output used for the psychophysical investigation (cf. Section 6.3).

Chapter 4

Liquid Crystal Display (LCD) characterisation

This chapter describes the processes of calibration and characterisation of the LCD, prior to it being used in the psychophysical experimentation. The LCD characterisation was carried out in terms of temporal stability, spatial uniformity, viewing angle, tone reproduction and colour reproduction.

4.1 LCD description

An EIZO CG210 LCD [178] was used in this work. The LCD was controlled by a graphic card (S3 Graphics Prosavage DDR (Microsoft Corporation)) in a personal computer running Windows XP professional.

To produce a colour image on the LCD, each pixel is divided into three sub-pixels, which are coloured red, green, and blue, respectively. The intensity of each sub-pixel can be controlled independently to yield several possible colours for each pixel. The graphic card was configured to display 24-bit colour. Figure 4-1 presents the sub-pixels forming a colour pixel on the LCD [176, 177].

The maximum resolution of the display was set to 1600 by 1200 at a frequency of 86 Hz [178]. However, the graphic card was configured at a resolution of 1600 by 1200 pixels at

a frequency of 60 Hz. This setting was applied throughout the experiment, because this is the native and maximum setting of the device. The specifications of the LCD are set out in Table 4-1.

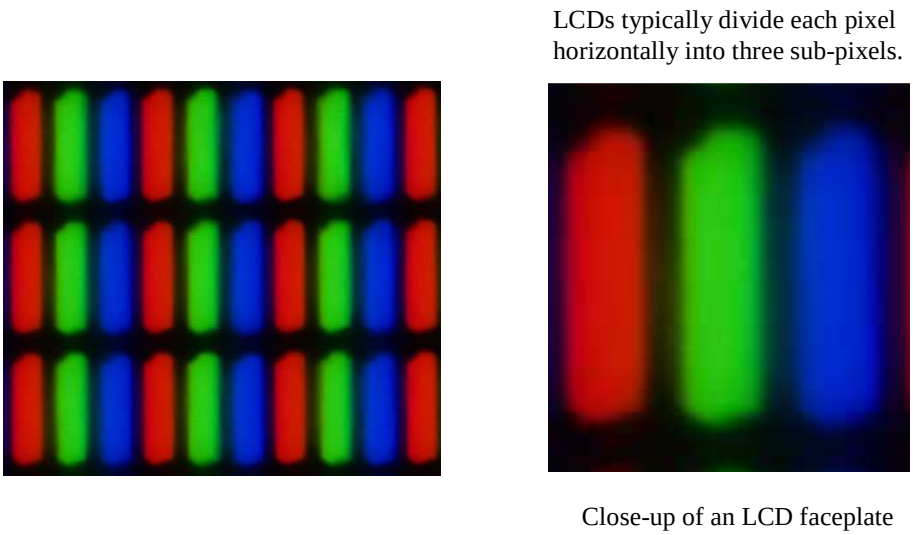


Figure 4-1. Example of sub-pixel on TFT-LCD.

(adapted from Wikipedia (viewed March 2010) [176] and Farrell *et al* [177] (pp. 25))

EIZO CG210	
Size and Type	54 cm (21.3") Thin film transistor (TFT) colour LCD panel
Viewing Angles (H, V)	170°, 170° (at contrast ratio of 10:1)
Luminance/ Contrast	250 cd/m^2 /550:1
Maximum Resolution	1600 × 1200
Active Display Size (H, V)	432 × 324 mm / 17.0 × 12.8"
Viewable Image Size	Diagonal: 540 mm / 21.3"
Pixel Pitch	0.270× 0.270 mm
Gamut Coverage	sRGB: 99%, Adobe RGB: 78%
Scanning Frequency (H, V)	Analogue :24 – 100 kHz, 49 – 86 Hz Digital : 31– 100 kHz, 59 – 61 Hz
Display Mode Options	sRGB, Custom

Table 4-1. The LCD technical specifications
(produced by Eizo [178]).

4.2 Conditions for measurements

Before the measurements, the LCD[1] was adjusted as close as possible to a white point of D65, a gamma of 2.2 and a white point luminance of 100 cd/m^2. These are the default values for photography/graphic design in Colour Navigator 4.1 Eizo software [179].

The LCD measurements were carried out in complete darkness, using a Minolta CS-200 hand held incident colorimeter (designed specifically to measure LCDs), which enables accurate measurement of luminance and chromaticity [180]. The colorimetry of the displayed samples was measured using the 0.2° measuring angle during a 30 second measuring period. The instrument's frequency was set at 60 Hz, the same frequency as that of the LCD as instructed by the manufacturer.

During measurements, the hand held incident colorimeter was mounted on a tripod with the colorimeter's optical axis being perpendicular to the faceplate of the LCD. The distance between the LCD's faceplate and the colorimeter was 128cm, as recommended by EN 61966-4: 2000, i.e. four times the display height [181] (pp. 9). The arrangement of the equipment is shown in Figure 4-2.

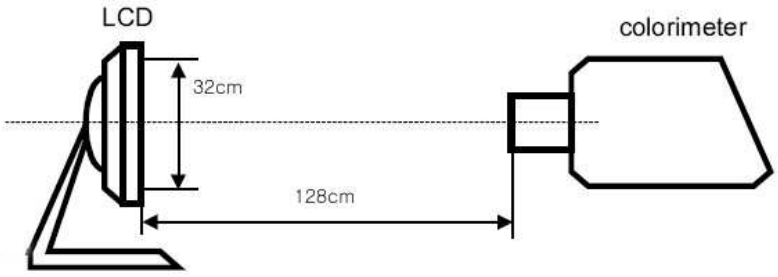


Figure 4-2. Equipment arrangement for the LCD measurements
(adapted from EN 61966-4: 2004 [181] (pp. 9)).

[1] sRGB reference display condition is a white point of D65, a gamma of 2.2 and a white point luminance of 80 cd/m^2(cf. Appendix A) [89] (pp. 9).

4.3 Temporal stability

The temporal stability of the LCD was estimated from a white patch displayed on a black background. The luminance *Y* and the CIE *x,y* chromaticity coordinates were measured three times for both short-term stability and mid-term stability [181] (pp. 28-31). The short-term stability was evaluated over 1 hour, every 2 minutes. The mid-term stability was evaluated over 4 hours at 10 minute intervals. During the experiment, the measuring point was monitored as temporal stability could be affected by lack of spatial non-uniformity and viewing angle of the display.

The short-term stability results are shown in Figure 4-3. According to EN 61966-4: 2000 [181] (pp. 29), the luminance *Y* was plotted against time (in minutes), where the luminance *Y* on the vertical axis was from 80 cd/m^2 (0.8% average luminance level) to 120 cd/m^2 (1.2% average luminance level). The output luminance level increased quickly early on and then reached a very stable level. However, after 20 minutes it actually started to decrease slightly. The luminance of the display changed slightly over the one-hour period. This luminance change was not perceptible by the human eye, i.e. a luminance difference of about 1% to 2% is not visible [106] (pp. 105), [182] (pp. 38-39).

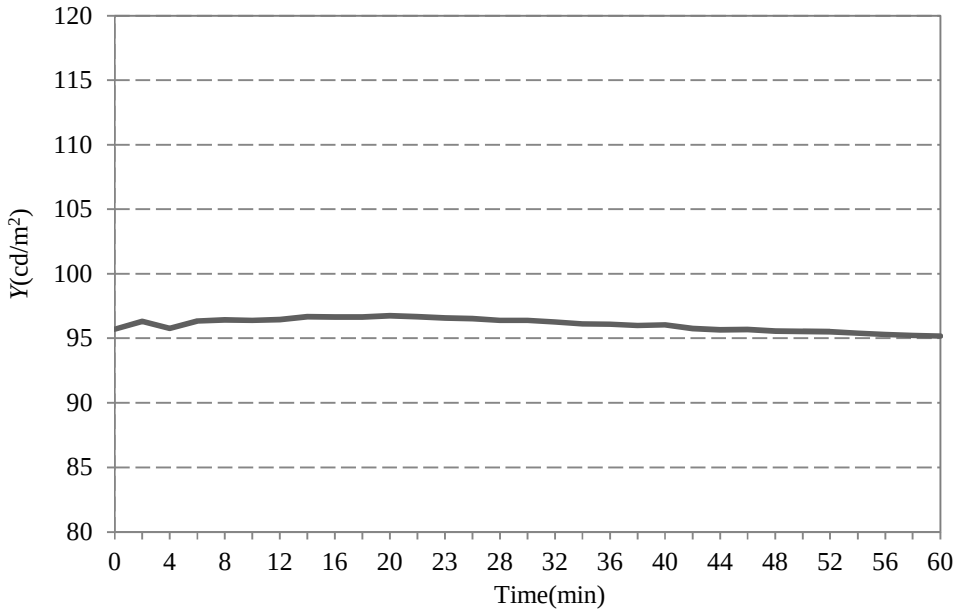


Figure 4-3. Short-term luminance stability on the LCD.

The CIE *x,y* chromaticity coordinates are plotted in Figure 4-4, where the vertical axis ranges from 0.25 to 0.35. The LCD stabilisation of the CIE *x,y* chromaticity coordinates seemed to be reached after about few seconds, and the LCD kept stable over the one-hour period.

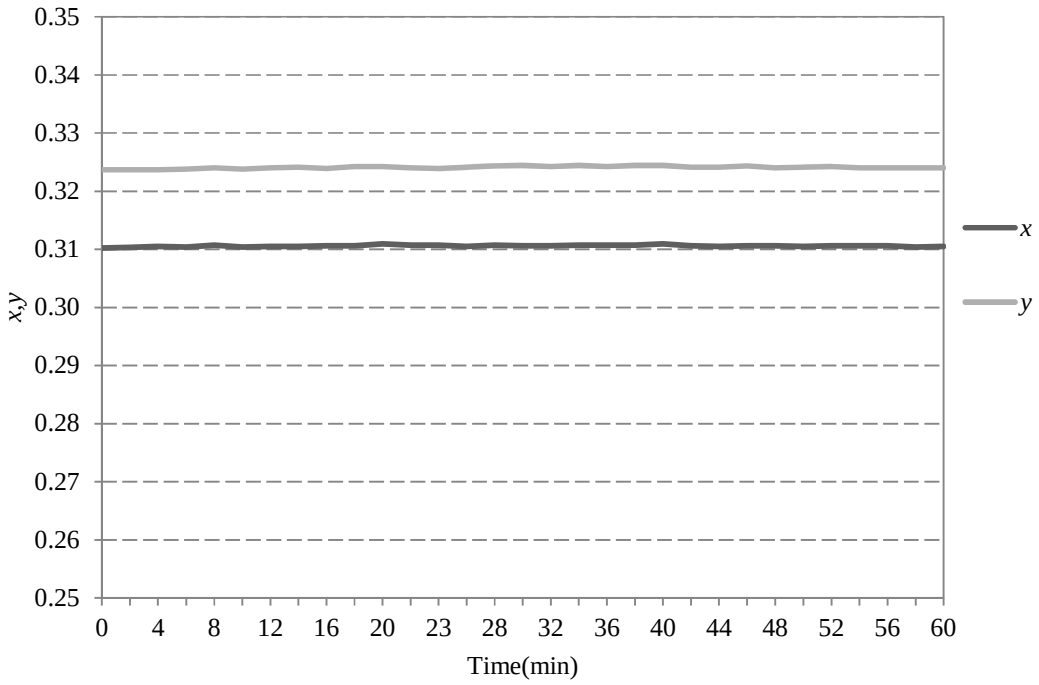


Figure 4-4. CIE *x,y* values of short-term stability on the LCD.

The results of mid-term temporal stability are shown in Figure 4-5 and Figure 4-6. Both the results of the luminance *Y* and the CIE *x,y* chromaticity coordinates presented a similar trend for short-term investigations, which are presented in Figure 4-3 and Figure 4-4, i.e. the luminance level increased quickly early on and then reached a remarkably stable level. After 20 minutes, it started to decrease slightly. The luminance of the display changed slightly over the measurement period. However, this luminance change, 93.01 (2.2%) at 240 minutes, was still not perceptible by the human eye. The chromaticity of the LCD reached a stable level after about few seconds, and the LCD stayed stable over the period.

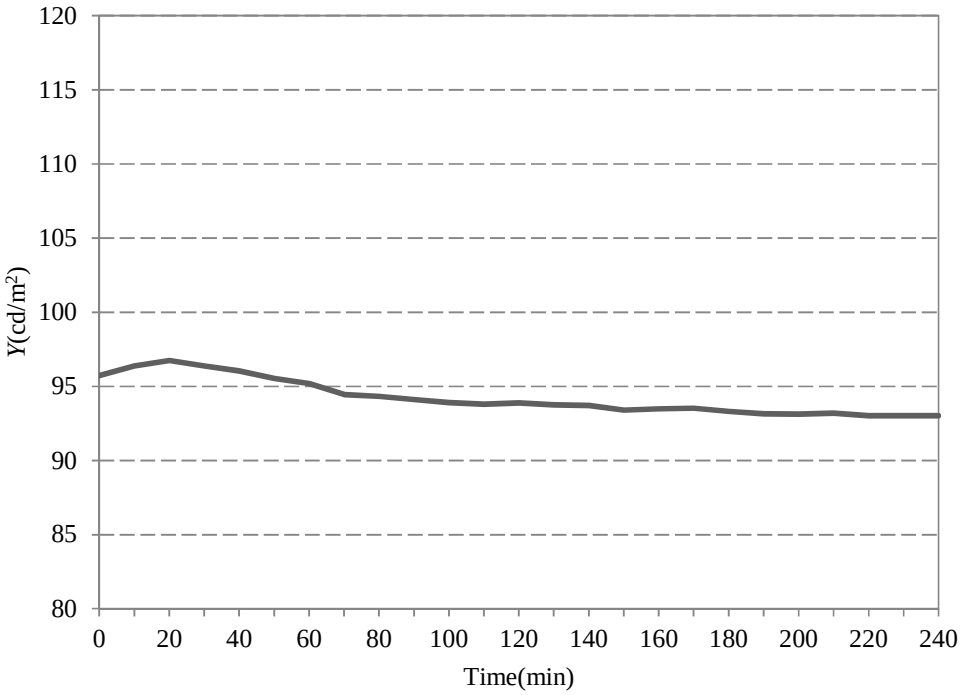


Figure 4-5. Mid-term luminance stability on the LCD.

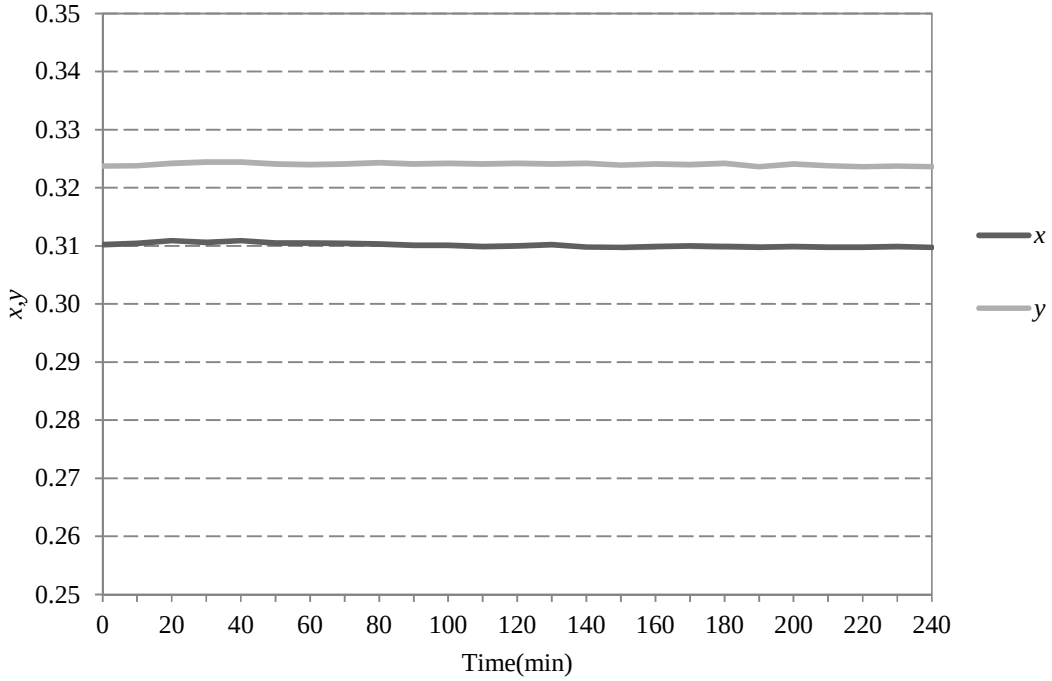


Figure 4-6. CIE *x,y* values of mid-term stability on the LCD.

Overall, the LCD temporal stability was remarkably good, when compared with previous researches on a CRT [106] (pp. 105-108) and LCDs [88] (pp. 6-8), [183] (pp. 18). The result confirmed that the LCD did not require significant warm up time (cf. Section 6.3).

4.4 Spatial uniformity

All measurements of the LCD were based on the assumption that the responses of all areas of the display were identical, for any input. In reality, the spatial characteristics of the display vary. The lack of uniformity of the display is often ignored, because it might be below the level of perception of the human eye.

The spatial uniformity of the LCD was investigated by measuring a white patch on 25 different areas of the display and comparing measurements between the middle area and across the display area. Figure 4-7 shows the measurement points on the LCD.

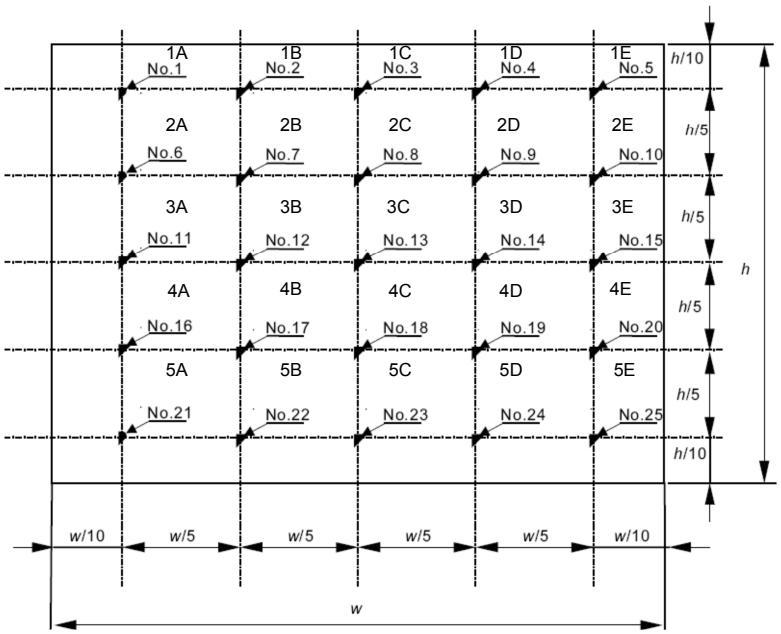


Figure 4-7. Measurement points for the LCD spatial uniformity (adapted from EN 61966-4: 2004 [181] (pp. 25)).

Figure 4-8 illustrates the uniformity of the LCD faceplate. The actual fluctuations of the 24 areas were ranged from 0.73 to 4.80 in ΔE^*_{ab} and from 0.62 to 3.45 in ΔE_{00}.

The lower-middle area as defined by 5C, showed the most non-uniform LCD measurements, 4.80 in ΔE^*_{ab}. The result is greater than the perceptibility colour difference in complex scenes, 3.00 in ΔE^*_{ab} [160]. A possible reason for the LCD non-uniformity is the backlight fall-off towards the faceplate [184] (pp. 35).

The correction of display spatial uniformity [185] (pp. 324) was not applied, because it could cause the creation of a contouring artefact. The contouring artefact would be more noticeable than lack of uniformity [32] (pp. 82).
The result led to the choice of the LCD's central area for the rest of the investigations (cf. Section 4.5, 4.6 & 4.7, section 5.1 & 5.2 and section 6.3).

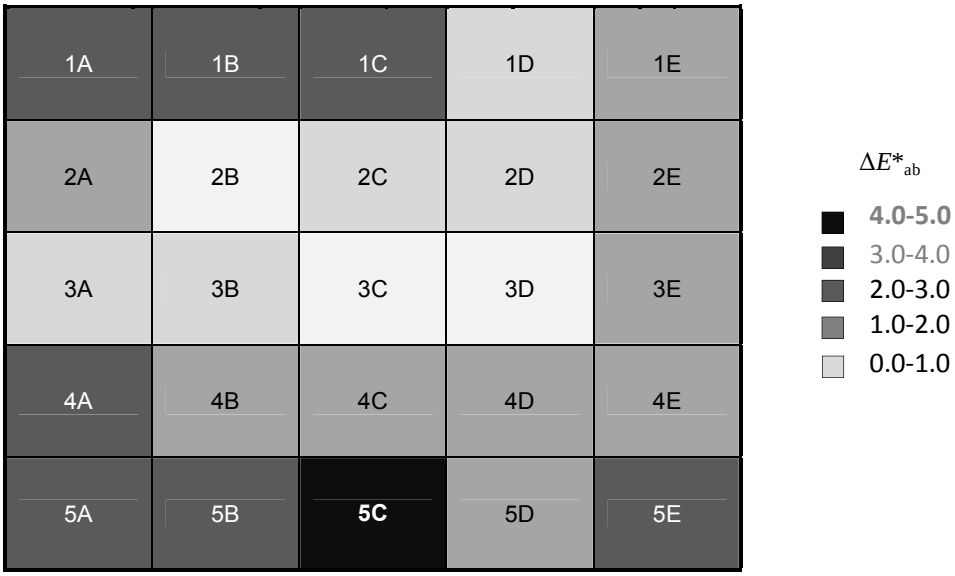


Figure 4-8. Spatial uniformity on the LCD, expressed in ΔE^*_{ab}.

4.5 Viewing angle

One of the major issues, when using an LCD, is the viewing angle. This is because LCDs usually have a limited viewing angle [151] (pp. 330). The display luminance variations with viewing angle were investigated by horizontal (right-left) and vertical (up-down) display rotations [181] (pp. 31-36). A number of patches were displayed in the centre of the display and measured by the Minolta CS-200. The arrangement of equipment is shown below in Figure 4-9.

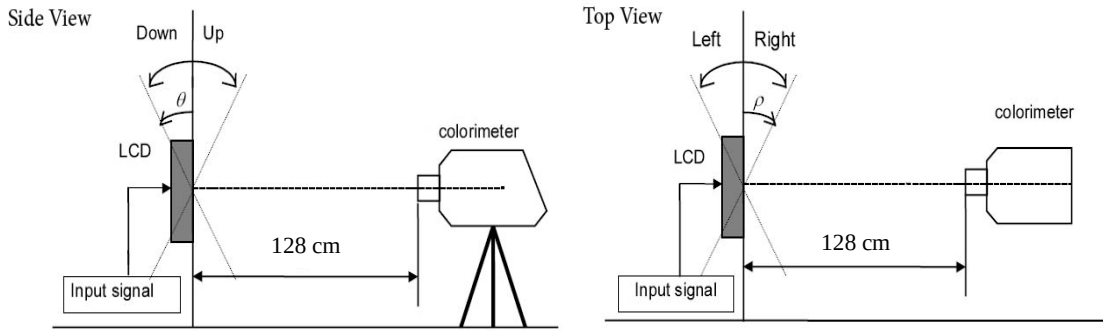


Figure 4-9. Equipment arrangement for the LCD viewing angle
(adapted from EN 61966-4: 2004 [181] (pp. 32)).

The luminance in peak red, green, blue and peak, medium and dark neutral scale signals are plotted in Figure 4-10 and Figure 4-11. The figures illustrate that the viewing angle characteristics have minimal effects on the lower luminance and a high impact on the higher luminance. Also, the figures show that the vertical viewing angle is impacted more than the horizontal viewing angle.

Table 4-2 presents the contrast ratio values at different viewing angles. The result shows that the manufacture's claim seems to be correct, which is the 10:1 contrast ratio of the 170° horizontal and vertical [178]. However, the 10:1 contrast ratio of the 170° viewing angle makes the image almost imperceptible, and as such the claim is not acceptable.

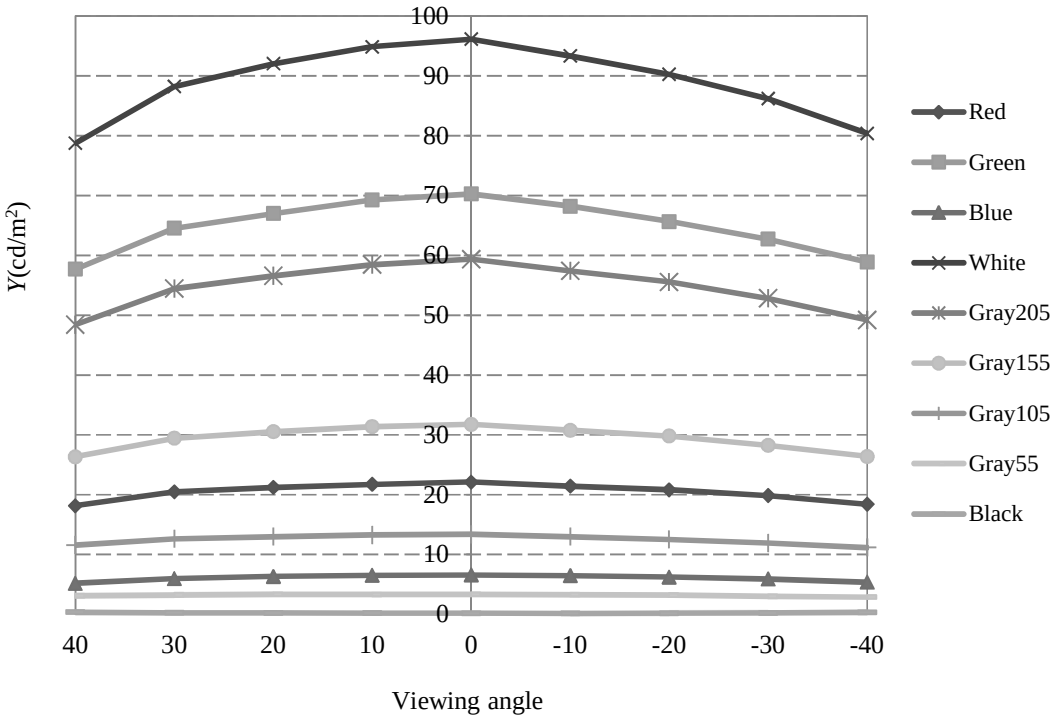


Figure 4-10. Horizontal viewing angle characteristics on the LCD.

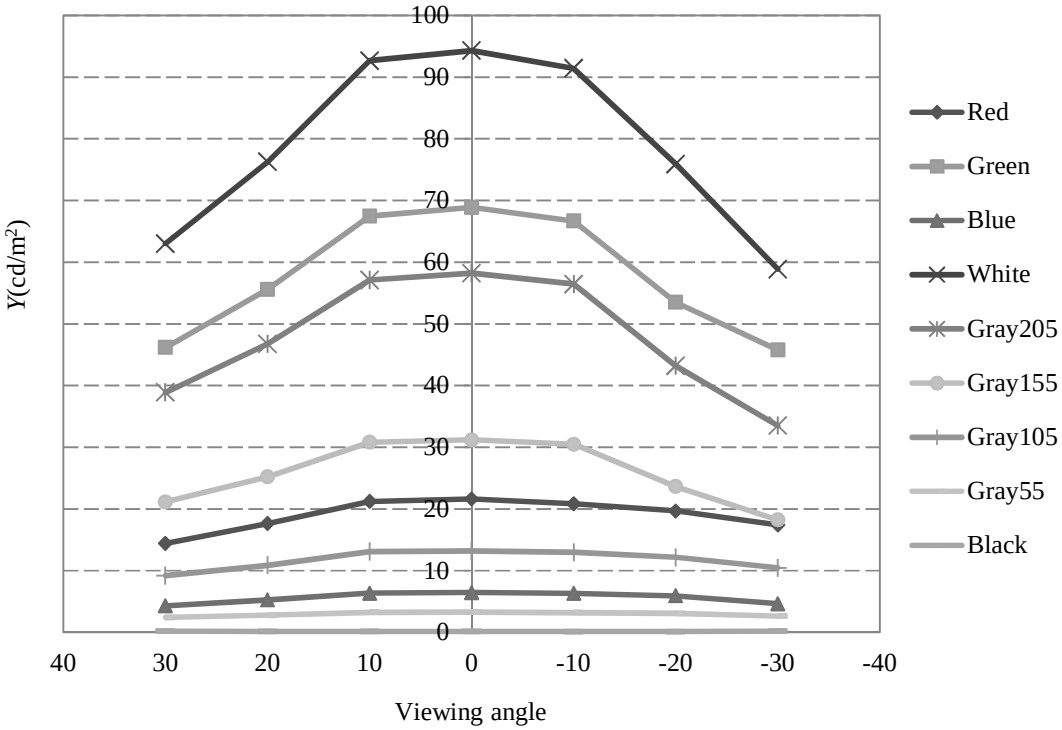


Figure 4-11. Vertical viewing angle characteristics on the LCD.

Viewing angle	Horizontal	Vertical
+30	344: 1	295: 1
+20	394: 1	440: 1
+10	474: 1	515: 1
0	534: 1	589: 1
-10	583: 1	596: 1
-20	492: 1	506: 1
-30	380: 1	268: 1

Table 4-2. Contrast ratio at various viewing angles on the LCD.

In total, the LCD appearance altered significantly with viewing angle. The result showed that the viewing angle of the LCD required careful control in the image quality study (cf. Section 6.3).

4.6 Tone reproduction

The LCD tone reproduction was assessed by the transfer function, described by the relationship between the input pixel values and the generated output luminance *Y* (in cd/m^2).

The transfer function was determined by measuring a number of colour patches. The colour patches were displayed one at a time, ranging from the system maximum to the minimum at 25 pixel value intervals. Each patch occupied 50 % of the LCD central area, with the surrounding area displaying black. The black background was suggested for LCD measurement by EN 61966-4: 2000 [181] (pp. 9). The interface of the software is presented in Figure 4-12.

The luminance of each patch was measured three times, using the Minolta CS-200, and then the measurements were averaged.

The LCD's transfer functions are illustrated in Figure 4-13. The result presented that the relationships between the input and output values appeared to be power functions. The LCD seemed to be corrected by built-in correction tables, to mimic the CRT response. This is because LCDs have a native sigmoid transfer function [85] (pp. 612), [87] (pp. 191-193).

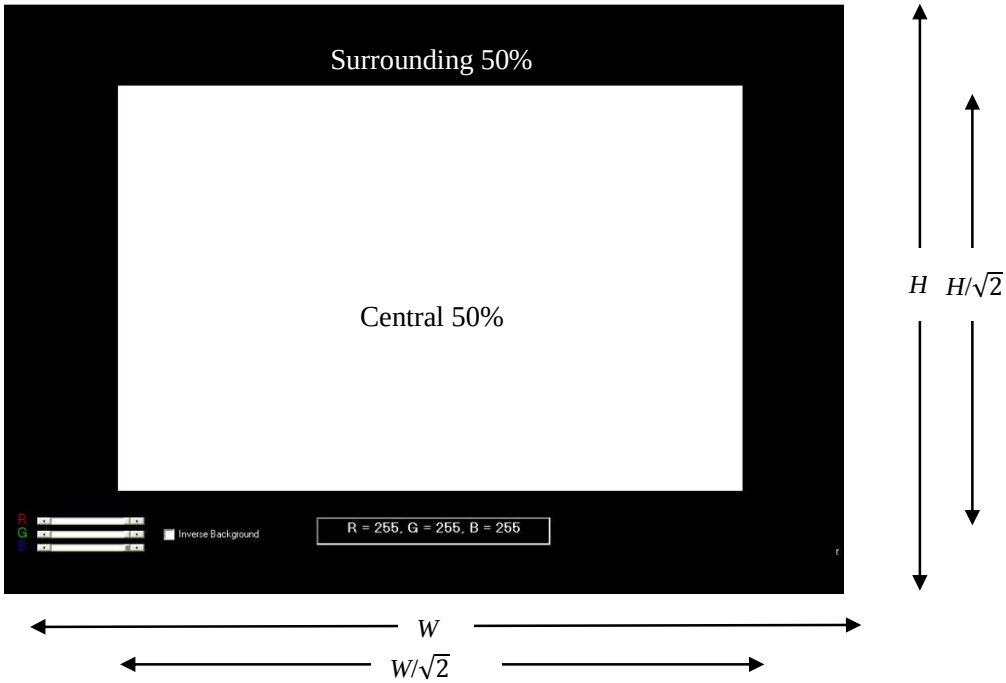


Figure 4-12. Interface for the LCD tone reproduction characterisation (produced by Bilissi [106] (pp. 113)).

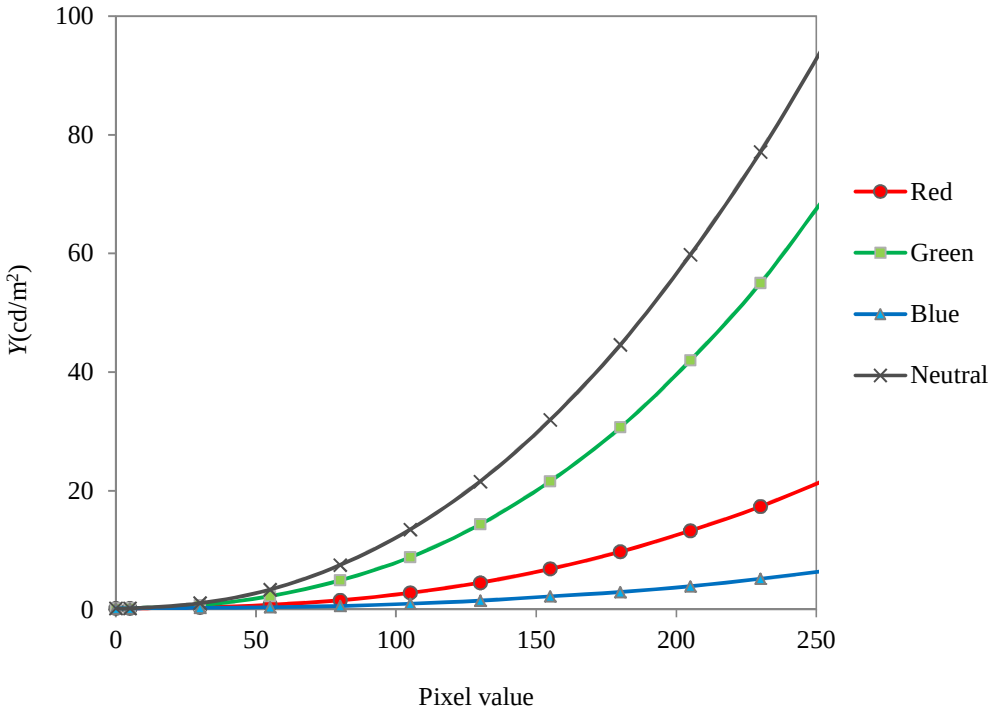


Figure 4-13. Transfer functions of the LCD at the R,G,B and neutral responses, plotted in linear-linear units.

The LCD's transfer functions in normalised units are also illustrated in Figure 4-14. In this figure, the transfer functions show appreciable neutral scale tracking, with the red, green and blue channel responses being similar.

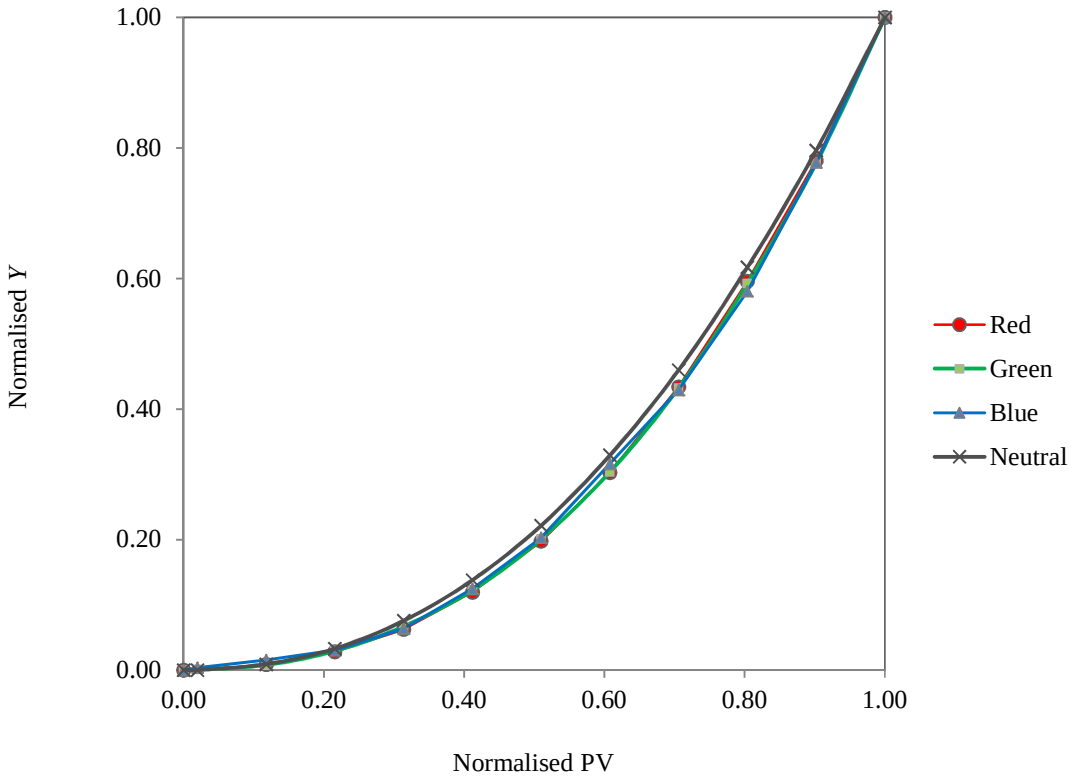


Figure 4-14. Transfer functions of the LCD at the R,G,B and neutral responses, plotted in linear-linear normalised units.

The LCD's transfer functions in log-log normalised units are also illustrated in Figure 4-15. The gamma and offset was derived with linear regression. Obtained data for each colour channel are listed in Table 4-3. It was observed that the gamma and offset of the linear regression was equal to 2.20 and 0.00 respectively for the neutral scale input signal.

The breakdown of the linear relationship was observed at the blue channel (r=0.992). In addition, the gamma error (γ=2.04) resulted at the blue channel, caused by low luminance. The phenomenon has also been found in a CRT [186] (pp. 147-148).

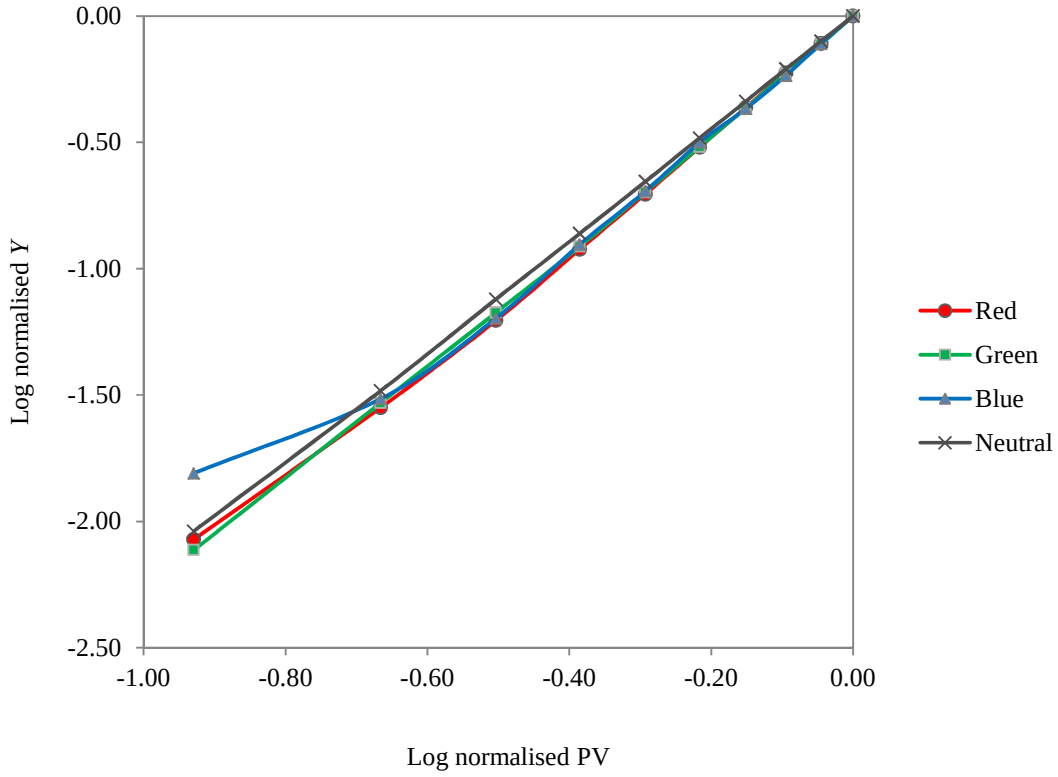


Figure 4-15. Transfer functions of the LCD at the R,G,B and neutral responses, plotted in log-log normalised units.

	Gamma (γ)	**Offset (o)**	**Correlation Coefficient (r)**
Red	2.22	0.02	0.999
Green	2.27	0.02	0.999
Blue	2.04	0.06	0.992
Neutral	2.20	0.00	0.999

Table 4-3. Gamma, offset and correlation coefficient of the LCD at the R,G,B and neutral colours.

In order to investigate channel additivity, the LCD neutral response and the combined RGB response were compared. The combined RGB response was obtained by equally weighting the three colour channels. The result of additivity is shown in Figure 4-16 and Table 4-4. The percentage errors between the neutral response and the combined RGB response were distributed through the different input signal levels (2.77%). The differences were small. The result indicates that the LCD is an additive colour system.

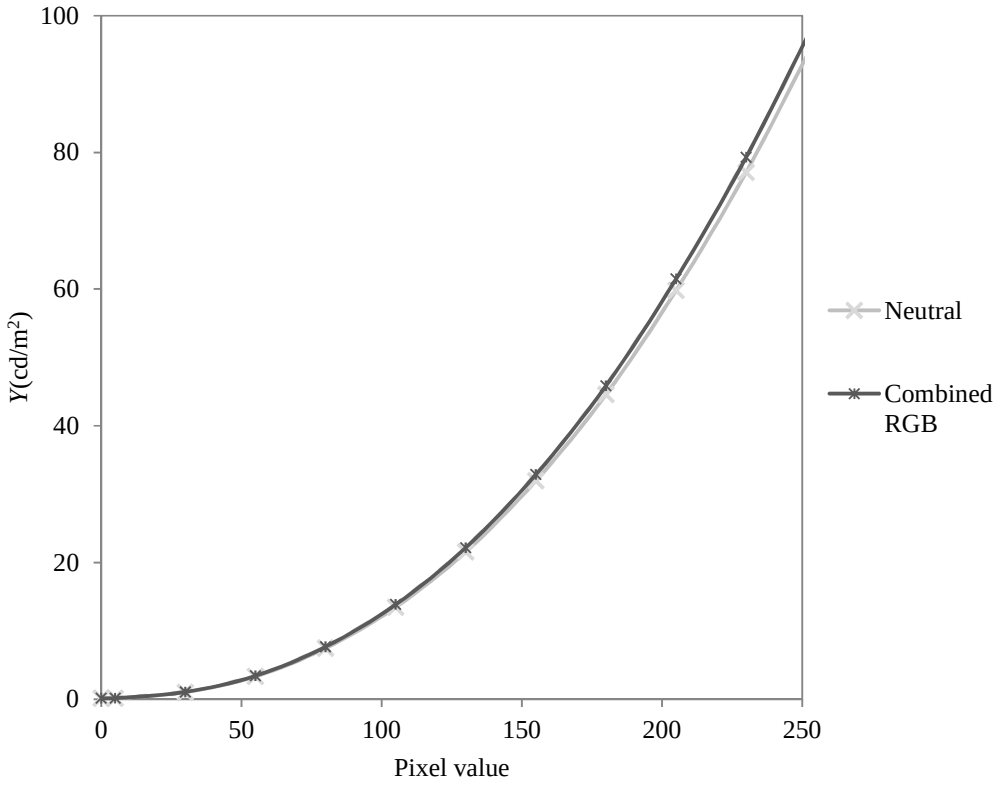


Figure 4-16. The LCD additivity, compared between the combined RGB response and the neutral response.

Input values	5	55	105	155	205	255
Combined RGB	0.19	3.45	13.87	32.89	61.50	99.54
Neutral	0.18	3.35	13.49	31.98	59.80	96.78
Percentage error	2.77%	2.77%	2.77%	2.77%	2.77%	2.77%

Table 4-4. The LCD additivity with percentage error.

4.7 Colour reproduction

The colour reproduction of the display was evaluated. Figure 4-17 shows the process for the LCD colour reproduction evaluation.

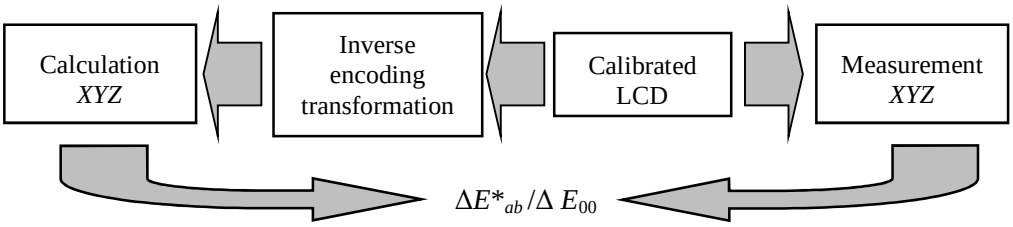


Figure 4-17. Process for the LCD colour reproduction characterisation.

Single, double and triple RGB colour combinations of digital colour signals were created one at a time, ranging from the system maximum to the minimum at 51 intervals. The displayed patches were measured three times, using the Minolta CS-200. The input RGB data on each colour patch were converted to the *XYZ* tristimulus values, using the sRGB inverse encoding transformation (cf. Appendix A). Finally, the differences between the calculation *XYZ* and the measurement *XYZ* were evaluated using two colour difference equations (ΔE^*_{ab} and ΔE_{00}). The reason of using two colour difference equations is explained in section 3.3.

Table 4-5 presents the result of the performance of the LCD's colour reproduction. The mean colour differences were 1.57 in ΔE^*_{ab} and 0.52 in ΔE_{00}. The median colour differences were 0.54 in ΔE^*_{ab} and 0.20 in ΔE_{00}. The maximum colour differences were 7.21 in ΔE^*_{ab} and 2.17 in ΔE_{00}. As expected, the maximum blue colour signal (R,G,B=0,0,255 at pixel value) had the most colour difference between the calculation *XYZ* and the measurement *XYZ*. Sharma [166] (pp. 29) noted that "the gamut of sRGB, is based on CRT-type of display and is very different and therefore not appropriate to use as a profile for an LCD flat-panel display. An LCD panel can display some colors, for example in the blue part of the color space. If an sRGB profile were used to represent this device, we would not get accurate colors, especially in the blue."

	Mean	**Median**	**Maximum**
ΔE^*_{ab}	1.57	0.54	7.21
ΔE_{00}	0.52	0.20	2.17

Table 4-5. Colour reproduction characterisation of the LCD, described in mean, median, maximum of colour differences (ΔE^*_{ab} and ΔE_{00}).

Figure 4-18 illustrates the LCD colour gamut and the sRGB colour gamut at the CIE *u'v'* chromaticity diagram.

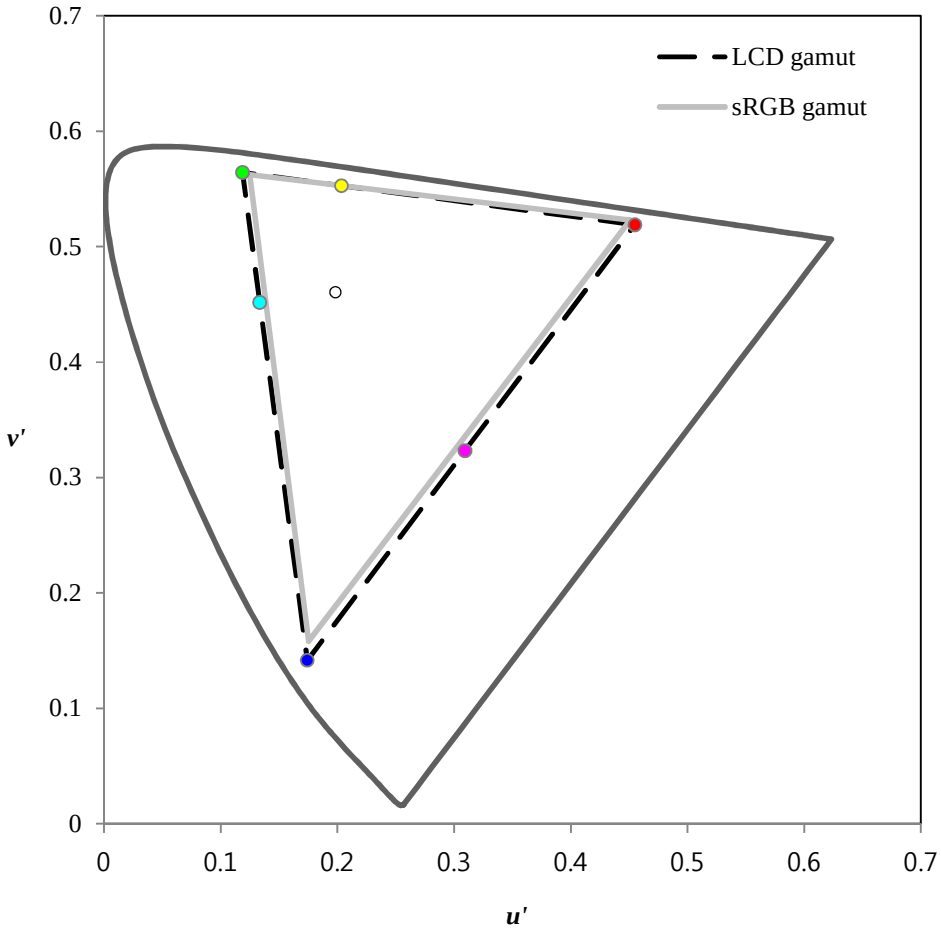


Figure 4-18. The LCD colour gamut and the sRGB colour gamut, plotted in the CIE *u'v'* chromaticity diagram.

4.8 Summary

The Eizo CG210 LCD was characterised in terms of temporal stability, spatial uniformity, viewing angle, tone reproduction and colour reproduction. The following summary lists the main findings of the LCD characteristics.

1. The LCD temporal stability was satisfactory. The result indicates that the LCD does not require significant warm up time in order to stabilise.
2. There was a spatial variation of the display, ranged from 0.73 to 4.80 in ΔE^*_{ab} and from 0.62 to 3.45 in ΔE_{00}. The highest colour difference was perceptible to the human eye and area located the lower-middle of the display. This result leads to the choice of the LCD centre area for the psychophysical investigation.
3. The LCD appearance altered significantly with viewing angle. The result indicates that the viewing angle of the LCD requires careful control in the image quality study.
4. The LCD transfer function was found to fit a power function model. The gamma and offset of the linear regression was equal to 2.20 and 0.00 respectively for the neutral scale input signal.
5. The LCD colour reproduction was relatively satisfactory. The average colour difference value was 1.57 in ΔE^*_{ab} and 0.52 in $\Delta E_{00.}$ The LCD can display accurately imaged colours for the psychophysical scaling.

The results were able to suggest the specific display conditions that were further studied in the subjective scaling of image quality (cf. Section 6.3) and the objective scaling (cf. Section 8.1).

Chapter 5

Modulation transfer and noise characterisation

This chapter presents image modulation transfer and noise characterisation of the imaging chain as part of assessing the performance of the imaging systems used in this study. Data from this investigation will be used later for the image quality metric implementation in chapter 8. The aim of the chapter is to quantify the modulation transfer function (MTF) and noise of the camera and the LCD, and to understand the effect of MTF associated with image down-sampling and of noise power spectrum (NPS) in relation to Gaussian noise.

5.1 Modulation Transfer Function (MTF)

5.1.1 Camera MTF

As mentioned in section 2.2.2.5, there are many methods for camera MTF measurement. For this project, the slanted edge technique was used for the camera MTF measurement, since it is quick and easy to implement and produces accurate and repeatable results [117] (pp. 235). It is a standard method recommended in ISO 12233: 2000 [110].

Experimental work was undertaken using the Canon EOS-1Ds digital camera (cf. Table 3-1). This digital camera operates with a CMOS sensor with approximately 8.8 μm

square pixel dimensions. The Nyquist limit of the CMOS sensor array, calculated from these pixel dimensions, is 56.8 cycles per millimetre.

Theoretically, *aliasing* occurs beyond the camera's Nyquist limit. Modern cameras generally prevent this aliasing by an anti-aliasing filter (low-pass filter) [187] (pp. 737). Alongside the positive adjustment, the filter also causes degradation in the optical MTF. Thus, digital cameras often include edge sharpening as part of their processing to compensate for the losses [187] (pp. 753).

A slanted edge test target, QA-62 [188], was carefully selected to determine the camera SFR and thus derive the MTF (cf. Section 2.2.2.5). The ratio of the maximum chart reflectance R_{max} to the minimum chart reflectance R_{min} was 74: 1. ISO 12233: 2000 [110] (pp. 6) recommends that the ratio should be not less than 40:1 and not greater than 80:1. The test target's modulation was 0.54.

The test target consists of 4 edges and 20 grey steps (Figure 5-1). The edges are intentionally slanted at 5-10 degree, and grey steps consist of different, spectrally neutral patches.

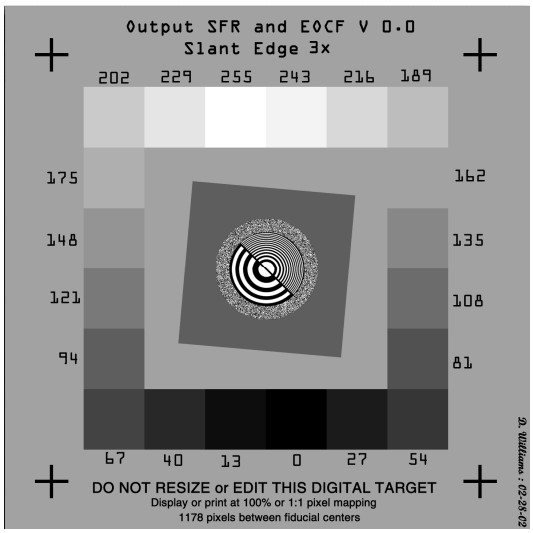


Figure 5-1. A slanted edge target, QA-62 (reproduced from Applied Image [188]).

The camera was mounted on a tripod with the optical axis of the lens being orthogonal to the plane of the target. The camera was placed 0.85 m away from the test target. This gave the desired low magnification, so that the frequency content of the target was constant for the range of spatial frequencies of interest [33] (pp. 187), [114] (pp. 135). Thus, the measured SFR was taken as the MTF of the result.

Two gas-filled tungsten lamps were placed either side of the target at 45° and 1m away, so that all areas of the target were uniformly illuminated. Even illumination was confirmed using a Kodak R-27 18% grey card, a SEKONIC L-308s light meter and a Minolta CL-200 colorimeter. Nine measurements were taken from different regions of the card. The non-uniformity of illumination was less than ± 5 % of average illumination (cf. Section 3.2).

The camera lens was covered with a black hood to reduce flare and was set to an aperture of *f* 11. This provided an increased depth of field as well as good lens performance, for the camera and the combined system (cf. Section 5.1.2). Keelan and Pagano [189] noted that the optimum MTF usually occurred two or three stops down from the wide open aperture. This is typically in the range of *f* 5.6 to *f* 11 (cf. Figure 8-1).

The camera was set to sRGB colour mode, setting 3 (cf. Section 3.4 and 3.5). It was operated in the self-timer mode to minimise distortion caused by camera shake.

The sharpness and noise settings of the camera were to keep a specific setting in the firmware. The aim of this setting was to use minimum sharpness and some noise reduction. However, both procedures are non-linear and may introduce distortions in the measured MTF.

The image was captured in the central area of the frame (cf. Section 3.3) and in correct focus. The equipment set-up is illustrated in Figure 5-2.

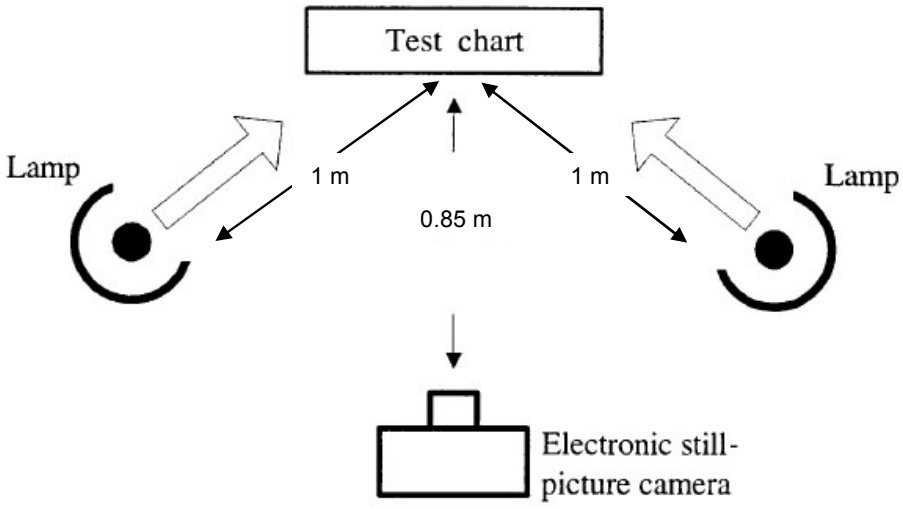


Figure 5-2. Equipment arrangement used to measure the camera MTF
(adapted from ISO 12233: 2000 [110] (pp. 10)).

The captured images were downloaded to a computer as 8-bit TIFF uncompressed images, and the ImCheck software [190] was used to calculate the grey scale MTF (cf. Section 2.2.2.5).

The averaged horizontal and vertical MTFs with variations in four measurements are illustrated in Figure 5-3. In accordance with ISO12233: 2000 [110] (pp. 15), the results are reported as the average of four measurements for the horizontal and vertical directions. The responses of the system are shown approximately up to its Nyquist limit. The same results are also illustrated in Figure 5-4, this time fitting them with a third degree polynomial function with a correlation coefficient, *r* (Table 5-1). The third degree polynomial is often used to represent the MTF of digital cameras [33] (pp. 180).

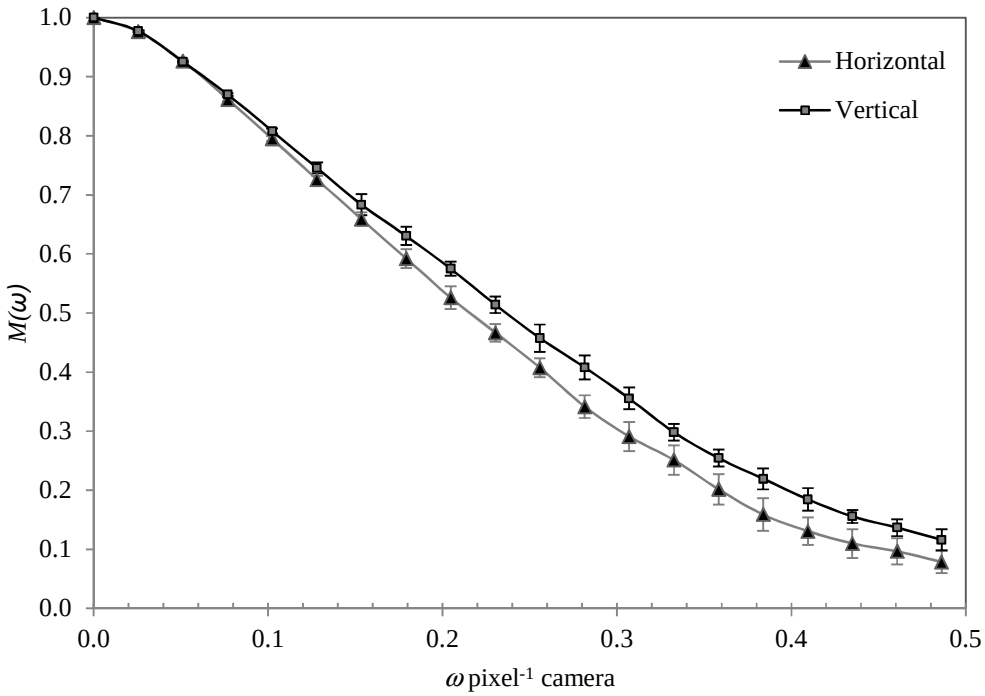


Figure 5-3. Measured MTF curves of the camera, including the variation in four measurements.

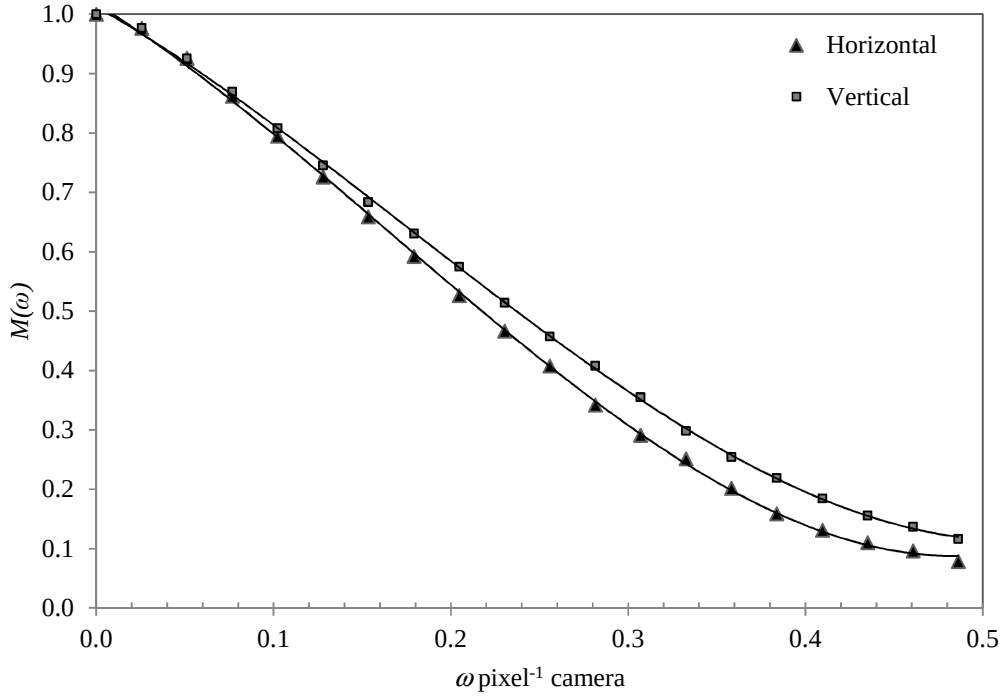


Figure 5-4. Polynomial functions representing the camera MTF curves.

Component	Polynomial function	*r*
Camera's horizontal edge	$M(\omega) = 8.483\omega^3 - 4.232\omega^2 - 1.861\omega + 1.017$	0.9998
Camera's vertical edge	$M(\omega) = 6.639\omega^3 - 3.487\omega^2 - 1.711\omega + 1.012$	0.9999

Table 5-1. Third degree polynomial functions representing the camera MTFs, $M(\omega)$. The spatial frequency, ω, is measured in cycles/pixel, and *r* represents the correlation coefficient of fitted functions to measured data.

The camera MTF 10% values were 58.5 cycles/mm in the horizontal direction and 61.3 cycles/mm in the vertical direction. The 10% MTF frequencies were approximately equivalent to that of the 10% MTF of medium-speed colour slide material [219]. The camera MTF 50% values were 26.1 cycles/mm in the horizontal direction and 29.5 cycles/mm in the vertical direction.

5.1.2 Camera-display MTF

This part of the work was based on a previous study [191], where the MTF of a CRT display system was measured with an SLR camera. The MTF of the display was evaluated by dividing the combined MTF by that of the acquisition system.

$$M_o(\omega)=M_a(\omega)\times M_d(\omega) \tag{5.1}$$

where the M_o is the MTF of the overall system, the M_a is the MTF of the acquisition system and the M_d is the MTF of the display system.

This method was adapted in the characterisation of the LCD MTF. It used relatively inexpensive equipment [191] (pp. 58) and has become a common way of measuring display MTFs [192] (pp. 6-8).

For this purpose, the measurement of the combined system MTF was initially carried out as follows:

1) A digitally constructed edge target was displayed in the central area of the LCD (cf. Section 4.4). The edge target had the same configuration as that of the camera target, i.e. it consisted of 4 edges and 20 grey steps (cf. Figure 5-1). Its modulation was 0.27. The different characteristic of the input modulation to that of the reflectance test target (cf. Section 5.1.1) tended to be insignificant, since similar spatial frequency responses were produced at different modulations, ranging from 0.2 to 0.7 [191] (pp. 64).

2) The camera was positioned approximately 0.9 m from the LCD faceplate, to minimise the effect of structural artefacts associated with the display [33] (pp. 186), [192] (pp. 7).

3) The camera was set to the same settings that were used to evaluate the camera MTF (cf. Section 5.1.1). The displayed edge was captured four times by the camera with its optical axis orthogonal to the LCD faceplate.

4) The captured images were downloaded to a computer as 8-bit TIFF uncompressed images and the ImCheck software [190] was implemented to calculate the combined grey scale MTF. The calculation was conducted using a high aspect ratio of the rectangular region-of-interest (ROI), to increase the signal-to-noise ratio of the MTF estimates [191] (pp. 62).

The equipment set-up is illustrated in Figure 5-5.

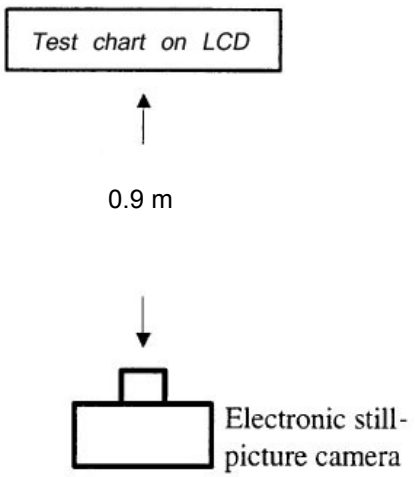


Figure 5-5. Set-up for measuring the combined camera-LCD MTF.

The responses of the combined system are shown up to the 30 (cycles/mm) frequency of the camera plane. Figure 5-6 illustrates the average frequency responses and includes the variation in the four measurements. The same results are also illustrated in Figure 5-7, this time fitted using a third degree polynomial function with a correlation coefficient, *r* (Table 5-2).

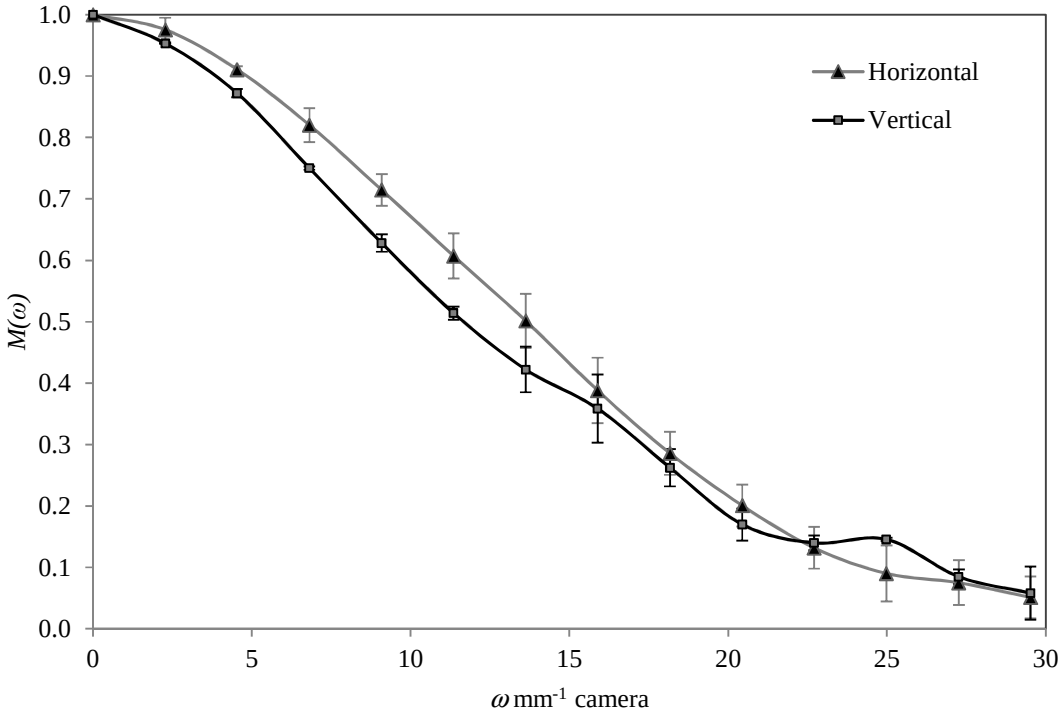


Figure 5-6. Measured MTF curves of the combined camera-LCD, including the variation in four measurements.

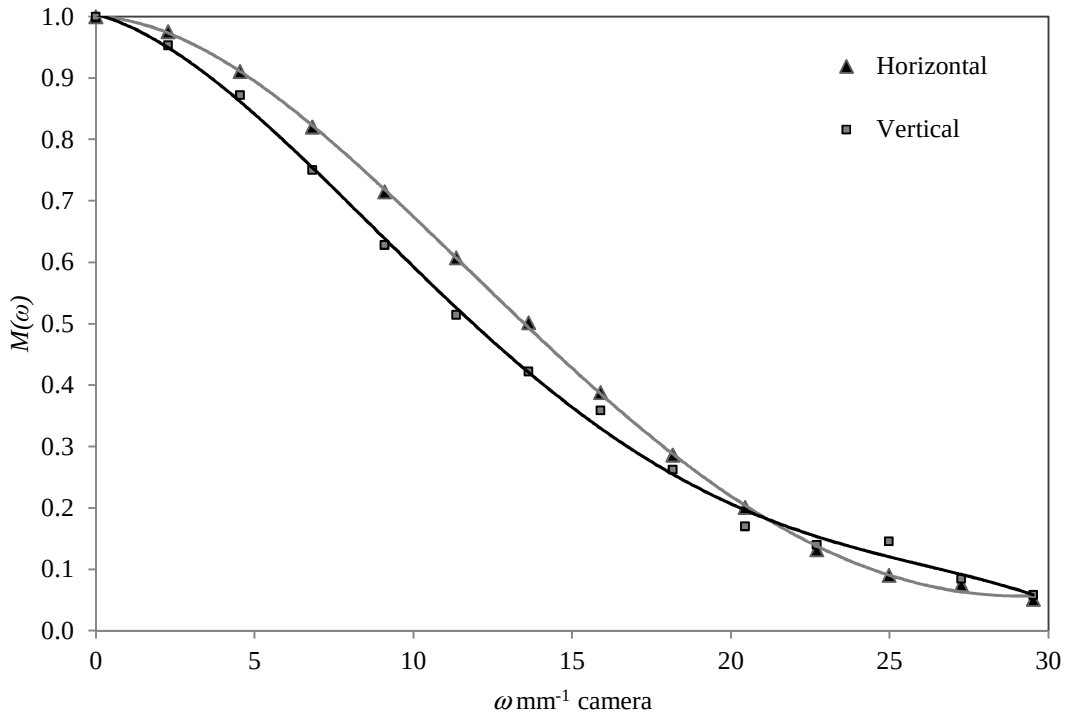


Figure 5-7. Polynomial functions representing the MTF curves of the combined camera-LCD.

Component	Polynomial function	*r*
Combined system's horizontal edge	$M(\omega) = 7E\text{-}05\,\omega^3 - 0.0028\,\omega^2 - 0.0131\,\omega + 1.0109$	0.9997
Combined system's vertical edge	$M(\omega) = 4E\text{-}05\,\omega^3 - 0.0011\,\omega^2 - 0.0362\,\omega + 1.0246$	0.9986

Table 5-2. Third degree polynomial functions representing the combined system MTFs, $M(\omega)$. The spatial frequency, ω, is measured in cycles/mm, and *r* represents the correlation coefficient of fitted functions to measured data.

5.1.3 Display MTF

The LCD MTF was finally calculated by dividing the combined MTF by the camera MTF. For this, it was assumed that each component was linear and that the MTF for each successive component was independent from that of the previous component [33] (pp. 183).

The responses of the system are shown approximately up to its Nyquist limit, which is 1.86 cycles per millimetre, calculated from 0.27 mm square pixel dimensions.

Figure 5-8 illustrates the LCD MTF results in cycles/mm on the display faceplate. The horizontal and vertical MTFs differ but not considerably. The average response points, fitted by the third degree polynomial functions, are also illustrated in Figure 5-9. The LCD MTF results will be useful for the calculation of the objective image quality metric (cf. Section 8.1).

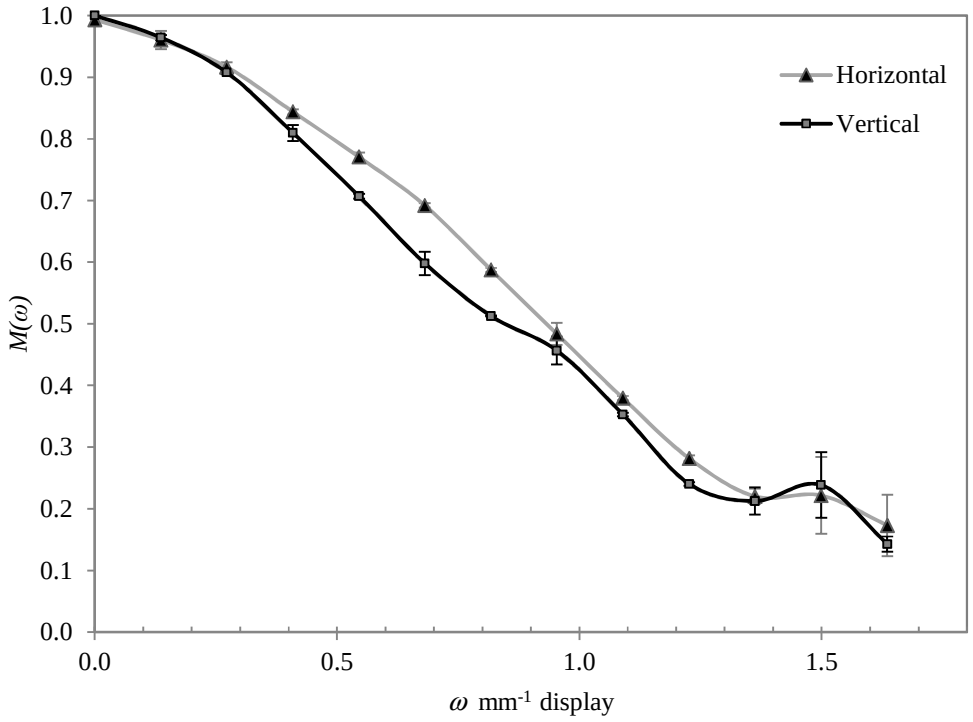


Figure 5-8. Calculated MTF curves of the LCD, including the variation in the measurements.

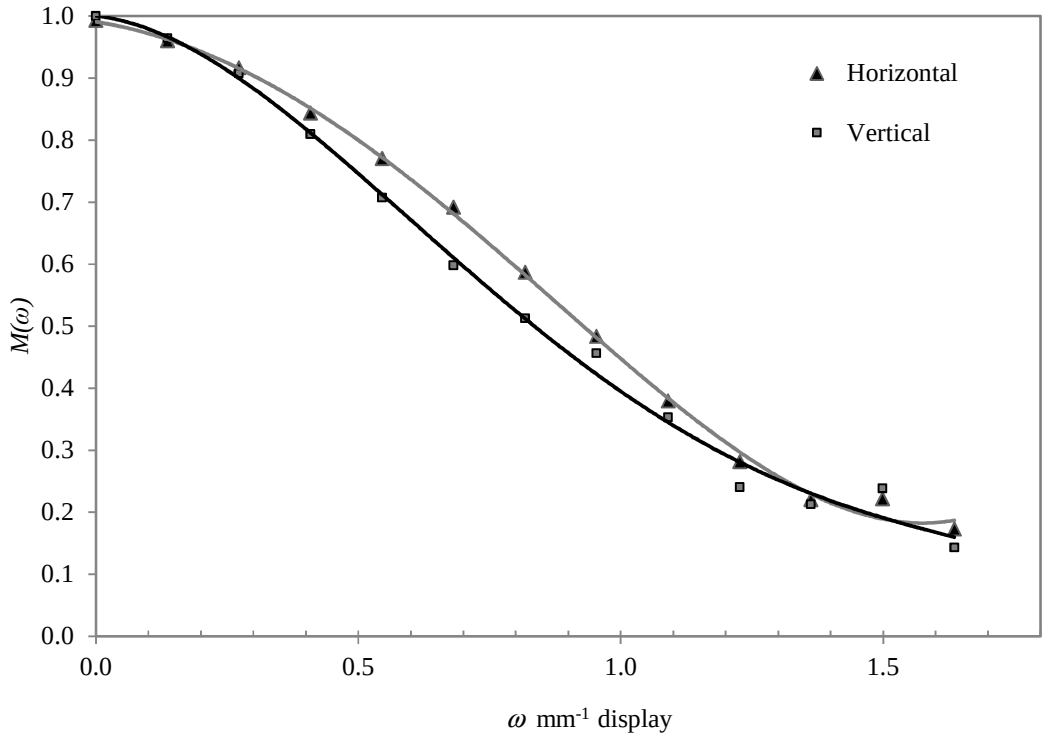


Figure 5-9. Polynomial functions representing the MTF curves of the LCD.

Third degree polynomial functions and exponential functions (often used to fit the display MTF [84] (pp. 25)) for the display are tabulated in Table 5-3. Both functions fit well to the measured data.

Component	Functions	*r*
LCD's horizontal edge in third polynomial function	$M(\omega) = 0.371\omega^3 - 0.903\omega^2 - 0.007\omega + 0.983$	0.9990
LCD's vertical edge in third polynomial function	$M(\omega) = 0.269\omega^3 - 0.557\omega^2 - 0.324\omega + 1.012$	0.9969
LCD's horizontal edge in exponential function	$M(\omega) = 1.264e^{-1.15\omega}$	0.9751
LCD's vertical edge in exponential function	$M(\omega) = 1.231e^{-1.2\omega}$	0.9783

Table 5-3. Third degree polynomial functions and exponential functions representing the LCD MTFs, $M(\omega)$. The spatial frequency, ω, is measured in cycles/mm, and *r* represents the correlation coefficient of fitted functions to measured data.

The LCD MTF 10% values were 1.93 cycles/mm in the horizontal direction and 1.73 cycles/mm in the vertical direction, which is the approximately LCD Nyquist limit, 1.86 cycles/mm. The LCD MTF 50% values were 1.14 cycles/mm in the horizontal direction and 0.95 cycles/mm in the vertical direction.

5.1.4 MTF associated with image down-sampling

The MTF associated with image down-sampling was investigated on three different image sizes, because the down-sampling affects the system MTF in the imaging chain [192] (pp. 8). The captured edges used for the determination of the camera MTF (cf. Section 5.1) further were down-sampled using bicubic interpolation at 1x, 1.5x and 2x decimation along both the x and y axes. After decimation the image sizes became 4064 by 2704 pixels for 1x, 2709 by 1802 pixels for 1.5x and 2032 by 1352 pixels for 2x. The reason for using bicubic interpolation is that it is used for the down-sampling of images employed in the subjective image quality examinations (cf. Section 6.2).

The MTF was derived using the ImCheck software [190]. The fitted third degree polynomial functions for the MTFs ($r \approx 0.99$) are illustrated in Figure 5-10 and Figure 5-11. The responses of the system, i.e. camera and interpolation MTFs, are shown approximately up to its Nyquist limit. This result shows that the image down-sampling affects significantly the image MTF. However, this difference does not necessarily correspond to the perceived sharpness of the different images. This is because the bicubic interpolation is a non-linear process. The MTF theory is strictly valid in linear, isotropic and spatially invariant systems (cf. Section 2.2.2.5).

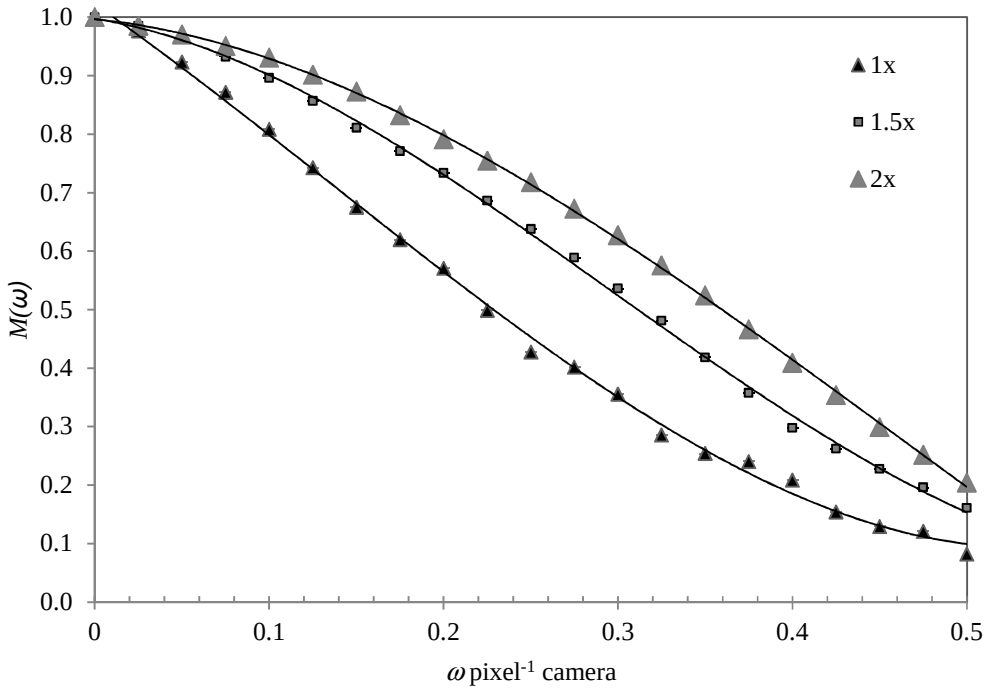


Figure 5-10. Polynomial functions representing the vertical MTFs associated with down-sampling, at three different image sizes.

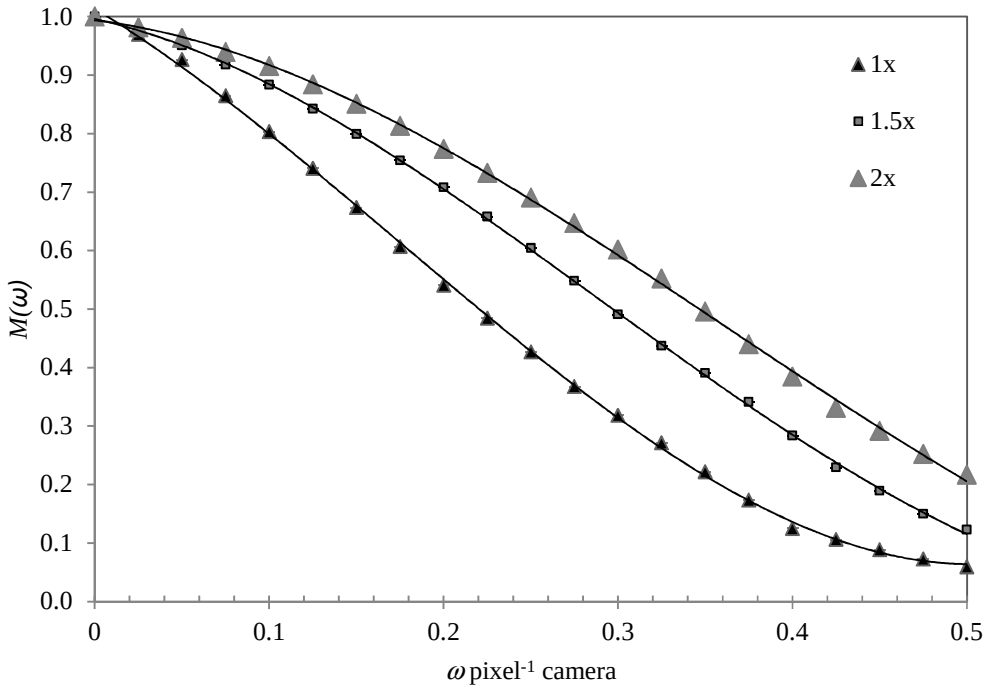


Figure 5-11. Polynomial functions representing the horizontal MTFs associated with down-sampling, at three different image sizes.

5.2 Noise

5.2.1 Camera noise

There are two methods for quantifying camera noise, uniform field measurement and test target measurement, as recommended in ISO 15739: 2003 [126] (pp. 5-8). The test target noise measurement was employed for the Canon EOS-1Ds digital camera noise measurement.

The camera can be set to ISO speed settings ranging from 100 to 1250 at 1/3 stop increments [153] (pp. 49). In addition, the camera can reduce image noise using a noise reduction function (cf. Table 3-1) [153] (pp. 129). In this study, the camera was set to an ISO speed of 100 in conjunction with noise reduction processing in the camera firmware. This setting was also employed for capturing the natural scenes used in the image quality assessment later (cf. Section 6.1).

The camera lens was set to an aperture of *f* 11 and focused correctly, as with the measurement of the camera MTF (cf. Section 5.1.1). The camera was again set to sRGB colour mode, setting 3 (cf. Section 3.4 and 3.5). The camera operated in self-timer mode to minimise distortion that could be caused by camera shake.

The slanted edge test target, QA-62 [188], was again selected to measure the camera noise. The target includes uniform 20 grey steps that can be used in noise evaluation (cf. Figure 5-1). The uniformity of the test chart was confirmed using density measurements taken at different spatial positions [126] (pp. 12).

Two gas-filled tungsten lamps produced the maximum unclipped level of illumination in the camera, checked by examining the image histogram. In addition, the non-uniformity of the illumination was confirmed to be less than ± 2 % of the average illumination.

Images were then captured ten times sequentially. ISO 15739: 2003 [126] (pp. 8) recommended that at least eight frames should be captured sequentially. During the acquisition of the images, the temperature was kept constant at around 21 degree Celsius at a distance of 0.2 m from the camera.

The captured images were downloaded and saved to a computer as 8-bit sRGB, TIFF uncompressed images using the software provided by Canon, via an IEEE 1394 connection.

The standard deviation of each image was then measured using the 100×100 pixels in the central area of each patch. The values were used to calculate total noise.

The total noise (σ_{total}) was calculated by:

$$\sigma_{\text{total}} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\sigma^{2}{}_{j}} \tag{5.2}$$

where σ_{total} is the total noise of the system, σ_j is the standard deviation of each individual exposed image j, and n is the number of exposed images.

The noise levels usually correspond to an image signal, thus it should describe the signal-to-noise ratio for the imaging system [127] (pp. 420-421).

The signal-to-noise ratio was calculated by:

$$\frac{S}{N} = \frac{L_{\text{sat}}\times 0.18\times \text{incremental gain}}{\sigma_{\text{total}}} \tag{5.3}$$

where L_{sat} is the luminance which gives the maximum unclipped output from the camera, e.g. for an eight bit system, this is 255. 0.18 is the 18% reflectance of the target of density of 0.9 with respect to a maximum level of 140%. Incremental gain is the first derivative of the OECF, which is the change of between the output and the input.

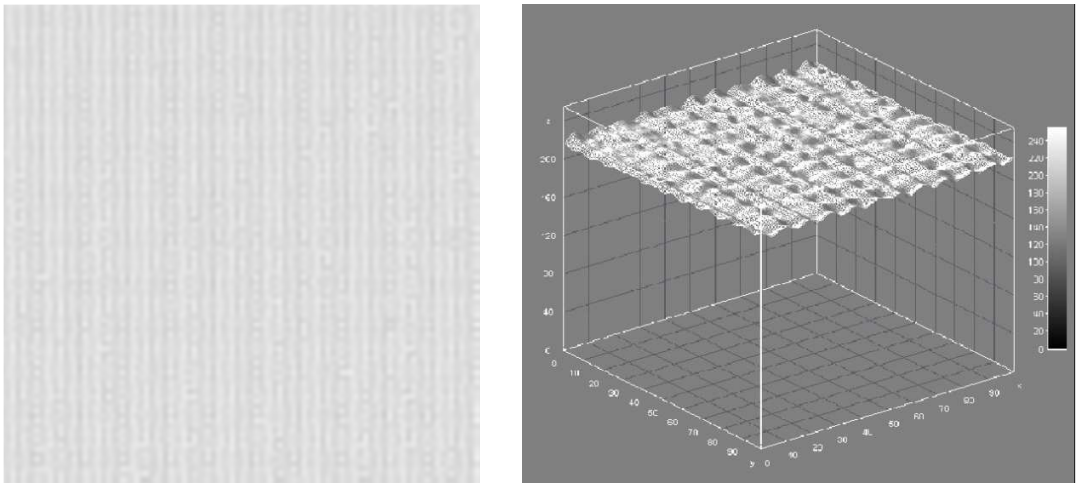
Figure 5-15 illustrates the signal-to-noise ratio of the camera, 47:1. The camera total noise value was 1.19. The camera noise characterisation was satisfactory, when compared with previous researches [193] (pp. 88-89), [194] (pp. 24). This result enabled a high-performance camera to be used for the LCD noise characterisation (cf. Section 5.2.2).

5.2.2 Display noise

The noise of the LCD was measured with the aid of the digital camera (cf. Table 3-1), as recommended by Roehrig *et al* [195, 196]. A suitable camera is essential for this method. Roehrig *et al* [196] (pp. 2) have noted that the available cameras for this method are characterised by low noise.

First, the camera was set to the same settings as those used to quantify the noise of the camera system (cf. Section 5.2.1). A digitally constructed QA-62 target was displayed on the LCD in a relatively small central area (cf. Section 4.4). Images were then captured ten times sequentially. The captured images were then downloaded and saved to a computer as uncompressed sRGB, TIFF files. The 100×100 pixels from the central area of the white patch were selected and saved them as separate uncompressed sRGB, TIFF files.

Second, the LCD structure removal was applied. This is because the captured images introduce the transmission differences of light through the liquid crystal cells. Figure 5-12 presents the transmission differences of light.



2D surface plot 3D surface plot

Figure 5-12. Transmission differences of light through the LCD, captured by the camera.

The display pixel structure leads to several large spikes in the power spectrum of the captured image. Figure 5-13 presents the one-dimensional power spectrum in Fourier space.

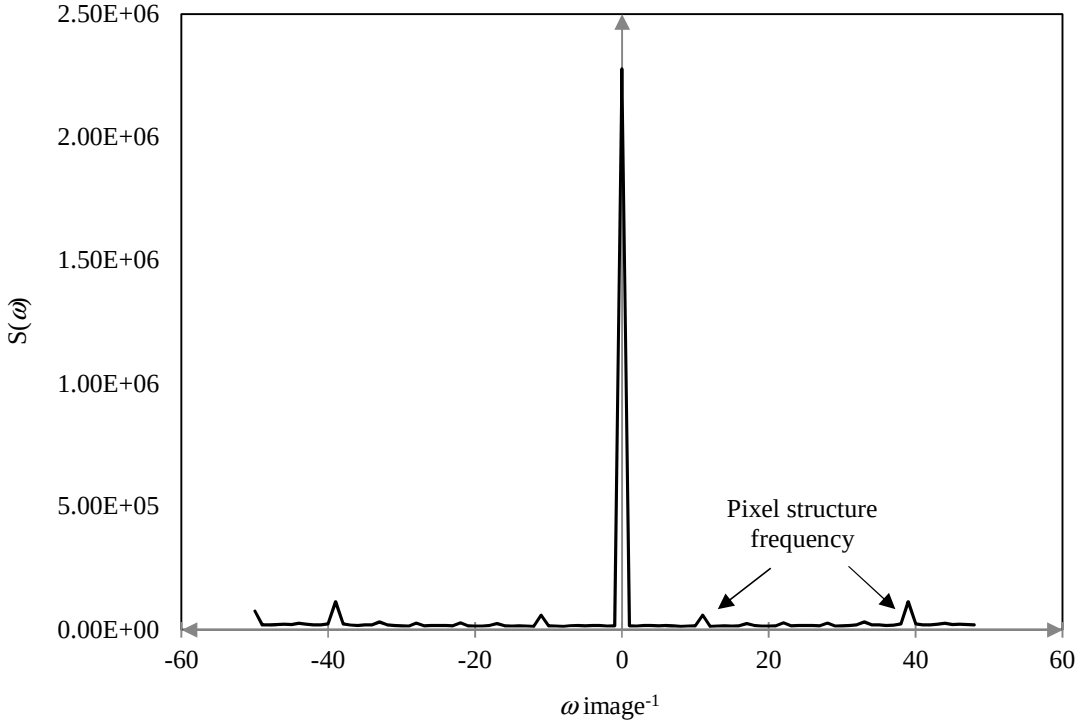
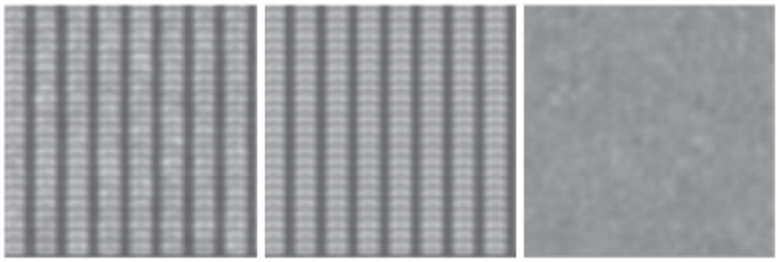


Figure 5-13. One-dimensional power spectrum in Fourier space.

This LCD structure removal was achieved by a *flat fielding* technique (Figure 5-14) [195] (pp. 167). According to Roehrig *et al* [196] (pp. 4), "the flat field for the structure removal is generated by precise super-positioning of CCD camera images of an LCD displaying a uniform field, taken at different spatial positions. Super-imposing many images precisely registered with respect to the pixel structure positioning provides a picture of the pixel structure only while the spatial noise has been averaged out – this is the desired flat field."



Noise with structure Flat field Noise without structure

Figure 5-14. Example of flat fielding technique. Removal of the pixel structure from left image using flat field image results in an image with spatial noise only and no structure (adapted from Roehrig [195] (pp. 167)).

After the removal of the LCD structure, the mean and standard deviation of the captured display image were measured. Ultimately this analysis led to the derivation of the LCD signal-to-noise ratio [196] (pp. 3).

$$SNR_{LCD} = \mu_{LCD} / \sigma_{LCD} \quad (5.4)$$

where μ_{LCD} is the sample mean of the LCD, and σ_{LCD} is the sample standard deviation of the LCD.

Figure 5-15 illustrates the signal-to-noise ratio of the LCD, 45:1. The LCD noise value was 2.13. The LCD noise result will be used in the calculation of the objective IQM (cf. Section 8.1).

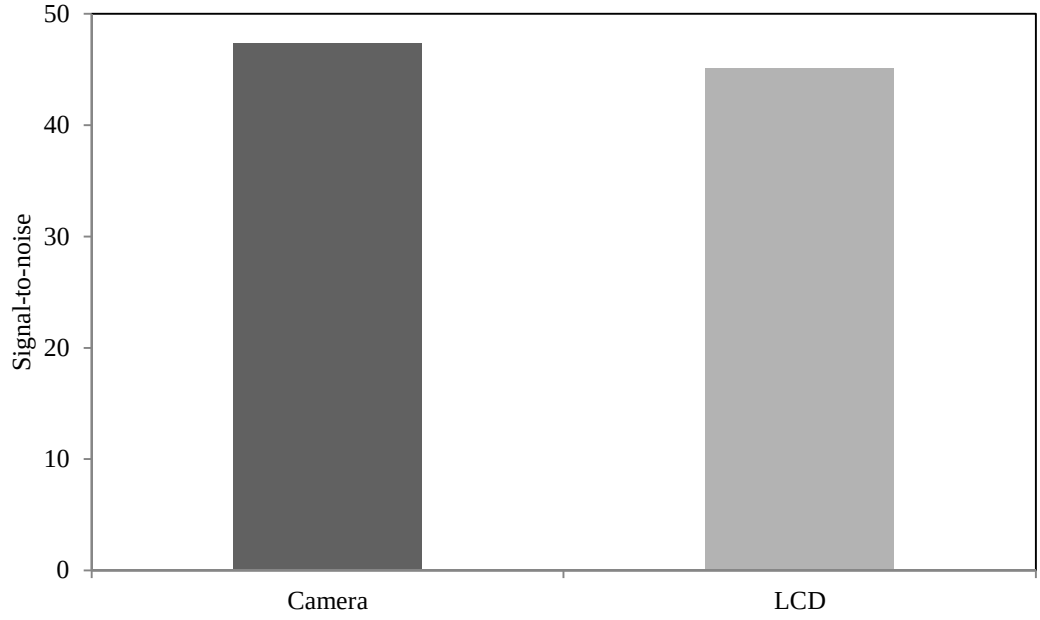


Figure 5-15. Signal-to-noise ratios of the camera and the LCD.

5.2.3 Noise power spectrum (NPS) associated with the addition of Gaussian noise

The NPS associated with adding Gaussian noise was investigated on three different standard deviations (σ). Gaussian noise was used to distort images employed in the image quality assessments later on (cf. Section 6.2 and 8.1).

First, uniformly exposed images of a Kodak R-27 18% grey card were acquired. The camera was set to the same settings used to evaluate the camera noise (cf. Section 5.2.1). Gaussian noise filters were then applied to the originals, using three different standard deviations (σ): 0.0, 0.1 and 0.2 (cf. Section 6.2). This created images with three different levels of noise.

Second, average pixel values of the pixel traces of 256 by 10 pixels that simulated a long thin slit were measured, and then the fluctuations were calculated by subtracting the mean value from each value of the pixel trace. The squared modulus of the Fourier transform of the fluctuations of each trace was then calculated. The measured NPS, N(u,v), was finally obtained by taking the ensemble average (cf. Section 2.2.2.6).

Third, the true NPS, N'(u,v), was calculated from the measured NPS, N(u,v), and the squared transfer function of the scanning system, T(u,v). The latter was produced using a 1 by 10 pixels simulated scanning aperture by a sinc function.

$$N'(u,v)=|T(u,v)|^2 N(u,v) \quad (5.5)$$

$$T(u,v)=\mathrm{sinc}(au)\mathrm{sinc}(lv) \quad (5.6)$$

where a is the scanning with, l is the scanning length and (u,v) is the spatial frequency at two-direction.

According to Jenkin [113] (pp. 440), "the noise field (i.e. a two-dimensional image containing noise alone) is scanned and sampled using a long thin slit to produce a one-dimensional trace for analysis."

The NPS results are illustrated in Figure 5-16. The low spatial frequencies of interest for the NPS are compared (up to ≈ 0.15 pixel^{-1}). They are due to the same image structure at high magnification when capturing the uniform test target. As expected, the area under the NPS increases with increasing standard deviation (σ). This is clearly presented in Figure 5-17, where NPS is plotted using the common logarithm units. The NPS obtained from these measurements were used in the calculation of the objective IQM (cf. Section 8.1).

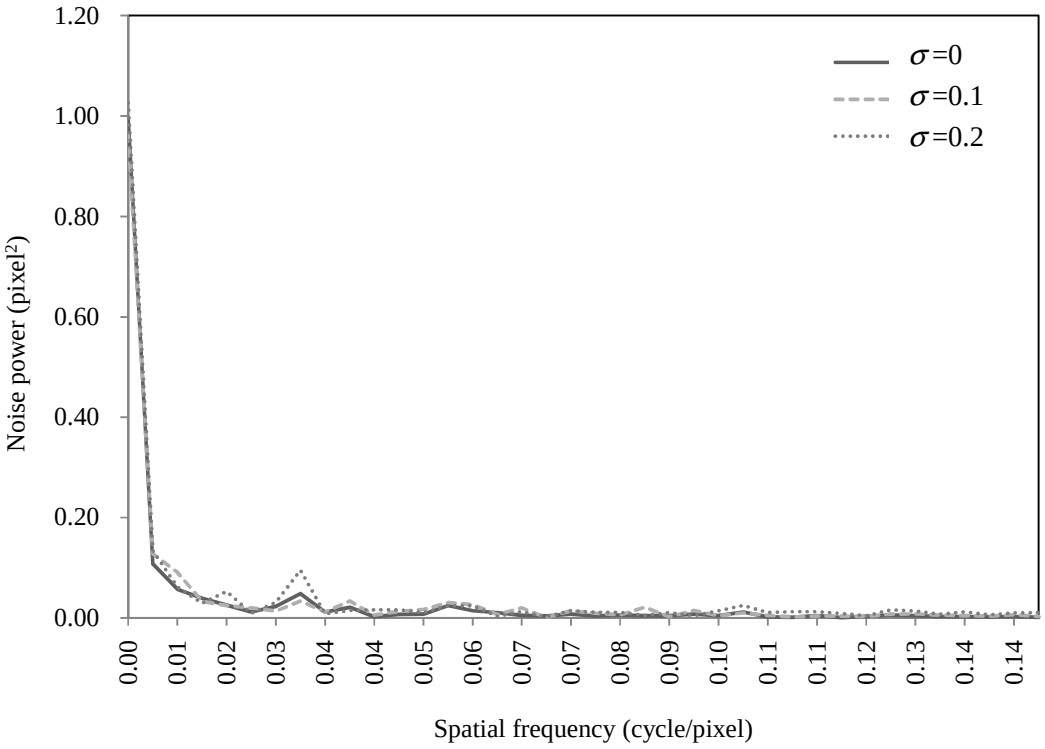


Figure 5-16. Comparison NPS with adding Gaussian noise for three different standard deviations.

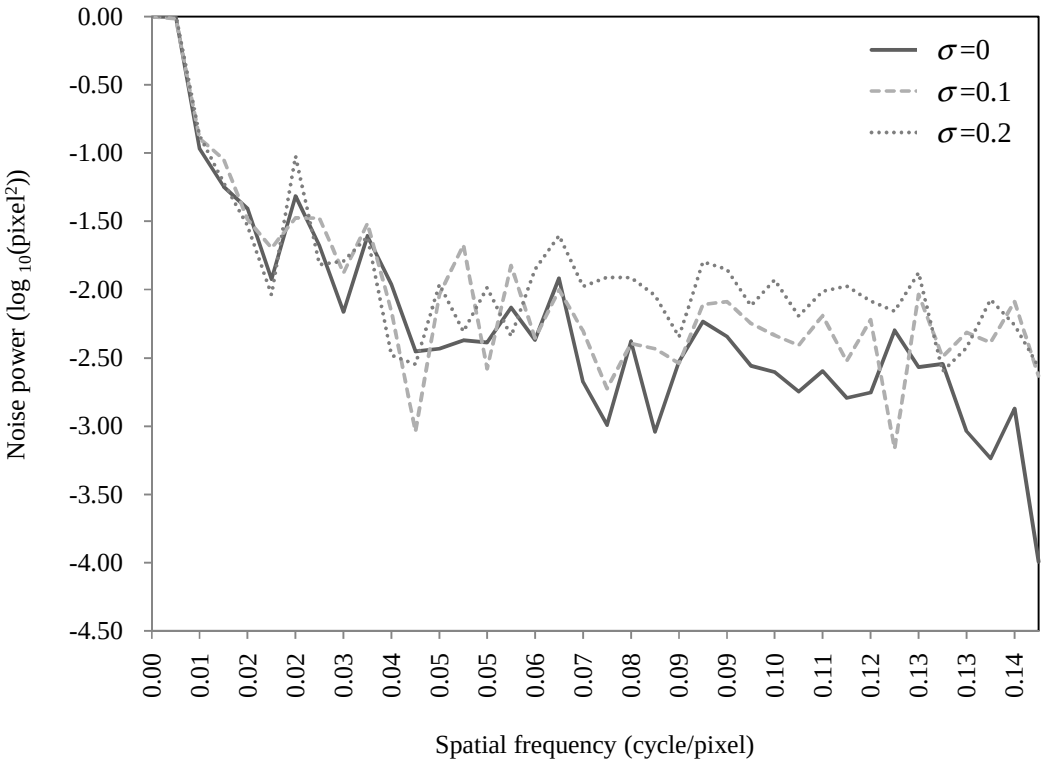


Figure 5-17. Comparison the common logarithm of NPS with adding Gaussian noise for three different standard deviations.

5.3 Summary

In this chapter, MTF and noise measurements were studied in the two imaging systems: a Canon EOS-1Ds camera and an Eizo CG210 LCD.

The camera MTF and noise were successfully quantified, and the results enabled the LCD MTF and noise quantification.

The following summary lists the main findings of the LCD MTF and noise characteristics.

1. The LCD MTF 50% values were 1.14 mm^{-1} in the horizontal direction and 0.95 mm^{-1} in the vertical direction.
2. The LCD MTF 10% values were 1.93 mm^{-1} in the horizontal direction and 1.73 mm^{-1} in the vertical direction, which is the approximately LCD Nyquist limit, 1.86 mm^{-1}.
3. The LCD noise value was 2.13, and the signal-to-noise ratio was 45: 1.

Further work could be carried out to investigate LCD MTF and noise measurements. It could ensure that the assumptions, made in quantifying the LCD MTF and noise, are appropriate.

The image down-sampling affects were significant. More serious investigations of MTF associated with image down-sampling are required for the MTF methods, since the MTF is strictly valid only in linear systems. As expected, the NPS in relation to Gaussian noise showed that as the level of standard deviation increased the area under the NPS increased.

These results enable further research into the objective (physical) scaling, which is the application of IQMs based on the MTF with noise metric to the image chain (cf. Chapter 8).

Chapter 6

Subjective image quality evaluation

This chapter presents a psychophysical scaling experiment and the derivation of subjective image quality scales. The aim of the work presented was to investigate the effect of scene content on the perceived image quality, specifically on the perceived sharpness and noisiness.

6.1 Image acquisition and selection

Images of natural scenes were acquired 1) by image capture, using a digital camera (cf. Table 3-1), and 2) from two Master Kodak Photo CDs [164]. These images covered a range of scene contents and a variety of image characteristics. The scenes represented a variety of subjects, such as portraits, natural scenes and buildings with plain and busy backgrounds. They were chosen to include various global and local illuminations, numerous colours, varying number and strength of lines, edges and spatial distribution of the subjects (cf. Section 2.2.1.1). The selected test scenes are illustrated in in Figure 6-1.

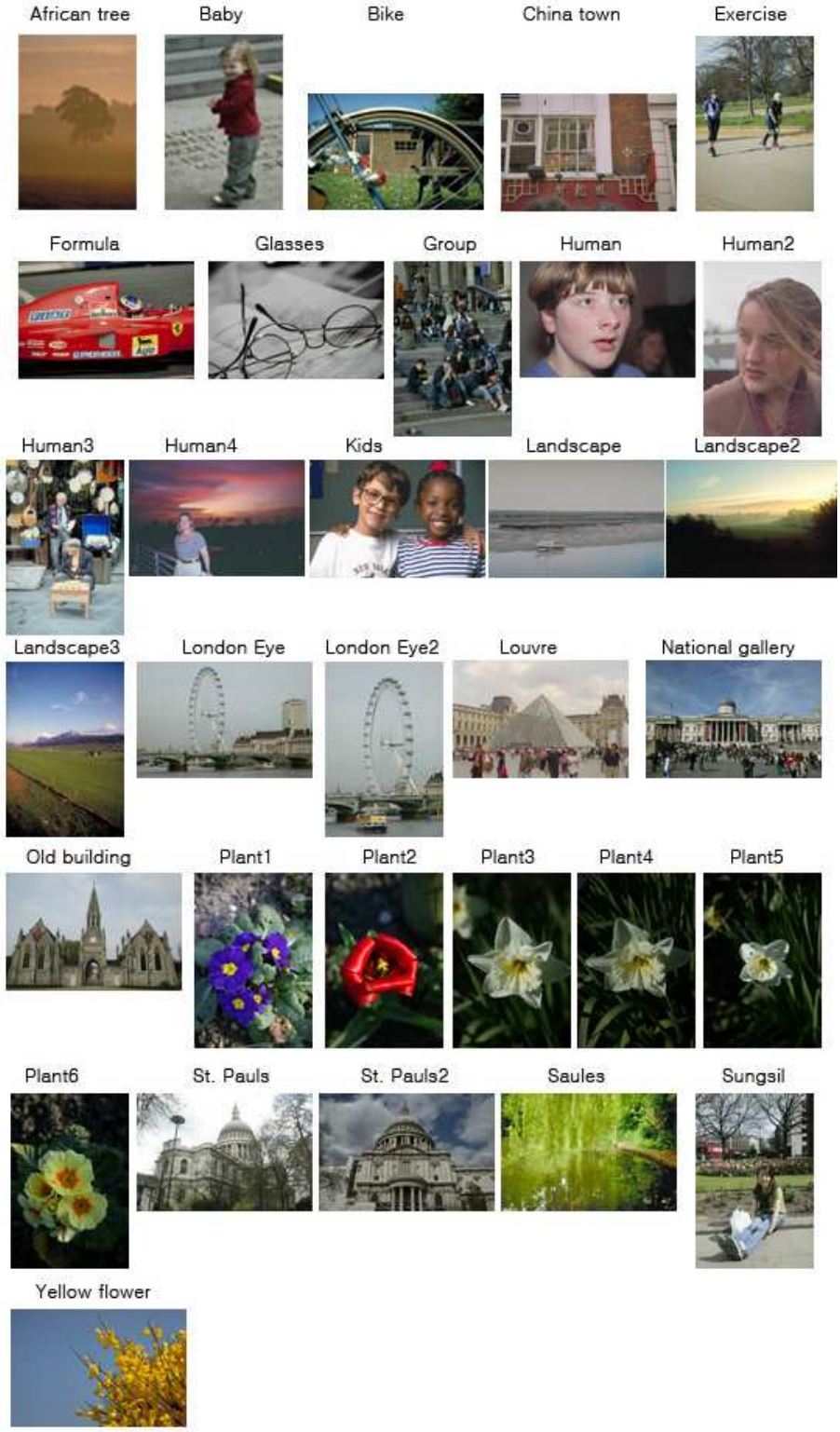


Figure 6-1. Thirty-two scenes used in the subjective quality scaling.

Sixteen natural scenes were captured using the Canon EOS-1Ds full frame digital SLR camera, equipped with a Canon EF 28-135mm *f* 3.5-5.6 IS USM zoom lens (cf. Table 3-1). The ISO speed 100 setting (cf. Section 5.1) with varying apertures of *f* 5.6-36 was used to capture all scenes (cf. Table 8-1). Camera exposure was determined by taking multiple reflection readings from various parts of the scene. This was achieved with the through-the-lens centre-weighting and spot metering modes of the camera[1]. The camera was set to auto colour balance mode and sRGB colour mode (cf. Section 3.4 and 3.5). The lens was focused manually. Scenes were recorded at about 11 megapixels (4064 by 2704 pixels). They were saved as 12-bit RAW files. They were then downloaded to a computer as 8-bit TIFF uncompressed images using the software provided by Canon, via an IEEE 1394 connection.

In addition, sixteen natural scenes were selected from two Master Kodak Photo CDs. The Master Kodak Photo CD images were opened at a resolution of 512 by 786 and at a colour resolution of 8 bits per channel in RGB colour space. After inspection to ensure that the colour reproduction was satisfactory, they were saved to sRGB colour space. The reason for using Master Kodak Photo CD images was to cover a various range of scene contents and characteristics that photographers, artists and consumers may wish record and reproduce faithfully. In addition, some of images were used a previous image quality study [10]. However, these images had limited information on the imaging system's characteristics, i.e. sharpness, noise and colour reproduction characteristics were not quantified in the laboratory.

All thirty-two images were then down-sampled to 317 by 476 pixels using bicubic interpolation and saved as uncompressed sRGB, TIFF files of approximately 440 Kb. The bicubic interpolation method is reported to be the best technique for maintaining image quality [198] (pp. 461), [199] (pp. 250).

[1] The centre-weighting measurement is adequate for average situations, and the spot metering is for precise measurement of the luminance of small areas of a large subject [197] (pp. 324).

6.2 Test stimuli

The thirty-two original images were manipulated by varying two physical parameters in MATLAB software [136] (Gaussian blurring and Gaussian noise) to obtain a large number of test stimuli with different levels of blur and noise.

Prior to deciding the ranges and levels of distortion, pilot studies were conducted on the calibrated 21 inch EIZO CG210 LCD. The *method of adjustment* [200] (pp. 129-130) was employed for this purpose, because it is the most straightforward method for determining observer thresholds for a given stimulus. In this technique, the experimenter had control over the magnitude of the stimulus itself across several trials [201] (pp. 59-60). The average adjustment across several trials was calculated as a 25:75 proportion of responses [45] (pp. 2), [200] (pp. 129). Each chosen distortion level corresponded to approximately one empirical JND when the images were viewed on the display used for further experimentation.

The following techniques were chosen for the manipulation of the original stimuli:

- Blurring: Firstly, Gaussian blurring was applied to the 32 originals. The standard deviation (σ) of the Gaussian low-pass kernel ranged from 0.01 to 1.24 at 0.3075 intervals. This created a total of 160 test images, i.e. 32 originals × 5 different levels of Gaussian blur.
- Noise: After blurring, the images were further distorted by adding Gaussian noise, using three different standard deviations (σ): 0.0, 0.1 and 0.2. This function created three different levels of uniform noise per blurring level. A total of 480 test stimuli were finally created, i.e. 32 originals × 5 different levels of Gaussian blur × 3 different levels of Gaussian noise.

The manipulation order is based on the fact that an application in the blurring after noise addition would suppress the visual response to noise. The blurring is a smoothing operation as a noise reduction.

6.3 Display, interface and viewing conditions

Psychophysical tests were carried out under dark viewing conditions to avoid display flare in the laboratory. ISO 3664: 2000 [51] (pp. 13) suggests that the level of ambient illumination should be less than or equal to 32 lx, when the monitor is switched off.

All the images were displayed on an EIZO CG210 LCD, controlled by a S3 Graphics Prosavage DDR graphic card in a personal computer running Windows XP professional. The graphic card was configured to display 24-bit colour, at a resolution of 1600 by 1200 pixels and a frequency of 60 HZ (cf. Table 4-1).

The display was switched on for fifteen minutes before the tests to allow stabilisation (cf. Section 4.3). The display interface that was generated in MATLAB, presented the images in the centre of the display area to minimise non-uniformity display effects (cf. Section 4.4). The test image was placed at a viewing distance of approximately 60 cm from the observers and subtended a visual angle of roughly 10° (cf. Section 4.5). The Eye-One Pro monitor calibrator was used to calibrate the display at a white point close to D65 (6500 K), a gamma of 2.2 and a white point luminance of 100 cd/m^2 – the default settings for contrast and luminance for this monitor (cf. Section 4.6 and 4.7).

The images were displayed at 100% pixel resolution (317 by 476 pixels). Each image occupied approximately 1/3 of the otherwise neutral grey display area. Figure 6-2 illustrates the graphic user interface (GUI) for the test. Prior to the investigations, the GUI manual was given to the observers to familiarise themselves with the software.

Figure 6-2. Graphic user interface (GUI) used in the categorical scaling.

Psychophysical tests were carried out in a quiet and comfortable laboratory space. The temperature of the viewing space was kept constant at around 20 degree Celsius. A wireless mouse was provided for easy selection of one of 5 quality categories, where 1 represented the worst quality and 5 the best quality. *Simple number categories* [203] (pp. 123) were used here, as they make the meaning of the scaling easy to understand with minimum observer effort, e.g. no-translation or definition of terms was given. As suggested in reference [11] (pp. 37-38), adequate personal space and a comfortable chair were also provided. Figure 6-3 shows a side view of the laboratory set-up and a front view from the observer's eye position.

Figure 6-3. Viewing room set-up at a side view (left) and a front view (right).

6.4 Observations

Subjective assessments were performed by a panel of fourteen selected observers, seven males and seven females. They were all familiar with the meaning and assessment of image quality. The age of observers ranged from 21 to 52 years old. All observers were previously tested and reported as those having normal colour vision and holding normal or corrected-to-normal visual acutance.

Each observer participated in the categorical scaling experiment 6 times, each time evaluating a different set of images. Each observation period was around 45 minutes. ISO 20462-1: 2005 [45] (pp. 7) suggests that the maximum observation period should be no more than 60 minutes to avoid tiredness or lack of concentration.

Before the test, observers were allowed several minutes to adapt to the dark viewing conditions of the laboratory, since the process of dark adaptation has a considerable impact on the perceived images [109] (pp. 94), [202] (pp. 367).

Images were displayed one at a time, in a random order. Observers were asked to place each test image according to its perceived image quality in one of 5 quality categories. They were provided with the following written instructions:

I would like to thank you for participating in this study. Please make yourself comfortable on the chair in front of the display. Please use the following the viewing conditions consistently, i.e. a viewing distance of approximately 60 cm from the faceplate, and subtended a visual angle of roughly 10°.

In this experiment, you will be evaluating the overall quality of a series of images using a psychophysical technique called the categorical method. Please remember that there are no right, or wrong answers because we are asking you about your perception of the quality of the images.

Here is how to evaluate the test images:

A single image will be presented at a time on the monitor. For each image, we ask you to score its perceived image quality. You can express your opinion from 1 to 5 (a score of 1 indicates the lowest image quality, and of 5 the highest image quality).

Once you express your view on a given test image, you are able to evaluate a new test image. When you are ready to evaluate a new image, you can click the "next" button, so that a new image is randomly accessed. When all images have been evaluated, a dialogue box will appear to inform you. Please let the experiment organiser know when you have finished going through the images.

The total observation period will be around 45 minutes.

Note. If you feel you made an incorrect response and would like to re-evaluate the image you are currently evaluating, press the "Reset" key in the pop-up menu. This will restart the evaluation sequence for the current image you are evaluating and record your new answer.

If you feel you want to re-evaluate a previous image that is not currently on the screen, press the "Re-run" key in the pop-up menu. This will allows you to re-commence the quality scaling test. In this case, previous recorded rating is ignored and the test starts from scratch.

6.5 Analysis of results

The indirect interval category scaling method [203] (pp. 123) was employed to give meaningful interval scales.

6.5.1 Scaling overall image quality

Interval scales were derived using the simplest condition D of Torgerson's Law of Categorical Judgement, which makes minimum assumptions regarding the category and sample variance: correlation coefficients and dispersions of both the sample and the category were constant. Category boundaries and sample values were obtained. This assumption means that all the elements under the square root sign in equation (6.1) are constant and equal to one.

$$t_g - S_j = z_{jg}\sqrt{\sigma^2{}_j + \sigma^2{}_g - 2\rho_{jg}\sigma_j\sigma_g} \tag{6.1}$$

$$g = 1,2,\cdots m+1; j = 1,2,\cdots n$$

Where t_g is the boundary value between categories, s_j is the scale value for each sample, z_{jg} is the unit normal deviate corresponding to the proportion, σ_g is the standard deviation of the category boundary, σ_j is the standard deviation of the sample scale value, ρ_{jg} is the correlation between sample scale value and category boundary, g is the category, j is the sample, m+1 is the number of categories and n is the number of samples.

In the condition D: $\sqrt{\sigma^2{}_\mathrm{j} + \sigma^2{}_\mathrm{g} - 2\rho_{\mathrm{jg}}\sigma_\mathrm{j}\sigma_\mathrm{g}} = 1$, the equation (6.1) changes to $\mathrm{t_g} - \mathrm{S_j} = \mathrm{z_{jg}}$ (6.2)

The *least square technique* was applied to prevent inaccurate scale values derived from zero and one elements in the proportion matrix [204, 205]. The key to the technique is to recognise that there is an incomplete set of equations that relate the scale value difference to the transformed values, or z-values, determined from the proportions, i.e. a linear relationship existed between the two unknowns (the boundary value between categories and the scale value for each sample) and the one known (the unit normal deviate

corresponding to the proportion). In addition, the last row in the matrix forces the sum of the scale values to be equal to zero. More details of the least square technique can be found in reference [203] (pp. 133-134).

The goodness-of-fit was examined using the chi-square (χ^2) test [203] (pp. 136-138). The chi-square value indicates that the condition D with the least square technique fits the subjective data well.

Figure 6-4 presents the interval scales for overall image quality. The solid lines represent subjective results from the combined (average of) thirty-two scenes; the grey square is the average from all scenes for each level of distortion. Each label on the x-axis represents a specific level of distortion - in blurring (B) and noise (N).

The results show that the modifications in the two attributes overall decreased image quality. This is not surprising, since the modifications in blur and noise generally degrade image quality [38] (pp. 26-28), [55] (pp. 180-181 and pp. 189-194). The original version of the images (B1N1) had an average scale value of 2.03, whilst the distorted images had lower scale values. This will be discussed in more detail in the individual attribute scales presented later (cf. Section 6.5.2).

The results also indicate considerable *scene dependency*. The broken lines in Figure 6-4 indicate the range of scale values derived from all scenes for each level of distortion.

The results also indicate some *observer variability*. The error bars in Figure 6-4 indicate the inter-observer agreement for the overall quality, which is calculated using the *coefficient of variation* (*CV*) across all observers [206] (pp. 42), [207] (pp. 2-3). The result of the 12 for the *CV* calculation indicates a good inter-observer agreement for the overall quality. The error bars also indicate a high variation around the end categories. A possible reason for this phenomenon was described by Cookingham [43] (pp. 90), i.e. the range effect at the categorical method (cf. Section 2.2.1.1).

$$CV = 100 \times \text{Standard deviation } (\sigma) / \text{ Mean } (\mu) \qquad (6.3)$$

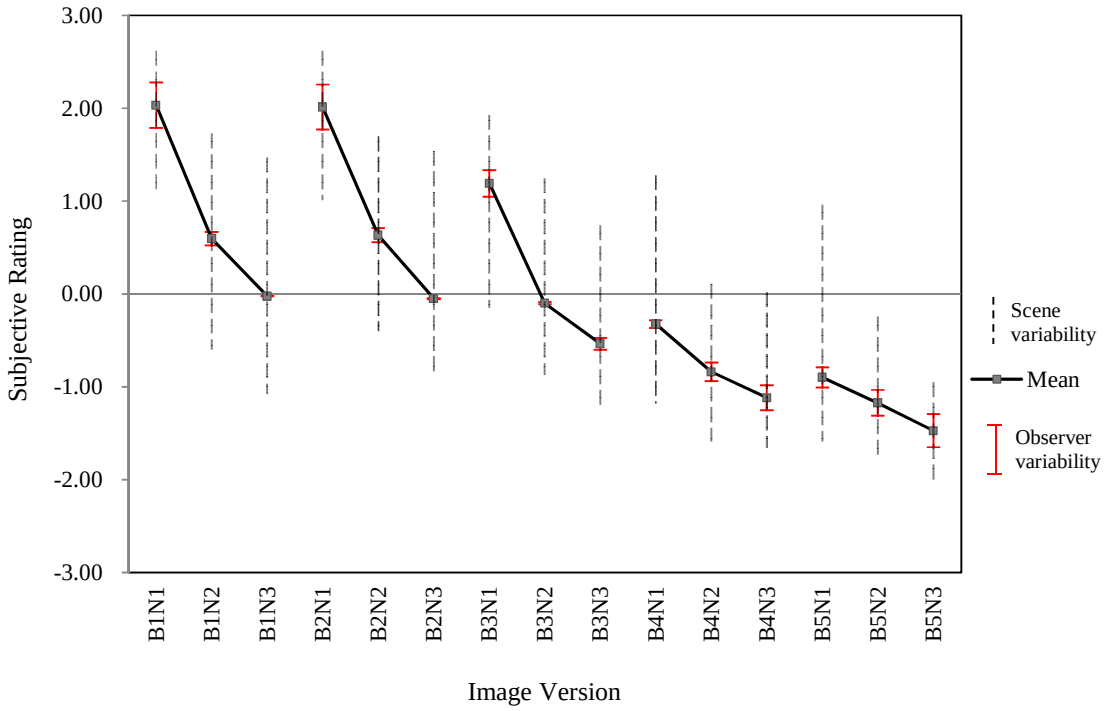


Figure 6-4. Interval scales of subjective image quality (solid line) with scene (broken line) and observer (error bar) variability, described as Case I (blur-noise).

In addition, individual *observer sensitivity* [208] (pp. 204) is presented in Figure 6-5. Its value ranged between 0.764 and 1.113. Observers 4, 6 and 9 were found to have relatively lower sensitivity, while the others have average sensitivity (similar to 1.00).

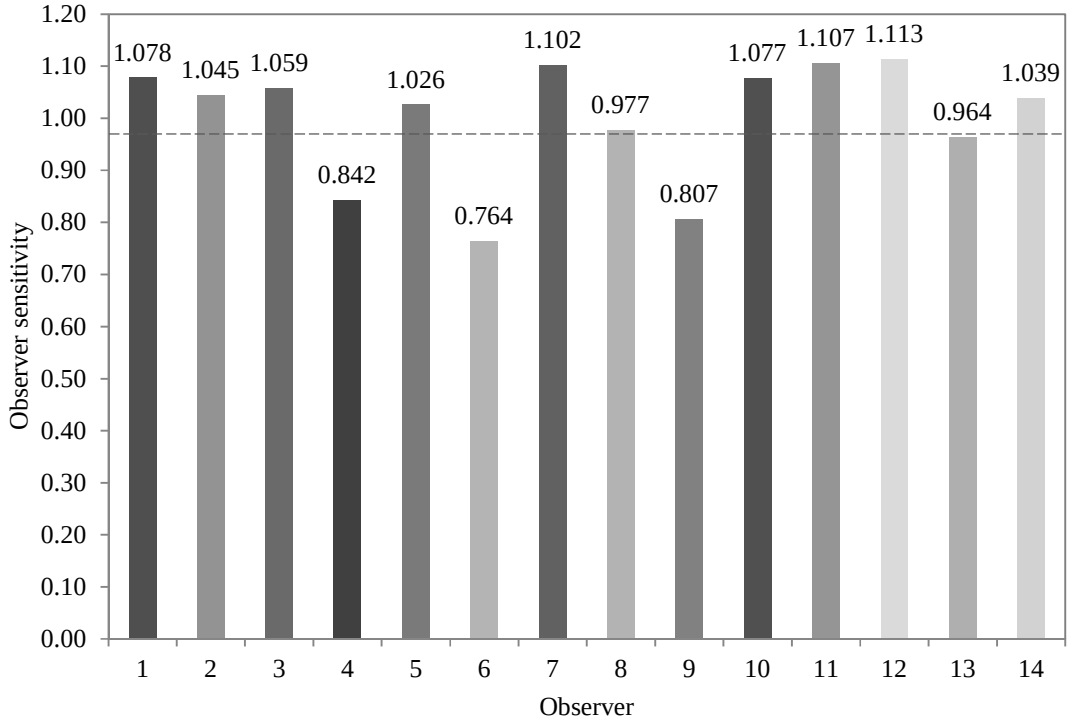


Figure 6-5. Individual observer sensitivity.

6.5.2 Scaling of individual attributes

The collected overall scale values were rearranged using *permutation* [209] (pp. 973) with an aim to derive individual attribute scales. Since overall scales represent perceived image quality that results from variations in two different image attributes (i.e. in blurring level and noise level), the attribute variations were first arranged as listed in Step1 of Table 6-1.

Permutation means the arrangement of item. The number of permutations is taken to be ${}^{n}P_{r}$, where n is a different item taken r at a position, e.g. the number of permutations of two stimuli taken two at a position is two, ${}^{2}P_{2} = 2! = 2$.

The total number of permutations is produced by the product of the total number of distorted images and the number of attribute permutations:

960 total permutations = 480 test images x 2 attribute permutations (6.4)

The full implementation of the method is illustrated in Table 6-1. In addition, examples of the implementation, i.e. representative subjective results from the combined (average of) thirty-two scenes, are shown in Table 6-2 and Table 6-3.

The scale values of the individual image attributes were examined across the average (calculated by the mean) scale values of the other attributes in process of Step 1 to Step 2. The first listed attribute in Step 1, in Table 6-1, is the *targeting attribute* in the permutation. Case I and Case II use the same data, but they are presented differently according to the targeting attribute.

Step 1: Arrangement		**Step 2: Average of last column in Step 1**	
Case	**Arrangement**	**Case**	**Arrangement**
I.	5 blur 3 noise	I.	5 blur
II.	3 noise 5 blur	II.	3 noise

Table 6-1. Individual attribute scaling. Case I and Case II use same data, but they are presented differently according to the *targeting attribute*.

	Step 1		**Step 2**	
	Attributes	**Scale value**	**Attributes**	**Scale value**
I	Blur1Noise1	2.03	Blur 1	0.87
	Blur1Noise2	0.60		
	Blur1Noise3	-0.02		
	Blur2Noise1	2.01	Blur 2	0.86
	Blur2Noise2	0.63		
	Blur2Noise3	-0.05		
	Blur3Noise1	1.19	Blur 3	0.18
	Blur3Noise2	-0.10		
	Blur3Noise3	-0.54		
	Blur4Noise1	-0.32	Blur 4	-0.76
	Blur4Noise2	-0.84		
	Blur4Noise3	-1.12		
	Blur5Noise1	-0.90	Blur 5	-1.18
	Blur5Noise2	-1.17		
	Blur5Noise3	-1.47		

Table 6-2. Example of individual attribute scaling for the combined (average of) thirty-two scenes, Case I.

	Step 1		Step 2	
	Attributes	**Scale value**	**Attributes**	**Scale value**
II	Noise1Blur1	2.03	Noise 1	0.80
	Noise1Blur2	2.01		
	Noise1Blur3	1.19		
	Noise1Blur4	-0.32		
	Noise1Blur5	-0.90		
	Noise2Blur1	0.60	Noise 2	-0.18
	Noise2Blur2	0.63		
	Noise2Blur3	-0.10		
	Noise2Blur4	-0.84		
	Noise2Blur5	-1.17		
	Noise3 Blur1	-0.02	Noise 3	-0.64
	Noise3Blur2	-0.05		
	Noise3Blur3	-0.54		
	Noise3Blur4	-1.12		
	Noise3Blur5	-1.47		

Table 6-3. Example of individual attribute scaling for the combined (average of) thirty-two scenes, Case II.

Overall quality scales, Case I and II from Step 1 in Table 6-1, are shown in Figure 6-4 (blur targeting attribute) and Figure 6-6 (noise targeting attribute). Each label on the x-axis represents a specific level of distortion - in noise (N) and blurring (B) i.e. according to the targeting attribute (listed first in the title of the graph), then the second attribute.

There is clearly a trade-off in image quality when varying the two attributes: high amounts of blur in the image significantly decreased the perception of noise and high noise in the image decreased the perception of blur [210] (pp. 290), i.e. the range of noise 1-3 in blur 1 varied 2.05, while the range of noise 1-3 in blur 5 varied 1.37 (Figure 6-4 and Step 1 in Table 6-2). The range of blur 1-5 in noise 1 varied 2.97, while the range of blur 1-5 in noise 3 varied 1.45 (Figure 6-6 and Step 1 in Table 6-3).

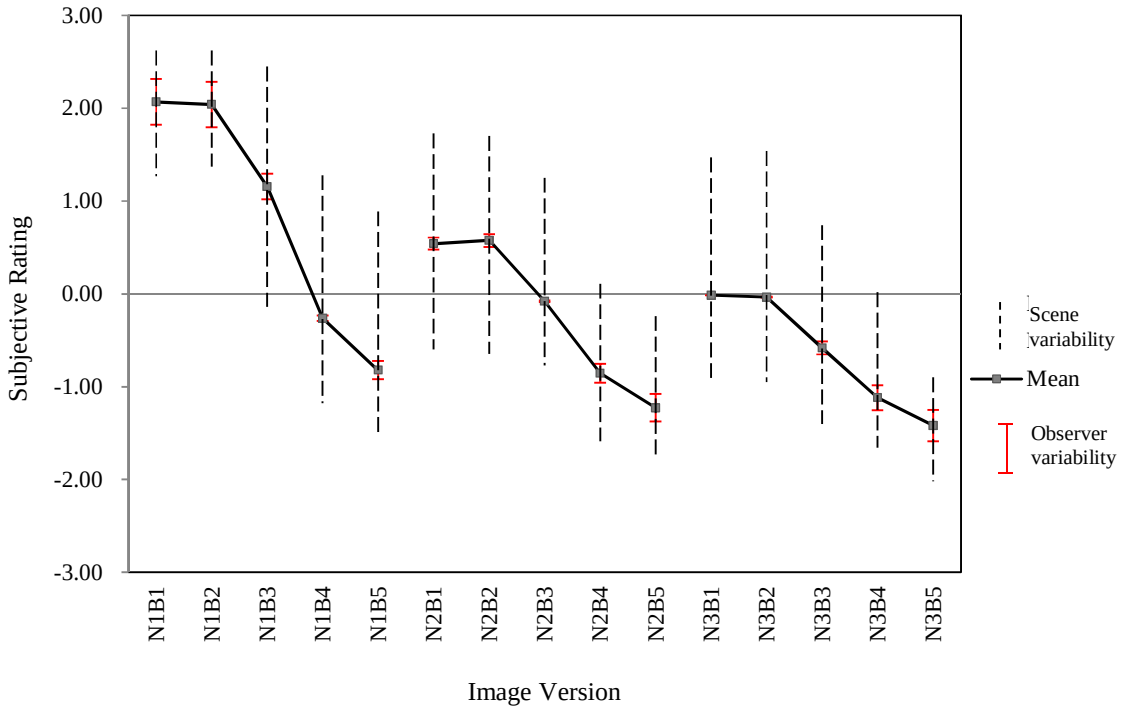


Figure 6-6. Interval scales of subjective image quality (solid line) with scene (broken line) and observer (error bar) variability, described as Case II (noise-blur).

Individual attribute scales, Case I and II from Step 2 in Table 6-1, are presented in Figure 6-7 and Figure 6-8. The solid lines represent subjective attribute scales from the combined (average of) thirty-two scenes; the grey square is the average from all scenes for each level of distortion. The broken lines indicate the range of attribute scales derived from all scenes for each level of distortion.

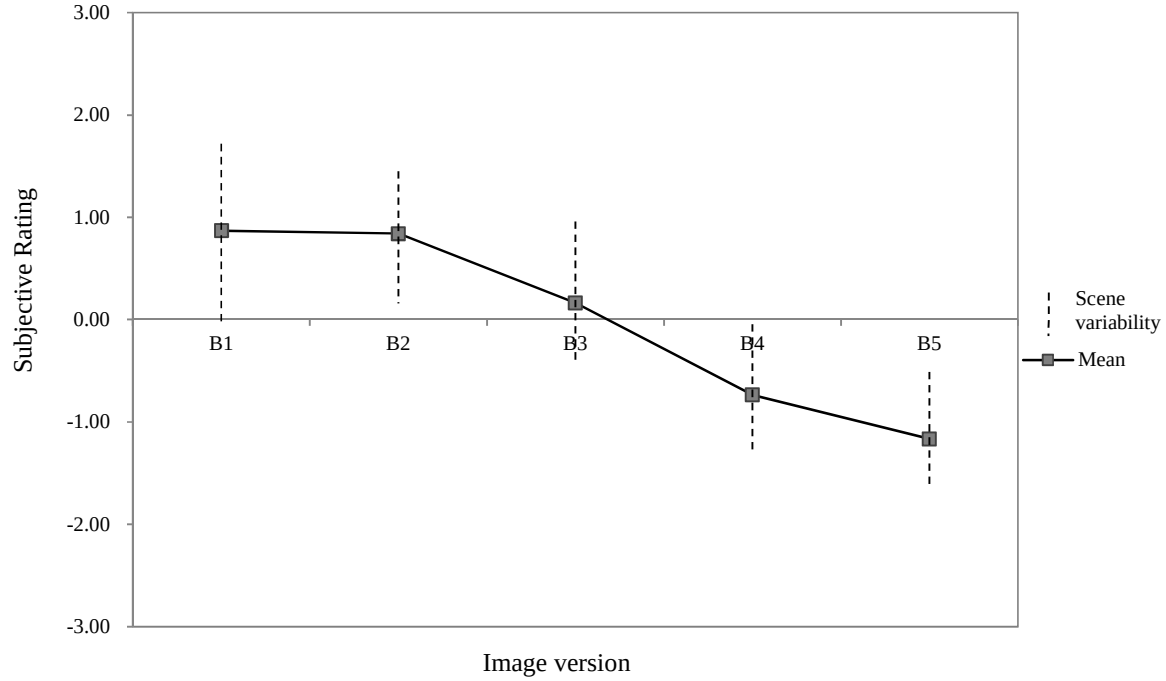


Figure 6-7. Scaled blur/sharpness attribute as a function of the sigma, analysed from interval scales of subjective image quality, Case I (blur).

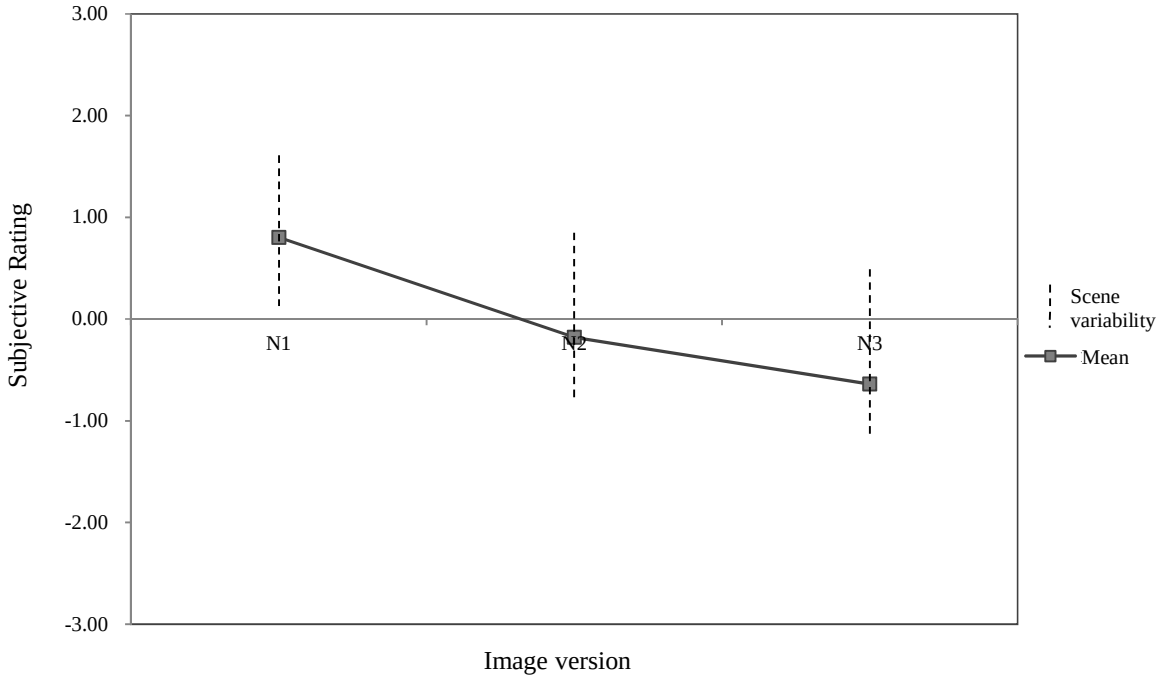


Figure 6-8. Scaled noise attribute, analysed from interval scales of subjective image quality, Case II (noise).

Figure 6-7 represents the blur targeting attribute from the Case I. In this figure, it is noticeable that the original and original + 1 empirical JND of blur versions were almost rated similarly. The quality is shown to decrease significantly for the next blur levels, and it reaches a point (at level B5) where more blur would not further reduce the quality, i.e. the image quality scale of blurring is an almost hyperbolic (S-shape) function (cf. Figure 2-2).

There was a reciprocal relationship between scene dependency and level of distortion, identified when considering image blurring, i.e. as the level of distortion increased the scene dependency decreased. This result indicates that low levels of blurring might affect different scenes in a different manner, but high levels of blurring tend to affect different scenes more equally [211] (pp. 158).

Figure 6-8 represents the noise targeting attribute from the Case II. Case (Noise) indicates that 2 levels of added noise (each separated by 1 JND) decreased equally image quality, but the 'toe' of the quality scale for noise was not reached (cf. Figure 2-3 and Figure 2-4). Also, the scene dependency variation appears similar in all levels of distortion. The results suggest that a larger range of distortions for added noise should be required to cover the full range of image quality.

The outcomes of these results (sharpness and noisiness) were in line with results from previous studies (cf. Section 2.2.1.3).

6.5.3 Measuring the effect of scene content

The effect of scene content on the perceived quality was examined by correlating the combined mean scale values from all scenes and the scale values of each individual scene [38] (pp. 28). This task was carried out for each individual attribute. In the example plot shown in Figure 6-9 for the scene 'Saules', the gradients represent the steepness of the lines fitting the data (one for each attribute) between the ratings for each individual scene and that of the combined scenes.

The gradient will be referred to as the *scene susceptibility* (cf. Section 2.2.1.2). If the gradient of the line is one, the scene susceptibility of the individual scene to the specific quality attribute is the same as that of the combined scenes. Thus, it represents an

'average' scene. If the gradient is larger than one, then the scene has larger susceptibility than that of the combined scenes for the specific attribute. The reverse is true when the gradient is smaller than one.

The constant in the linear relationship indicates whether the individual scene got a better rating than the combined scenes, i.e. the overall ratings for each individual scene is better than the overall ratings of the combined scenes (positive offset) or the opposite (negative offset).

The *coefficient of determination*, r^2, indicates the strength of the relationship between them. The coefficients of determination on blurring and on added noise were all close to 1.0, confirming that this method for deriving scene susceptibility is satisfactory.

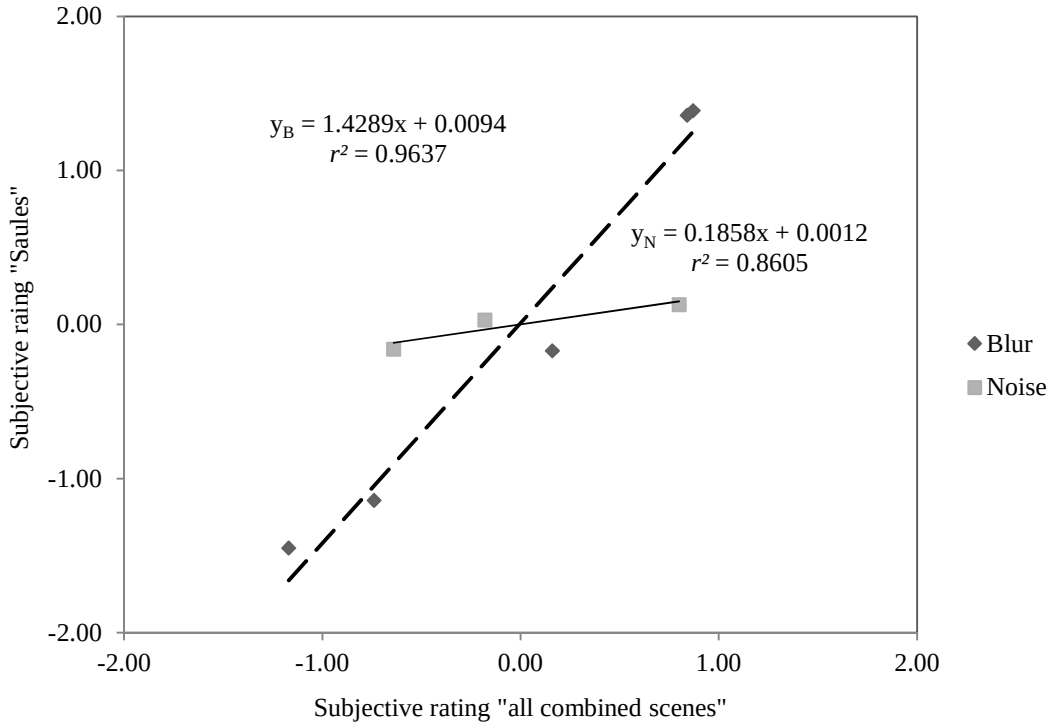


Figure 6-9. Scene dependency on the "Saules."

Figure 6-10 and Figure 6-11 present examples of scenes having different gradients: "African tree" and "Baby."

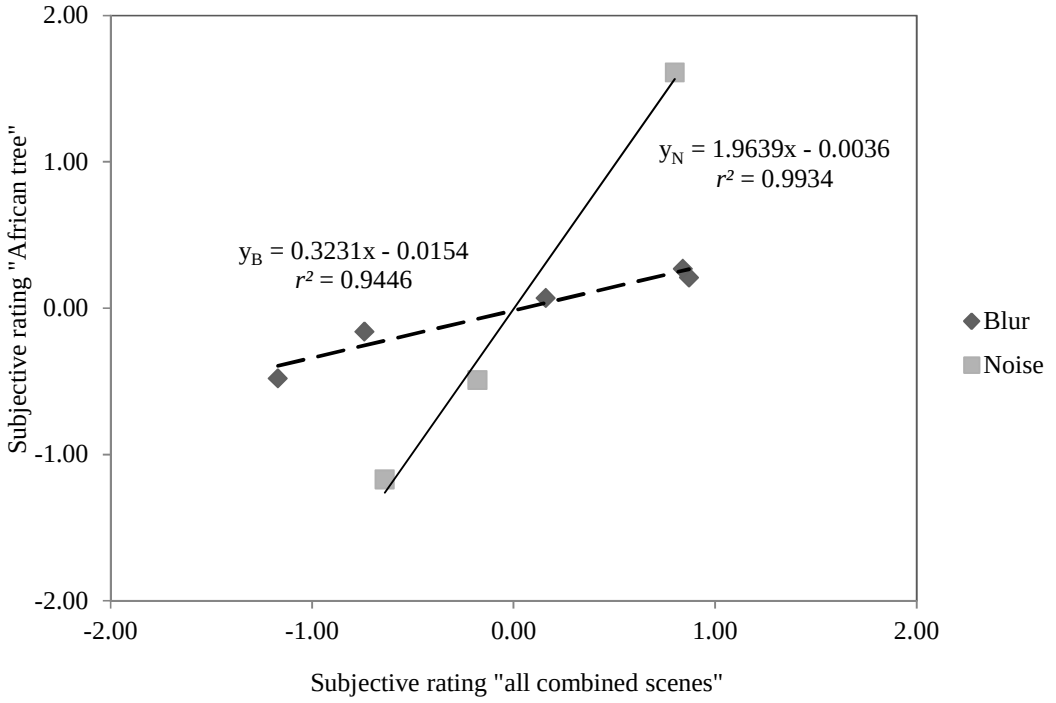


Figure 6-10. Scene dependency on the “African tree.”

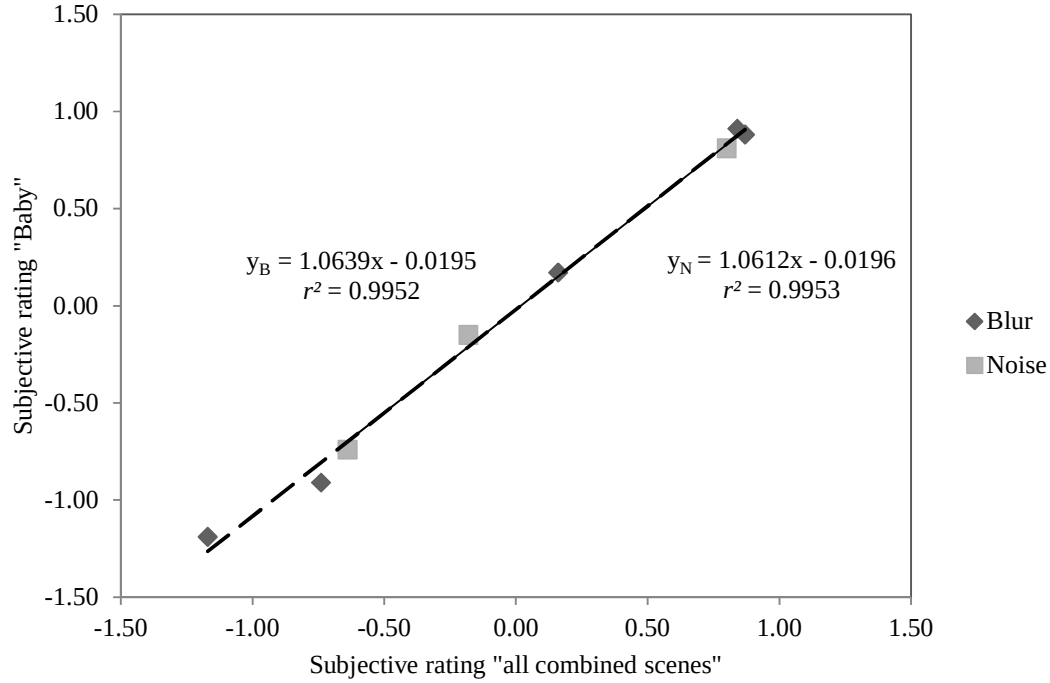


Figure 6-11. Scene dependency on the “Baby.”

Table 6-4 presents the gradient, constant and coefficient of determination (r^2), for each test scene and for blur/sharpness and noise attributes.

	Blur			Noise		
	Gradient	**Constant**	r^2	**Gradient**	**Constant**	r^2
African tree	0.32	-0.02	0.945	1.96	0.00	0.993
Baby	1.06	-0.02	0.995	1.06	-0.02	0.995
Bike	1.21	0.30	0.984	0.72	0.29	0.994
China town	0.95	0.04	0.995	0.97	0.04	0.984
Exercise	1.15	-0.20	0.966	0.51	-0.21	0.989
Formula	1.00	0.53	0.997	1.14	0.53	0.978
Glasses	0.86	0.10	0.974	1.17	0.11	1.000
Group	1.07	-0.38	0.980	0.66	-0.38	1.000
Human	0.91	0.07	0.990	1.10	0.06	1.000
Human2	0.32	0.31	0.948	1.07	-0.30	0.998
Human3	1.24	-0.20	0.965	0.55	-0.21	0.998
Human4	1.08	-0.43	0.991	1.11	-0.43	0.977
Kids	1.18	0.27	0.996	1.15	0.26	1.000
Landscape	0.86	-0.20	0.967	1.44	-0.19	1.000
Landscape2	0.63	0.06	0.936	1.82	0.06	0.976
Landscape3	1.05	0.07	0.987	1.31	0.07	0.999
London Eye	0.86	-0.37	0.996	1.09	-0.38	0.989
London Eye2	0.93	-0.42	0.999	1.28	-0.41	0.987
Louvre	1.16	-0.35	0.998	1.03	-0.36	1.000
National gallery	1.07	-0.13	0.996	0.96	-0.13	1.000
Old building	1.29	0.08	0.997	0.92	0.08	1.000
Plant1	1.15	0.79	0.985	0.34	0.79	0.841
Plant2	0.79	0.53	0.969	0.85	0.53	0.993
Plant3	0.87	0.22	0.992	1.13	0.21	0.997
Plant4	0.80	-0.12	0.994	1.12	-0.12	0.991
Plant5	1.09	0.08	0.998	0.99	0.08	0.988
Plant6	0.97	0.13	0.997	1.04	0.13	1.000
St. Pauls	1.40	0.03	0.991	0.50	0.02	0.999
St. Pauls2	1.10	0.03	0.996	0.87	0.03	0.999
Saules	1.43	0.01	0.964	0.19	0.00	0.861
Sungsil	1.24	-0.31	0.981	0.83	-0.31	0.990
Yellow flower	0.91	0.23	0.996	1.14	0.22	0.989

Table 6-4. The gradient, constant, and coefficient of determination for thirty-two scenes at blur/sharpness and noise attributes.

Regarding blurring: the scenes “Human2”, “African tree” and “Landscape2” were found to have the lowest gradient values, indicating the very low susceptibility to blurring. On the other hand, “Saules” has the highest gradient value, indicating the highest susceptibility to blurring. Figure 6-12 shows the thirty-two scenes ranked from lowest to highest, the scene susceptibility with respect to blur/ sharpness.

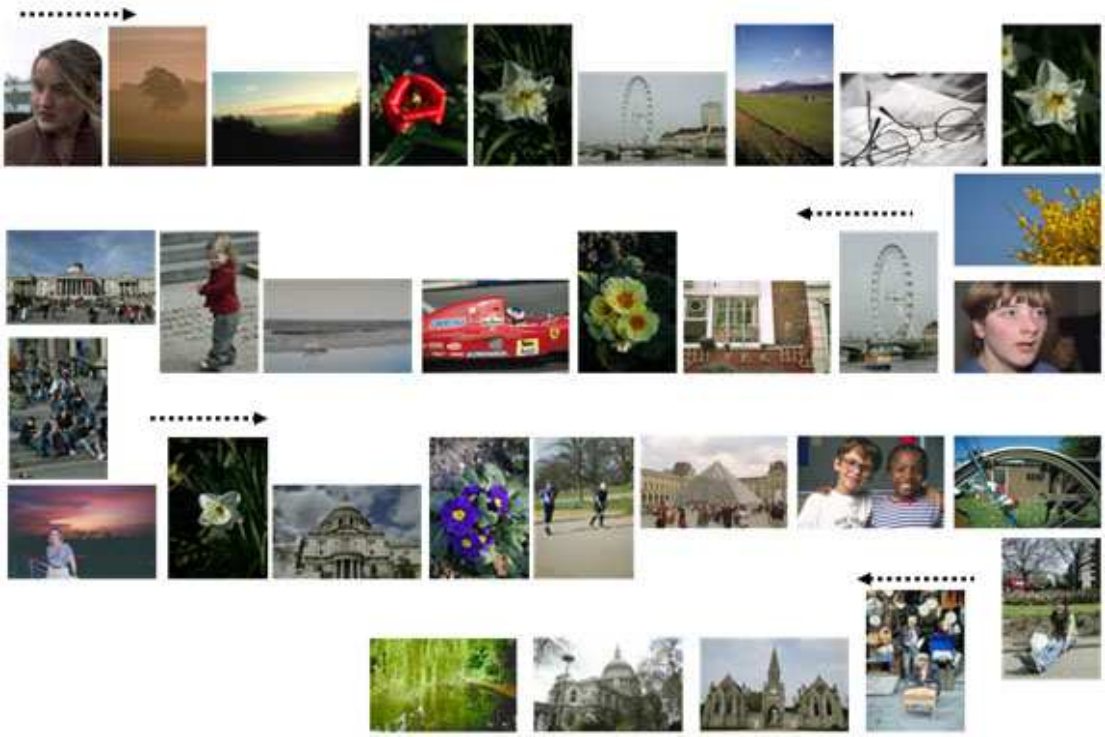


Figure 6-12. Scene ordered according to scene susceptibility parameter on blur/ sharpness.

Regarding noisiness: an extreme result was produced by the “Saules” scene (cf. Figure 6-9), having an extremely low gradient of 0.19. This is the busiest scene in the set [10] (pp. 266-267). The noise was largely masked by the high-frequency information in this scene and thus very low susceptibility to added noise (cf. Section 7.1). Figure 6-13 shows the thirty-two scenes ranked from lowest to highest scales, according to the scene susceptibility parameter on noisiness.

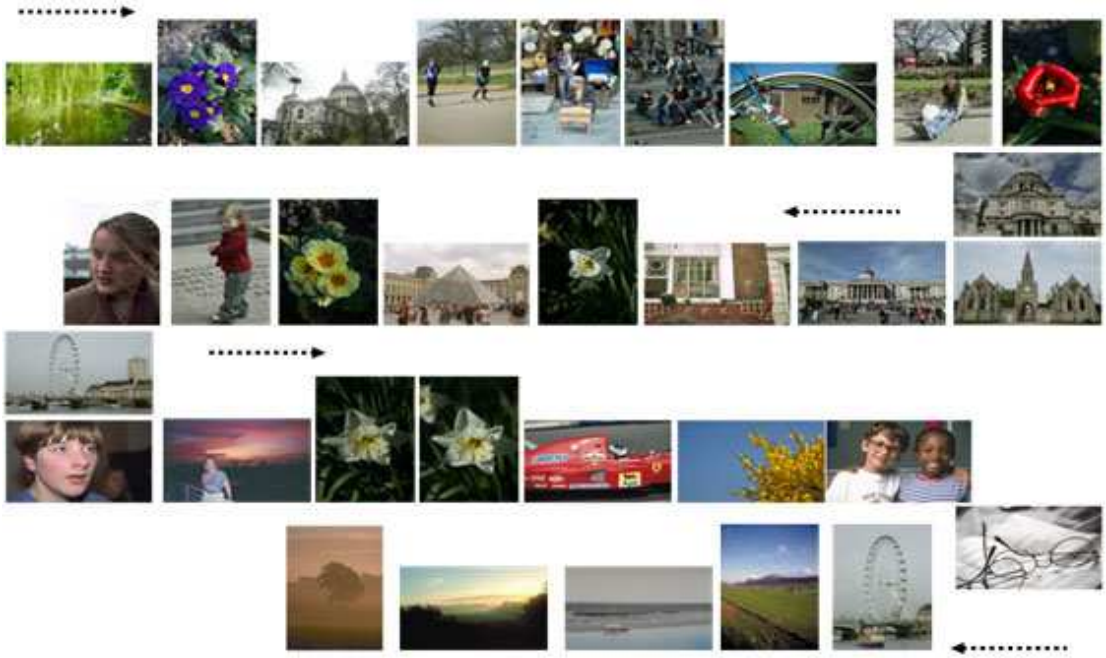


Figure 6-13. Scene ordered according to scene susceptibility parameter on noisiness.

In Figure 6-12 and Figure 6-13, the scenes are almost ranked in reverse, which indicates that scenes with high susceptibility to sharpness have low susceptibility to noisiness, and the reverse. This will be discussed in more detail later, in the section that discusses correlation between scene susceptibility parameters and scene descriptors (cf. Section 7.2).

6.6 Summary

A large-scale categorical judgement experiment was conducted. In all fourteen observers performed a total of 6720 observations. Psychometric scaling was used to create interval scales indicating the overall quality of images subjected to variations in blur and noise.

From the overall image quality assessments, individual sharpness and noisiness attribute scales were derived. The following summary lists the main findings of subjective image quality evaluation.

1. The level of distortion that was introduced by blurring covered the entire image quality scale. However, the level of added noise was too small for investigating the consequences of these variations on the full range of image quality.
2. Blurring decreased significantly the perception of noise, while added noise decreased the perceived blurring.
3. The perceived image quality depends on scene content.

Further the scene content variation in individual sharpness and noisiness was quantified. This was achieved by correlating the combined mean scale values from all scenes and the scale values of each individual scene, for each attribute. The scene dependency parameters were successfully derived.

The results from this chapter will enable further research into objective scene classification (cf. Section 7.2 and 7.3) and take forward the study of objective (physical) image quality scaling (cf. Chapter 8).

Chapter 7

Objective scene classification with respect to image quality

The aim of this chapter is to objectively classify the thirty-two original test scenes that were used in the psychophysical investigation in chapter 6. Objective scene classification will enable further study into application of physical image quality predictions.

The classification involves: 1) *feature generation*: investigation of various scene descriptors derived to describe properties that influence image quality (or individual attribute quality), 2) *feature selection*: investigation of the degree of correlation between scene descriptors and scene susceptibility parameters (cf. Section 6.5.3) and 3) k-means clustering for scene grouping.

7.1 Feature generation to describe scene descriptors

The first step towards the objective scene classification was a feature generation step (cf. Figure 2-22). This step aims to investigate a number of *scene descriptors*, derived to describe the original scene properties/features.

A number of scene descriptors related to texture and spatial image properties were derived using second-order statistical measures and measurement from edge detection. This was because the observer's preference should be affected by the spatial frequency properties of subjects. They describe how 'busy' or how 'flat' a scene is. These measures

were applied to the grey scale version of the image, which was obtained from the 8 bits per channel sRGB image by [136]:

$$\text{Grey scale image} = 0.2989R + 0.5870G + 0.1140B. \quad (7.1)$$

where R, G, B correspond to the pixel value of the R, G and B channels, respectively.

7.1.1 Second-order statistical measures

Second-order statistical measurements, which relate to textural information in images, were calculated from the grey level co-occurrence matrix (GLCM).

Implementation was carried out in MATLAB [136] using the default angle and distance values: 0° angle and 1 pixel distance. The parameters investigated in this work are listed below (cf. Section 2.2.2.8):

- Contrast (or inertia): relates to the scene texture variation. Contrast is 1 for a "complex" imaged scene.
- Homogeneity: relates to the scene texture variation. Homogeneity is 1 for a "uniform" imaged scene. Contrast (or inertia) and homogeneity are strongly, but inversely, correlated.
- Correlation (or linearity): relates to the scene texture variation with linearity.
- Energy: relates to the disorders in scene textures. The highest energy values occur when the grey level distribution has a constant or periodic form.

Table 7-1 lists the result of second-order statistical measurements.

	Contrast	Homogeneity	Correlation	Energy
African tree	0.05	0.98	0.98	0.26
Baby	0.15	0.93	0.95	0.17
Bike	0.37	0.87	0.92	0.11
China town	0.19	0.93	0.94	0.19
Exercise	0.32	0.88	0.91	0.15
Formula	0.22	0.94	0.95	0.16
Glasses	0.15	0.94	0.96	0.24
Group	0.31	0.89	0.90	0.15
Human	0.11	0.95	0.98	0.17
Human2	0.09	0.95	0.98	0.21
Human3	0.35	0.88	0.95	0.11
Human4	0.11	0.96	0.97	0.21
Kids	0.18	0.94	0.98	0.20
Landscape	0.05	0.97	0.96	0.31
Landscape2	0.07	0.97	1.00	0.18
Landscape3	0.14	0.94	0.97	0.16
London Eye	0.16	0.94	0.95	0.33
London Eye2	0.16	0.94	0.95	0.39
Louvre	0.26	0.90	0.94	0.16
National gallery	0.42	0.89	0.87	0.18
Old building	0.23	0.92	0.95	0.27
Plant1	0.21	0.91	0.94	0.16
Plant2	0.07	0.97	0.96	0.34
Plant3	0.10	0.96	0.98	0.25
Plant4	0.13	0.94	0.96	0.30
Plant5	0.11	0.95	0.96	0.40
Plant6	0.11	0.95	0.97	0.24
St. Pauls	0.57	0.82	0.94	0.09
St. Pauls2	0.27	0.90	0.90	0.17
Saules	0.51	0.81	0.87	0.10
Sungsil	0.66	0.82	0.88	0.10
Yellow flower	0.21	0.92	0.81	0.43

Table 7-1. Scene descriptor values from the second-order statistical measurements.

Regarding textural information: the scenes “St. Pauls”, “Saules” and “Sungsil” were found to have the lowest values in Homogeneity, Correlation and Energy, indicating complexity of texture in the imaged scene (there are the highest values in Contrast). On the contrary, “African tree” and “Landscape2” have the highest value in Homogeneity, Correlation and Energy, indicating that the imaged scene consists of mainly uniform areas.

7.1.2 Measurement from edge detection

The Sobel, Prewitt and Laplacian of Gaussian (LoG) edge detection algorithms were used to quantify the presence and strength of edges in the grey scale image. The Sobel and Prewitt edge detection algorithms are representative first-order methods. The LoG operator is a typical second-order method. The reason for this implementation is that the perception of the sharpness of an image is directly related to edge profiles of the image.

The Sobel and Prewitt edge detectors were applied using a 3 × 3 kernel size and 0.04 for sigma, and the LoG edge detector was applied using a 5 × 5 kernel size and 0.5 for sigma, which is the default value employed in MATLAB [136]. All edge detectors were operated with the ‘replicate’ boundary option in MATLAB [136], where the boundaries were assumed to be equal to the nearest border value. During the edge detection, the magnitude of the edge (G) was computed by [138] (pp. 577-580):

$$G = \sqrt{G_x^2 + G_y^2} \tag{7.2}$$

where G_x and G_y are the horizontal and vertical edge gradients of the image, respectively.

Then all edge gradients were averaged. Figure 7-1 (a) and (b) illustrate two original images and the corresponding threshold images after Sobel edge detection with the average edge gradient, related to the edges’ strength as well as the amount of edge information in the image. Figure 7-2 illustrates an original image and the corresponding images after Sobel, Prewitt and LoG edge detection.

(a) Average Sobel edge gradient: 11.68

(b) Average Sobel edge gradient: 66.19

Figure 7-1. Average edge gradient value with its original image (left) and the corresponding threshold image after Sobel edge detection (right), described in “African tree” (top) and “Kids” (bottom).

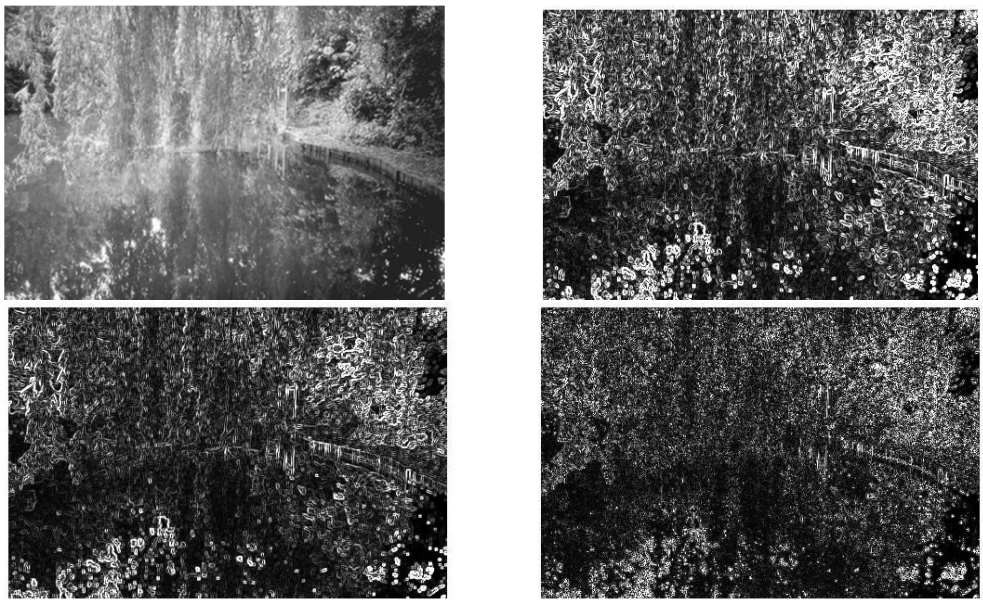


Figure 7-2. The “Saules” image (top-left) and its edge images after Sobel (top-right), Prewitt (bottom-left) and Laplacian of Gaussian (bottom-right) edge detection.

Table 7-2 lists the result from edge detection.

	Sobel	**Prewitt**	**Laplacian of Gaussian (LoG)**
African tree	11.68	8.45	8.40
Baby	58.03	42.59	27.98
Bike	113.12	83.04	61.87
China town	62.89	46.69	33.52
Exercise	78.02	56.61	49.90
Formula	66.28	49.26	34.96
Glasses	55.27	40.84	23.39
Group	85.00	62.74	47.01
Human	28.72	21.01	15.62
Human2	27.74	20.12	17.21
Human3	89.64	66.07	48.97
Human4	31.19	22.95	17.95
Kids	66.19	49.13	33.37
Landscape	22.16	16.31	14.01
Landscape2	17.36	12.69	11.16
Landscape3	37.88	27.36	25.92
London Eye	47.95	35.38	30.48
London Eye2	45.78	33.68	28.38
Louvre	57.57	41.90	37.14
National gallery	88.48	65.47	51.83
Old building	54.00	39.67	31.62
Plant1	68.10	50.31	24.57
Plant2	24.55	18.24	8.27
Plant3	30.18	22.44	10.19
Plant4	41.06	30.44	17.29
Plant5	31.43	23.38	10.49
Plant6	36.19	26.90	11.86
St. Pauls	111.65	80.85	75.46
St. Pauls2	82.46	60.80	50.00
Saules	101.69	73.25	68.69
Sungsil	125.09	90.91	80.53
Yellow flower	51.19	37.24	28.48

Table 7-2. Scene descriptor values from edge detection.

Regarding edge information: the scenes “African tree” and “Landscape2” have low *G* values, indicating the low presence and strength of edges in the imaged scene. In contrast, “Bike”, “St. Pauls”, “Saules” and “Sungsil” indicate the high presence and strength of edges in the imaged scene.

7.2 Feature selection using correlation between scene descriptors and scene susceptibility parameters

The second step toward the objective scene classification was a feature selection step (cf. Figure 2-22). This step aims to investigate the degree of correlation between scene descriptors and scene susceptibility parameters, i.e. represented by the gradient values in Table 6-4. Scene descriptors that successfully correlated with scene susceptibility in sharpness and noisiness provided a means toward the objective scene classification.

Spearman’s correlation coefficient, r_s, was used to investigate the correlation between various scene descriptors and scene susceptibility to noisiness and to sharpness. Spearman’s correlation coefficient is useful when data have a ranking but no clear numerical interpretation, such as when assessing preferences for data on an ordinal scale [144] (pp. 212). It is thus an appropriate measure for the purpose. The correlation coefficients range between -1.0 (indicating perfect anti-correlation) and 1.0 (indicating perfect correlation), with 0 indicating no correlation at all [145] (pp. 80-81).

$$r_s = 1 - 6 \sum_{1}^{n^2} d^2 \Big/ n(n^2 - 1) \tag{7.3}$$

where *d:* the difference in ranks, and *n*: the number of items in the sample.

Successful correlations were obtained between noisiness susceptibility parameters and second-order statistical measures, as well as measures derived from edge detection. When a correlation coefficient is larger than the level of significance at a 1% probability level,

i.e. the critical value, it indicates statistical significance. The critical value of the Spearman's correlation coefficient is 0.452 for a sample size of thirty-two [144] (pp. 214).

Successful correlations were also obtained between sharpness susceptibility parameters and, again, second-order statistical measures and measures derived from the edge detection. Table 7-3 shows the successful correlation coefficients for noisiness and sharpness.

Scene descriptors	**Correlation coefficient (r_s) for scene susceptibility to noisiness**	**Correlation coefficient (r_s) for scene susceptibility to sharpness**
Contrast (Inertia)	-0.694	0.802
Homogeneity	0.738	-0.781
Correlation (Linearity)	0.644	-0.550
Energy	0.577	-0.647
Average Sobel gradient	-0.701	0.786
Average Prewitt gradient	-0.701	0.786
Average LoG gradient	-0.593	0.747

Table 7-3. Successful correlation coefficients for noisiness and sharpness.

There were several interesting relationships between scene content and scene susceptibility to noisiness and sharpness. The results confirmed that the higher the texture in the scene content, the lower the susceptibility to noisiness and the higher the susceptibility to sharpness. For example, the correlation coefficients between the homogeneity and the scene susceptibility to noisiness and to sharpness were 0.738 and -0.781 respectively. In addition, the presence of a high number of strong edges in the image significantly decreased the susceptibility to noisiness and increased the susceptibility to sharpness. For example, the correlation coefficients between the average Sobel metric and the scene susceptibility to noisiness and to sharpness were -0.701 and 0.786 respectively. The results agreed with previous researches [10] (pp. 261), [16] (pp. 663), i.e. the spatial frequency properties of subjects were important scene dependence parameters in the observer's preferences and the noise was probably masked by the mainly high-frequency information in the scene.

It was also confirmed that the relationship between sharpness and noisiness is complimentary (cf. Figure 6-4 and Figure 6-6).

Overall, the results indicate that there is a clear relationship between selected scene descriptors and scene susceptibility parameters. Thus, the scene descriptors that correlated with sharpness and noisiness scene susceptibility can be used to objectively classify scenes.

7.3. Clustering for natural scenes

Finally, k-means clustering was implemented to objectively group the thirty-two test scenes according to their susceptibility to both sharpness and noisiness. The k-means clustering allows for simple and speedy grouping of large data sets [148] (pp. 526-528).

As mentioned before (in the section 2.2.2.8), k-means clustering consists of several steps [21] (pp. 406-408). The first step is to define a fixed number of clusters, k. Possible methods for choosing k include empirical and numerical methods. The empirical method is usually preferred. In relevant image quality investigations, k is chosen to be equal to 3.0 [208] (pp. 204). Once k is chosen, then modifications of the distances between all points in the n^{th} cluster (n varying from 1 to 3) and the centre of the cluster are applied (step 2). The main idea for their modifications is that the average distances between all points in each cluster and the central point are minimal. During these modifications, new cluster centres are allocated using Euclidean distances (step 3). The modification stops when the average distances from all points in the n^{th} cluster and the new central point have reached the minimum (step 4). Figure 7-3 illustrates the k-means clustering process.

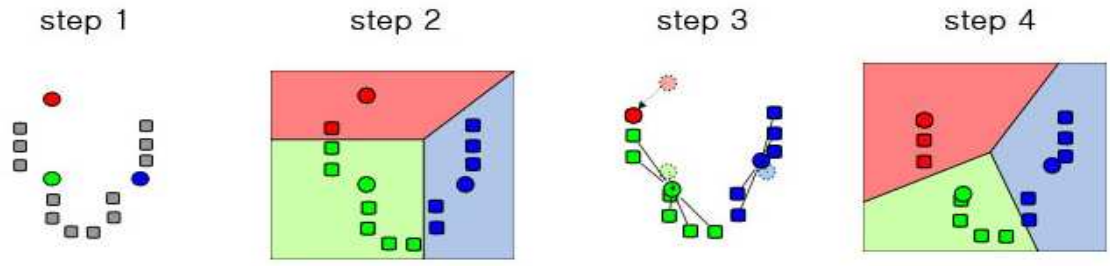


Figure 7-3. Simplified diagram of k-means clustering (produced by Wikipedia (viewed May 2009) [149]).

The two scene descriptors that correlated most successfully with both noisiness and sharpness susceptibility, i.e. the *homogeneity* and *average Sobel edge gradient* descriptors, were used to implement the clustering. Clustering was implemented in the SPSS programming environment [212]. Figure 7-4 presents the three clusters, with the initial and final centres of each cluster. The images corresponding to each of the three clusters (or groups) are shown in Figure 7-5.

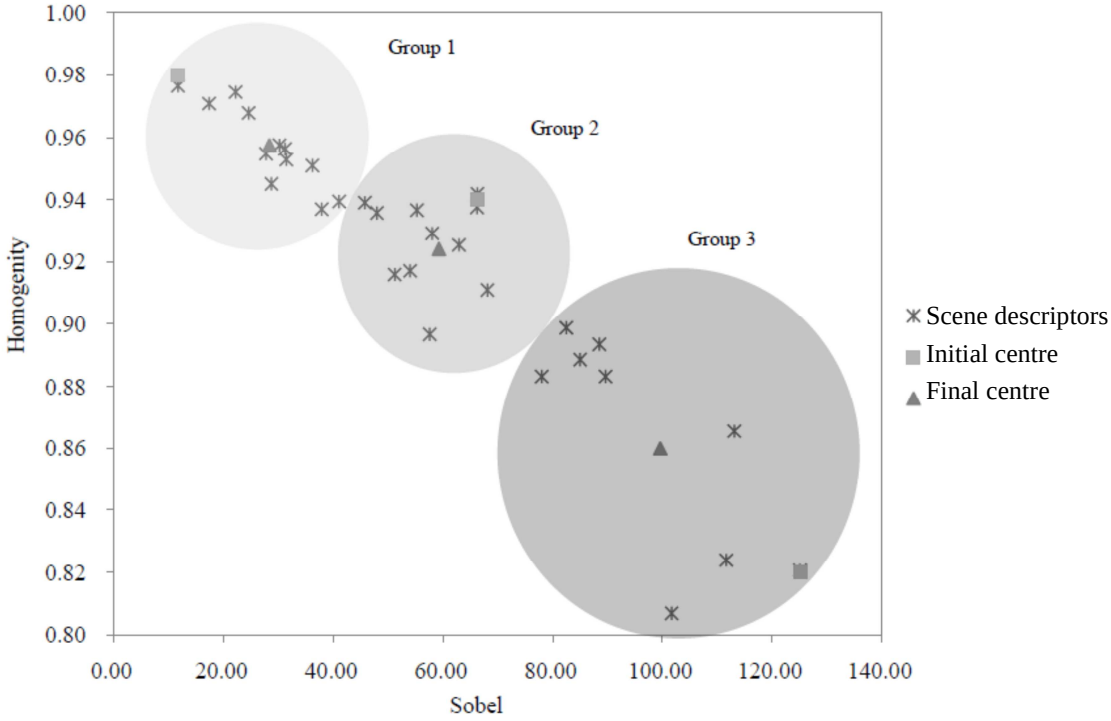
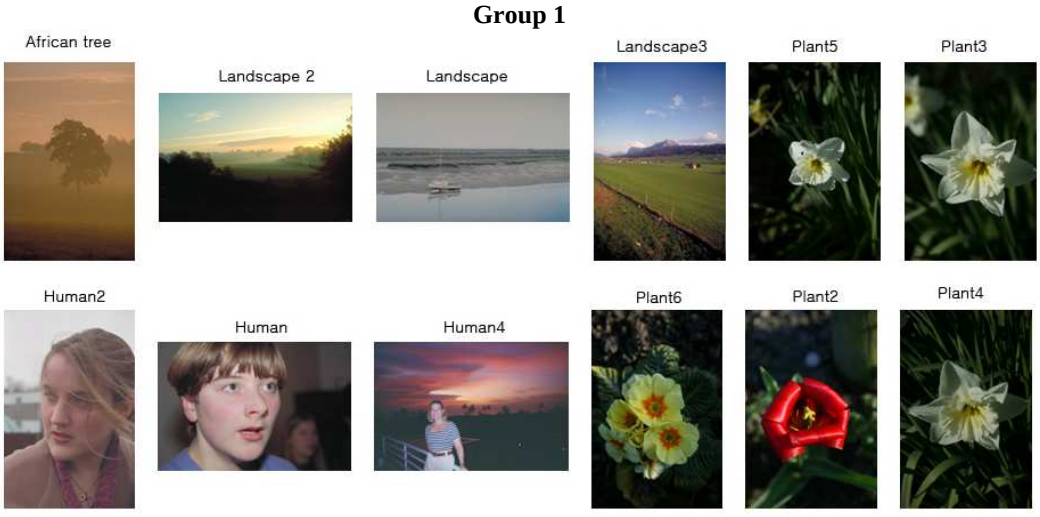


Figure 7-4. Three clusters (groups) in the scatter diagram, measured using the *homogeneity* and *average Sobel edge gradient* descriptors.



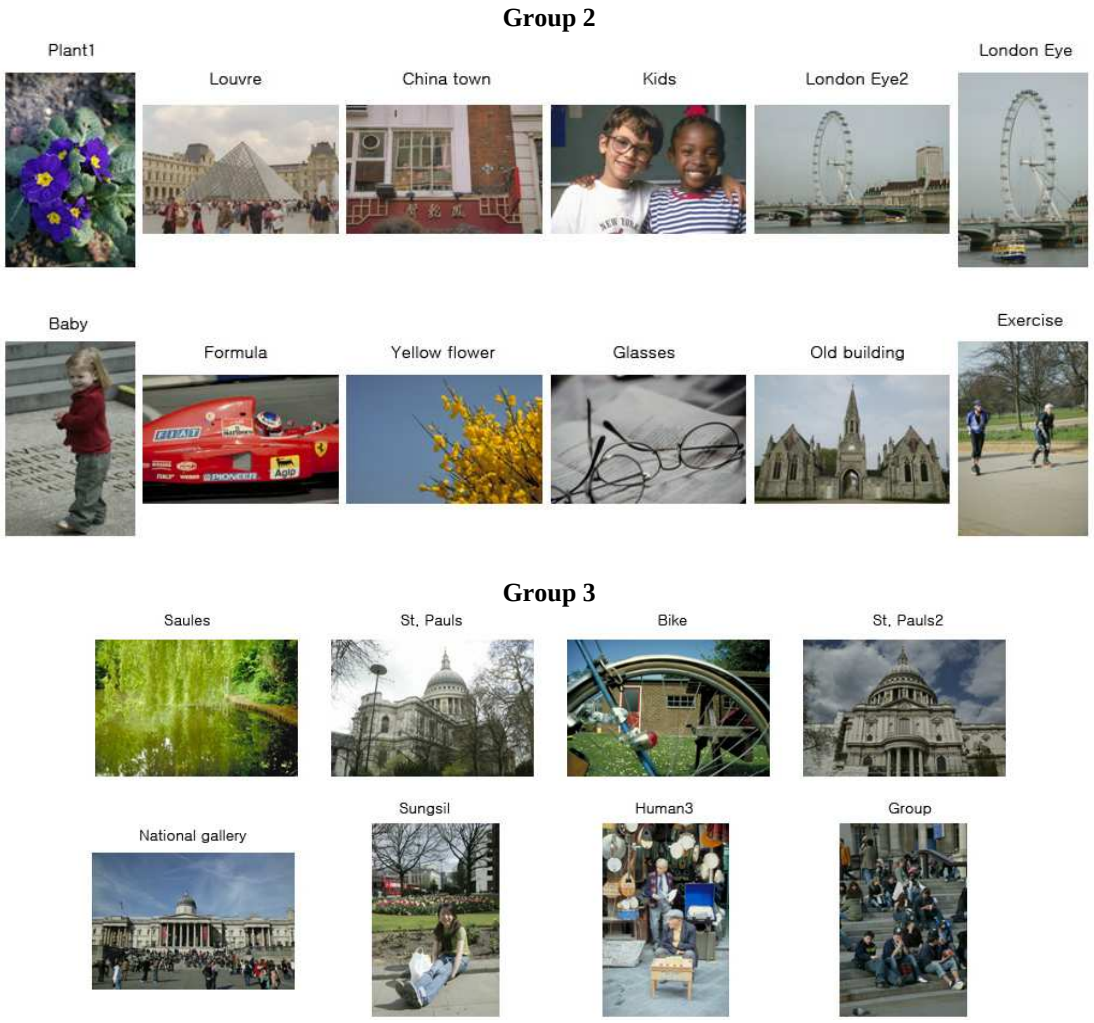


Figure 7-5. Images in three clusters (groups).

The k-means clustering classifies images into three groups, using the selected scene descriptors. Each group might reflect the spatial configurations in the imaged scene, which are well-known "context" factors [11] (pp. 25).

This k-means clustering involves the old rule “preferences do not occur in a vacuum, they are always formed relative to a context [213] (pp. 76)”, and no-requirement was needed of the assumption that average spatial configurations of scene elements are considered to generate quality scales.

The validation of the objective scene classification was examined using visual inspection, i.e. visual observation [18, 19, 20]. The images in the same group by k-means clustering tend to have common visual features, e.g. the “Landscape” seems to have similar structural features to “Human 4”, “Landscape 3”, “Plant 2” (Group 1), while it seems to have different structural features to “Plant 1” (Group2) and “Saules” (Group3). On the other hand, the classification of images around the boarder is an arguable point, e.g. does “plant 4” visually belong to the scene group as “Landscape” (Group 1) or is it visually closer to “Baby” (Group2)?

Overall, there was a reasonable match between inspection and objective scene classification. Thus, it can be concluded that the three groups of scenes were effectively derived using the objective scene classification by k-means clustering, i.e. scenes with: 1) low susceptibility to sharpness distortions and high susceptibility to noisiness, 2) average susceptibility to sharpness distortions and noisiness, 3) high susceptibility to sharpness distortions and low susceptibility to noisiness.

7.4. Summary

Firstly, a number of scene descriptors were successfully derived from second-order statistical measurements, as well as measures derived from edge detection. They were concerned with the extraction of image features that affect sharpness and noisiness judgements.

Secondly, the degree of correlation between scene descriptors and scene susceptibility parameters that were derived in section 6.5.3 was investigated, using the Spearman's correlation coefficient. Successful correlations were obtained between: sharpness and noisiness susceptibility parameters and second-order statistical measures, as well as measures derived from edge detection. These correlations indicate that the selected scene descriptors successfully represent sharpness and noisiness susceptibility, and can be used to objectively and automatically classify the test scenes used in image quality investigations.

Thirdly, using the selected scene descriptors and applying k-means clustering, three groups of scenes were effectively derived, i.e. scenes with: 1) low susceptibility to sharpness distortions and high susceptibility to noisiness, 2) average susceptibility to sharpness distortions and noisiness, 3) high susceptibility to sharpness distortions and low susceptibility to noisiness.

The results from this chapter enable further research into the application of physical image quality predictions using the objective scene classification (cf. Section 8.3).

Chapter 8

Calculation of image quality metrics

This chapter sets out the details for objective (physical) image quality scaling. This scaling was carried out using two different device-dependent metrics: the Effective Pictorial Information Capacity (EPIC) [74] and the Perceived Information Capacity (PIC) [71]. The performance of the metrics was assessed using subjective results obtained in chapter 6, before the implementation of modifications that account for scene content. From the assessment, the modified metrics were considered to be an improvement over the original metrics. A validation experiment was then carried out to test the improvement in the metrics that account for scene content.

8.1. EPIC and PIC implementation

An example of a device-dependent model is the Effective Pictorial Information Capacity (EPIC) [214]. It is based on signal transfer theory [75] and has been relatively successful in predicting the quality of compressed images and images that vary in sharpness and noisiness [74] (pp. 7).

EPIC values were calculated by combining two system variables: the effective pixel dimension in the image (n) and the number of effective distinguishable levels for each recording cell (dynamic range of a system) (m). These variables were cascaded from the input, processing, output and visual system (cf. Section 2.2.2.1).

For the EPIC metric calculation, first the measurement of the effective pixel dimension in the image (n) was carried out as follows:

1) The effective pixel dimension (ω) was taken as the width of line spread function (LSF) at which the MTF falls to 50% [214] (pp. 3). The model accounted for the variations in sharpness of the captured test stimuli (input), the blur introduced by Gaussian blur filtering (processing), the sharpness of the 21 inch LCD system (output) and the CSF-Barten's model- which was employed as the model for the human eye (HVS). The imaging chain was that employed in the psychophysical scaling of image quality (cf. Section 6.1 and 6.3).

   In more detail:

   Input: The MTF of the capturing system was calculated across the range of focal lengths and *f*-numbers used to capture each test image. These are presented in Table 8-1.

Image	**Focal length**	***f*-number**	**Image**	**Focal length**	***f*-number**
Baby	135	6.3	Exercise	112	5.6
Group	135	11	London Eye	100	22
London Eye2	135	20	National gallery	28	5.6
Old building	28	8	Plant1	135	5.6
Plant2	135	5.6	Plant3	100	5.6
Plant4	100	36	Plant5	135	5.6
Plant6	135	5.6	St. Pauls	28	11
St. Pauls2	28	22	Sungsil	35	8

Table 8-1. Lens focal lengths and *f*-numbers for each of the captured images.

Figure 8-1 shows the fluctuation of the measured MTFs with varying focal lengths 28-135mm and *f*-numbers 5.6-36, which were used to capture all the different test scenes.

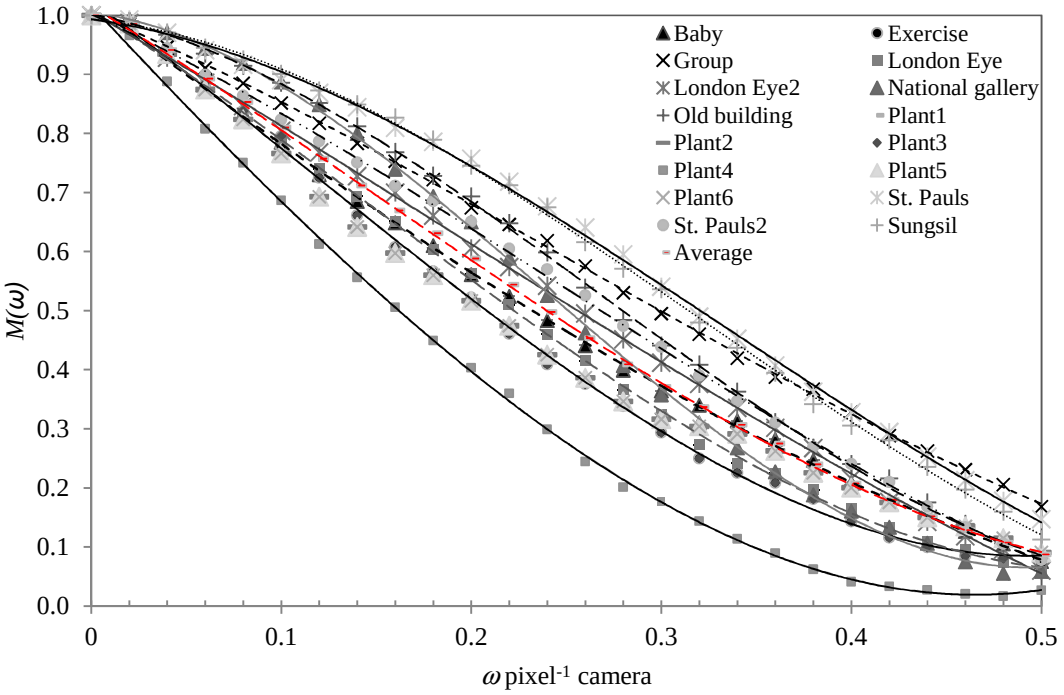


Figure 8-1. Polynomial functions representing the variation of the camera MTFs for the captured images using different lens focal lengths and *f*-numbers (refer to Table 8-1). The average measurement appears with a red line.

Since the image input systems and experimental conditions for the Kodak Photo CD images [74] (pp. 5) were unknown (cf. Section 6.1), the MTF of an 'average input' system was assumed. Figure 8-1 includes the MTF of an 'average input' system, based on the Canon EOS-1Ds system. There are likely to be variations between the assumed and the actual Photo CD image MTFs (cf. Figure 8-1). Although such differences in sharpness for the Kodak Photo CD images exist, including them in the test set was, in my view, worthwhile for assessing the image quality of scenes with large variations in scene content. Jenkin *et al* [74] (pp. 5) noted that "scenes were chosen purposely to contain different amounts of detail, low varying areas, various degrees of global and local intensities and colourfulness, a variety of dominant colours and strong and weak edges." This was the aim for my project too. In addition, Barten [220] (pp. 158) noted that "image quality metrics are usually generic measures.

This means that they are independent from actual pictorial content. They do not contain the modulation of different spatial frequency components of an actual image, but only the MTF by which these modulations are multiplied. This is almost remarkable, but in practice, it appears that the real amplitude of these components does not play an important role in the judgement of image quality." In this project, a various range of different scenes are captured by the input system. Based on Barten's statement, we can envisage that, potentially, the MTF of input system using test target may not play a significant role in the MTF metrics at different scenes. This was the problem for my project. Nevertheless, the unspecified MTFs of the Photo CD images are expected to render results related to these images prone to errors.

Processing: Captured images of edges were manipulated by varying Gaussian blurring in MATLAB software [136]. The standard deviation (σ) of the Gaussian low-pass kernel ranged from 0.01 to 1.24 at 0.3075 intervals. This was the same range as that implemented the test images in the psychophysical scaling of image quality (cf. Section 6.2).

The error range was calculated using the statistical error [33] (pp. 134), [215] (pp. 296). Its result produced up to $\pm$8%.

$$SE \approx \frac{1}{\sqrt{X\delta\omega}} \tag{8.1}$$

where *SE* is the standard error, *X* is the length of data in pixels and $\delta\omega$ is the effective bandwidth of the measurement.

Figure 8-2 illustrates the MTF curves at various amounts of Gaussian blur.

The MTF of the 10x interpolation was not taken into account, as the bicubic interpolation is a highly non-linear process (cf. Section 5.1.4). Some researchers take into account the MTF of non-linear process [192] (pp. 8). In some other works it is being excluded [32] (pp. 58-62), [74] (pp. 6).

Output: The MTF of the 21 inch LCD system was calculated by dividing the combined MTF by the camera MTF in section 5.1.3. The vertical LCD MTF, fitted by a third degree polynomial function, is presented in Figure 8-3.

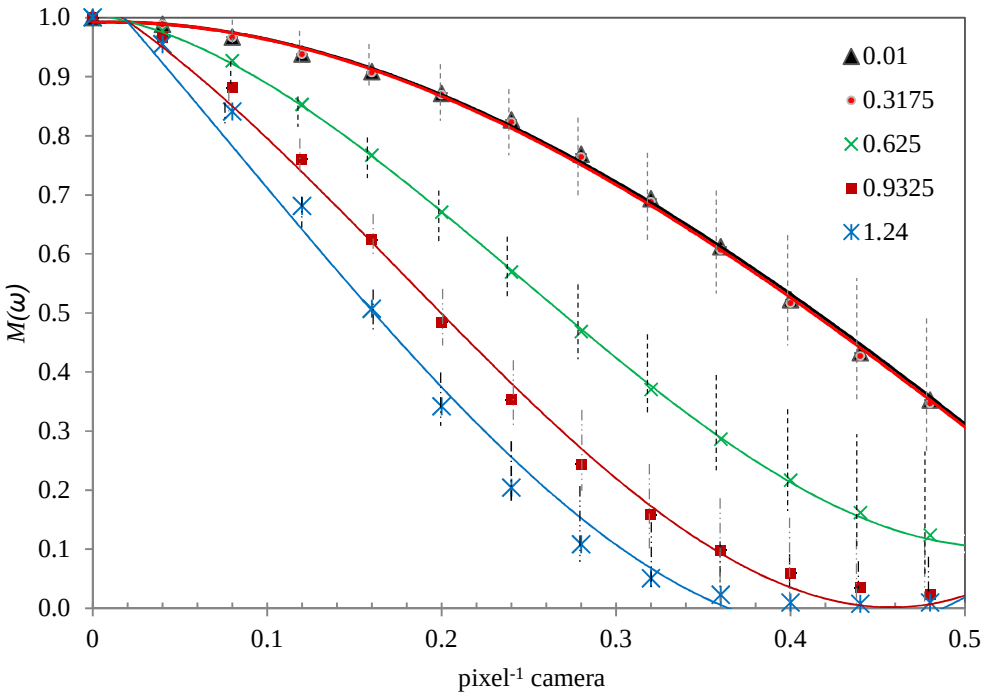


Figure 8-2. Polynomial functions representing the variation of the MTFs with varying the σ using Gaussian blur.

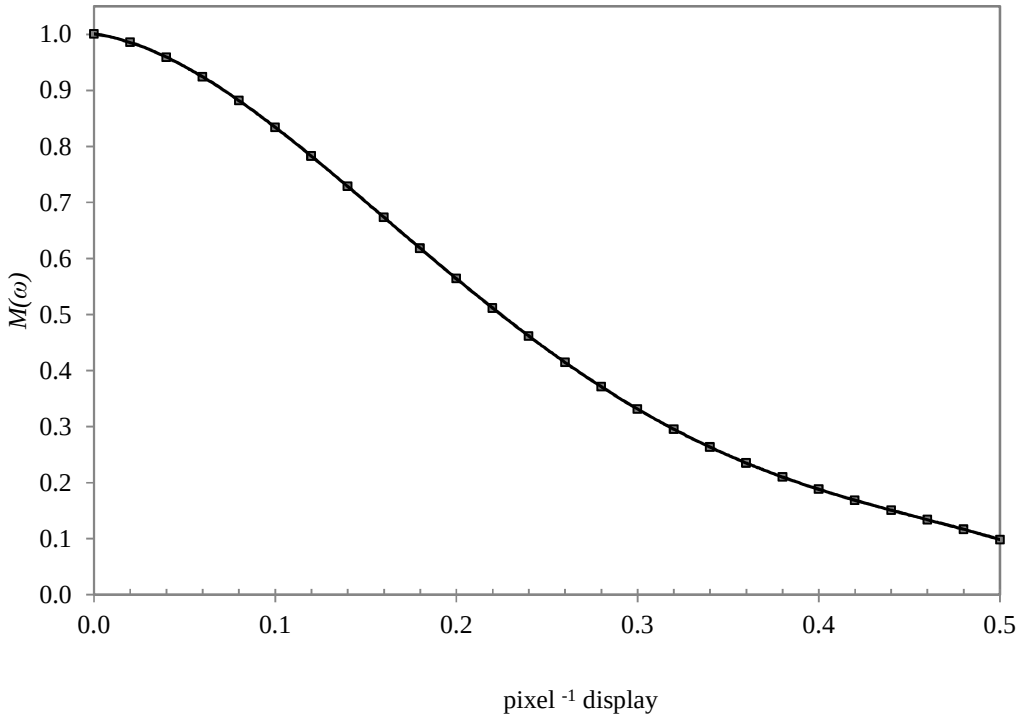


Figure 8-3. Polynomial function representing the vertical MTF curve of the LCD.

HVS: The Barten CSF was used as the eye's MTF model, because it takes into account various viewing conditions and has been successfully implemented in metrics that predict sharpness and noisiness [1] (pp. 377) (cf. Section 2.2.2.7). The specific viewing conditions employed were a luminance of 100 cd/m^2, a viewing distance of approximately 0.6 m, and a visual angle of 90◦ (cf. Section 6.3). Figure 8-4 illustrates the Barten CSF calculated with respect to the LCD plane.

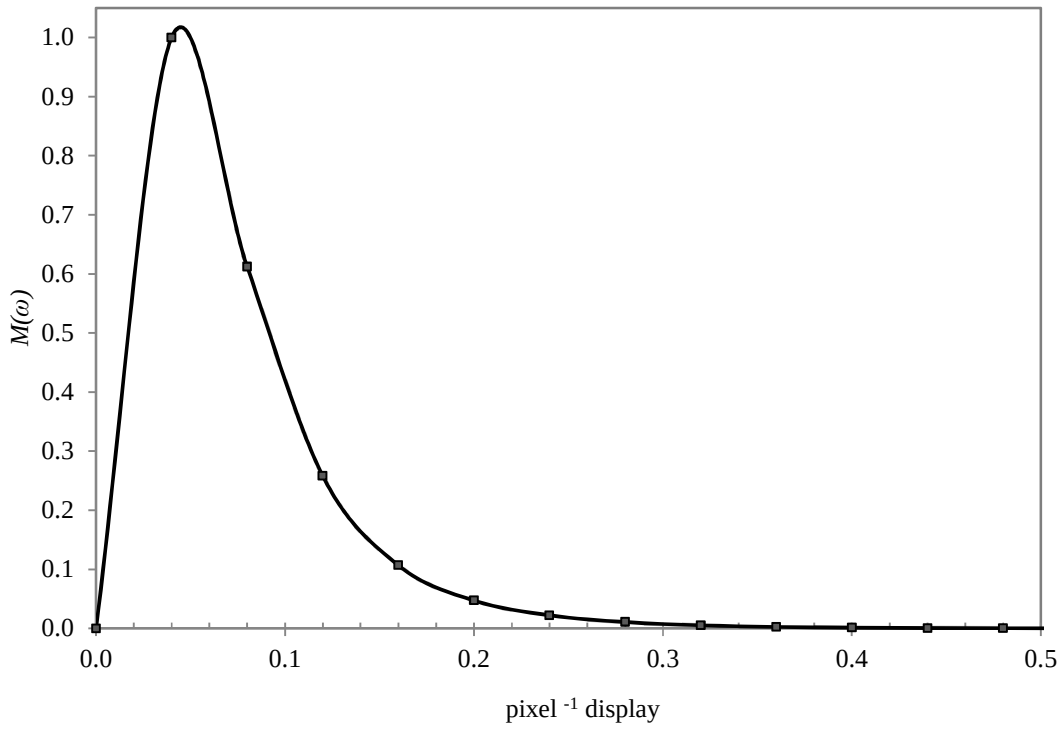


Figure 8-4. Contrast sensitivity function (CSF) at a luminance of 100 cd/m^2, a viewing distance of approximately 0.6 m and a visual angle of 90^o on the LCD plane.

Total system: The total system MTF was obtained by multiplying the individual system MTFs. For this, it was assumed that each component was linear and that the MTF for each successive component was independent from that of the previous component [33] (pp. 183). All measurements were considered in the LCD plane. Figure 8-5 shows examples of total system MTFs, using the average capture MTF (input), all different Gaussian blur MTFs (processing), the MTF of the LCD (output) and the CSF (HVS).

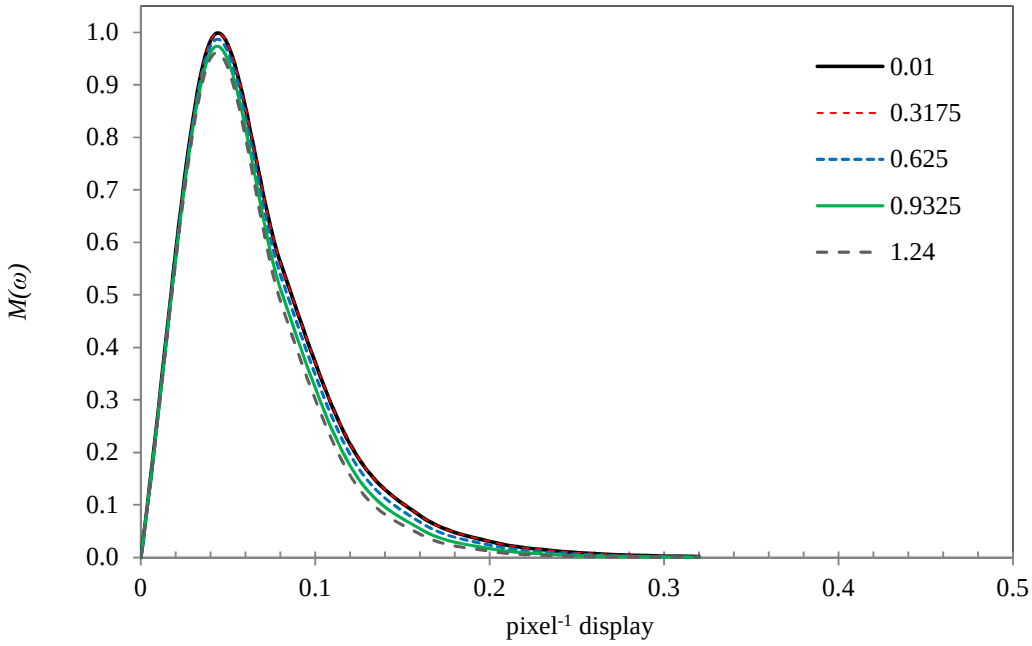


Figure 8-5. Examples of total system MTFs, obtained by multiplying the MTFs of input, output and HVS and the MTFs of different levels of blur (σ values).

The LSFs for each captured scene were then calculated using the relationship between LSF and MTF (cf. Figure 2-17), i.e. the MTF of an imaging system is calculated as the normalised modulus of the Fourier transform of the LSF. An example of the implementation, i.e. the calculation of one LSF, is shown in Table 8-2, and the results of the LSFs for each image are presented in Table 8-3.

Relative Distance (in pixels)	**FT^{-1} of MTF**	**Modulus of FT^{-1} of MTF**	**Normalised Modulus of FT^{-1} of MTF**
0.0000	0.1363	0.1363	1
0.1708	0.0884 + 0.0854i	0.1229	0.858941
0.3415	0.016 + 0.0963i	0.0976	0.592414
0.5123	-0.0255 + 0.0713i	0.0757	0.361673
0.6830	-0.0408 + 0.0437i	0.0598	0.193738
0.8538	-0.0434 + 0.0233i	0.0493	0.082814
1.0246	-0.0422 + 0.0099i	0.0433	0.020503
1.1953	-0.0414	0.0414	0
1.3661	-0.0422 - 0.0099i	0.0433	0.020503
1.5368	-0.0434 - 0.0233i	0.0493	0.082814
1.7076	-0.0408 - 0.0437i	0.0598	0.193738
1.8784	-0.0255 - 0.0713i	0.0757	0.361673
2.0491	0.016 - 0.0963i	0.0976	0.592414
2.2199	0.0884 - 0.0854i	0.1229	0.858941

Table 8-2. The calculation of LSF for “Exercise” scene.

Image	Half widths of 0.5 LSF at Blur 1 (0.01)	Half widths of 0.5 LSF at Blur 2 (0.3175)	Half widths of 0.5 LSF at Blur 3 (0.625)	Half widths of 0.5 LSF at Blur 4 (0.9325)	Half widths of 0.5 LSF at Blur 5 (1.24)
Baby	0.48749	0.48833	0.49078	0.49462	0.50087
Exercise	0.47651	0.47736	0.47981	0.48365	0.48990
Group	0.44989	0.45075	0.45319	0.45703	0.46328
London Eye	0.48168	0.48254	0.48498	0.48882	0.49507
London Eye2	0.47377	0.47463	0.47708	0.48091	0.48717
National gallery	0.47677	0.47763	0.48007	0.48391	0.49016
Old building	0.43160	0.43246	0.43491	0.43875	0.44499
Plant1	0.52999	0.53085	0.53330	0.53714	0.54339
Plant2	0.52999	0.53085	0.53330	0.53714	0.54339
Plant3	0.52999	0.53085	0.53330	0.53714	0.54339
Plant4	0.53822	0.53908	0.54153	0.54537	0.55162
Plant5	0.52999	0.53085	0.53330	0.53714	0.54339
Plant6	0.52999	0.53085	0.53330	0.53714	0.54339
St. Pauls	0.42922	0.43008	0.43252	0.43636	0.44261
St. Pauls2	0.46133	0.46218	0.46463	0.46847	0.47472
Sungsil	0.42488	0.42574	0.42819	0.43203	0.43828
Average	0.47178	0.47264	0.47508	0.47892	0.48517

Table 8-3. Calculated half width of individual scene LSFs at height 0.5.

2) The area of the image (A_{im}) was calculated as 167 × 111 mm. This was formed by the combination of the image pixels, 476 by 317 pixels, and the common SXGA pixel dimension of the LCD [216], [217] (pp. 13). This is because the EPIC is designed with the specific area of image condition (162 × 105 mm) [74] (pp. 5). As mentioned before, metrics are simpler, more efficient computationally, and sometimes are designed with specific application in mind, e.g. they could apply to a specific area of the image on display devices [218] (pp. 943).

3) The effective pixel dimension at image area *(n*) was then calculated by:

$$n = \frac{A_{\text{im}}}{\omega^2} \tag{8.2}$$

where A_{im} is the area of the image, and ω is the effective pixel dimension.

The number of effective distinguishable levels for each recording cell (dynamic range of a system) (m) was calculated by:

$$m \approx \frac{DS}{2k\sigma} + 1 \tag{8.3}$$

where DS is the difference between the maximum and minimum possible levels of the recording system (256 available levels in an 8-bit system), k is a constant (2) and σ represents the system's diminishing ability to distinguish independent levels, calculated from each individual component of the imaging chain: the noise in the test stimuli (processing), the measured noise of a 21 inch LCD system with the aid of the high-performance digital camera (input-output (cf. Section 5.2.2)) and the human eye parameter (1).

The value of σ varies with the scanning aperture (or window), thus the actual standard deviation of the noise was approximated by $\sigma_A \approx \sigma/L$ for square recording cells via an implementation of the Selwyn granularity, where L is the width of the LSF. All measurements were performed in the display plane and the pixel^{-1} units. Jenkin [74] (pp. 4) noted that "care must be taken to ensure that all measurements are performed in the same plane (eye or display typically) and the units and measurement aperture of the variance are understood in order that they may be combined properly."

The number of effective distinguishable levels of the system was considered to be 64. This number has been found to be a typical value for the intensity levels distinguishable on a CRT faceplate under normal display viewing conditions [33], [74] (pp. 4). This result is of course valid for the specific conditions. It was also empirically tested and found true on the 21 inch LCD. Changing the LCD luminance, angle of subtense and viewing conditions can change the number of effective distinguishable levels of the system.

The perceived information capacity (C) was then calculated, based on the effective pixel dimension in the image (n) and the number of effective distinguishable levels for each recording cell (m).

$$C = n\log_2(m) \tag{8.4}$$

EPIC values were finally derived in bits per steradian by:

$$\mathrm{EPIC} = \frac{C}{\Omega} \tag{8.5}$$

The visual solid angle (Ω) is given by:

$$\Omega = A_{im}/r^2 \tag{8.6}$$

where A_{im} is the area of the image, and r is the viewing distance.

Another example of a device-dependent model is the Perceived Information Capacity (PIC) [71] (pp. 8). It is based on SNR [70] (pp. 59) and is derived from the system's parameters, including MTF and NPS. These variables are cascaded from the input, processing, output and visual system.

For this PIC metric calculation, the measurement of the MTF and NPS was combined as follows:

$$\mathrm{PIC} = \mathrm{k}_1 \left[\int_0^{u_{max}} \ln\left(1 + \frac{S_0(u) M_c^2(u) M_d^2(u) M_{eye}^2(u)}{N(u) M_{eye}^2(u) + N_{eye}(u)}\right) \frac{\mathrm{du}}{\mathrm{u}} \right]^{0.5} + \mathrm{k}_2 \tag{8.7}$$

The signal spectrum, $S_0(u)$, was obtained from the power spectrum of an 'average' scene. $M_c(u)$ was the MTF of each individual capture (for each scene) multiplied by the MTF of the Gaussian blur that was introduced to the captured test stimuli. $M_d(u)$ was the measured MTF of the display system, and $M_{eye}(u)$ used the Barten's CSF model. For image noise, $N(u)$, was the area under the noise power spectrum of the imaging chain. $N_{eye}(u)$, the eye's noise, the suggested parameter 1 was used. The constants k_1 and k_2 were set to 1 and 0, respectively. These constants were determined from a comparison with Barten's parameters [69] (pp. 7). However, these constants have no physical meaning [220] (pp. 161).

The approximation for the power spectrum of an 'average' scene was calculated according to the equations given by Barten [69] (pp. 5):

$$S_0(u) = \frac{a^2}{2\left(\frac{1}{XY}+u^2\right)} \tag{8.8}$$

where

$$a^2 = \frac{\Delta(\log I)^2}{2}\frac{2}{\pi}\frac{1}{\ln(XYu_{\max}^2)} \tag{8.9}$$

where I is the intensity, in this case the pixel value for a signal channel, and the value of X is obtained from:

$$\frac{1}{X^2} = \frac{1}{X_0^2} + \frac{1}{X_e^2} + \left(\frac{u}{N_e}\right)^2 \tag{8.10}$$

where X_0 is the angular size of the object, and X_e is the maximum angular size over which the eye is able to integrate the information (X_e=12°). N_e is the maximum of cycles over which the eye is able to integrate the information (N_e=15 cpd). In addition, a similar relationship is used to find Y.

Figure 8-6 and Figure 8-7 present the average EPIC and PIC values. Each label on the x-axis represents a specific level of distortion in blurring (B) and noise (N).

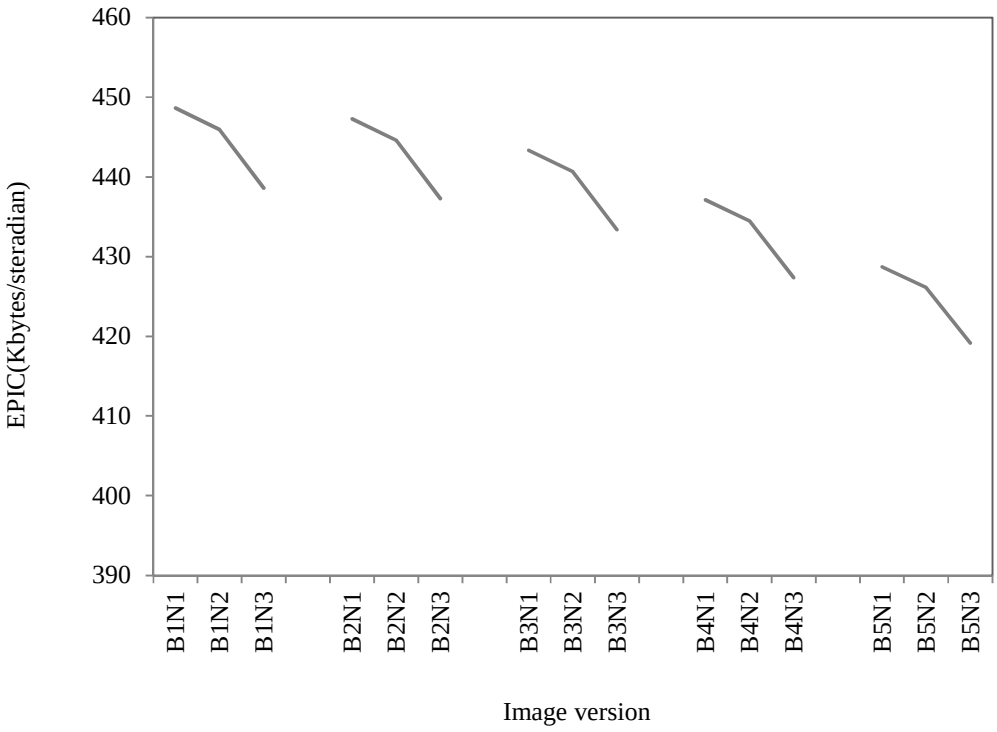


Figure 8-6. EPIC values for all levels of distortion in blurring (B) and noise (N).

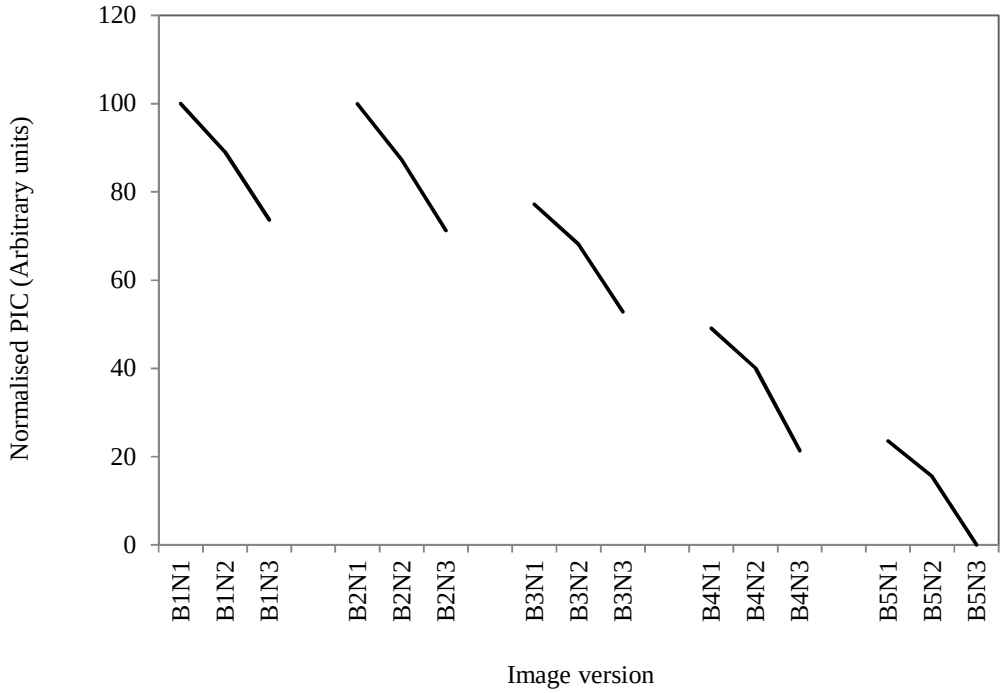


Figure 8-7. Normalised PIC values for all levels of distortion in blurring (B) and noise (N).

Figure 8-8 presents the comparison between EPIC and PIC. Their values were normalised and rescaled as percentages, for comparison purposes, i.e. the EPIC unit is Kbytes/steradian, and the PIC unit is arbitrary.

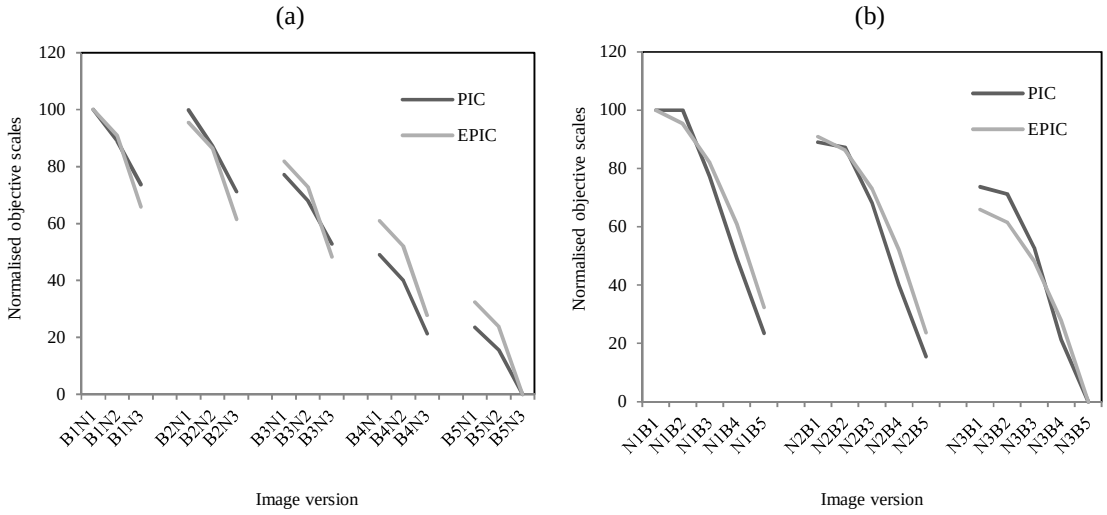


Figure 8-8. Comparison between normalised EPIC and normalised PIC values for all levels of distortion in blurring (B) and noise (N); (a) targeting attribute-blur (b) targeting attribute-noise. (a) and (b) use same data, but they are presented differently according to the *targeting attribute*.

The results demonstrated that the EPIC metric is more sensitive to noisiness at the same level of blurring than the PIC (in the Figure 8-8 (a)), while the PIC metric is more sensitive to sharpness at the same level of noise than EPIC (in the Figure 8-8 (b)). This can explain why the EPIC metric correlates relatively well with noisiness for flat scenes and the PIC metric correlates relatively well with sharpness for complex scenes, when the two metrics are compared. Jenkin *et al* [221] (pp. 69) demonstrated that cascading many MTF and NPS curves within the $SQRI_n$ calculation heavily biases the metric towards low-frequencies in the system. Topfer and Jacobson [71] (pp. 8) described that the SNR type's metrics as correlating well with perceived image quality for complex scenes and noisy images. This phenomenon was found to be the case in the EPIC and PIC evaluation of the individual scenes described later (cf. Section 8.2).

8.2. EPIC and PIC evaluation as device-dependent quality predictors

A validation experiment was carried out to test the success of the EPIC and PIC metrics in predicting perceived image quality, using results from subjective tests involving thirty-two test scenes, replicated with various degrees of sharpness and noisiness (cf. Chapter 6).

The Pearson's correlation coefficient (r) was used to investigate the degree of correlation between objective and subjective scales. A degree of correlation larger than a level of significance at 1% probability level indicates that the degree of correlation between the two variables is statically significant. The degree of correlation should exceed 0. 570 for a sample size of thirty-two. However, this critical degree may be no practical, due to the large sample size [144] (pp. 219). Therefore, the *coefficient of determination* (r^2) was also calculated as the third way of measuring statistical significance [222] (pp. 196). The reason for using the coefficient of determination is that r^2 is often used in image quality metric validation.

Figure 8-9 and Figure 8-10 show the r and r^2 coefficients for the thirty-two test scenes. Table 8-4 and Table 8-5 list the detailed results. Relatively successful correlations were obtained, i.e. the average of 32 coefficients is r =0.865 (r^2=0.748) in EPIC and r =0.870 (r^2=0.757) in PIC. The result confirmed that the metrics were reliable predictors of image quality. Other studies have indicated similar correlations [71] (pp. 22), [74] (pp. 7).

In general, device-dependent image quality metrics have been shown to be unequally successful in predicting the quality of individual scenes [10] (pp. 259), [74] (pp. 6). As expected, the results confirmed that the metrics could not predict the quality of different images with varying scene content consistently, i.e. the metrics performed well on most of the scenes, but predicted less successfully the perceived quality of some individual, non-standard looking scenes, e.g. “African tree”.

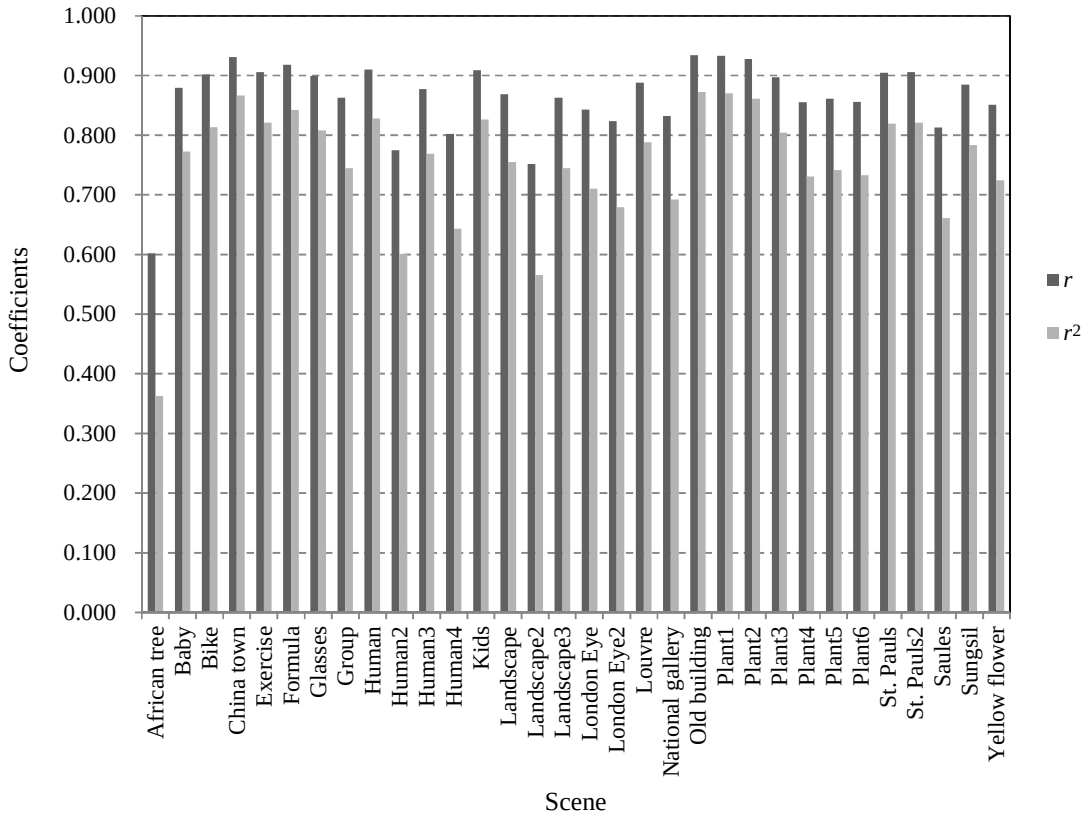


Figure 8-9. Evaluation of EPIC as an image quality prediction, r and r^2 coefficients, for thirty-two test scenes.

Image	*r*	r^2	Image	*r*	r^2	Image	*r*	r^2	Image	*r*	r^2
African tree	0.602	**0.362**	Baby	0.879	0.773	Bike	0.902	0.814	China town	0.931	0.867
Exercise	0.906	0.821	Formula	0.918	0.843	Glasses	0.899	0.808	Group	0.863	0.745
Human	0.910	0.828	Human2	0.775	0.601	Human3	0.877	0.769	Human4	0.802	0.643
Kids	0.909	0.826	Landscape	0.869	0.755	Landscape2	0.752	**0.566**	Landscape3	0.863	0.745
London Eye	0.843	0.711	London Eye2	0.824	0.679	Louvre	0.888	0.789	National gallery	0.832	0.692
Old building	0.934	0.872	Plant1	0.933	0.870	Plant2	0.928	0.861	Plant3	0.897	0.805
Plant4	0.855	0.731	Plant5	0.861	0.741	Plant6	0.856	0.733	St. Pauls	0.905	0.819
St. Pauls2	0.906	0.821	Saules	0.813	0.661	Sungsil	0.885	0.783	Yellow flower	0.851	0.724

Table 8-4. The *r* and r^2 coefficients for thirty-two scenes, compared between EPIC and perceived quality. The bold letters indicate the scene dependency in EPIC metric, “African tree” and “Landscape2.”

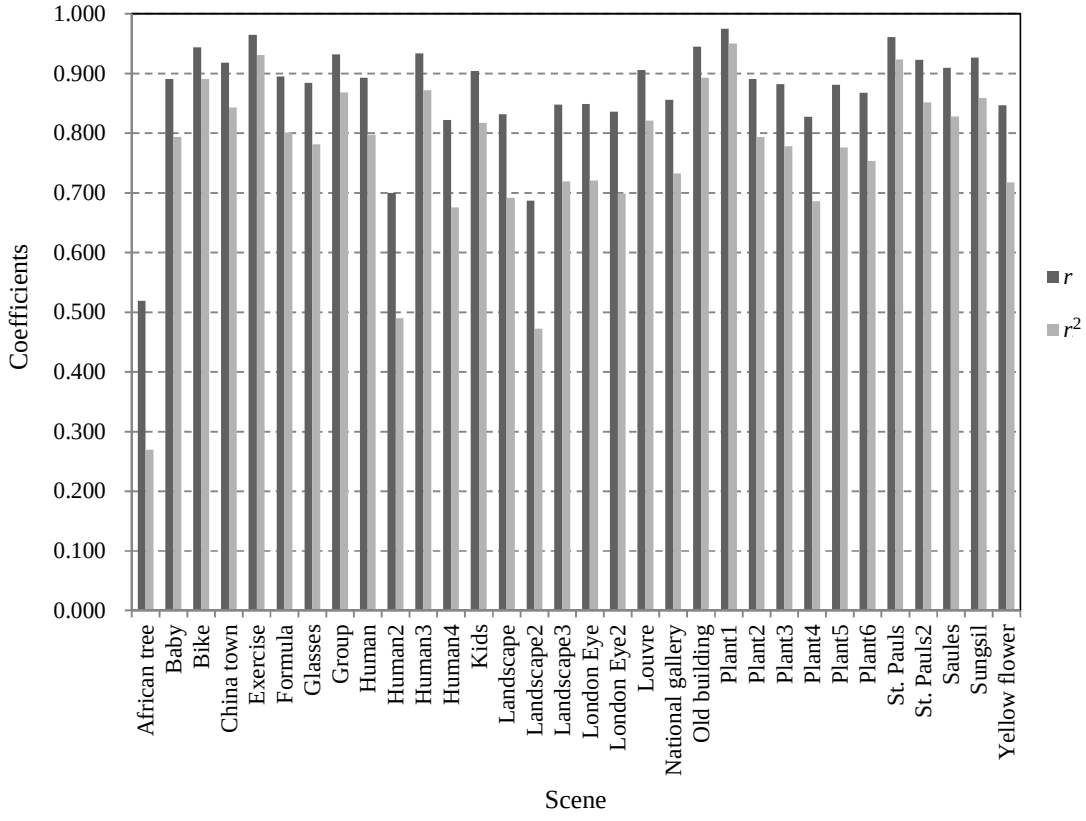


Figure 8-10. Evaluation of PIC as an image quality prediction, *r* and r^2 coefficients, for thirty-two test scenes.

Image	*r*	r^2	Image	*r*	r^2	Image	*r*	r^2	Image	*r*	r^2
African tree	**0.519**	**0.269**	Baby	0.891	0.794	Bike	0.944	0.891	China town	0.918	0.843
Exercise	0.965	0.931	Formula	0.895	0.801	Glasses	0.884	0.781	Group	0.932	0.869
Human	0.893	0.797	Human2	0.700	**0.490**	Human3	0.934	0.872	Human4	0.822	0.676
Kids	0.904	0.817	Landscape	0.832	0.692	Landscape2	0.687	**0.472**	Landscape3	0.848	0.719
London Eye	0.849	0.721	London Eye2	0.836	0.699	Louvre	0.906	0.821	National gallery	0.856	0.733
Old building	0.945	0.893	Plant1	0.975	0.951	Plant2	0.891	0.794	Plant3	0.882	0.778
Plant4	0.828	0.686	Plant5	0.881	0.776	Plant6	0.868	0.753	St. Pauls	0.961	0.924
St. Pauls2	0.923	0.852	Saules	0.910	0.828	Sungsil	0.927	0.859	Yellow flower	0.847	0.717

Table 8-5. The *r* and r^2 coefficients for thirty-two scenes, compared between PIC and perceived quality. The bold letters indicate the scene dependency in PIC metric, “African tree,” “Human2” and “Landscape2.”

8.3. EPIC and PIC with objective scene classification

In order to improve the reliability of the metrics, the objective scene classification model [10] (pp. 269) and the *integrated hyperbolic increment function* (IHIF) regression [223] (pp. 124) were considered. Figure 8-11 illustrates a modular image quality framework to tackle the problem of scene dependency in the metrics. There are three modules in the framework.

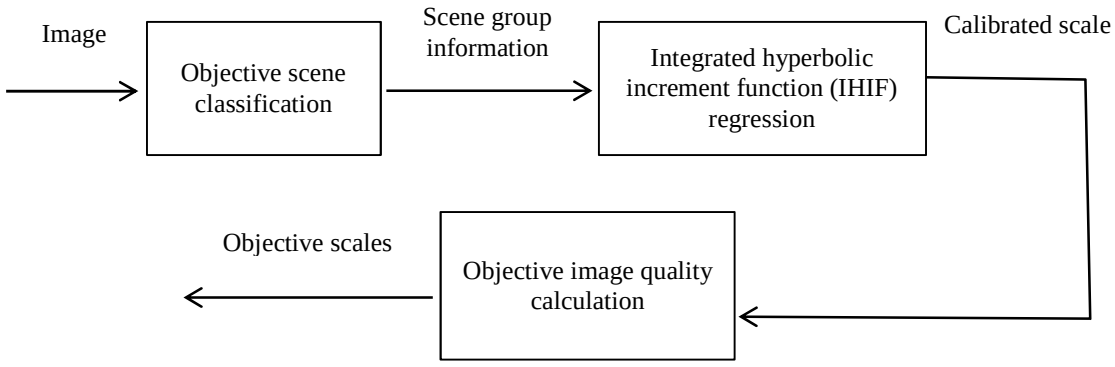


Figure 8-11. Basic stages of quality calculation within the objective scene classification and the IHIF regression.

1) Objective scene classification was firstly considered in order to tackle the problem of scene dependency in the metrics. Triantaphillidou *et al* [10] (pp. 269) proposed a method for scene classification, which uses objective scene descriptors that correlate with subjective criteria on scene susceptibility to image quality attributes. Scene descriptors are derived to describe inherent scene properties that human observers refer to when they judge the quality of images. This classification process enabled fully automatic classification of scenes into ‘standard’ or ‘non-standard’ groups, instead of being classified by inspection. The objective scene classification undertaken in this study is described in detail in chapter 7.

   All of the thirty-two scenes in this study were previously classified into three groups as follows: 1) low susceptibility to sharpness and high susceptibility to noisiness 2) average susceptibility to sharpness and noisiness, 3) high susceptibility to sharpness and low susceptibility to noisiness (cf. Figure 7-5).

2) This scene group information was used to attempt to improve predictions, using a fitting operation. According to Keelan [14] (pp. 147), “one way to characterize the variability associated with observer sensitivity and scene susceptibility is to classify scenes and observers into small numbers of groups and to form subsets of assessments based on combinations of the groups. The data from the different subsets can be separately fit with integrated hyperbolic increment function (IHIF) regressions. ”

   The IHIF [223] (pp. 124) was employed to account for the quality changes arising for scene susceptibility. The reason for using the IHIF is to quantify the image quality automatically and to obtain reliable results that correlate well with subjective assessments, i.e. the metric produces immediate results without human involvement [210] (pp. 289).

   The IHIF value was calculated with free parameters (Table 8-6) [208] (pp. 204). The IHIF value ranged between 0.73 and 1.36, i.e. for the intermediately susceptible group (i.e. middle), the IHIF value was approximately equal to 1.0; for the least susceptible group (i.e. least susceptibility), the IHIF value was 0.73; for the most susceptible group (i.e. most susceptibility), the IHIF value was 1.36.

$$\Delta Q(\Omega) = \frac{R_r}{\Delta\Omega_\infty^2} \ln\left(1 + \frac{\Delta\Omega_\infty(\Omega - \Omega_r)}{R_r}\right) - \frac{\Omega - \Omega_r}{\Delta\Omega_\infty} \quad (8.11)$$

where ΔQ is quality difference, Ω is objective metric (units variable), Ω_r is objective metric at reference threshold position, $\Delta\Omega_\infty$is asymptotic objective metric just noticeable difference (JND) increment and R_r is radius of curvature of quality loss function at Ω_r [223] (pp. 124 and 471).

Observer Sensitivity	**Scene Susceptibility**	$\Delta\Omega_\infty$	Ω_r	Ω_{-1}
Less Sensitivity 50%	Least Susceptibility 25%	5.500	20.76	39.9
Less Sensitivity 50%	Mean	4.364	16.43	36.7
Less Sensitivity 50%	Middle 50%	4.078	17.11	37.2
Less Sensitivity 50%	Most Susceptibility 25%	3.980	11.94	34.4
Mean	Least Susceptibility 25%	5.298	18.72	36.3
Mean	Mean	3.902	15.07	35.2
Mean	Middle 50%	3.708	15.81	35.5
Mean	Most Susceptibility 25%	3.256	10.95	34.3
More Sensitivity 50%	Least Susceptibility 25%	5.088	16.62	33.7
More Sensitivity 50%	Mean	3.470	13.70	33.7
More Sensitivity 50%	Middle 50%	3.351	14.50	33.8
More Sensitivity 50%	Most Susceptibility 25%	2.654	9.81	35.9

Table 8-6. IHIF fit parameters to predict quality change arising from an attribute, in all cases R_r=151.6. The quantity Ω_{-1} is the objective metric value Ω at which one just noticeable difference (JND) of quality loss occurs
(produced by Keelan [208] (pp. 204)).

The IHIF value was shown to relatively match the perceived quality loss variables for scene susceptibility for the three groups (Table 8-7). Thus, it was demonstrated that the IHIF value could be used to calibrate scenes for automatic calculation of metric values, i.e. Pearson's correlation coefficients (r) were 0.999 for noisiness and 0.969 at blur/sharpness for the three groups. The scene susceptibility of the three groups was also valid for the weighting values. The scene susceptibility to noisiness or blur/sharpness was averaged from the perceived scene susceptibility parameters (cf. Table 6-4) in three groups (cf. Figure 7-5).

Scene Susceptibility	IHIF value	Scene susceptibility to noisiness	Scene susceptibility to blur/sharpness
Least Susceptibility	0.73	0.72	0.81
Middle	0.98	0.97	1.05
Most Susceptibility	1.36	1.25	1.15

Table 8-7. IHIF values and scene susceptibility to noisiness or blur/sharpness in three groups.

The IHIF calibration was considered only for noise, since it is implemented for each attribute separately and it was shown that the relationship between noisiness and sharpness was complimentary (cf. Section 6.5). Therefore, a high IHIF value represents a high susceptibility to noisiness and low susceptibility to sharpness/blur.

An example of the use of the IHIF value for calibrating noise in group 1 is illustrated in Table 8-8. The calibrated noise scale in the last column is obtained by multiplying the noise scale with the IHIF value.

Varying Attribute	σ values of the Gaussian noise kernel	IHIF value	Calibrated σ values of the Gaussian noise kernel in Group 1
Noise1	0	1.36	0
Noise2	0.1		0.136
Noise3	0.2		0.272

Table 8-8. Example of using IHIF for calibrating noise.

3) The calibrated scale was then used to improve the reliability of the metrics. The calculation was the same as that of the EPIC and PIC metrics, as set out in section 8.1.

The EPIC and PIC scales were normalised and rescaled as percentages, since a scale with the maximum unit of 100 is easy to understand. 100 is set as the original quality, which means perceived absence for artefacts.

The normalisation was applied here as a linear process. It could reflect the nature of the variation (other than scene content) among the test stimuli [45] (pp. 17). Osberger [224] (pp. 21) has noted that “care must be taken in assuming linearity when applying psychophysical results from tests using simple, artificial stimuli to complex, natural images,” since “the HVS is highly adaptive to the diverse and complex range of stimuli which it deals with in the natural world.”

Figure 8-12 and Figure 8-13 present the calibrated EPIC and PIC values for the three groups. Each label on the x-axis represents a specific level of distortion in blurring (B) and noise (N).

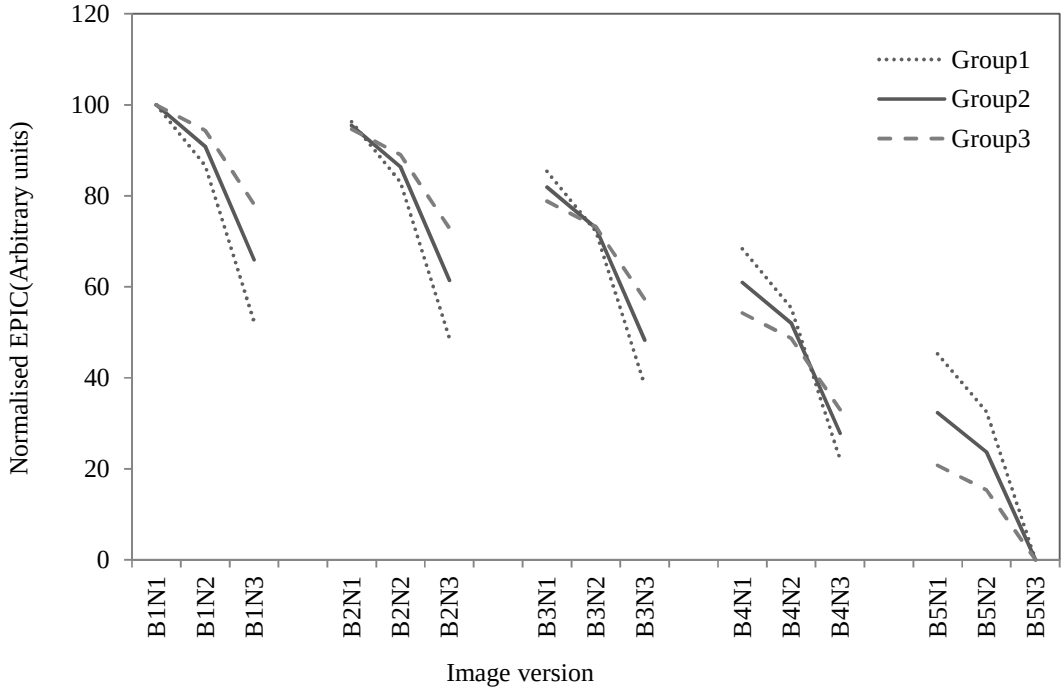


Figure 8-12. Calibrated EPIC values in three clusters (groups).

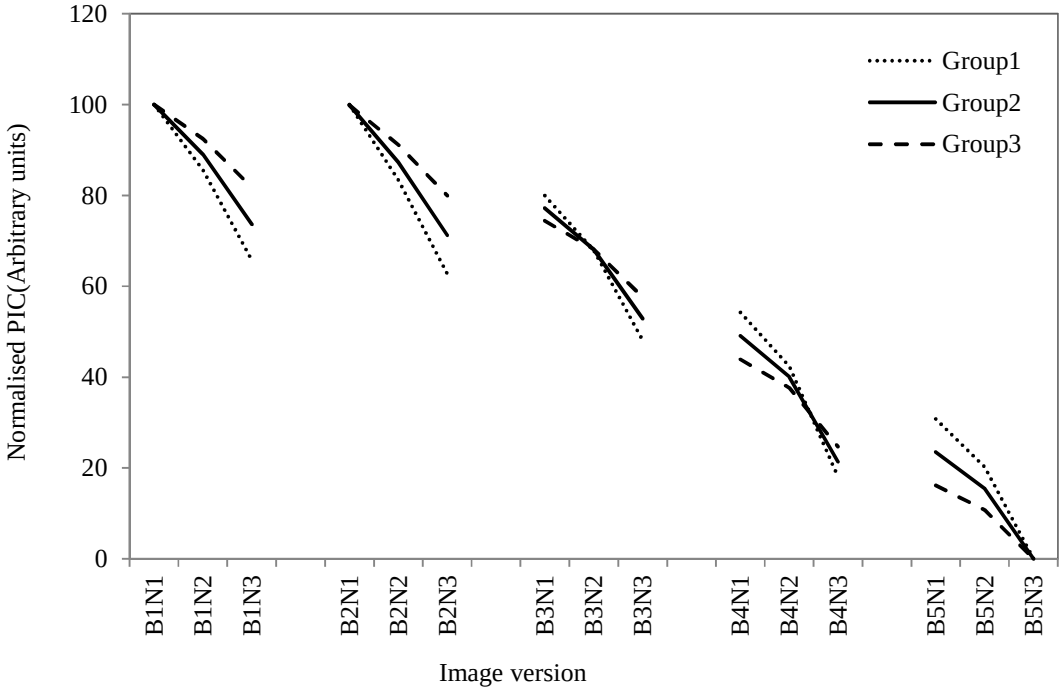


Figure 8-13. Calibrated PIC values in three clusters (groups).

8.4. Calibrated EPIC and PIC evaluation

The success of the implementation was evaluated by correlating the subjective scaled values from section 6.5.1, and the calibrated quality scales obtained in section 8.3. The Pearson's correlation coefficient (r) and the coefficient of determination (r^2) were again used to examine the success of the correlation.

Figure 8-14 and Figure 8-15 show the r and r^2 coefficients for the thirty-two test scenes. Table 8-9 and Table 8-10 list the detailed results. Results from the implementation showed that generally the quality predictions were improved. The average of 32 coefficients increased from 0.865 (r^2=0.748) to 0.882 (r^2=0.778) in EPIC and from 0.870 (r^2=0.757) to 0.886 (r^2=0.785) in PIC.

Improvements were particularly shown in scenes belonging to group 1, with low susceptibility to sharpness and high susceptibility to noisiness. For example, the coefficients at "African tree" scene increased from 0.602 (r^2=0.362) to 0.740 (r^2=0.548) in EPIC and from 0.519 (r^2=0.269) to 0.618 (r^2=0.382) in PIC. In the "Human 2" scene, the coefficients increased from 0.775 (r^2=0.601) to 0.874 (r^2=0.764) in EPIC and from 0.700 (r^2=0.490) to 0.782 (r^2=0.612) in PIC. The coefficients at "Landscape2" scene increased from 0.752 (r^2=0.566) to 0.838 (r^2=0.702) in EPIC and from 0.687 (r^2=0.472) to 0.763 (r^2=0.582) in PIC.

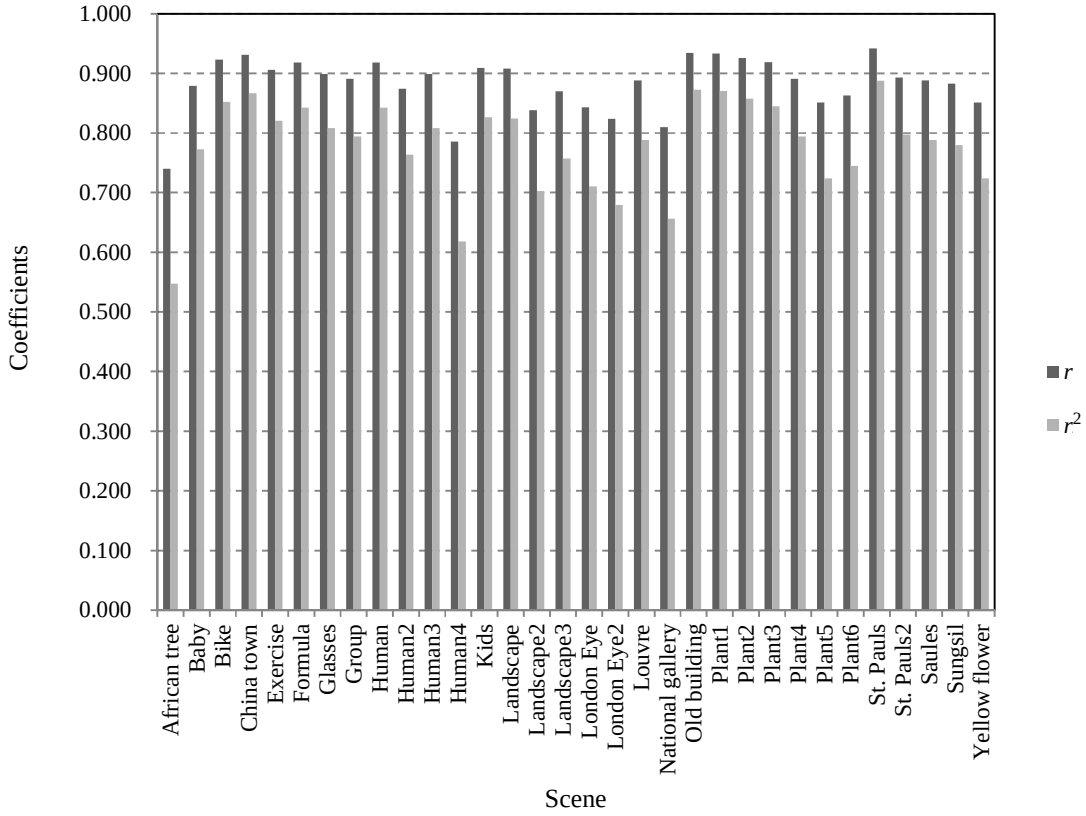


Figure 8-14. Evaluation of calibrated EPIC as an image quality prediction, r and r^2 coefficients, for thirty-two test scenes in three groups.

	Image	***r***	***r²***	**Image**	***r***	***r²***	**Image**	***r***	***r²***
Group 1	African tree	0.740	**0.548**	Human	0.918	0.843	Human2	0.874	0.764
	Human4	0.786	0.618	Landscape	0.908	0.824	Landscape 2	0.838	0.702
	Landscape3	0.870	0.757	Plant2	0.926	0.857	Plant3	0.919	0.845
	Plant4	0.891	0.794	Plant5	0.851	0.724	Plant6	0.863	0.745
Group 2	Baby	0.879	0.773	China town	0.931	0.867	Exercise	0.906	0.821
	Formula	0.918	0.843	Glasses	0.899	0.808	Kids	0.909	0.826
	London Eye	0.843	0.711	London Eye2	0.824	0.679	Louvre	0.888	0.789
	Old building	0.934	0.872	Plant1	0.933	0.870	Yellow flower	0.851	0.724
Group 3	Bike	0.923	0.852	Group	0.891	0.794	Human3	0.899	0.808
	National gallery	0.810	0.656	St. Pauls	0.942	0.887	St. Pauls2	0.893	0.797
	Saules	0.888	0.789	Sungsil	0.883	0.780			

Table 8-9. The r and r^2 coefficients for thirty-two scenes in three groups, compared between calibrated EPIC and perceived quality. The bold letter indicates the scene dependency in calibrated EPIC metric, “African tree.”

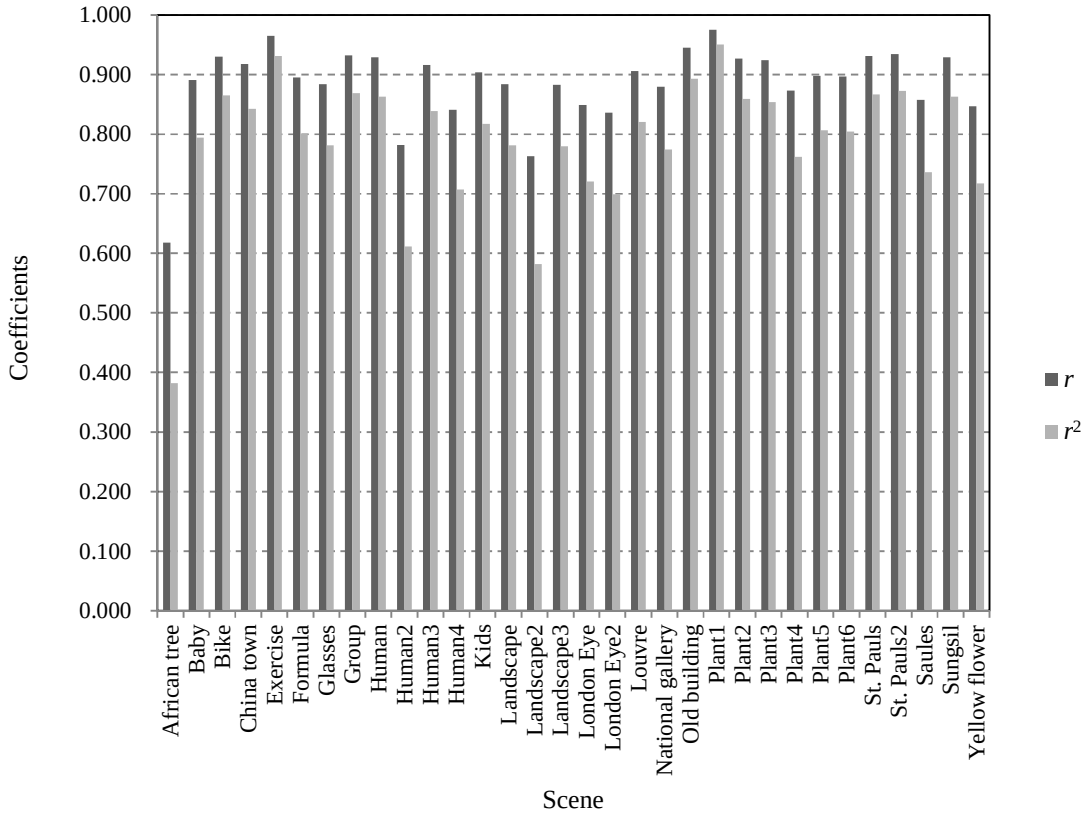


Figure 8-15. Evaluation of calibrated PIC as an image quality prediction, r and r^2 coefficients, for thirty-two test scenes in three groups.

	Image	***r***	***r²***	**Image**	***r***	***r²***	**Image**	***r***	***r²***
Group 1	African tree	0.618	**0.382**	Human	0.929	0.863	Human2	0.782	0.612
	Human4	0.841	0.707	Landscape	0.884	0.781	Landscape 2	0.763	0.582
	Landscape3	0.883	0.780	Plant2	0.927	0.859	Plant3	0.924	0.854
	Plant4	0.873	0.762	Plant5	0.898	0.806	Plant6	0.897	0.805
Group 2	Baby	0.891	0.794	China town	0.918	0.843	Exercise	0.965	0.931
	Formula	0.895	0.801	Glasses	0.884	0.781	Kids	0.904	0.817
	London Eye	0.849	0.721	London Eye2	0.836	0.699	Louvre	0.906	0.821
	Old building	0.945	0.893	Plant1	0.975	0.951	Yellow flower	0.847	0.717
Group 3	Bike	0.930	0.865	Group	0.932	0.869	Human3	0.916	0.839
	National gallery	0.880	0.774	St. Pauls	0.931	0.867	St. Pauls2	0.934	0.872
	Saules	0.858	0.736	Sungsil	0.929	0.863			

Table 8-10. The r and r^2 coefficients for thirty-two scenes in three groups, compared between calibrated PIC and perceived quality. The bold letter indicates the scene dependency in calibrated PIC metric, “African tree.”

In total, objective scene classification was considered in order to tackle the problem of scene dependency in the metrics. This implementation allows automatic grouping of scenes into ‘standard’ or ‘non-standard’ groups, so that it allows automatic calculation of the metric values.

Results from the implementation showed that generally the quality predictions were improved. Most importantly, they were shown to correlate equally well with subjective quality scales for standard and non-standard scenes.

In addition, more extensive investigations into objective scene classification, e.g. optimal clustering numbers (k=4 or 5), could help to provide an even better metric calibrations for objective image quality predictions. However, an extensive number of clusters or a continuous fitting of all scenes might not be required, since there may not be any significant advantage with regard to accurate image quality predictions, i.e. the metrics already performed well on most of the scenes, and it is a slight increase in correlation for some scenes. For example, the coefficients at “Plant1” scene had a slight increase from 0.933 (r^2=0.870) to 0.944 (r^2=0.851) in EPIC, using 1.15 scene susceptibility. Using a large number of clusters could make the calculations time-consuming.

8.5. Summary

A validation experiment was first carried out to test the success of the metrics; EPIC and PIC, using results from subjective tests that involving the thirty-two test scenes replicated with various degrees of sharpness and noisiness (cf. Chapter 6). In general, the metrics were found to be reliable predictors of image quality. However, as expected of device-dependent metrics, they were less successful at predicting the perceived quality of some 'non-standard' scenes which had atypical spatial and structural content.

Objective scene classification was considered in order to tackle the problem of scene dependency in the metrics. The scene classification employed for the purpose used objective scene descriptors, which correlated with subjective criteria on scene susceptibility (cf. Chapter7). The implementation thus allowed the automatic grouping of scenes into 'standard' or 'non-standard' groups, instead of being classified by inspection. From this scene group information, the quality change arising for an attribute in the scene group was obtained using the integrated hyperbolic increment function (IHIF) regression, as described by Keelan.

The classification and metric calibration performance was quite encouraging, not only because it improved mean image quality predictions from all scenes, but also because it catered for non-standard scenes, which originally produced low correlations. Also, it is because of the efficiency of computation. The findings indicate that the scene classification method has a great potential for tackling the problem of scene dependency, when modelling device-dependent image quality.

Chapter 9

Discussion

In this chapter, a discussion of the results from the practical work conducted during the investigation is presented. The evaluations of the input (camera) and the output (LCD) system characterisation are firstly discussed. A detailed discussion on the results from the subjective tests follows. Finally, this chapter presents a discussion of EPIC and PIC as quality predictors.

9.1. Systems characterisation and calibration

Both input and output systems were characterised for the aspects of image quality attributes, i.e. tone, colour, resolution, sharpness and noise. The characterisation provided a means for producing accurate and reproducible results for image quality investigations. It also determined the limitations of the devices for image quality investigations.

Digital camera system

The Canon EOS-1Ds camera system was identified as having limitations with regard to spatial uniformity. The top-right corner of the capturing frame was shown to be the least uniform. The result was greater than the perceptibility colour difference in complex scenes, 3.00 in ΔE^*_{ab} [160]. Therefore, the rest of the investigations were carried out using the central area of the image.

The tone and colour reproduction results demonstrated variation with respect to the camera colour settings in the process profiles. The results showed that a reproduced image's tone and colour were affected by the camera colour settings in the process profiles, and the colour setting was influential in optimising colour images. An sRGB setting of 3 was used to optimise colour images, i.e. the smallest colour difference between original *XYZ* and calculated *XYZ* was achieved with the sRGB colour setting of 3.

The colour reproduction also demonstrated that there was a lack of colour accuracy between scene and image. Thus, the custom profile was applied to a polynomial regression model, which was able to provide equivalent colour reproduction with high polynomial terms. However, there is a significant scientific challenge in calibrating the camera's colour reproduction in relation to real-life photography. The process is complicated and time-consuming. Further extensive investigations are required for accurate colour reproduction in the real scene images.

The camera's MTF was measured for its ability to depict picture details. The camera's 10% MTF values were 58.5 cycles/mm in the horizontal and 61.3 cycles/mm in the vertical. The camera may depict picture detail up to those frequencies.

The camera's signal-to-noise ratio was 47:1. The camera's noise characterisation was satisfactory, thus this result led to the choice of a high-performance camera for the LCD noise characterisation.

LCD

There was a slight variation in the luminance on the Eizo CG210 LCD temporal characteristic. This change in luminance, however, was not perceptible to the human eye. The LCD did not require significant warm up time in order to reach stabilisation.

The LCD system showed limitations with regard to spatial uniformity. The actual fluctuations of the 24 areas ranged from 0.73 to 4.80 in ΔE^*_{ab} and from 0.62 to 3.45 in ΔE_{00}. The lower-middle LCD areas showed the most non-uniformity, 4.80 in ΔE^*_{ab}. The result was greater than the perceptibility colour difference in complex scenes, 3.00 in ΔE^*_{ab} [160]. The result showed that a reproduced image was affected by the displayed

spatial point. This led to the choice of the LCD's central area for the psychophysical investigation of image quality.

The LCD had a limited viewing angle. This characteristic was of considerable significance in this study, where a differing viewing angle significantly influences the assessment of subjective image quality. It was also confirmed that the vertical viewing angle had more impact than the horizontal. The LCD required significant careful viewing angle control during the image quality study.

There was a variation in colour reproduction on the LCD. The average colour difference value was 1.57 in ΔE^*_{ab} and 0.52 in $\Delta E_{00.}$ The LCD's reproduction of colours was acceptable for the psychophysical scaling.

In the tone reproduction investigations, the optimal overall gamma in the imaging chain, camera-LCD, was confirmed. The optimal overall gamma was 1.32 from 2.20 and 0.60 in the LCD and the camera, respectively. It was identified that these settings were intentionally rendered by the manufacture based on the observer's preference, i.e. the optimal overall gamma was generally between 1.0 and 1.5 [83] (pp. 379).

The ability of the LCD system to depict picture detail was measured. The LCD 10% MTF values were 1.93 cycles/mm in the horizontal and 1.73 cycles/mm in the vertical. The LCD 50% MTF values were 1.14 cycles/mm in the horizontal and 0.95 cycles/mm in the vertical. The LCD MTF result was used in the calculation of the objective IQM.

The noise of the LCD was measured with the aid of a high-performance digital camera. The signal-to-noise ratio of the LCD was 45:1. The LCD outcome was significant, when compared to the previous results and ranged from 31.5:1 to 63.5:1 [195] (pp. 166), [196] (pp. 6). The LCD noise result was used in the calculation of the objective IQM.

Two points were emphasised by the characterisation of the systems. First, there was the effect of imaging system characteristics on the image quality, which was a crucial factor in the image quality. Second, imaging systems have inherent limitations in reproducing the original scene. This should be a tremendous challenge for imaging system

development, since successful systems often confine the inevitable error to those that are least noticeable visually [226] (pp. 436).

9.2. Subjective image quality evaluation

A novel subjective scaling method was carried out in order to derive individual attribute scales from the overall image quality assessments. This approach does not require scaling of individual attributes and does not require the assumption that the attribute is one-dimensional, i.e. observers can see the quality attribute independently.

In this scaling study, the image quality is considered as an overall impression, not individual attribute scaling. Although the individual attribute scaling is commonly used, it provides a limited correlation with actual perceived image quality, and therefore it is logical to consider image quality as an overall. Engeldrum [11] (pp. 22-23) has noted that it is necessary to avoid pitfall of individual scaling and to take much care of quality studies.

Furthermore, the results showed that the five-ranges of distortion introduced by blurring covered the entire image quality scale. However, the three-levels of added Gaussian noise were too small for investigating the consequences in the full range of image quality.

There was interesting a trade-off between noise and blur. High amounts of blurring in the image significantly decreased the perception of noise, and high noise in the image decreased the perceived blur.

Relationship between scene dependency parameters and scene descriptors

This work was carried out to investigate the degree of correlation between scene descriptors and scene dependency parameters. A number of scene descriptors were derived from second-order statistical measures and measurement from edge detection.

The outcomes showed that there was a successful correlation between the sharpness and noisiness susceptibility parameters and the spatial frequency properties of subjects, i.e.

second-order statistical measures, as well as measures derived from edge detection. The higher the frequency in the scene content, the lower the susceptibility to noisiness and the higher the susceptibility to sharpness. The result was in line with results from previous studies [10] (pp. 261), [16] (pp. 663).

Three points were emphasised by the subjective image quality evaluation. First, the perceived subjective quality is dependent upon the pictorial content of the test images. Second, the spatial frequency properties of subjects are pivotal scene dependence parameters in the observer's preferences. Third, there is a correlation between the spatial frequency properties of subjects and the observer's scene susceptibility to sharpness and noisiness.

9.3. Objective image quality evaluation

Imaging system characterisation and subjective image quality evaluation have repeatedly shown that the perceived quality is dependent both upon the device and scene characteristics.

Subjective overall image quality was predicted using two device-dependent IQMs, EPIC [74] and PIC [71]. The metrics were found to be reliable predictors of image quality. However, they were not equally successful in predicting the quality of different images with varying scene content. This was in line with findings from previous researches [71] (pp. 22), [74] (pp. 7). Device-dependent image quality metrics tend to perform well on standard looking scenes, but perform less well on some of non-standard looking scenes with atypical spatial and structural content: for example, scenes of very low busyness with many flat areas [10] (pp. 261). The problem of scene dependency when employing device-dependent image quality metrics was confirmed.

Improvements to the performance of the metrics were then considered using objective scene classification. The objective scene classification employed for the purpose utilised objective scene descriptors, which correlated with subjective criteria on scene

susceptibility. This process enabled fully automatic classification of scenes into ‘standard’ or ‘non-standard’ groups, instead of being classified by inspection. This objective scene classification was highlighted as one of the key areas in this work.

A validation experiment was then carried out to test the improvement in the metrics. Results from the implementation showed that the quality predictions were improved successfully, after according for individual scene classification. Most importantly, the metric scales were shown to correlate equally well with subjective quality scales of standard and non-standard looking scenes.

The classification and metric calibration performance was quite encouraging, not only because it improved mean image quality predictions from all scenes, but also because it catered for non-standard scenes, which originally produced low correlations. Also, it is because of the efficiency of computation. The findings indicate that the automatic scene classification method has a great potential for tackling the problem of scene dependency, when modelling device-dependent image quality. It can provide a tool for fully automatic and quick derivation of image quality rating.

This is a practical consideration of scene classification with respect to image quality, a possible way of overcoming the problem of scene dependency in image quality modelling, suggested by Keelan [14] (pp. 147). Based on this suggestion, this work is further expended in scene classification, which uses objective *scene descriptors* that correlate with subjective criteria on *scene susceptibility* to image quality attributes [10] (pp. 269).

This finding could form the basis for a new direction of research, simultaneously predicting perception according to imaging system parameters, observer and scene properties, taking into account classification.

9.4. Fundamental comments

The scientific originalities of this research are summarised as follows:

- The provision of an experimental paradigm for measuring the effect of scene content on image quality with respect to sharpness and noisiness.

- The development of a methodology in subjective image quality assessment, in order to derive individual attribute scales from the overall image quality assessments. This approach does not require scaling of individual attributes and does not require the assumption that the attribute is one-dimensional, i.e. observers can see the quality attribute independently. The method might be ideal for investigating the relationship between attributes.
- The demonstration that the specific, selected scene properties correlate within perceived image quality of sharpness and noisiness and the derivation of various scene descriptors for scene properties.
- The application of the objective scene classification to resolve the problem of scene dependency in device-dependent image quality metrics. This allows a quick and simple objective quality measurement utilizing imaging system characteristics and scene properties.

A successful objective scaling study, based on the subjective image quality scaling and the system properties, could provide a quick and easy to implement method for the evaluation of image quality in industrial laboratories. It could also assist fundamental academic research concerning the evaluation of image quality, such as algorithm development and imaging system development.

Chapter 10

Conclusions & Recommendations for further work

10.1. Conclusions

The following conclusions are drawn from the research conducted in this thesis.

- Perceived image quality is dependent upon the imaging system characteristics.

  Characterisation of the imaging system showed that the perception of image quality was dependent on the imaging system. For example, the tone and colour reproduction results were demonstrated to have variation with respect to the camera colour settings in the process profiles.

- There are limitations on imaging systems with regard to reproducing scenes.

  It was demonstrated that there was a lack of colour accuracy in the reproduction of colour between scene and image. The MTF and noise measurements also demonstrated the same finding.

- Perceived image quality is also dependent upon the scene properties.

  Psychophysical scaling was used to demonstrate that the perception of image quality was dependent on scene properties. In addition, successful scene dependency quantification was found.

- There are relationships between sharpness/blur and noise.

High amounts of blurring in an image were shown to significantly decrease the perception of noise, and high noise was shown to decrease the perceived blur.

- Spatial frequency properties of subjects are crucial scene dependence parameters and influence observers' perception of image quality in blurred and noisy images.

  It was demonstrated that there was a correlation between the sharpness and noisiness susceptibility parameters and the spatial frequency properties of subjects.

- Device-dependent metrics are unequally successful in predicting the quality of different images with varying scene content.

  It was found that two device-dependent image quality metrics, Effective Pictorial Information Capacity (EPIC) and Perceived Information Capacity (PIC), predicted less successfully the perceived quality of non-standard scenes with atypical spatial and structural content.

- An effective device-dependent metric is required to predict human perception of image quality, based on both imaging system properties, as well as scene properties.

- Scene classification has the potential for tackling the problem of sharpness and noisiness scene susceptibility when modelling device-dependent image quality.

  The classification and metric calibration performance was quite encouraging, not only because it improved mean image quality predictions from all scenes, but also because it catered for non-standard scenes, which originally produced low correlations.

10.2. Recommendations for further work

Several recommendations for further work in this area can be made.

- Extensive investigations of the standard conditions for the camera would be required in order to achieve an accurate colour reproduction. Fairchild *et al* [227] (pp. 1) demonstrated that careful and simple characterisation of digital SLR cameras could result in visually equivalent colour reproduction, 4.0 and 6.9 in ΔE^*_{ab}.

- Extensive investigations of various conditions would be required by custom profile (polynomial regression model) for the accuracy of colour reproduction, not only illumination in D65. Possible implementation in the camera is in common illuminations, such as tungsten light (3200 K), fluorescent light (4500 K), electronic flash (5000 K) and overcast daylight (7000 K).
- More investigations of MTF associated with image down-sampling are required at different MTF methods and a wide range of down-sampling levels.
- More information is required on the MTF and noise characteristics of the display device. This is necessary for the improvement of the image quality metric. Furthermore, the accurate characterisation of the device would assist in calibrating the system, such as noise reduction.
- More serious effort could be put into quantifying a JND. The *method of limits* could be utilised to provide more precise threshold data than the method of adjustment [200] (pp. 130-131).
- Extending the range of distortion would be necessary in future work, since the range of added Gaussian noise was too small to investigate the consequences on the full range of image quality, e.g. 5 or 7 JNDs [200] (pp. 131) in added Gaussian noise.
- The quality ruler method could be used to further study subjective quality scaling, since the implementation includes the scene-dependent ruler calibration [17] (pp. 11).
- Further work is required into identification, selection and classification of scenes, as this is an unresolved challenge within the science community, until now [10] (pp. 269).
- More extensive investigations of scene descriptors would help to understand scene properties. The combination of various scene descriptors, e.g. the integration of colour–texture descriptors [228], may describe more successfully the susceptibility of test scenes to noisiness and sharpness.
- Further investigation is required to derive scene descriptors from regions of interest within the image (via local application of the algorithm). A possible example of this is

a using the central part of the image as a sharpness-critical region and the periphery of the image as a noisiness-critical region [229] (pp. 5).

- More consideration is given to the determination of the number of clusters in a data set. This is a distinct issue from the process of solving the clustering problem. A possible way of dealing with this issue is to use a mathematical approach, such as a rule of thumb [149], an elbow method or an information criterion approach.
- Different clustering algorithms could be used to apply objective scene classification [147, 148], such as hierarchical clustering and Fuzzy c-means clustering. More investigations of objective scene classification could assist with better grouping of scene susceptibility and lead to better objective image quality investigations.
- It is hoped that, in the near future, an image-dependent and device-dependent metric can be created that takes into account colour and spatial characteristics. Topfer [225] (pp. 303) noted that “preference is an important element in the evaluation of the effects of colour and tone on image quality. Despite the preferential aspect of these attributes, the framework of image quality modelling developed for artifactual attributes is still applicable, with certain extensions.”

Appendix A

sRGB encoding transformation

The standard RGB colour space encoding, sRGB, was originally designed by HP and Microsoft as the default colour space encoding for the Internet [89]. The sRGB is based on a typical CRT display primaries and transfer function. Table A-1 lists the CIE 1931 chromaticity coordinates of the sRGB reference primaries. In addition, Table A-2 shows the reference display, the reference viewing condition and the reference observer.

	Red	**Green**	**Blue**
x	0.6400	0.3000	0.1500
y	0.3300	0.6000	0.0600
z	0.0300	0.1000	0.7900

Table A-1. The CIE 1931 chromaticity coordinates for primaries.

Reference display	
Luminance level	80 cd/m^2
White point	CIE D65 (x=0.3127, y=0.3291)
Model offset(R, G, B)	0.055
Model gamma(R, G, B)	2.2
Reference viewing condition	
Background- surrounding the image	20% of the reference display white
Surround- area surrounding the display	20% reflectance
Proximal field	20% of the reference display white
Ambient illuminance level	64 lx
Ambient white point	x=0.3457, y=0.3585 (D50)
Viewing gamma	1.125
Reference observer	
Observer	CIE 1931 2° standard observer

Table A-2. sRGB reference display, reference viewing condition and reference observer.

The sRGB encoding transformation is described briefly as follows [89]:

1) Transform from *XYZ* to RGB

$$\begin{bmatrix} R_{\text{sRGB}} \\ G_{\text{sRGB}} \\ B_{\text{sRGB}} \end{bmatrix} = \begin{bmatrix} 3.2406 & -1.5372 & -0.4986 \\ -0.9689 & 1.8758 & 0.0415 \\ 0.0557 & -0.2040 & 1.0570 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{A.1}$$

if $R_{\text{sRGB}}, G_{\text{sRGB}}, B_{\text{sRGB}} \leq 0.0031308$

$$\begin{aligned} R'_{\text{sRGB}} &= R_{\text{sRGB}} \times 12.92 \\ G'_{\text{sRGB}} &= G_{\text{sRGB}} \times 12.92 \\ B'_{\text{sRGB}} &= B_{\text{sRGB}} \times 12.92 \end{aligned} \tag{A.2}$$

or if $R_{\text{sRGB}}, G_{\text{sRGB}}, B_{\text{sRGB}} > 0.0031308$

$$\begin{aligned} R'_{\text{sRGB}} &= 1.055 \times R_{\text{sRGB}}^{(1.0/2.4)} - 0.0055 \\ G'_{\text{sRGB}} &= 1.055 \times G_{\text{sRGB}}^{(1.0/2.4)} - 0.0055 \\ B'_{\text{sRGB}} &= 1.055 \times B_{\text{sRGB}}^{(1.0/2.4)} - 0.0055 \end{aligned} \tag{A.3}$$

and

$$\begin{aligned} R_{\text{8bit}} &= round(((WDC - BDC) \times R'_{\text{sRGB}}) + KDC) \\ G_{\text{8bit}} &= round(((WDC - BDC) \times G'_{\text{sRGB}}) + KDC) \\ B_{\text{8bit}} &= round(((WDC - BDC) \times B'_{\text{sRGB}}) + KDC) \end{aligned} \tag{A.4}$$

2) Transform from RGB to *XYZ*

$$\begin{aligned} R'_{\text{sRGB}} &= (R_{\text{8bit}} - BDC)/(WDC - BDC) \\ G'_{\text{sRGB}} &= (G_{\text{8bit}} - BDC)/(WDC - BDC) \\ B'_{\text{sRGB}} &= (B_{\text{8bit}} - BDC)/(WDC - BDC) \end{aligned} \tag{A.5}$$

BDC is a black digital count, and WDC is a white digital count (8 bits/channel).

if $R'_{\text{sRGB}}G'_{\text{sRGB}}B'_{\text{sRGB}} \leq 0.04045$

$$\begin{aligned} R_{\text{sRGB}} &= R'_{\text{sRGB}} \div 12.92 \\ G_{\text{sRGB}} &= G'_{\text{sRGB}} \div 12.92 \\ B_{\text{sRGB}} &= B'_{\text{sRGB}} \div 12.92 \end{aligned} \tag{A.6}$$

or if $R'_{\text{sRGB}}G'_{\text{sRGB}}B'_{\text{sRGB}} > 0.04045$

$$\begin{aligned} R_{\text{sRGB}} &= \left[(R'_{\text{sRGB}} + 0.0055) \Big/ 1.055 \right]^{2.4} \\ G_{\text{sRGB}} &= \left[(G'_{\text{sRGB}} + 0.0055) \Big/ 1.055 \right]^{2.4} \\ B_{\text{sRGB}} &= \left[(B'_{\text{sRGB}} + 0.0055) \Big/ 1.055 \right]^{2.4} \end{aligned} \tag{A.7}$$

and

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R_{\text{sRGB}} \\ G_{\text{sRGB}} \\ B_{\text{sRGB}} \end{bmatrix} \tag{A.8}$$

Adobe RGB encoding transformation

The Adobe RGB colour space encoding was developed by Adobe Systems, as an RGB working space suited for print production [90]. The Adobe RGB has been used widely by the photographic industry. The CIE 1931 chromaticity coordinates of the green primary are slightly different to those of the sRGB (Table A-3).

	Red	**Green**	**Blue**
x	0.6400	0.2100	0.1500
y	0.3300	0.7100	0.0600
z	0.0300	0.0800	0.7900

Table A-3. The CIE 1931 chromaticity coordinates for Adobe RGB primaries.

The reference display, the reference viewing condition and the reference observer are shown below Table A-4.

Reference display	
Luminance level	160 cd/m^2
White point	CIE D65 (x=0.3127, y=0.3291)
Black point	X_K =0.5282, Y_K = 0.5557, Z_K=0.6052
Reference viewing condition	
Surround- area surrounding the display	20% reflectance
Ambient illuminance level	32 lx
Ambient white point	CIE D65 (x=0.3127, y=0.3291)
Reference observer	
Observer	CIE 1931 2° standard observer

Table A-4. Adobe RGB reference display, reference viewing condition and reference observer.

The Adobe RGB encoding transformation between 8-bit RGB values and CIE *XYZ* values involves the following steps [90]:

1) Normalisation of the RGB pixel values to a colour space value ranging from 0.0 to 1.0.

2) Linearization of the RGB signal using the reference display transfer function.

$$\begin{aligned} R' &= (R_{8bit}/255)^{2.199} \\ G' &= (G_{8bit}/255)^{2.199} \\ B' &= (B_{8bit}/255)^{2.199} \end{aligned} \tag{A.9}$$

3) Transformation (i.e. 3 x 3 matrix multiplication) from linear RGB to *XYZ*.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.57667 & 0.18556 & 0.18823 \\ 0.29735 & 0.62736 & 0.07529 \\ 0.02703 & 0.07069 & 0.99133 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \tag{A.10}$$

The inverse transformation is obtained by reversing steps 1-3.

Appendix B

The following publications and presentations produced by the author during this project. The refereed and published conference proceedings are reproduced in this thesis. The unpublished conference contributors are not reproduced in this thesis.

Publications & Presentations

Refereed and published conference proceedings

- Kyung Hoon Oh, Sophie Triantaphillidou, and Ralph E. Jacobson, Perceptual image attribute scales derived from overall image quality assessments, *SPIE Proceeding: Image quality and system performance VI* , Vol. 7242, San Jose, USA (2009).
- Kyung Hoon Oh, Sophie Triantaphillidou, and Ralph E. Jacobson, Scene classification with respect to image quality measurements, *SPIE Proceeding: Image quality and system performance VII*, Vol. 7529. , San Jose, USA (2010).
- Kyung Hoon Oh, Sophie Triantaphillidou, and Ralph E. Jacobson, Device dependent, scene dependent quality predictions using Effective Pictorial Information Capacity, *SPIE Proceeding: Image quality and system performance VIII*, San Francisco, USA (2011).

Other refereed and/or unpublished conference contributions

- sRGB colour errors in a digital camera- LCD imaging chain, Institute of Physics conference 2007, Swansea University, Swansea, Wales, UK (2007).
- The accuracy of sRGB colour reproduction of a camera and a LCD imaging chain, The Royal Photographic Society Imaging Science Group: Digital Futures 2007, Royal Institute of British Architects, London, UK (2007). (http://www.rps-isg.org/docs/DF2007programme.pdf)

- Scene classification with respect to image quality measurements, The Royal Photographic Society Imaging Science Group: Digital Futures 2009, Institute of Physics, London, UK (2009).
  (http://www.rps-isg.org/DF2009_presentations.php)

Awards

1. The Royal Photographic Society G.I.S. and A.R.P.S qualification (2008).
2. The Royal Photographic Society Imaging Science Group Travel Grant (2009).

Perceptual image attribute scales derived from overall image quality assessments

Kyung Hoon Oh, Sophie Triantaphillidou, Ralph E. Jacobson
Imaging Technology Research Group, University of Westminster, Harrow, UK

ABSTRACT

Psychophysical scaling is commonly based on the assumption that the overall quality of images is based on the assessment of *individual attributes* which the observer is able to recognise and separate, i.e. sharpness, contrast, etc. However, the assessment of individual attributes is a subject of debate, since they are unlikely to be independent from each other.

This paper presents an experiment that was carried to derive individual perceptual attribute interval scales from overall image quality assessments, therefore examine the weight of each individual attribute to the overall perceived quality. A psychophysical experiment was taken by fourteen observers. Thirty two original images were manipulated by adjusting three physical parameters that altered image blur, noise and contrast. The data were then arranged by *permutation*, where ratings for each individual attribute were averaged to examine the variation of ratings in other attributes.

The results confirmed that one JND of added noise and one JND of added blurring reduced image quality more than did one JND in contrast change. Furthermore, they indicated that the range of distortion that was introduced by blurring covered the entire image quality scale but the ranges of added noise and contrast adjustments were too small for investigating the consequences in the full range of image quality. There were several interesting tradeoffs between noise, blur and changes in contrast. Further work on the effect of (test) scene content was carried out to objectively reveal which types of scenes were significantly affected by changes in each attribute.

Keywords: Psychophysical scaling, image quality, overall quality assessment, perceptual quality attributes.

1. INTRODUCTION

Image quality can be defined as the overall impression of image excellence. Many psychophysical investigations have been conducted on the assessment of *individual image attributes*. This individual assessment has been the subject of discussion since a single image quality attribute is unlikely to be independent from other attributes [1, 2]. It creates a problem in simplifying image quality measurements since it does not consider the complicated relationships between them [3].

This paper describes experimental work that was carried out to derive individual perceptual attribute interval scales from overall image quality assessments. This approach does not require scaling of individual attributes and does not require the assumption that the attribute is one dimensional. This research aims: 1) to investigate the perceptual constraints that determine image quality 2) to determine the weight of each individual attribute to the overall image quality.

2. IMPLEMENTATION OF PSYCHOPHYSICAL SCALING

2.1 Image acquisition and selection

Images of natural scenes were acquired i) by image capture, using a digital camera and ii) from two Master Kodak Photo CDs, to cover a range of images with differing content and characteristics. The scenes represented a variety of subjects, such as portraits, natural scenes, buildings with plain and busy background, etc. They were chosen to include various global and local illuminations, numerous colours, varying number and strength of lines, edges and spatial distribution of the subjects. The test scenes are included in Appendix.

Fifteen natural scenes were captured using a Canon EOS-1Ds full frame digital SLR camera, equipped with a Canon EF 28-135mm f 3.5-5.6 IS USM zoom lens. The ISO 100 setting was used to capture all scenes. Camera exposure was determined by taking multiple reflection readings from various parts of the scene. This was achieved with the through-the-lens centre-weighting and spot metering modes of the camera. The camera was set to auto colour balance mode and sRGB colour mode. The lens was focused manually. Scenes were recorded at about 11 mega pixels resolution (4064× 2704) in a CMOS sensor (with approximately 8.8 μm square pixel dimensions). The scenes were saved as 12-bit RAW files and then downloaded to a computer as 8-bit TIFF uncompressed images by using the software provided by Canon, via an IEEE 1394 connection.

In addition seventeen natural scenes were selected from two Master Kodak Photo CDs. The Master Kodak Photo CD images were opened at a spatial resolution of 512 by 786 and at a colour resolution of 8 bits per channel.

All thirty two images were down-sampled to 317 by 476 pixels using spline interpolation and saved as TIFF files of approximately 400 KB.

2.2 Test Stimuli

The thirty two original images were manipulated by altering three physical parameters: Gaussian blurring, Gaussian noise and contrast adjustment to obtain a large number of test stimuli with different levels of blur, noise and contrast. Prior to deciding the ranges and levels of distortion, pilot studies were conducted on the calibrated 21 inch EIZO CG-210 LCD, which was then used for the investigation. Each chosen distortion level corresponded to approximately one JND when images were viewed on the display.

The following techniques were chosen for the manipulation of the original stimuli:

Blurring: Firstly, *Gaussian blurring* was applied on the thirty two originals. The standard deviation (σ) of the *Gaussian low-pass kernel* ranged from 0.01 to 1.24 at 0.3075 intervals. This created a total of one hundred and sixty test images.

Additive Noise: After blurring, the images were further distorted by *Gaussian noise* filtering, using three different standard deviations (σ): 0.0, 0.1 and 0.2. This function created three different levels of uniform noise and provided a total of four hundred and eighty distorted images.

Contrast adjustment: After blurring and adding noise, contrast adjustment was applied to all distorted images at five different levels. This included the original level, two levels for contrast enhancement and two levels for contrast reduction. The five levels of contrast were achieved by varying contrast (γ) from 0.9 to 1.1, at 0.05 intervals. A total of two thousand four hundreds test stimuli were finally created.

2.3 Psychophysical display, interface and viewing conditions

Psychophysical tests were carried out under dark viewing conditions. All images were displayed on an EIZO CG-210 LCD, controlled by the S3 Graphics Prosavage DDR graphic card in a personal computer. The graphic card was configured to display 24-bit colour, at a resolution of 1600 by 1200 pixels and a frequency of 60 HZ.

The display was switched on for fifteen minutes before the tests to allow stabilization. It was placed at a viewing distance of approximately 60 cm from the observers, and subtended a visual angle of roughly 10°. The Eye-One Pro monitor calibrator was used to calibrate the display at a white point close to D65, a contrast of 2.2 and a white point luminance of 100 cd/m^2 – the default settings for contrast and luminance for this monitor.

2.4 Observations

Subjective assessments were performed by a panel of fourteen selected observers, seven males and seven females. They were all familiar with the meaning and assessment of image quality. The age of observers ranged from 21 to 52 years old. All observers had normal colour vision and normal or corrected-to-normal visual acutance.

Each observer undertook a *categorical scaling* experiment 6 times, each time evaluating a different set of images. Each

individual observation period was around 45 minutes. The standard ISO20462-1 suggests that the observation periods should be a maximum of 60 minutes to avoid tiredness or lack in concentration [4]. Before starting the test, observers were allowed several minutes to adapt to the dark viewing conditions of the laboratory [5].

Images were displayed one at a time, randomly, in the centre of the display area to minimise non-uniformity display effects. Observers were asked to place each test image according to the perceived image quality in one out of 5 quality categories, with 1 indicating the worst quality and 5 the best quality.

3. ANALYSIS OF THE ASSESSMENTS

3.1 Scaling overall image quality

Interval scales were derived using the simplest condition, *D*, of *Torgerson's Law of Categorical Judgements* [6], which makes minimum assumptions regarding the category and sample variance: correlation coefficients and dispersions of both the sample and the category are constant. Category boundaries and sample values were obtained. The least square technique was applied to prevent inaccurate scale values from zero-one elements in proportion matrix [6].

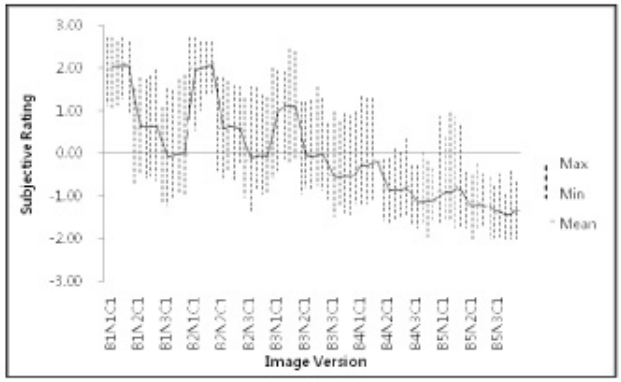


Figure 1. Interval scales of combined scenes

Figure 1 presents interval scales of overall image quality from the combined (average of) 32 scenes. Each label in the x axis represents a specific level of distortion - in blurring (B) and noise (N). Within each of these levels there are 5 different variations in contrast - represented by the individual points in the graph. Each label corresponds to the point which indicates the 1st variation in contrast (i.e. therefore C1 in the label).

The results confirmed that the modifications in all three image attributes, in most cases, decreased image quality. The original versions of the images had an average scale value of 2.03, whilst most of the distorted images had a lower scale value. The specific changes in contrast did not significantly affect image quality (in a few cases they improved it), while noisiness and sharpness were found to be by far the most influential attributes on perceived image quality. This will be discussed in more detail in the individual attributes scales presented later.

The results also indicate scene dependency. The broken lines in Figure 1 indicate the range of scale values derived from all scenes for each level of distortion; the grey square is the average from all scenes.

3.2 Scaling of individual attributes

The collected scale values were rearranged using permutation[1] [7]. Since we have variations in three different image attributes (i.e. in blurring level, noise level, and contrast) the attributes were first arranged as listed in Step1 of Table 1.

The total number of permutations is produced by the product of the total number of distorted images and the number of attribute permutation:

14400 total permutations = (2400 test images) x (6 attribute permutations)

The full implementation of the method is illustrated in Table 1. In addition, an example of the implementation is shown in Table 2. The scale values of the individual image attributes were examined across the average (calculated by the mean) scale values of the other attributes [8]. The first listed attribute in Step 1, in Table 1, is the *targeting attribute* in the permutation. The *mean scale value* for this attribute is calculated by Steps 2 and 3. Step two calculates the average scale value across the last attribute listed in Step 1, Table 1. Step 3 calculates the average scale value across the last listed attribute in Step 2, Table 1.

Step 1: Arrangement				Step 2: Average of last column in Step 1			Step 3. Average of last column in Step 2
I.	5 blur	3 noise	5 contrast	I.	5 blur	3 noise	I. 5 blur
II.	5 blur	5 contrast	3 noise	II.	5 blur	5 contrast	II. 5 blur
III.	3 noise	5 contrast	5 blur	III.	3 noise	5 contrast	III. 3 noise
IV.	3 noise	5 blur	5 contrast	IV.	3 noise	5 blur	IV. 3 noise
V.	5 contrast	5 blur	3 noise	V.	5 contrast	5 blur	V. 5 contrast
VI.	5 contrast	3 noise	5 blur	VI.	5 contrast	3 noise	VI. 5 contrast

Table 1. Individual attribute scaling

	Step 1		Step 2		Step 3	
	Attributes	Scale value	Attributes	Scale value	Attributes	Scale value
I	Blur1Noise1Contrast1	1.47	Blur1Noise1	1.64	Blur1	1.39
	Blur1Noise1Contrast2	1.79				
	Blur1Noise1Contrast3	1.79				
	Blur1Noise1Contrast4	1.73				
	Blur1Noise1Contrast5	1.42				
	Blur1Noise2Contrast1	1.15	Blur1Noise2	1.42		
	Blur1Noise2Contrast2	1.79				
	Blur1Noise2Contrast3	0.96				
	Blur1Noise2Contrast4	1.68				
	Blur1Noise2Contrast5	1.52				
	Blur1Noise3Contrast1	1.08	Blur1Noise3	1.13		
	Blur1Noise3Contrast2	1.22				
	Blur1Noise3Contrast3	1.38				
	Blur1Noise3Contrast4	1.15				
	Blur1Noise3Contrast5	0.8				

Table 2. Example of individual attribute scaling

[1] Permutation means *arrangement* of items. The word *arrangement* is used, if the order of items *is considered*. In general, the number of permutation is taken by ${}^{n}P_{r}$, where n is different item of r at a position. Three of stimuli and three at a time forms six permutations, ${}^{3}P_{3}$=3!= 6.

Results from cases I to VI from Step 1, Table 1, are shown in Figure 2. Each resulting figure uses the same data, but they are presented differently i.e. according to the targeting attribute (listed first in the title of the graph), then the second and third attributes. For example the top left graph in Figure 2 presents the data in the same fashion as Figure 1 – case I. In the graph next to it, the data are presented according to the same targeting attribute but the second and third attributes are interchanged – case II. The middle row in Figure 2 shows cases III and IV and the last row cases V and VI, as listed in Table 1. Similarly to Figure 1, on the graphs on Figure 2 the labels in the x axis represent a specific level of distortion for the targeting and second attributes, where as each point represents the scale value of the targeting, second and third attributes.

Further results from Step 2, Table 1, are shown in Figure 3. Finally, the individual attribute scales are presented in Figure 4, which are derived from the final Step 3 in the process. The results present the mean scale value of quality for each attribute.

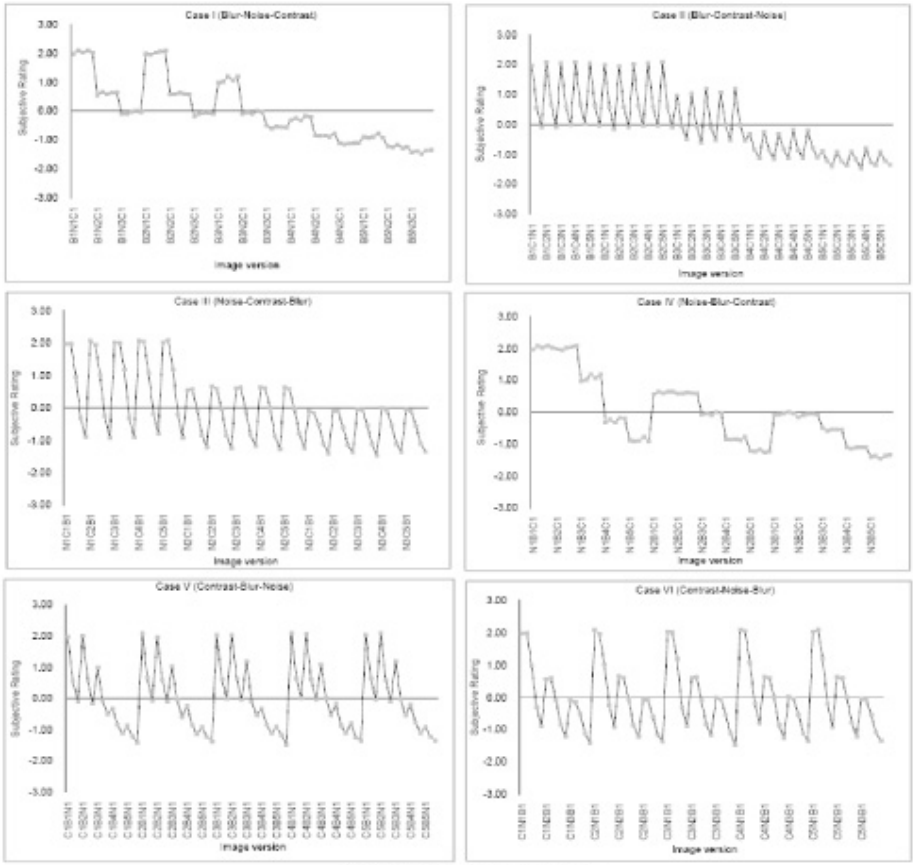


Figure 2. Resulting scales from Step 1

There are several interesting tradeoffs in image quality when varying the three attributes, i.e. blur, noise and contrast (Figure 3). The results agrees with previous research which found that the higher the sharpness the higher the graininess [8, 9]. This is observed in Figure 3, cases I and IV (Blur-Noise & Noise-Blur) which indicate that high amount of blur in the image significantly decreased the perception of noise (case I) and high noise decreased perceived blur (case IV). The relationship between the contrast and the other attributes is not significant for the specific levels of contrast modification that were used for this experiment, which appear not to have altered image quality. This is seen in Figure 4 which indicates that one JND of added noise and one JND of blur reduced image quality much more than one JND in contrast changes.

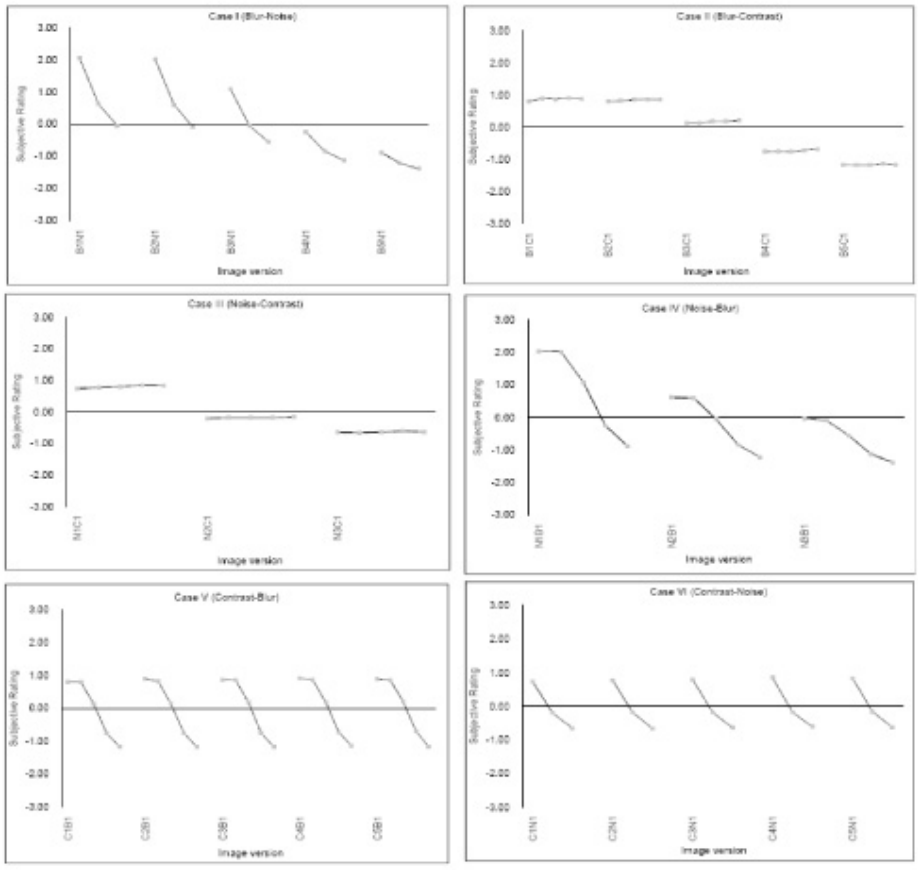


Figure 3. Interval scales of combined scenes (Step 2)

From Figure 4, adding noise reduces the quality scale values by an average of 0.48, adding blur also reduced it by an average of 0.41, whereas increasing contrast increased it very slightly by 0.01.

Case (Blur) we notice that the original and original + 1 JND of blur were almost rated similarly. The quality is shown to decrease significantly for the next blur levels and it reaches a point (at level B5) where more blur would not further reduce it. I.e. the image quality scale of blurring is a hyperbolic (S-shape) function. The ranges of added noise (2 levels) and contrast adjustments (4 levels) were too small for investigating the consequences in the full range of image quality. Case (Noise) indicates that 2 levels of added noise (each separated by 1 JND) decreased equally image quality but the 'toe' of the quality scale for noise was not reached. Finally, as mentioned above, 2 modifications in contrast around the 'optimum' contrast did not alter image quality – Case(Contrast). The results are of course valid for the specific display and viewing conditions.

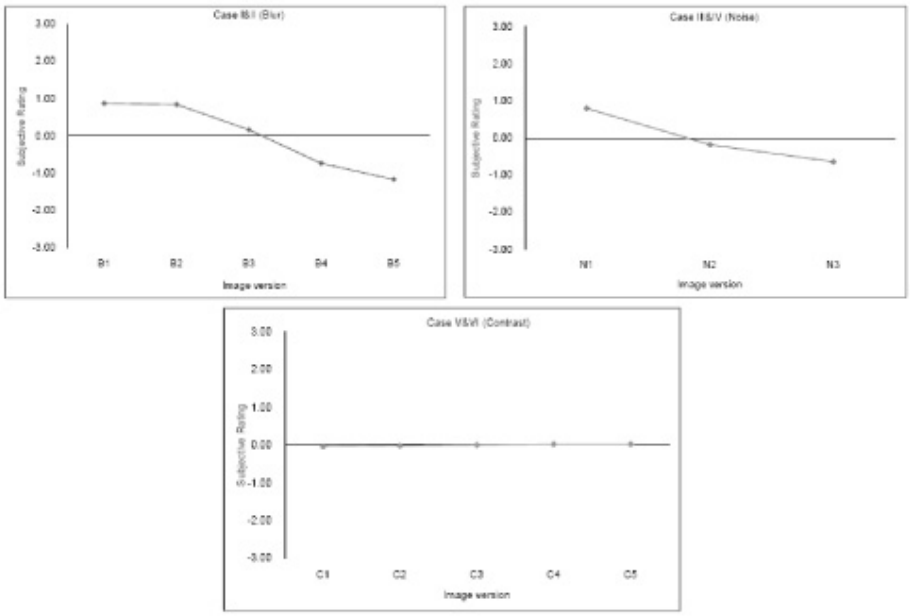


Figure 4. Individual attribute scales

3.3 Scene dependency in the observations

The results show scene dependency caused by the nature of scenes and their inherent properties. This is shown in Figures 5, 6 and 7, where the mean scale value of the distortion level is indicated by the grey square and the range covered by all scenes with the broken lines. When considering image blurring, as the level of distortion increased the scene dependency decreased. Figure 5 indicates that low levels of blurring might affect different scenes in a different manner, but high levels of blurring tend to affect different scenes more equally. On the other hand, Figures 6 and 7 show that the variation from the mean scale value for added noise and variations in contrast appears similar in all levels of distortion. This result suggests that as image quality decreases we found less and less scene dependency but high quality images tend to be rated differently according to their individual scene content.

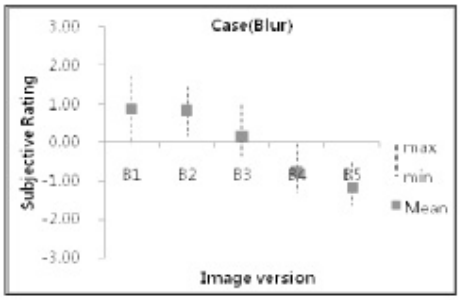


Figure 5. Scene dependency on blur scale

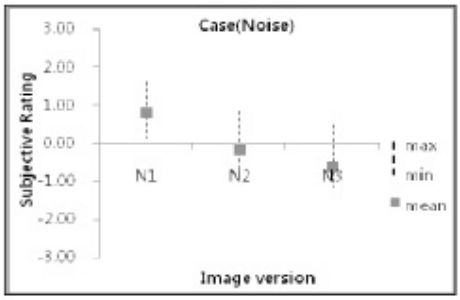


Figure 6. Scene dependency on noise scale

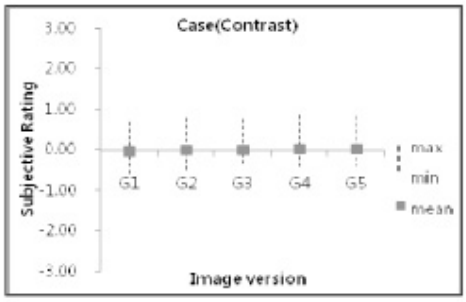


Figure7. Scene dependency on contrast scale

The effect of individual scenes was examined. To determine the comparability of the results for each individual scene with the average combined rating for all scenes, the scale values of each scene were plotted against the combined mean ratings from all scenes, for each individual attribute [8].

In the example plot shown in Figure 8 for the scene 'Saules', the gradients of the lines fitting the data (one for each attribute) represent the compatibility between *the range of ratings* for each individual scene and that of the combined scenes. If the gradient of the line is one, the range of scale values for the scene is the same as that of the combined scenes for the specific attribute. If the gradient is larger than 1, then the scene has a larger quality range meaning that it is more sensitive than the 'average scene' to the changes in the specific attribute. The reverse is true when the gradient is smaller than one.

The correlation coefficient lines fitting the data indicates the strength of relationship between them. The constant in the linear relationship indicates whether, overall, the individual scene got better average quality ratings than the average ratings of the combined scenes (positive offset) or the opposite (negative offset).

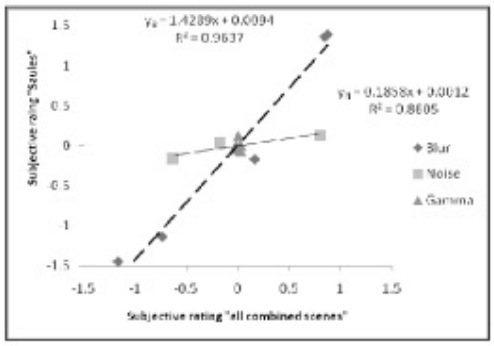


Figure 8. Scene dependency on the "Saules"

	Blurring			Noisiness			Contrast		
	Gradient	Constant	r^2	Gradient	Constant	r^2	Gradient	Constant	r^2
African tree	0.32	-0.02	0.945	1.96	0.00	0.993	0.75	-0.02	0.394
Baby	1.06	-0.02	0.995	1.06	-0.02	0.995	-1.62	-0.04	0.460
Bike	1.21	0.30	0.984	0.72	0.29	0.994	2.60	0.29	0.954
China town	0.95	0.04	0.995	0.97	0.04	0.984	2.60	0.04	0.764
Exercise	1.15	-0.20	0.966	0.51	-0.21	0.989	1.37	-0.21	0.442
Formula	1.00	0.53	0.997	1.14	0.53	0.978	0.03	0.52	0.000
Glasses	0.86	0.10	0.974	1.17	0.11	1.000	-2.81	0.09	0.564
Group	1.07	-0.38	0.980	0.66	-0.38	1.000	2.28	-0.38	0.565
Human	0.91	0.07	0.990	1.10	0.06	1.000	1.44	0.06	0.517
Human2	0.32	0.31	0.948	1.07	-0.30	0.998	2.13	-0.30	0.589
Human3	1.24	-0.20	0.965	0.55	-0.21	0.998	2.09	-0.21	0.764
Human4	1.08	-0.43	0.991	1.11	-0.43	0.977	4.28	-0.42	0.946
Kids	1.18	0.27	0.996	1.15	0.26	1.000	3.06	0.27	0.589
Landscape	0.86	-0.20	0.967	1.44	-0.19	1.000	2.03	-0.20	0.761
Landscape2	0.63	0.06	0.936	1.82	0.06	0.976	1.10	0.05	0.871
Landscape3	1.05	0.07	0.987	1.31	0.07	0.999	-0.07	0.06	0.001
London Eye	0.86	-0.37	0.996	1.09	-0.38	0.989	-0.09	-0.38	0.003
London Eye2	0.93	-0.42	0.999	1.28	-0.41	0.987	-1.12	-0.43	0.444
Louvre	1.16	-0.35	0.998	1.03	-0.36	1.000	0.10	-0.37	0.014
National gallery	1.07	-0.13	0.996	0.96	-0.13	1.000	2.40	-0.12	0.469
Old building	1.29	0.08	0.997	0.92	0.08	1.000	-0.78	0.07	0.093
Plant1	1.15	0.79	0.985	0.34	0.79	0.841	2.40	0.80	0.844
Plant2	0.79	0.53	0.969	0.85	0.53	0.993	5.07	0.54	0.986
Plant3	0.87	0.22	0.992	1.13	0.21	0.997	-1.50	0.20	0.353
Plant4	0.80	-0.12	0.994	1.12	-0.12	0.991	-1.53	-0.13	0.607
Plant5	1.09	0.08	0.998	0.99	0.08	0.988	0.59	0.07	0.168
Plant6	0.97	0.13	0.997	1.04	0.13	1.000	-1.32	0.11	0.350
St. Pauls	1.40	0.03	0.991	0.50	0.02	0.999	3.06	0.03	0.851
St. Pauls2	1.10	0.03	0.996	0.87	0.03	0.999	-1.47	0.01	0.480
Saules	1.43	0.01	0.964	0.19	0.00	0.861	0.09	0.00	0.001
Sungsil	1.24	-0.31	0.981	0.83	-0.31	0.990	2.66	-0.31	0.857
Yellow flower	0.91	0.23	0.996	1.14	0.22	0.989	-0.01	0.22	0.001

Table 3. The gradient, constant, and correlation coefficients for each test image for all 3 attributes

Table 3 presents the gradient, constant and correlation coefficient of the regression lines for each test scene and for all three attributes.

Regarding blurring: the scenes "Human2", "African tree" and "Landscape2" were found to be the three less sensitive scenes to blurring, with the lower gradient values. On the other hand, "Saules" has the highest gradient value, indicating high sensitivity to blurring.

Regarding added noise: An extreme result was produced by the "Saules" scene (Figure 8), which was shown the most insensitive scene to added noise, having an extremely low gradient of 0.19. This is the busiest scene in the set [10] and the noise was probably masked by the high frequency information in this scene.

The most unusual results regarding both blurring and noise were produced from the "African tree" scene. The scene was found to be the most sensitive to added noise (gradient = 1.96) and the second most insensitive to blurring (gradient = 0.32). This result was also confirmed by previous research [11].

The correlation coefficients of the linear regression on blurring and added noise were all close to 1.0 whereas on contrast they were variable (0.001~0.9861).

4. CONCLUSION

A large-scale categorical judgment experiment was conducted. In all 14 observers performed a total of 2400 observations. Psychometric scaling was used to create interval scales indicating the quality of images subjected to blur, noise and variations in contrast.

This work successfully derived individual perceptual attribute interval scales from overall image quality assessments and the following summary lists the main effects of the individual attributes.

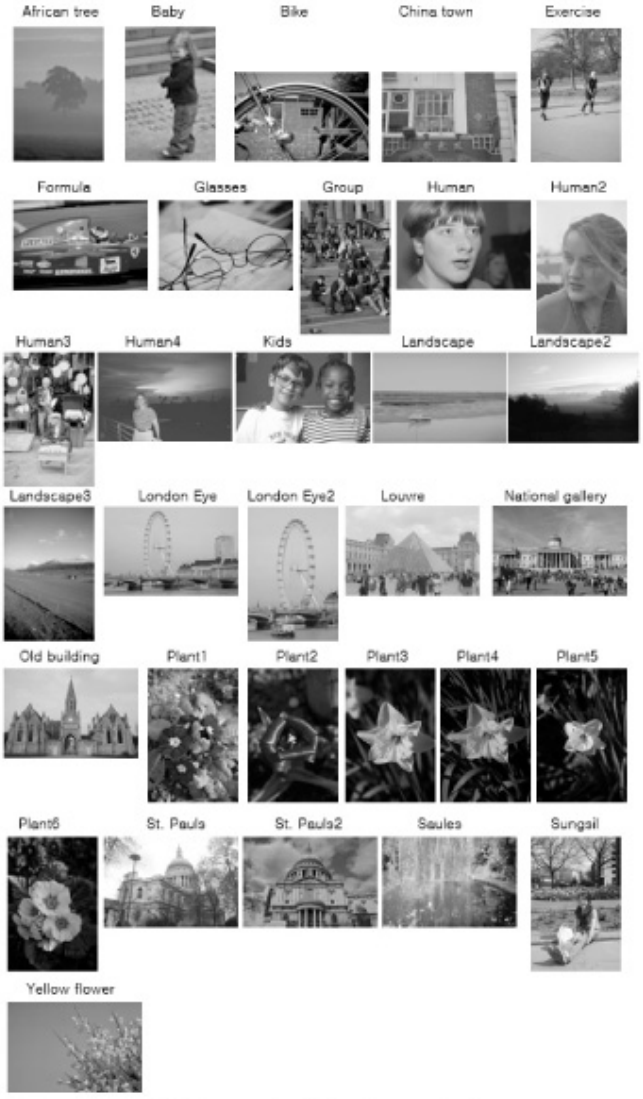
1. One JND of added noise and one JND of added 'blurring' reduced image quality more than one JND in contrast change - in the range of distortion applied in this experiment.
2. The range of distortion was introduced by blurring to cover the entire image quality scale; however, the ranges of added Gaussian noise and contrast adjustment were too small for investigating the consequences on the full range of image quality.
3. Blurring decreased significantly the perception of noise.
4. Adding noise decreased perceived blurring.
5. Image quality assessments showed scene dependency.
6. As image quality decreases we found less and less scene dependency but high quality images tend to be rated differently according to their individual scene content.

Further work on the effect of scene content will be carried out to objectively reveal which types of scenes are significantly affected by changes in each attribute.

REFERENCES

[1] Triantaphillidou, S., "Aspects of Image Quality in the Digitisation of Photographic Collection," University of Westminster PhD Thesis, London, (2001).
[2] Roufs, A., "Perceptual image quality: concept and measurement," Philips Journal of Research. 30, 33-38 (1992).
[3] Engeldrum, P., "The image quality circle," [Psychometric Scaling], Imcotek Press, Winchester, 5-18. ch 2 (2000).
[4] ISO:20462-1., "Photography-psychophysical experimental methods for estimating image quality," International Organization : ISO, (2005).
[5] Hunt, R., "The effect of daylight and tungsten light adaptation on colour perception," Journal of the Optical Society of America. 40, 362-371 (1950).
[6] Engeldrum, P., "Indirect interval scaling-category scaling methods," [Psychometric Scaling], Imcotek Press, Winchester,123-138. ch 10 (2000).
[7] Stroud, K A (Kenneth Arthur)., "Probability," [Engineering mathematics], Palgrave, Basingstoke, 1155-1197 (2001).
[8] Johnson, G and Fairchild, M., "Sharpness Rule," IS&T Color imaging Conference, 24-30 (2000).
[9] Sawyer, J., "Effect of graineiness and sharpness on perceived print quality," symposium on photographic image quality, 222-231 (1980).
[10] Triantaphillidou, S, Allen, E and Jacobson, R., "Image quality Compression Between JPEG and JPEG 2000 II: Scene dependency, Scene analysis and Classification," Jornal of Imaging Science and Technology. 51(3), 259-270 (2007).
[11] Jenkin, R, Triantaphillidou, S. and Richardson, M., "Effective Pictorial Information Capacity as an Image Quality Metric," Proc: SPIE/IS&T Electronic imaging 2007: Image quality and system performance. 6494, 0-9 (2007).

Appendix



Test images selected for the experiment

Scene classification with respect to image quality measurements

Kyung Hoon Oh, Sophie Triantaphillidou and Ralph E. Jacobson
Imaging Technology Research Group, University of Westminster, HA1 3TP, Harrow, UK

ABSTRACT

Psychophysical image quality assessments have shown that subjective quality depended upon the pictorial content of the test images. This study is concerned with the nature of *scene dependency*, which causes problems in modeling and predicting image quality. This paper focuses on scene classification to resolve this issue and used K-means clustering to classify test scenes. The aim was to classify thirty two original test scenes that were previously used in a psychophysical investigation conducted by the authors, according to their *susceptibility* to sharpness and noisiness. The objective scene classification involved: 1) investigation of various *scene descriptors*, derived to describe properties that influence image quality, and 2) investigation of the degree of correlation between scene descriptors and scene *susceptibility parameters*. Scene descriptors that correlated with scene susceptibility in sharpness and in noisiness are assumed to be useful in the objective scene classification. The work successfully derived three groups of scenes. The findings indicate that there is a potential for tackling the problem of sharpness and noisiness scene susceptibility when modeling image quality. In addition, more extensive investigations of scene descriptors would be required at global and local image levels in order to achieve sufficient accuracy of objective scene classification.

Keywords: scene dependency (scene susceptibility) of image quality, scene classification, scene descriptors (image analysis tools)

1. INTRODUCTION

Image quality can be defined as the overall impression of image excellence and depends upon the pictorial content of the test images [1, 2]. This study is concerned with the nature of scene dependency, which causes problems in modeling and predicting image quality, especially in device dependent image quality measures. This is because objective quality measures tend to perform relatively well on individual average-looking scenes, but they provide lower correlation with subjective assessments when working with non-standard looking scenes.

There are several ways of overcoming the problems caused by scene dependency [3]. One commonly employed is to exclude results obtaining from 'odd scenes' in quality measurements. These, however, do not effectively represent the range and variety of different scenes that photographers, artists and consumers may wish to record and reproduce faithfully [3]. Furthermore, scenes that deviate in content from a representative set (e.g. ISO set of test scenes [4]) may not be reproduced appropriately, since they are not in accordance with the 'average' reproduction derived from image quality results.

Keelan [5] suggests test scene classification with respect to image quality. The classification he proposes which is based on test scene content and it's impact of quality attributes, is as follows; a) most susceptible scenes 25%, b) least susceptible scenes 25% and c) intermediately susceptible scenes 50%. In addition, Triantaphillidou *et al* [3] propose test scene classification, using objective *scene descriptors* that correlate with subjective criteria on scene susceptibility to image quality attributes. Scene descriptors are derived to describe basic inherent scene properties that human observers refer to when they judge the quality of images.

The aim of the research describe here was to classify thirty two original test scenes that were previously used in a psychophysical investigation conducted by the authors [2], according to their *susceptibility* (see section 3.1) to sharpness and noisiness. The objective scene classification involved: 1) investigation of various scene descriptors, derived to describe properties that influence image quality, and 2) investigation of the degree of correlation between scene descriptors and scene susceptibility parameters.

2. SCENE DESCRIPTORS

The first step in the objective scene classification was to investigate scene descriptors, derived to describe a number of scene properties. The algorithms deriving these descriptors were implemented in global and local image regions. The reason for local region implementation was that some researchers [6, 7] believe that a local measure of image quality is probably more useful than a global one. A Kadir and Brade's saliency model [8] was applied in MATLAB [9] for this purpose. The implementation involved in following:

- the division of a 20x20 grid on the image
- the calculation of the local entropy in each grid, using a radius from 3 to 70 pixels
- the detection of 30 high in saliency points
- the erosion of the non-saliency areas to amplify the saliency areas

Figure 1 illustrates the saliency process for one test image and presents local regions derived from the saliency model implementation for another four test scenes.

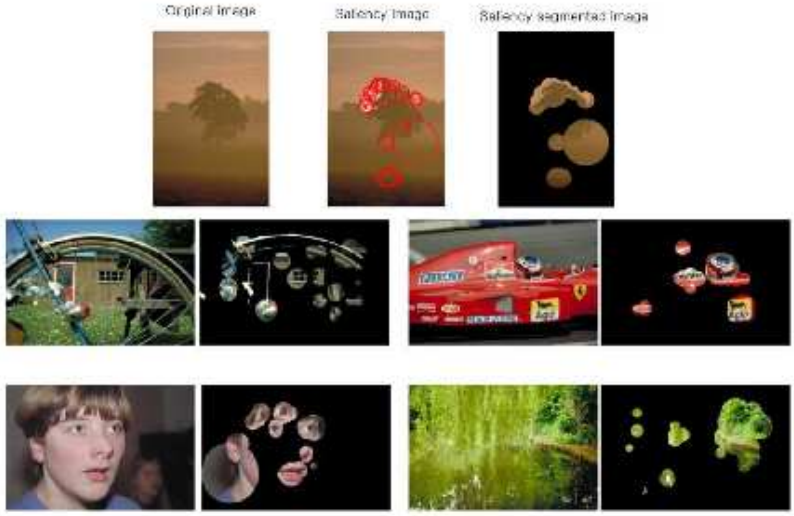


Figure 1. Saliency process for one test image (top row) and local regions for another four test scenes.

A number of scene descriptors were derived using first-order and second-order statistical measures as well as edge detection. Some of these measures (sections 2.1, 2.2 and 2.3) were applied to the *grayscale version* of the image, which was obtained from the 8-bit per channel sRGB image by [9]:

$$Grayscale\_image = 0.2989R + 0.5870G + 0.1140B.$$

where R, G, B correspond to the pixel value of the R,G and B channels, respectively.

Further, first-order statistical measures were employed to derived measures from the image represented in CIELAB coordinates (section 2.4).

2.1 First-order statistical measures

First-order statistical measures were derived from the Probability Density Functions (PDF) of the grayscale image. The ones investigated in this work are listed below:

- Mean: is the average value in PDF.
- Median: is that value of the middle term of PDF when all the observations are arranged is ascending or descending order.
- Mode: is the value that occurred most often in PDF.
- Variance: is a measure of contrast in PDF, the second power of standard deviation.
- Skewness: is a measure of imbalance of the PDF. We get a value close to zero when the distribution of grey level is balanced (symmetric PDF).
- Entropy: is a measure of information content of the PDF.

2.2 Second-order statistical measures

Second-order statistical measurements, which reveal textural information in images, were calculated from the gray-level co-occurrence matrix (GLCM) [9, 10]. Implementation was carried out in MATLAB [9] using default angle and distance values: 0 and 1 in pixels, respectively. The ones investigated for this work are listed below:

- Inertia (or contrast (Co)):

$$Co = \sum_{i=0} \sum_{j=0} |i-j|^2 P(i,j)$$

- Homogeneity (H):

$$H = \sum_{i=0} \sum_{j=0} \frac{P(i,j)}{1+|i-j|}$$

- Correlation (or linearity (Cor)):

$$Cor = \sum_{i=0} \sum_{j=0} \frac{(i-m_i)(j-m_j)P^2(i,j)}{\sigma_i \sigma_j}$$

- Energy (Ene):

$$Ene = \sum_{i=0} \sum_{j=0} P(i,j)^2$$

Where $P(i,j)$ is the the joint probability distribution of pairs of pixels (i, j). m_i and m_j are the mean values of the pair of gray levels i and j. σ_i and σ_j are the standard deviation value of the pair of gray levels i and j [10].

2.3 Measurement from edge detection

The Sobel, Prewitt and LOG (Laplacian of Gaussian) edge detection algorithms were used to quantify the presence and strength of edges in the grayscale image [11]. The Sobel and Prewitt edge detectors performed using a 3 × 3 kernel size and 0.04 for sigma [3]. The LOG edge detector was set to a 5 × 5 kernel size and 0.5 for sigma, which is the default value employed in MATLAB [9]. All edge detectors were operated with the 'replicate' boundary option in MATLAB, where the boundaries were assumed to be equal the nearest border value. During the edge detection, the magnitude of edge (G) was computed by the square-root operation [11]:

$$G = \sqrt{G_x^2 + G_y^2}$$

where G_x and G_y are the horizontal and vertical edge gradients of the image respectively.

Then all individual edge gradients were averaged. Figure 2 illustrates two original images and the corresponding threshold images after Sobel edge detection with the average edge gradient, related to the edges' strength as well as the amount of edge information in the image.

Average edge gradient: 11.60 Average edge gradient: 66.15

Figure 2. Example of average edge gradient

2.4 Measurement from the CIELAB image

The variance in chroma and saturation were considered as measures of color information. They have been shown to correlate successfully with the perceived image colorfulness [3] and perceived color strength, respectively [12]. The variance in chroma *(VC*)* was calculated [3]:

$$VC^* \cong \sqrt{\sigma_{a^*}^2 + \sigma_{b^*}^2}$$

In addition, color strength metric *(VS*)*, based on the definition of saturation: Saturation = Chroma/Lightness, derived by [12]:

$$VS^* = VC^*/L^*$$

where the lightness (L^*) is $L^* = L^*_{mid} + \|L^*_{mid} - L^*_i\|$

where L^*_{mid}=50.

3. CORRELATION BETWEEN SCENE DESCRIPTORS AND SCENE SUSCEPTIBILITY PARAMETERS

The second step in the objective scene classification was to investigate the degree of correlation between scene descriptors and scene susceptibility parameters, described in reference [2]. Scene descriptors that successfully correlated with scene susceptibility in sharpness and in noisiness provided means toward the objective scene classification.

3.1. Scene susceptibility parameters

The scene susceptibility parameters were collected from previous experimental work on 'Perceptual image attribute scales derived from overall image quality assessments' [2] (Table 1). They were based on the visual quality loss that occurred to *individual test scenes* with sharpness and noisiness distortions.

	Susceptibility to sharpness	Susceptibility to noisiness		Susceptibility to sharpness	Susceptibility to noisiness
African tree	0.32	1.96	Baby	1.06	1.06
Bike	1.21	0.72	China town	0.95	0.97
Exercise	1.15	0.51	Formula	1.00	1.14
Glasses	0.86	1.17	Group	1.07	0.66
Human	0.91	1.10	Human2	0.32	1.07
Human3	1.24	0.55	Human4	1.08	1.11
Kids	1.18	1.15	Landscape	0.86	1.44
Landscape2	0.63	1.62	Landscape3	1.05	1.31
London Eye	0.86	1.09	London Eye2	0.93	1.28
Louvre	1.16	1.03	National gallery	1.07	0.96
Old building	1.29	0.92	Plant1	1.15	0.34
Plant2	0.79	0.85	Plant3	0.87	1.13
Plant4	0.80	1.12	Plant5	1.09	0.99
Plant6	0.97	1.04	St. Pauls	1.40	0.50
St. Pauls2	1.10	0.87	Saules	1.43	0.19
Sungail	1.24	0.83	Yellow flower	0.91	1.14

Table 1. Subjective scene susceptibility parameters for sharpness and noisiness

A scene susceptibility parameter was identified for each test scene, by calculating the gradient of the straight line connecting average subjective quality ratings (calculated from the entire test-set) and individual quality ratings for the test scene. When the gradient of the line is one, the subjective scale values for the individual scene are the same with these of the combined scenes - for the specific attribute. When the gradient is larger than one the individual scene is more susceptible than the 'average scene' to changes in the specific attribute. The reverse is true when the gradient is smaller than one. An example is shown in Figure 3 for the test scene 'Saules', with gradients for scene susceptibility to noisiness and sharpness equal to 0.1858 and 1.4289 respectively.

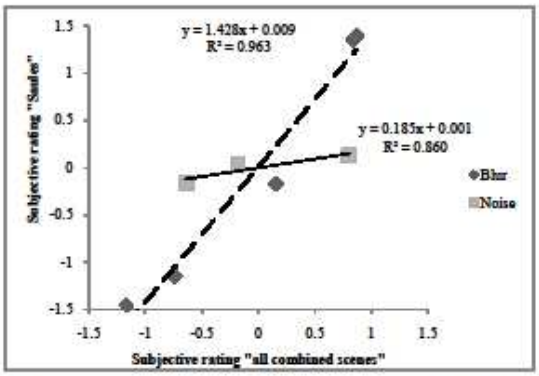


Figure 3. Scene susceptibility parameters for the test scene "Saules" (shown in Figure 7)

3.2. Scene descriptors versus susceptibility parameter for noisiness and sharpness

The Spearman's correlation coefficient, r_s, was derived to investigate correlation between scene descriptors and scene susceptibility to noisiness and to sharpness. The coefficient is useful when data have a ranking but no clear numerical interpretation, such as when assessing preferences for data on an ordinal scale [13]. The correlation coefficients range between -1.0 (indicating perfect anti-correlation) and 1.0 (indicating perfect correlation), with 0 denoting no correlation at all.

Successful correlations[1] were obtained between noisiness susceptibility parameters and most second order statistical measures, as well as measures derived from edge detection. Table 2 shows the successful correlation coefficients for noisiness. An example of correlating susceptibility with a scene descriptor is shown in Figure 4.

Scene descriptors	Correlation coefficient (r_s) for scene susceptibility to noisiness	Correlation coefficient (r_s) for scene susceptibility to sharpness
Inertia (Contrast):	-0.694	0.802
Homogeneity	0.738	-0.781
Correlation (Linearity)	0.644	-0.550
Energy	0.577	-0.647
Average Sobel gradient	-0.701	0.786
Average Prewitt gradient	-0.701	0.786
Average LOG gradient	-0.593	0.747

Table 2. Successful correlation coefficients

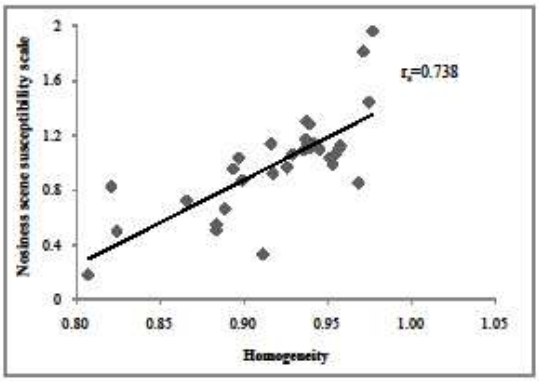


Figure 4. Relationship between the homogeneity descriptor and the susceptibility parameter for noisiness

Successful correlations were also obtained between sharpness susceptibility parameters and, again, most second order statistical measures and measures derived from the edge detection[2]. Table 2 shows also the successful correlation coefficients obtained for sharpness. An example is shown in Figure 5.

[1] When a correlation coefficient is larger than a level of significance at 1% probability level, it indicates statically significant [13].
[2] For both sharpness and noisiness susceptibility predictions, correlations were more successful when the measures were applied at global image level.

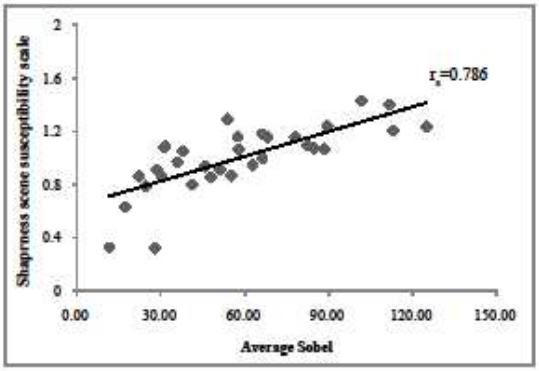


Figure 5. Relationship between the average Sobel descriptor and the susceptibility parameter for sharpness

There were several interesting relationships between scene content and scene susceptibility to noisiness and sharpness. The results confirmed that the higher the texture in the scene content, the lower the susceptibility to noisiness and the higher the susceptibility to sharpness. For example, the correlation coefficients between the homogeneity and scene susceptibility to noisiness and to sharpness were 0.738 and -0.781 respectively. In addition, high presence and strength of edges in the image significantly decreased the perception of noise and increased the susceptibility to sharpness. For example, the correlation coefficients between the average Sobel metric and scene susceptibility to noisiness and to sharpness were -0.701 and 0.786 respectively. It is also evident and confirmed that the relationship between sharpness and noisiness is complimentary i.e. high amount of blur in the image significantly decreased the perception of noise, and high noise decreased perceived blur [2, 7].

Correlations were more significant when the descriptors were derived from the entire image (algorithms were applied globally). Further investigation is required for the derivation of scene descriptors from specific image regions of interest (algorithm application locally). For example, using the central part of the image, as by Keelan and Jin have suggested [7] as a sharpness-critical region and the periphery of the image as a noisiness-critical region. Also, further investigation is required toward the combination of various scene descriptors to derive scene metrics that may describe more successfully the susceptibility of test scenes to noisiness and sharpness.

Overall the results indicated that there is association between selected scene descriptors and scene susceptibility parameters. Thus, the scene descriptors that correlated with sharpness and noisiness scene susceptibility can be used to objectively classify scenes.

4. CLUSTERING FOR NATURAL SCENES

Finally, k-means partitional clustering was implemented to objectively group the 32 test scenes according to their susceptibility to both sharpness and noisiness.

The k-means partitional clustering consists of several steps [14]. The first step of is to define a fixed number of clusters, k. The choice of k is exceedingly important in clustering: an inappropriate choice of k may yield poor results while the correct choice of k is often ambiguous. Possible methods for choosing k include empirical and numerical methods [15]. The empirical method is usually preferred [15]. In relevant image quality investigations k is usually chosen to be equal to 3.0 [5, 16]. Once k is chosen, then modifications of the distances between all points in n^{th} cluster (n varying from 1 to k) and the centre of the cluster are applied. The main idea for modification is that the average distances between all points in cluster and the central point is minimal. During these modifications, new cluster centers are allocated using Euclidean distances. The modification stops when the averages distance from all points in n^{th} cluster and the new central point is minimized.

Two scene descriptors that correlated successfully with both noisiness and sharpness susceptibility, i.e. the *homogeneity* and *average Sobel edge gradient descriptor*, were used for testing the clustering. Clustering was implemented in SPSS programming environment [17]. Figure 6 presents the three clusters with the initial and final centres of the cluster, and then the images corresponding to each of the three clusters (or groups) are shown in Figure 7.

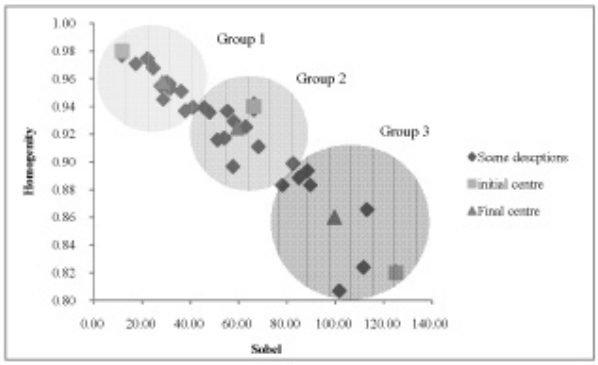


Figure 6. Initial and final centre in three groups

5. CONCULSION

A number of scene descriptors were successfully derived from first-order and second-order statistical measurements as well as edge detection. They were concerned with the extraction of image features, such as brightness, contrast, texture, edges, color contrast etc.

The degree of correlation between scene descriptors and scene susceptibility parameters was investigated using Spearman's correlation coefficient. Successful correlations were obtained between: scene susceptibility parameters for *noisiness* and the *homogeneity descriptor*, and scene susceptibility parameters for *sharpness* and the *average edge gradient descriptors*. These correlations indicated that the selected scene descriptors successfully represented sharpness and noisiness susceptibility and can be used to classify the test scenes used in image quality investigations.

Using the selected scene descriptors and applying K-mean clustering, three groups of scenes were successfully derived, i.e. scenes with: 1) low susceptibility to sharpness distortions and high susceptibility to noisiness 2) average susceptibility to sharpness distortions and noisiness, 3) high susceptibility to sharpness distortions and low susceptibility to noisiness.

The findings indicate that there is a potential for tackling the problem of sharpness and noisiness scene susceptibility when modeling image quality. More extensive investigations of scene descriptors with respect to both global and local image features will help further toward objective scene classification of test scene used in image quality investigations.

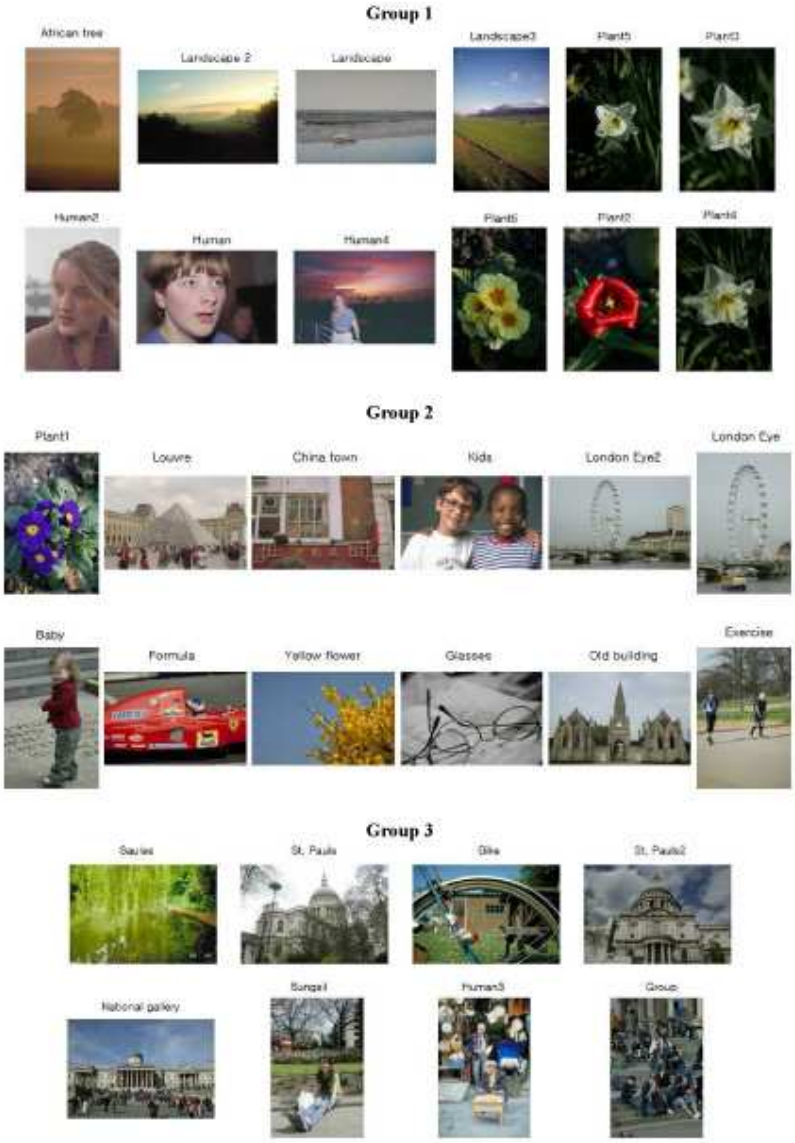


Figure 7. Images in three clusters (groups)

REFERENCES

[1] Keelan, B., [Handbook of Image Quality: Scene and Observer Variability], Marcel Dekker INC, New York, ch 10 (2003).
[2] Oh, K. H., Triantaphillidou, S. and Jacobson, R. E. "Perceptual image attribute scales derived from overall image quality," Proc. SPIE 7242-11, (2009).
[3] Triantaphillidou, S. Allen, E. and Jacobson, R. E., "Image Quality of JPEG vs JPEG 2000, Part 2: Scene Dependency, Scene Analysis and Classification," Jornal of Imaging Science and Technology, Vol 51(3), 259-270 (2007).
[4] ISO 20462-3, [Photography-psychophysical experimental methods for estimating image quality], International Organization : ISO (2005).
[5] Keelan, B., [Handbook of Image Quality: A detailed example of objective metric design], Marcel Dekker INC, New York, ch 14 (2003).
[6] Westen, S. J. P, Lagendijk, R. L. and Biemond, J., "Perceptual image quality based on a multiple channel HVS mode," Proceedings of the 1995 International Conference on Acoustics Speech and Signal processing 4, (1995).
[7] Keelan, B. and Jin, E. W., "Weighting of field heights for sharpness and noisiness," Proc. SPIE 7242, (2009).
[8] Kadir, T. and Brady, M., "Scale, Saliency and Image Description," International Journal of Computer Vision, Vol 45(2), 83-105 (2001).
[9] MathWorks., [MATLAB 2007a], The MathWorks Inc (2007).
[10] Pratt, W. K., [Digital image processing: Image feature extraction], John Wiley & Sons, New York & Chichester, 557-596 (1991).
[11] Gonzalez, R. C. and Woods, R. E., [Digital Image Processing: Image segmentation], Prentice Hall, London, ch 10 (2002).
[12] Othman, A. and Martinez, K., "Colour appearance descriptors for image browsing and retrieval," Proc. SPIE 6820, (2008).
[13] Siegel, F. A., [Statistics and data analysis], John Wiley & Sons, New York & Chichester (1996).
[14] Yendrikhovskij, S., [Colour Image Science: Image Quality and Colour Categorisation], John Wiley & Sons. London, ch.16 (2002).
[15] Catherine, A. S. and Gareth, M. J., "Finding the number of clusters in a data set: An information theoretic approach," Journal of the American Statistical Association. Vol 98 (January), 750–763 (2003).
[16] Keelan, B., [Handbook of Image Quality: Monte Carlo Simulation of System Performance], Marcel Dekker INC, New York, ch 28 (2003).
[17] SPSS., [SPSS 14.0], SPSS Inc (2005).

Device dependent, scene dependent quality predictions using Effective Pictorial Information Capacity

Kyung Hoon Oh, Sophie Triantaphillidou and Ralph E. Jacobson

Imaging Technology Research Group, University of Westminster
Watford Road, Harrow, Middlesex, UK, HA1 3TP

ABSTRACT

This study aims to introduce improvements in the predictions of device-dependent image quality metrics (IQMs). A validation experiment was first carried out to test the success of such a metric, the Effective Pictorial Information Capacity (EPIC), using results from subjective tests involving 32 test scenes replicated with various degrees of sharpness and noisiness. The metric was found to be a good predictor when tested against average ratings but, as expected by device-dependent metrics, it predicted less successfully the perceived quality of individual, non-standard scenes with atypical spatial and structural content. Improvement in predictions was attempted by using a modular image quality framework and its implementation with the EPIC metric. It involves modeling a complicated set of conditions, including classifying scenes into a small number of groups. The scene classification employed for the purpose uses objective scene descriptors which correlate with subjective criteria on scene susceptibility to sharpness and noisiness. The implementation thus allows automatic grouping of scenes and calculation of the metric values. Results indicate that model predictions were improved. Most importantly, they were shown to correlate equally well with subjective quality scales of standard and non-standard scenes. The findings indicate that a device-dependent, scene-dependent image quality model can be achieved.

Keywords: Scene dependency of image quality, objective scene classification, Effective Pictorial Information Capacity (EPIC), Monte Carlo simulation, Integrated hyperbolic increment function (IHIF)

1. INTRODUCTION

Many image quality models and metrics capable of predicting the perception of quality have been proposed over the last fifty years [1]. They can be classified into two groups: the device-independent models and the device-dependent models [2].

Device-independent models attempt to predict the visual perception of the image itself, without knowledge of the imaging system's characteristics. These models have an actual benefit in dealing with *scene dependency* in image quality, i.e. the effect of scene content the perceived quality. However, they are unable to measure quality based on the imaging system variables [3].

Device-dependent models attempt to predict image quality from an 'average signal', usually embodied on test targets, the effect of which may be removed from the measured result to obtain the system performance. They focus on the imaging system's characteristics. Device-dependent models are applied in a straightforward fashion for the measurement of image quality of various system variables, such as sharpness, noise, and contrast. Such models are extremely powerful tools for measuring and predicting quality when the properties of the test scene (i.e. signal) are available [3]. However, if the characteristics of the input signal are not in hand (and not accounted for), these models are poor quality predictors when applied to individual test scene judgments, which are usually shown to be scene dependent [1, 4]. Scene dependency is a major reason why these models perform well on individual 'average looking' scenes and test targets, but provide lower

correlation with subjective assessments when working with a variety of scenes with different than 'average signal' characteristics.

An example of such a device-dependent model is the Effective Pictorial Information Capacity (EPIC) [5, 6]. It is based on signal transfer theory and has been relatively successful in predicting the quality of compressed images and images that vary in sharpness and noisiness. It tends to perform well on 'average' type of scenes, but metric results correlate less successfully with subjective assessments of non-standard scenes with atypical spatial and structural content [5].

This work is a case study on the calibration of a device-dependent image quality metric, such as EPIC, to account for the input signal characteristics. A validation experiment was first carried out to test the success of the EPIC, using results from subjective tests involving 32 test scenes replicated with various degrees of sharpness and noisiness. Improvement in predictions was then attempted by using Keelan's image quality framework [7] and its implementation with the EPIC metric.

2. EFFECTIVE PICTORIAL INFORMATION CAPACITY

2.1 EPIC implementation

The EPIC metric is based on Shannon's information theory [8] and has been proposed as a sharpness and noisiness image quality metric (IQM) [5]. EPIC values were calculated by combining two system variables: the effective pixel dimension in the image and the number of effective distinguishable levels for each recording cell (dynamic range of a system). These variables are cascaded from the input, output and visual system.

For the EPIC metric calculation, first the measurement of the effective pixel dimension in the image (n) was carried out [5, 6]:

1) The effective pixel dimension (n) was taken as the width of line spread function (LSF) at which the MTF falls to 50%. The model accounted of the varying sharpness of the test stimuli (input), the sharpness of the 21' LCD system (output) and employed Barten's contrast sensitivity function (CSF) as the model for the human eye. The imaging chain was that employed in a previous study dealing with the psychophysical scaling of image quality [9]. The Barten's CSF was used as the eye's MTF model because it takes into account various viewing conditions [10]. The specific viewing conditions employed were a luminance of 100 cd/m^2, a viewing distance of approximately 0.6 m, and an image size of 167 x 111 mm.

2) The effective pixel dimension at image area (n) was then calculated by:

$$n = \frac{A_{im}}{\omega^2} \quad (1)$$

where, A_{im} is area of image and ω is the effective pixel dimension.

The number of effective distinguishable levels for each recording cell (dynamic range of a system) (m) were calculated by [5, 6]:

$$m = \frac{DS}{2k\sigma} + 1 \quad (2)$$

where, DS is the difference between the maximum and minimum possible levels of the recording system (256 available levels in an 8-bit system) and k is a constant (2) and σ represents the system's diminishing ability to distinguish independent levels, calculated from each individual component of the imaging chain: the noise in the test stimuli (input), the measured noise of a 21' LCD system (output) and the human eye parameter.

The number of effective distinguishable levels of the system was approximately 64, empirically tested on the 21' LCD. The 64 has been found to be a typical value for the intensity levels distinguishable on a CRT faceplate under normal display viewing conditions [11].

The perceived information capacity (C) was then calculated, based on the effective pixel dimension in the image (n) and the number of effective distinguishable levels for each recording cell (m) [5, 6].

$$C = n\log_2(m) \tag{3}$$

EPIC values were finally derived in bits per steradian by:

$$EPIC = \frac{C}{\Omega} \tag{4}$$

Ω is given by:

$$\Omega = A_{im}/r^2 \tag{5}$$

A_{im} is the displayed area of the image and r viewing distance.

2.2 EPIC as an overall quality predictor

A validation experiment was carried out to test the success of the EPIC metric in predicting perceived image quality, using results from subjective tests involving 32 test scenes, replicated with various degrees of sharpness and noisiness [9].

To investigate the degree of correlation between objective (i.e. EPIC values) and subjective scales, the Pearson's correlation coefficient (r) was used [12, 13]. A degree of correlation larger than a level of significance at 1% probability level indicates that the degree of correlation between the two variables is statically significant. The correlation coefficient ranges between -1.0 (indicating perfect anti-correlation) and 1.0 (indicating perfect correlation), with 0 denoting no correlation at all.

Figure 1 presents a scatter plot of average subjective ratings versus EPIC values obtained from all scenes. A relatively successful correlation (r = 0.808) was obtained. The result confirmed that the metric is a reliable predictor of image quality when tested against average ratings. Other studies have indicated similar correlations (r = 0.91) [5].

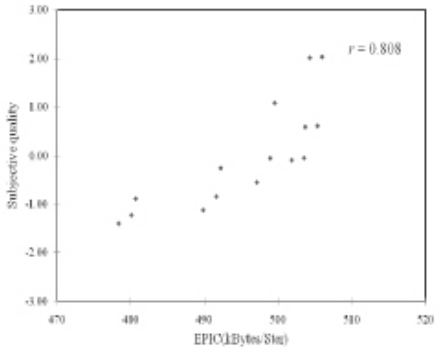


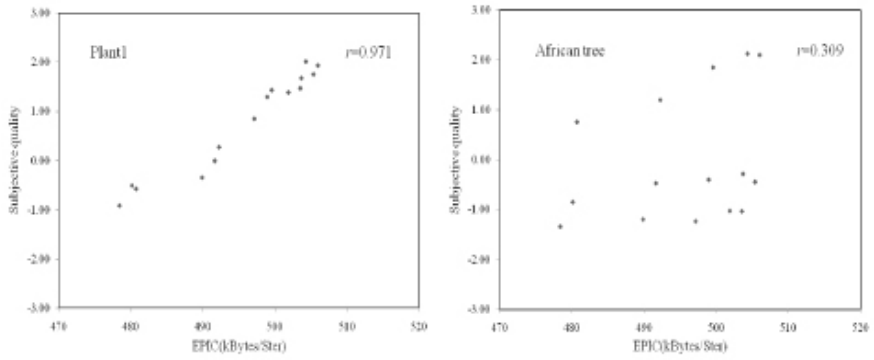
Figure 1. Comparison between EPIC values and average subjective rating from all scenes.

In general, device-dependent IQMs are shown to be unequally successful in predicting the quality of individual scenes [1, 4, 13]. Table 1 shows individual correlation coefficients for thirty two test scenes (cf. Appendix). As expected, the EPIC metric was unequally successful in predicting the quality of different images with varying scene content, i.e. the metric performed well on most of the scenes, but predicted less successfully the perceived quality of individual, non-standard looking scenes.

Image	r	Image	r	Image	r	Image	r
African tree	0.309	Baby	0.788	Bike	0.893	China town	0.814
Exercise	0.899	Formula	0.794	Glasses	0.743	Group	0.859
Human	0.753	Human2	0.447	Human3	0.883	Human4	0.753
Kids	0.817	Landscape	0.681	Landscape2	0.498	Landscape3	0.740
London Eye	0.708	London Eye2	0.691	Louvre	0.816	National gallery	0.796
Old building	0.859	Plant1	0.971	Plant2	0.811	Plant3	0.762
Plant4	0.709	Plant5	0.814	Plant6	0.782	St. Pauls	0.923
St. Pauls2	0.851	Saules	0.915	Sungsil	0.853	Yellow flower	0.743

Table 1. Correlation coefficients for each scene (the average of 32 coefficients is 0.771).

Figure 2 presents examples scenes having very different correlation coefficients: (a) Plant1 (r = 0.971), (b) African tree (r = 0.309).



(a) (b)

Figure 2. Comparison between mean EPIC values and subjective rating of scenes (a) Plant 1 (b) African tree.

3. EPIC AND OBJECTIVE SCENE CLASSIFICATION

In order to improve the performance of the EPIC, we combined Keelan's IHIF model, proposed for calibrating device-dependent metrics [14, 15], and objective scene classification, as proposed by Triantaphillidou *et al* [4]. The test scenes in this study were previously objectively classified into three groups with different susceptibility to noisiness and sharpness [16].

3.1 An image quality framework that includes EPIC and objective scene classification

Monte Carlo simulation has been adopted in image quality modelling by Keelan [7] and includes basic factors such as i) an objective metric associated with the image quality attributes in question, ii) observer sensitivity, and iii) scene susceptibility. A diagram of the proposed framework, involving the EPIC as the objective IQ metric and accounting for scene susceptibility is illustrated in Figure 3.

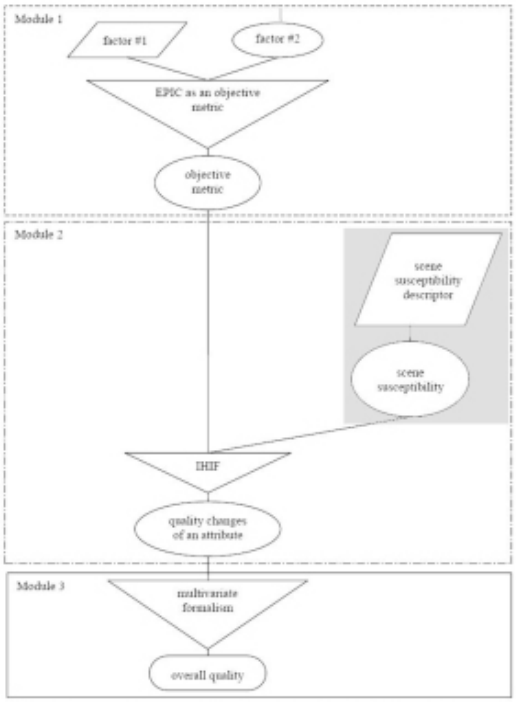


Figure 3. Diagram of the IQ framework proposed by Keelan [7], using the EPIC as the IQ metric and objective scene classification to account for scene susceptibility (part highlighted in grey).

There are three modules in the framework. An example of the implementation of the first module is illustrated in Table 2 and Table 3. EPIC values[1] for each degree of sharpness/blur and noisiness (i.e. IQ attributes - factor #2 in Figure 3) were derived, as listed in Step 1 of Table 2. The resulting data was then analyzed using the top-down method[2] [9] to derive EPIC values for each individual attribute (i.e. blur, noise), as listed in Step 2 of Table 2. All derived EPIC values for each step in blur and noise are listed in the "Derived EPIC values" in Table 3.

The "Derived EPIC values" was then used to derive the "Quality differences" for the individual attributes, as listed in Table 3, i.e. $\Delta Q = Q - Q_r$, where Q_r was the reference quality for each attribute. Values for Blur1 and Noise1 in Table 3 were used as the reference quality values for each attribute.

Step 1		Step 2	
Varying Attributes	EPIC values	Varying Attribute	Derived EPIC values for blur only
Blur1Noise1	505.99	Blur1	497.79
Blur1Noise2	500.61		
Blur1Noise3	486.79		
Blur2Noise1	504.31	Blur2	496.15
Blur2Noise2	498.95		
Blur2 Noise3	485.19		

Table 2. Example of individual attribute scaling in the EPIC.

Varying Attribute	Derived EPIC values	Quality differences	Varying Attribute	Derived EPIC values	Quality differences
Blur1	497.79	0.00	Noise1	496.56	0.00
Blur2	496.15	-1.64	Noise2	491.30	-5.27
Blur3	491.48	-6.31	Noise3	477.78	-18.79
Blur4	484.30	-13.49			
Blur5	473.00	-24.80			

Table 3. Derived EPIC values and quality differences scaling.

The second module used the Integrated Hyperbolic Increment Function (IHIF) to account for scene susceptibility.

Objective scene classification previously classified the test scenes used in this work into three groups with: 1) low susceptibility to sharpness and high susceptibility to noisiness 2) average susceptibility to sharpness and noisiness, 3) high susceptibility to sharpness and low susceptibility to noisiness (cf. Appendix) [16]. The classification first employed objective scene descriptors (objective measures derived from scenes to describe individual scene properties), which were found to correlate with subjective criteria on scene susceptibility to sharpness/blur and noisiness; subsequently k-means clustering was used to group the scenes. In Figure 3 the grey box denotes the objective scene classification as proposed in [4] and carried out in [16]. This implementation allows the fully automatic calculation of the metric values, i.e. values are calibrated for individual scenes that are classified objectively, instead of being classified by inspection.

[1] Keelan [7] suggests the quality changes in JNDs arising from individual attributes of image quality.
[2] It is the process of deriving individual attribute scales from overall quality assessments [9].

The IHIF was then employed to account for the quality changes arising for sharpness/blur and for noisiness in each group of scenes. The IHIF value was calculated with free parameters, as described in [14, 15].

For the group of scenes having 'middle' scene susceptibility [15], i.e. average looking scenes as in group 2) above, the IHIF value was approximately equal to 1.0. For the 'least susceptible' group, i.e. groups 1) or 3) above depending on the attribute, the IHIF value was 0.73. For the 'most susceptible' group, i.e. group 3) or 1) above depending on the attribute, the IHIF value was 1.36.

An example of the use of the IHIF value for calibrating the quality differences for the group with least susceptibility to sharpness/blur is illustrated in Table 4. The calibrated quality differences in the last column are obtained by multiplying the quality differences with the IHIF value.

Varying Attribute	Quality differences	IHIF value	Calibrated quality differences in Group 1
Blur1	0.00	0.73	0.00
Blur2	-1.64		-1.20
Blur3	-6.31		-4.62
Blur4	-13.49		-9.87
Blur5	-24.80		-18.14

Table 4. Example of using IHIF value for calibrating quality differences.

The calibrated quality differences for each attribute and each group is shown in Table 5.

	Varying Attribute	Calibrated quality differences	Varying Attribute	Calibrated quality differences
Group 1	Blur1	0.00	Noise1	0.00
	Blur2	-1.20	Noise2	-7.16
	Blur3	-4.62	Noise3	-25.53
	Blur4	-9.87		
	Blur5	-18.14		
Group 2	Blur1	0.00	Noise1	0.00
	Blur2	-1.60	Noise2	-5.14
	Blur3	-6.16	Noise3	-18.33
	Blur4	-13.16		
	Blur5	-24.19		
Group 3	Blur1	0.00	Noise1	0.00
	Blur2	-2.23	Noise2	-3.85
	Blur3	-8.58	Noise3	-13.74
	Blur4	-18.33		
	Blur5	-33.70		

Table 5. Quality differences values at each group after applying IHIF value.

The last module is a multivariate formulation. The quality changes from individual quality attributes in each group integrate into the overall quality rating. The multivariate formulation is expressed as a root-mean-square (RMS) sum [17]:

$$\Delta Q_m = -\left(\sum_i (-\Delta Q_i)^{n_m}\right)^{1/n_m} \quad (6)$$

where, the ΔQ_i is the quality change arising from the i^{th} attribute (i.e. sharpness/blur and noisiness), ΔQ_m is the overall quality change and n_m is 2.

Figure 4 presents the overall quality changes in each group. Each label in the x axis represents a specific level of distortion in blur (B) and noise (N).

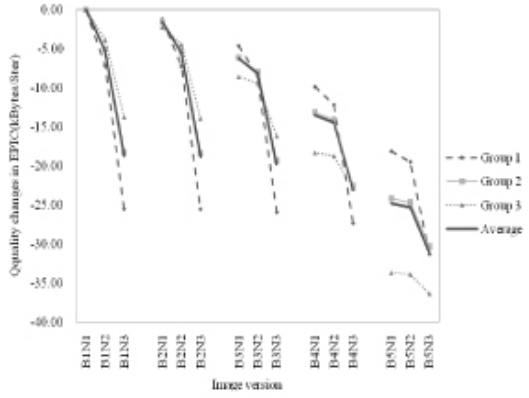


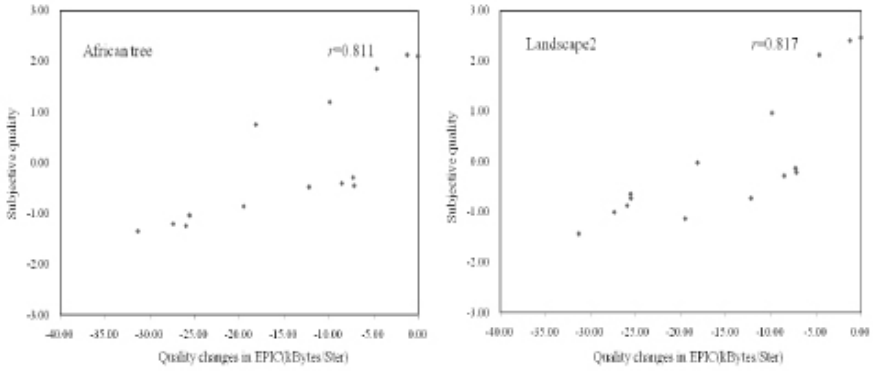
Figure 4. Overall quality changes for each group and average.

3.2. Objective vs subjective quality

The success of the IQ framework implementation was evaluated by correlating subjective scaled values from [9], this time with the calibrated quality differences obtained in section 3.1. The Pearson's correlation coefficient (r) was again used to examine the success of the correlation. Table 6 lists the correlation coefficients for all scenes. Results from the implementation indicated that, generally the quality predictions of the metric were improved. The mean correlation coefficient between subjective and objective ratings for the 32 scenes increased from 0.771 (in Table 1) to 0.869 (in Table 6). Most importantly, the calibrated objective values were shown to correlate equally well with subjective quality scales of standard and non-standard looking scenes. Improvements in correlations were particularly shown in scenes belonging to Group 1, with low susceptibility to sharpness and high susceptibility to noisiness. For example, the correlation coefficient for scene 'African tree' increased from 0.309 to 0.811and of 'Landscape 2' from 0.498 to 0.817 (Figure 5). Overall, the findings indicated that a device-dependent, scene-dependent IQM can be achieved.

	Image	r	Image	r	Image	r
Group 1	African tree	0.811	Human	0.830	Human2	0.854
	Human4	0.742	Landscape	0.862	Landscape2	0.817
	Landscape3	0.828	Plant2	0.845	Plant3	0.860
	Plant4	0.872	Plant5	0.806	Plant6	0.811
Group 2	Baby	0.907	China town	0.931	Exercise	0.835
	Formula	0.943	Glasses	0.912	Kids	0.913
	London Eye	0.856	London Eye2	0.859	Louvre	0.896
	Old building	0.891	Plant1	0.855	Yellow flower	0.885
Group 3	Bike	0.923	Group	0.897	Human3	0.892
	National gallery	0.871	St. Pauls	0.916	St. Pauls2	0.912
	Saules	0.872	Sungsil	0.896		

Table 6. Correlation coefficients in each class (the mean of 32 coefficient is 0.869).



(a) (b)

Figure 5. Comparison between objective scales and subjective rating of scenes (a) African tree, (b) Landscape2.

4. CONCLUSION

This study puts forward the idea of device-dependent and scene-dependent IQ metric, where metric values are calibrated to account for the susceptibility of the individual scenes for which metric values are calculated. Further, the scene susceptibility is estimated using objective scene classification that relates it to scene (image) features and inherent properties and thus metric calculations can be fully automated.

In a case study we used the EPIC as the device dependent metric. First an EPIC validation experiment was carried out, using results from subjective image quality tests that involved thirty two test scenes, replicated with various degrees of sharpness and noisiness. An objective quality scale was derived which represented the quality of the sharpness/blur and noise variations. The metric was found to be a good predictor of image quality when tested against average ratings from all scenes. However, as expected, the metric predicted less successfully the perceived quality of individual, non-standard scenes with different susceptibilities to sharpness and noisiness.

Then a modular image quality framework was implemented using the EPIC metric. The aim was to simultaneously predict the quality of images that vary in sharpness/blur and noise levels, while accounting for the susceptibility of different images with varying scene content to sharpness and noisiness. The framework was originally derived using Monte Carlo simulation and accounts for observer sensitivity and scene susceptibility. The later is done by classifying scenes into a small number of groups. In our implementation only the scene susceptibility was considered. The scene classification that was employed in the framework uses objective scene descriptors that are found to correlate with subjective criteria on scene susceptibility to sharpness and noisiness and k-mean clustering to derive three groups.

Results from framework implementation indicate that the quality predictions of the EPIC metric were improved. The mean correlation coefficient between subjective and objective ratings increased from 0.771 to 0.869. Most importantly, the metric scales were shown to correlate equally well with subjective quality scales of standard and non-standard scenes. The findings indicated that a device-dependent and scene-dependent IQM can be achieved.

REFERENCES

[1] Trinataphillidou, S. and Jacobson, R. E. [Colour Image Science: Metric approaches to image quality], Wiley and sons, London, Ch.18 (2002).

[2] Fairchild, M. D. "Image quality measurement and modeling for digital photography," International Congress on Imaging Science, Tokyo, (2002).

[3] Johnson, G. M. [Measuring images: differences, quality and appearance], PhD Thesis, New York, RIT, (2003).

[4] Triantaphillidou, S., Allen, E. and Jacobson, R. E. "Image Quality of JPEG vs JPEG 2000 Part 2: Scene Dependency, Scene Analysis and Classification," Journal of Imaging Science and Technology, Vol. 51(3), 259-271 (2007).

[5] Jenkin, R., Triantaphillidou, S. and Richardson, M. "A Effective Pictorial Information Capacity as an Image Quality Metric", Proc. SPIE 6494, (2007).

[6] Jenkin, R. and Richardson, M. "Comparison between the Effective Pictorial Information Capacities JPEG 6b and 2000", Opto-Ireland 2005: Imaging and Vision, Vol. 13, 5823, (2005).

[7] Keelan, B. [Handbook of Image Quality: Monte Carlo Simulation of System Performance], Marcel Dekker INC, New York, Ch.28, (2003).

[8] Shannon, C.E. "A Mathematical Theory of Communication," The Bells System Technical Journal, Vol. 27, 623-656 (1948).

[9] Oh, K. H., Triantaphillidou, S. and Jacobson, R. E. "Perceptual image attribute scales derived from overall image quality", Proc. SPIE 7242, (2009).

[10] Barten, P. G. J. "Physical model for the contrast sensitivity of the human eye," Proceeding SPIE: Human Vision, Visual Processing, and Digital Display III, Vol. 1666, 57-72 (1992).

[11] Triantaphillidou, S. [Aspects of Image Quality in the Digitisation of Photographic Collection], PhD Thesis, London, University of Westminster, (2001).

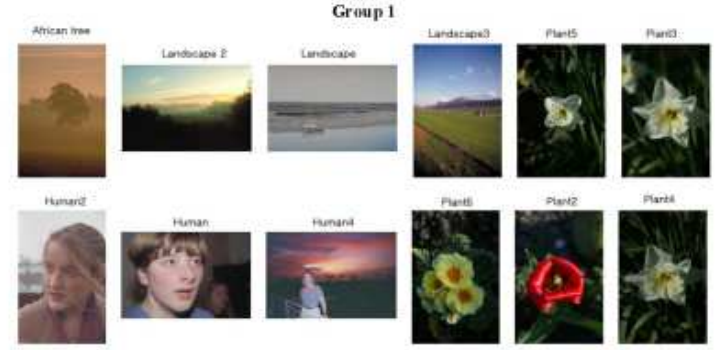
[12] De Freitas Zampolo, R and Seara, R. "A comparison of image quality metric performances under practical conditions", IEEE International Conference on Image Processing 2005, Vol. 3, 1192-1195 (2005).

[13] Ford, A. M. [The relationship Between Image quality and Still image compression], PhD Thesis, London, University of Westminster, (1997).

[14] Keelan, B. [Handbook of Image Quality: A General Equation to Fit Quality Loss Functions], Marcel Dekker INC, New York, Ch 9 (2003).
[15] Keelan, B. [Handbook of Image Quality: A detailed example of objective metric design], Marcel Dekker INC, New York, Ch.14 (2003).
[16] Oh, K. H., Triantaphillidou, S. and Jacobson, R. E. "Scene classification with respect to image quality measurements", Proc. SPIE 7629, (2010).
[17] Keelan, B. [Handbook of Image Quality: Predicting Overall Quality from Image Attributes], Marcel Dekker INC, New York, Ch 11 (2003).

APPENDIX – TEST IMAGES IN THREE GROUPS

Group 1



Group 2

Group 3

References

1. Trinataphillidou, S. and Jacobson, R. E. Metric approaches to image quality, in: MacDonald, L. W. and Luo, M. R. (editors), *Colour image science*, Chichester, John Wiley & sons, 2002, pp. 371-392.

2. LEE, Hsien-Che. Image quality, in: *Introduction to color imaging science*, Cambridge, Cambridge Press, 2005, pp. 564-584.

3. Yendrikhovskij, S. Towards perceptually optimal colour reproduction of natural scenes, in: MacDonald, L. W. and Luo, M. R. (editors), *Colour imaging- vision and technology*, Chichester, John Wiley & sons, 1999, pp. 363-382.

4. Wang, Z. Why is image quality assessment so difficult? *IEEE International Conference on Acoustics, Speech & Signal Processing*, Vol. 5, pp. 3313-3316, 2002.

5. Fairchild, M. D. Image quality measurement and modeling for digital photography, [Presentation Slides] *International Congress on Imaging Science*, Tokyo, 2002.

6. Engeldrum, P. G. Image quality modeling: Where are we? *IS&T PICS: Image processing, image quality, image capture, systems conference*, Vol. 52, pp. 251-255, 1999.

7. Daly, S. The visible differences predictor: An algorithm for the assessment of image fidelity, in: Watson, A. B. (editor), *Digital images and human vision*, Cambridge, MIT Press, 1993, pp. 179-206.

8. Johnson, G. M. Measuring images: differences, quality and appearance, PhD Thesis, New York, Rochester Institute of Technology, 2003.

9. Fairchild, M. D. and Johnson, G. The iCAM framework for image appearance, image differences, and image quality, *Journal of Electronic Imaging*, Vol. 13, pp. 126-138, 2004.

10. Triantaphillidou, S., Allen, E. and Jacobson, R. E. Image quality of JPEG vs JPEG 2000 part 2: Scene dependency, scene analysis and classification, *Journal of Imaging Science and Technology*, Vol. 51, pp. 259-271, 2007.

11. Engeldrum, P. The process of scaling and some practical hints, in: *Psychometric scaling*, Winchester, Imcotek Press, 2000, pp. 19-42.

12. Corey, G. P., Clayton, M. J. and Cuprey, K. N. Scene dependence of image quality, *Photographic Science and Engineering*, Vol. 27, pp. 9-13, 1982.

13. Frieser, H. and Biedermann, K. Experiments on image quality in relation to the modulation transfer function and graininess of photographs, *Photographic Science and Engineering*, Vol. 7, pp. 28-34, 1963.

14. Keelan, B. W. Scene and observer variability, in: Keelan, B. W. (editor), *Handbook of image quality*, New York, Marcel Dekker, 2003, pp.129-148.

15. Hunt, R. W. G. Image structure in colour photography, in: *The reproduction of colour* 6th, Chichester, John Wiley & sons, 2004, pp. 300-334.

16. Jones, L. A. and Condit, H. R. The brightness scale of exterior scenes and the computation of correct photographic exposure, *Journal of the Optical Society of America*, Vol. 31, pp. 651-678, 1941.

17. ISO 20462-3: 2005. Photography-psychophysical experimental methods for estimating image quality, International Organization for Standardization, 2005.

18. Allen, E., Triantaphillidou, S. and Jacobson, R. E. Image quality compression between JPEG and JPEG 2000 I: Psychophysical measurements, *Journal of Imaging Science and Technology*, Vol. 51, pp. 248-258, 2007.

19. Teeselink, I., Bilmmaert, F. and Ridder, H. Image categorization, *Journal of Imaging Science and Technology*, Vol. 44, pp. 552-559, 2000.

20. Mojsilovic, A. and Rogowitz, B. Capturing image semantics with low-level descriptors, *International Conference on Image Processing*, pp. 18-21, 2001.

21. Yendrikhovskij, S. Image quality and colour categorisation, in: MacDonald, L. W. and Luo, M. R. (editors), *Colour image science*, Chichester, John Wiley & sons, 2002, pp. 393-419.

22. Szummer, M. and Picard, R. W. Indoor-outdoor image classification, *IEEE Intl Workshop on Content-based Access of Image and Video Databases*, pp. 42-51, 1998.

23. Vailaya, A., Jain, A. and Zhang, Hong Jiang. On image classification: City images vs landscapes, *International Journal of Pattern Recognition*, Vol. 31, pp. 1921-1935, 1998.

24. Vailaya, A., Figueiredo, M., Jaina, A. and Zhang, Hong Jiang. H. A Bayesian framework for hierarchical semantic classification of vacation images, *IEEE International Conference on Multimedia Computing and Systems*, Vol. 3656, pp. 415-426, 1998.

25. ISO 12231: 2005. Photography-electronic still picture imaging-vocabulary, International Organization for Standardization, 2005.

26. Keelan, B. W. Can image quality be usefully quantified? in: Keelan, B. W. (editor), *Handbook of image quality*, New York, Marcel Dekker, 2003, pp. 3-18.

27. Jacobson, R. E. An evaluation of image quality metrics, *Journal of Photographic Science*, Vol. 43, pp. 7-16, 1995.

28. Silverstain, D. A. and Farrell, J. E. The relationship between image fidelity and image quality, *Proceeding ICIP-96*, Vol. 1, pp. 881-884, 1996.

29. Klein, S. A. Image quality and image compression, in: Watson, A. B. (editor), *Digital images and human vision*, Cambridge, MIT Press, 1993, pp. 73-88.

30. Berns, R. S. Measuring color quality, *Billmeyer and Saltzman's principles of color technology* , New York, Wiley-Interscience, 2000, pp. 107-130.

31. Engeldrum, P. G. A framework for image quality models, Massachusetts, *Journal of Imaging Science and Technology*, Vol. 39, pp. 312-318, 1995.

32. Ford, A. M. The relationship between image quality and still image compression, PhD Thesis, London, University of Westminster, 1997.

33. Triantaphillidou, S. Aspects of image quality in the digitisation of photographic collection, PhD Thesis, London, University of Westminster, 2001.

34. Jacobson, R. E. Approaches to total quality for the assessment of imaging systems. *Information Services & Use*, Vol. 13, pp. 235-246, 1993.

35. Bartleson, C. The combined influence of sharpness and graininess on the quality of colour prints, *Journal Photographic Science*, Vol. 30, pp. 33-38, 1982.

36. Engeldrum, P. G. The Image Quality Circle, in: *Psychometric scaling*, Winchester, Imcotek Press, 2000, pp. 5-18.

37. Biedermann, H. and Frieser, K. Experiments on image quality in relation to the modulation transfer function and graininess of photographs, *Photographic Science and Engineering*, Vol. 7, pp. 28-33, 1963.

38. Johnson, G. M. and Fairchild, M. Sharpness rule, *IS&T Color Imaging Conference*, Vol. 8, pp. 24-30, 2000.

39. Stevens, S. S. On the theory of scales and measurements, *Science*, Vol. 103, pp. 677-680, 1946.

40. Besore, C. Testing and evaluation, in: Woodlief, T. (editor), *SPIE handbook of photographic science and engineering*, New York, Wiley-Interscience, 1973, pp. 1023-1132.

41. ISO 20462-2: 2005. Photography-psychophysical experimental methods for estimating image quality, International Organization for Standardization, 2005.

42. ITU-R. BT 500-12. Methodology for subjective assessment of the quality of television pictures, International Telecommunication Union - Radiocommunication Sector, 2009.

43. Cookingham, R. E. Calibrated psychometrics using quality rulers, in: Keelan, B. W. (editor), *Handbook of image quality*, New York, Marcel Dekker, 2003, pp. 87-100.

44. MacDonald, L. and Jacobson, R. E. Assessing image quality, in: MacDonald, L. W. (editor), *Digital heritage: Applying digital imaging to cultural heritage*, Oxford, Elservier, 2006, pp. 351-373.

45. ISO 20462-1: 2005. Photography-psychophysical experimental methods for estimating image quality, International Organization for Standardization, 2005.

46. Bartleson, C. J. Measuring differences, *Optical Radiation Measurements*, Academic Press, Vol. 5, pp. 441-489, 1984.

47. Westerink, J. H. D. M. and Roufs, J. A. J. Subjective image quality as a function of viewing distance, resolution and picture size, *SMPTE Journal*, Vol. 98, pp. 113-119, 1989.

48. Van der Zee, E. and Boesten, M. H. W. A. The influence of luminance and size on the image quality of complex scenes, *IPO Annual Progress Report*, pp. 69-75, 1980.

49. Myers, K. J. Addition of a channel mechanism to the ideal observer model, *Journal of the Optical Society of America A*, Vol. 4, pp. 2447–2457, 1987.

50. Sankaran, S., Frey, E., Gilland, K. and Tsui, B. Tsui. Optimum compensation method and filter cut off frequency in myocardial SPECT: A human observer study, *The Journal of Nuclear Medicine*, Vol. 43, pp. 432-438, 2002.

51. ISO 3664: 2000. Viewing conditions for graphic technology and photography, International Organization for Standardization, 2000.

52. Wallace, G. K. The JPEG still picture compression standard, *IEEE Transactions on Consumer Electronics*, Vol. 38, pp. xviii - xxxiv, 1992.

53. Usevitch, B. E. A tutorial on modern lossy wavelet image compression: Foundations of JPEG 2000, *IEEE Signal Processing Magazine*. Vol. 18, pp. 22–35, 2001.

54. Nijenhuis, M. R. M. Sampling and interpolation of static images: A perceptual view, PhD Thesis, Eindhoven, Eindhoven University of Technology, 1993.

55. Barten, P. G. J. Effect of various parameters on image quality, in: *contrast sensitivity of the human eye and its effects on image quality*, Bellingham, SPIE Press, 1999, pp. 175-200.

56. Kayargadde, V. Feature extraction for image quality prediction, PhD Thesis, Eindhoven, Eindhoven University of Technology, 1995.

57. Kominek, J. Introduction to fractal compression (long), [Online] 1996, http://www.faqs.org/faqs/compression-faq/part2/section-8.html.

58. Ahmada, A. J. and Null, C. H. Image quality: multidimensional problem, in: Watson, A. B. (editor), *Digital images and human vision*, Cambridge, MIT Press, 1993, pp. 141-148.

59. Farrell, J. E. Image quality evaluation, in: MacDonald, L. W. and Luo, M. R. (editors), *Colour imaging - vision and technology*, Chichester, John Wiley & sons, 1999, pp. 285-313.

60. Jacobson, R. E. Image quality: meanings, minefields and mastery, [Presentation Slides] Good Picture 2003, London, The Royal Photographic Society, 2003.

61. Crane, E. M. An objective method for rating picture sharpness: SMT acutance, *Journal of the SMPTE*, Vol. 73, pp. 643-647, 1964.

62. Gendron, R. G. An improved objective method for rating picture sharpness: CMT acutance, *Journal of the SMPTE*, Vol. 82, pp. 1009-1012, 1973.

63. Crane, E. M. Acutance and granulance, *SPIE Proceeding: Image quality*, Vol. 310, pp. 125-132, 1981.

64. ANSI/HFS 100: 1988. American national standard for human factors engineering of visual display terminal workstations, Human Factors Society, 1988.

65. Charman, W. N. and Olin, A. Image quality criteria for aerial camera system, *Photographic Science and Engineering*, Vol. 9, pp. 385-397, 1965.

66. Gerfelder, N. and Muller, W. Objective quality estimation for digital images in multimedia environment, in: MacDonald, L. W. and Luo, M. R. (editors), *Colour imaging - vision and technology*, Chichester, John Wiley & sons, 1999, pp. 339-361.

67. Beaton, R. J. and Farley, W. W. Comparative study of The MTFA, ICS, and SQRI image quality metrics for visual display systems. Armstrong Lab, Air Force Systems Command, Report AL-TR-1992–0001, 1991.

68. Barten, P. G. J. Evaluation of subjective image quality with the square-root integral method, *Journal of the Optical Society of America A*, Vol. 7, pp. 2024-2031, 1990.

69. Barten, P. G. J. Evaluation of the effect of noise on subjective image quality, *SPIE Proceeding: Human vision, visual processing and digital display II*, Vol. 1453, pp. 2-15, 1991.

70. Higgins, G. C. Image quality criteria, *Journal of Applied Photographic Engineering*, Vol. 3, pp. 53-60, 1977.

71. Topfer, K. and Jacobson, R. E. The relationship between objective and subjective image quality criteria, *Journal of Information Recording Materials*, Vol. 21, pp. 5–27, 1993.

72. Pointer, M. R. Measuring colour reproduction, *Journal of Photographic Science*, Vol. 34, pp. 81-90, 1986.

73. Zhang, X. M. and Wandell, B. A. A spatial extension to CIELAB for digital color image reproduction, *Journal of the Society for Information Display*, Vol. 5, pp. 61-63, 1997.

74. Jenkin, R., Triantaphillidou, S. and Richardson, M. A. Effective Pictorial Information Capacity as an Image Quality Metric, *SPIE Proceeding: Image quality and system performance IV*, Vol. 6494, 2007.

75. Shannon, C. E. A Mathematical Theory of Communication, *The Bells System Technical Journal*, Vol. 27, 623-656, 1948.

76. Hurter, F. and Driffield, V. C. Photochemical investigations and new method of determination of the sensitiveness of photographic plates, *The Journal of Society of Chemical Industry*, pp. 455-469, 1890.

77. Jones, L. A. On the theory of tone reproduction with graphic methods for the solution of the problem, *Journal of Franklin Institute*, Vol. 190, pp. 39-90, 1920.

78. Jones, L. A. Photographic reproduction of tone, *Journal of the Optical Society of America*, Vol. 2, pp. 232-258, 1921.

79. Jones, L. A. Psychophysics and photography, *Journal of the Optical Society of America*, Vol. 35, pp. 66-88, 1944.

80. Nelson, C. N. Tone and colour reproduction. James, T. H (editor), *The theory of photographic process*, New York, The Macmillan Company, 1966, pp. 536-577.

81. Attridge, G. G. Sensitometry, in: Jacobson, R. E. *et al* (editors), *The manual of photography* 9th, Oxford, Focal Press, 2000, pp. 218-246.

82. ISO 14524: 1999. Photography-electronic still-picture cameras-Methods for measuring opto-electronic conversion functions (OECFs), International Organization for Standardization, 1999.

83. Triantaphillidou, S. Tone reproduction, in: Allen, E. and Triantaphillidou, S. (editors), *The manual of photography* 10th, Oxford, Focal Press, 2010, pp. 377-392.

84. Ford, A. M., Jacobson, R. E. and Attridge, G. G. Characteristics of CRT monitors for the display of still digital images, *Journal of Photographic Science*, Vol. 44, pp. 23-26, 1996.

85. Sharma, G. LCDs Versus CRTs - color calibration and gamut considerations, *Proceedings of the IEEE*, Vol. 90, pp. 605-622, 2002.

86. Tamura, N., Tsumura, N. and Miyake, Y. Masking model for accurate colorimetric characterisation of LCD, *Journal of the Society for Information Display*, Vol. 11, pp. 1-7, 2003.

87. Kwak, Y. and MacDonald, L. Characterisation of a desktop LCD projector, *Displays*, Vol. 21, pp. 179-194, 2001.

88. Fairchild, M. D. and Wyble, D. Colorimetric characterisation of the Apple studio display (Flat Panel LCD), Munsell Color Science Laboratory Technical Report, Rochester Institute of Technology, 1998.

89. EN 61966-2-1:2000. Multimedia systems and equipment-colour measurement and management-part 2-1: colour management-Default RGB colour space- sRGB, European standard: IEC, 2000.

90. Adobe_system. Adobe® RGB (1998) Color image encoding, Version 2005-05, Adobe system, 2005.

91. Triantaphillidou S., Jacobson R. E. and Ford A. M. Preferred tone reproduction of images on soft displays, *ICPS'98 Proceeding: International congress on imaging science*, pp. 204-208, Belgium, 1998.

92. Fairchild, M. D. Colorimetry, in: *Color appearance models* 2nd, Chichester, John Wiley & sons, 2004, pp. 53-82.

93. Hunt, R. W. G. Objectives in colour reproduction, The *Journal of Photographic Science*, Vol. 18, pp. 205-215, 1970.

94. Guild, J. The colorimetric properties of the spectrum, *Philosophical transactions of the Royal Society of London*, Ser A, Vol. 230, pp. 149-187, 1932.

95. Wright, W. D. A re-determination of the trichromatic coefficients of the spectral colours, *Transactions of the optical society*, Vol. 30, pp. 141-164, 1928-29.

96. CIE, Proc. 8th session, Cambridge, Commission Internationale de l'Eclairage, 1931.

97. CIE Publication 15.2. Colorimetry, Commission Internationale de l'Eclairage, 1986.

98. Hunt, R. W. G. Colour standard and calculation, in: *The reproduction of colour* 6th, Chichester, John Wiley & sons, 2004, pp. 92-125.

99. CIE Supplement No.2 to CIE Publication 15: Recommendations on uniform color spaces, color-difference equations, psychometric color terms, Commission Internationale de l'Eclairage, 1978, now Part of CIE 15.2-1986: Colorimetry.

100. Fairchild, M. D. Color appearance terminology, in: Fairchild, M. D. (editor), *Color appearance models* 2nd, Chichester, John Wiley & sons, 2004, pp. 83-93.

101. Triantaphillidou, S. Colorimetry and Colour Difference Formulae, [Presentation Slides] *Lecture note for MSC Digital imaging and Photographic Imaging*, University of Westminster, London, 2004.

102. Clarke, F. J. J., McDonald, R. and Rigg, B. Modification to the JPC79 colour-difference formula, *Journal of the Society of Dyers and Colourists*, Vol. 100, pp. 128-132, 1984.

103. CIE Publication 116. Industrial colour-difference evaluation, Commission Internationale de l'Eclairage, 1995.

104. CIE Publication 142. Improvement to industrial colour-difference evaluation, Commission Internationale de l'Eclairage, 2001.

105. Luo, M. R., Cui, G. and Rigg, B. The development of CIE 2000 colour-difference formula: CIEDE2000, *Color Research and Application*, Vol. 26, pp. 340-350, 2001.

106. Bilissi, E. Aspects of image quality and the internet, PhD Thesis, London, University of Westminster, 2004.

107. Wyszecki, G. and Stiles, W. S. Colorimetry, in: *Color science: concepts and methods, quantitative data and formulae*, Chichester, John Wiley & sons, 1982, pp. 117-248.

108. Witt, K. CIE color difference metrics, in: János Schanda (editor), *Colorimetry: understanding the CIE system,* New York, Wiley-Interscience, 2007, pp. 79-100.

109. Triantaphillidou, S. Introduction to colour science, in: Allen, E. and Triantaphillidou, S. (editors), *The manual of photography* 10th, Oxford, Focal Press, 2010, pp. 77-102.

110. ISO 12233: 2000. Photography-electronic still picture cameras-Resolution measurements, International Organization for Standardization, 2000.

111. Ray, S. Optical aberrations and lens performance, in: Jacobson, R. E. *et al* (editors), *The Manual of Photography* 9th, Oxford, Focal Press, 2000, pp. 72-82.

112. Burns, P. Evaluating digital scanner and cameras imaging performance for digital collections, [Presentation Slides] Half-day short course, London, The Royal Photographic Society, 2006.

113. Jenkin, R. Noise, sharpness resolution and information, in: Allen, E. and Triantaphillidou, S. (editors), *The manual of photography* 10th, Oxford, Focal Press, 2010, pp. 433-456.

114. Jenkin, R. On the application of the modulation transfer function to discrete imaging system, PhD Thesis, London, University of Westminster, 2002.

115. Jenkin, R. Image and image formation, in: Allen, E. and Triantaphillidou, S. (editors), *The manual of photography* 10th, Oxford, Focal Press, 2010, pp. 119 - 138.

116. Lehmbeck, D. R. and Urbach, J. C. Image quality for scanning, in: Marshall, G. F. (editor), *Handbook of optical and laser scanning*, New York, Marcel Dekker, 2001, pp. 139 -264.

117. Triantaphillidou, S., Jacobson, R. E. and Jenkin, R. An evaluation of MTF determination methods for 35mm scanners, *IS&T PICS: Image Processing, image quality, image capture, systems conference*, Vol. 52, pp. 231-235, 1999.

118. Dainty, J. C. and Shaw, R. The modulation transfer function, in: *Image science: principles, analysis and evaluation of photographic-type imaging processes*, London, Academic Press, 1974, pp. 232-275.

119. Dainty, J. C. Methods of measuring the modulation transfer function of photographic emulsions. *Optica Acta*, Vol. 18, pp. 795-813, 1971.

120. Sensiper, D., Boreman, G. D. and Ducharme, A. D. Modulation transfer function testing of detector arrays using narrow-band laser speckle, *Optical Engineering*, Vol. 32, pp. 395-400, 1993.

121. Van Metter, R. Measurement of MTF by noise power analysis of –dimensional white noise patterns. *The Journal of Photographic Science*, Vol. 38, pp. 144-147, 1990.

122. Cao, F., Guichard, F. and Hornung, Hervé. Measuring texture sharpness of a digital camera, *SPIE Proceeding: Digital photography V*, Vol. 7250, 2009.

123. Axford, N. R. Theory of image formation, in: Jacobson, R. E. *et al* (editors), *The manual of photography* 9th, Oxford, Focal Press, 2000, pp. 393-412.

124. Reichenbach, S., Park, S. and Narayanswamy, R. Characterizing digital image acquisition devices, *Optical Engineering*, Vol. 30, pp. 170-177, 1991.

125. Burns, P. Slanted-edge MTF for digital camera and scanner analysis, *IS&T PICS: Image processing, image quality, image capture, systems conference*, Vol. 53, pp. 135-138, 2000.

126. ISO 15739: 2003. Photography-electronic still-picture imaging-Noise measurements, International Organization for Standardization, 2003.

127. Axford, N. R. Image and information, in: Jacobson, R. E. *et al* (editors), *The manual of photography* 9th, Oxford, Focal Press, 2000, pp. 413-427.

128. Georgeson, M. A. and Sullivan, G. D. Contrast constancy: deblurring in human vision by spatial frequency channels, *The Journal of Physiology*, Vol. 252, pp. 627–656, 1975.

129. Barten, P. G. J. Physical model for the contrast sensitivity of the human eye, *SPIE Proceeding: Human vision, visual processing and digital display III*, Vol. 1666, pp. 57-72, 1992.

130. Barten, P. G. J. Model for the spatial contrast sensitivity of the eye, in: *contrast sensitivity of the human eye and its effects on image quality*, Bellingham, SPIE Press, 1999, pp. 27-66.

131. Keelan, B. W. Monte Carlo Simulation of system performance, in: Keelan, B. W. (editor), *Handbook of image quality*, New York, Marcel Dekker, 2003, pp. 413-422.

132. Theodoridis, S. and Koutroumbas, K. Introduction, in: *Pattern recognition* 2nd, London, Elsevier, 2003, pp. 1-12.

133. Theodoridis, S. and Koutroumbas, K. Feature generation II, in: *Pattern recognition* 2nd, London, Elsevier, 2003, pp. 269-320.

134. Gonzalez, R. C. and Woods, R. E. Representation and description, in: *Digital image processing* 2nd, London, Prentice Hall, 2002, pp. 643-692.

135. Pratt, W. K. Image feature extraction, in: *Digital image processing*, Chichester, John Wiley & sons, 1991, pp. 557-596.

136. MathWorks. MATLAB, The MathWorks Inc, 2007.

137. Pratt, W. K. Edge detection, in: *Digital image processing*, Chichester, John Wiley & sons, 1991, pp. 491-556.

138. Gonzalez, R. C. and Woods, R. E. Image segmentation, in: *Digital image processing* 2nd, London, Prentice Hall, 2002, pp. 567-642.

139. Umbaugh, S. E. Segmentation and edge/line detection, in: *Computer imaging: digital image analysis and processing*, London, CRC Press, 2005, pp. 121-200.

140. Duda, R. O., Hart, P. E. and Stork, D. G. Introduction, in: *Pattern classification*, Chichester, John Wiley & sons, 2001, pp. 1-19.

141. Friedman, M. and Kandel, A. Feature selection, in: *Introduction to pattern recognition: Statistical, structural, neural, and fuzzy logic approaches*, London, Imperial College Press, 1999, pp. 141-166.

142. Theodoridis, S. and Koutroumbas, K. Feature selection, in: *Pattern recognition* 2nd, London, Elsevier, 2003, pp. 163-206.

143. Rees, D. G. Hypothesis testing, in: *Essential statistics*, London, Chapman & Hall, 2001, pp. 139-160.

144. Rees, D. G. Correlation of quantitative variables, in: *Essential statistics*, London, Chapman & Hall, 2001, pp. 211-230.

145. Cohen, S. S. One sample with two variables, in: *Practical statistics*, London, Edward Arnold, 1988, pp. 62-85.

146. Jain, A. K. Image analysis and computer vision. in: *Fundamentals of digital image processing*, Prentice hall, 1989, pp. 342-430.

147. Theodoridis, S. and Koutroumbas, K. Clustering: basic concepts, in: *Pattern recognition* 2nd, London, Elsevier, 2003, pp. 397-428.

148. Duda, R. O., Hart, P. E. and Stork, D. G. Unsupervised learning and clustering, in: *Pattern classification*, Chichester, John Wiley & sons, 2001, pp. 517-600.

149. Wikipedia, the free encyclopaedia. [Online] Wikimedia Foundation Inc, May 2009, http://en.wikipedia.org/wiki/K-means_clustering.

150. Catherine, A. S. and Gareth, M. J. Finding the number of clusters in a data set: An information theoretic approach, *Journal of the American Statistical Association*, Vol. 98, pp. 750-763, 2003.

151. Bala, R. Device characterisation, in: Gaurav Sharma (editor), *Digital color imaging handbook*, London, CRC Press, 2003, pp. 269-384.

152. LEE, Hsien-Che. Device calibration, in: *Introduction to color imaging science*, Cambridge, Cambridge Press, 2005, pp. 387-414.

153. Canon. EOS-1Ds digital instruction manual, Tokyo, Canon, 2002.

154. Ray, S. Types of camera, in: Jacobson, R. E. *et al* (editors), *The manual of photography* 9th. Oxford, Focal Press, 2000, pp. 104-130.

155. Canon. [Online] 2006, http://www.canon.co.uk.

156. Ray, S. Photographic light sources, in: Jacobson, R. E. *et al* (editors), *The manual of photography* 9th, Oxford, Focal Press, 2000, pp. 16-38.

157. EN 61966-9:2004. Multimedia systems and equipment-colour measurement and management-part 9: digital camera, European standard: IEC, 2004.

158. Orava, J. and Jaaskelainen, T. Color errors of digital camera, *Color Research and Application*, Vol. 29, pp. 217-221, 2004.

159. GretagMacbeth. Color-Eye® 7000A spectrophotometer, GretagMacbeth.

160. Stokes, M., Fairchild, M. D. and Berns, R. S. Colorimetrically quantified visual tolerances for pictorial images, *Proceeding TAGA conference*, Vol. 2, pp. 757-777, 1992.

161. Uroz, J., Luo, M. R. and Morovic, J. Perception of colour differences in large printed images, in: MacDonald, L. W. and Luo, M. R. (editors), *Colour image science*, Chichester, John Wiley & sons, 2002, pp. 49-73.

162. Green, P. Colorimetry and colour difference, in: Green, P. and MacDonald, L. W. (editors), *Colour engineering: achieving device independent colour*, Chichester, John Wiley & sons, 2002, pp. 49-78.

163. Bilissi, E., Jacobson, R. E. and Attridge, G. G. Just noticeable differences and acceptability of sRGB images displayed on a CRT monitor, *Imaging Science Journal*, 2008, Vol. 56, pp. 189-200.

164. Kodak. [Online] 2006, http://www.kodak.com.

165. Scion Corporation, [Online] 2006, http://www.scioncorp.com.

166. Sharma, A. Introduction, in: *Understanding color management*, New York, Thomson/Delmar Learning, 2004, pp. 1-48.

167. Hung, P. C. Colorimetric calibration in electronic imaging devices using a look-up tables model and interpolations, *Journal of Electronic Imaging*, Vol. 2, pp. 53-61, 1993.

168. Kang, H. R. Color scanner calibration, *Journal of Imaging science and Technology*, Vol. 36, pp. 162-170, 1992.

169. Hong, G., Luo, R. and Rhodes, P. A study of digital camera colorimetric characterisation based on polynomial modeling, *Color Research and Application*, Vol. 26, pp. 76-84, 2001.

170. Cheung, V. and Westland, S. A comparative of characterisation of colour cameras by means of neural networks and polynomial transform, *Journal of Coloration Technology*, Vol. 120, pp. 19-25, 2004.

171. Pointer, M., Attridge, G. G. and Jacobson, R. E. Practical camera characterisation for colour measurement, *The Imaging Science Journal*, Vol. 43, pp. 63-80, 2001.

172. Howell, D. C. Describing and exploring data. *Statistical Methods in Psychology*, Duxbury, 2001, Vol. 2, pp. 15-72.

173. Kang, H. R. RGB color space, in: *Computational Color Technology*, Bellingham, SPIE Press, 2006, pp. 77-101.

174. Kang, H. R. Color input devices, in: *Color Technology for Electronic Imaging Devices*, Bellingham, SPIE Press, 1997, pp. 272-294.

175. ISO 17321-1: 2006. Graphic technology and photography-Colour characterisation of digital still camera, International Organization for Standardization, 2006.

176. Wikipedia, the free encyclopaedia. [Online] Wikimedia Foundation Inc, March 2010, http://en.wikipedia.org/wiki/Pixel.

177. Farrell, J. E., Parmar, M., Catrysse, P. and Wandell, B. Digital camera simulation, in *handbook of digital imaging*, John Wiley & sons (in preparation) [Online] 2006, http://www.imageval.com/public/Products/ISET/ResearchPapers.html.

178. EIZO. ColorEdge CG210-N Specifications, EIZO, 2008.

179. EIZO. ColorNavigator Ver 4.1. : EIZO NANAO, 2005.

180. KONICA-MINOLTA, [Online] 2006, http://www.konicaminolta.com.

181. EN: 61966-4: 2000. Multimedia systems and equipment-colour measurement and management-part 4: Equipment using liquid crystal display panels, European standard: IEC, 2000.

182. Pearson, D. E. *Transmission and display of pictorial information*, London, Pentech Press, 1975.

183. Gibson, J. and Fairchild, M. Colorimetric characterisation of three computer displays (LCD and CRT), Munsell Color Science Laboratory Technical Report, Rochester Institute of Technology, 2000.

184. Kimpe, T. and Sneyders, Y. Effect of non-uniformity on DICOM GSDF compliance of medical displays, *International Journal of Computer Assisted Radiology and Surgery*, Vol. 1, pp. 35-36, 2006.

185. Berns, R. S., Gorynski, M. E. and Motta, R. J. CRT colorimetry. Part II: Metrology, *Color Research and Application*, Vol. 18, pp. 315-325, 1993.

186. Ford, A. M., Jacobson, R. E. and Attridge, G. G., Assessment of a CRT Display System, *The Journal of Photographic Science*, Vol. 44, pp. 147 – 154, 1996.

187. Parulski, K. and Spaulding, K. Color image processing for digital cameras, in: Gaurav Sharma (editor), *Digital color imaging handbook*, London, CRC Press, 2003, pp. 727-758.

188. Applied Image. [Online] Applied Image, 2007, http://www.aig-imaging.com.

189. Keelan, B. W. and Pagano, D. M. External projection lens aperture, US Patent 5-537-166, 1996.

190. LosBurns, [Online] 2009, http://losburns.com/imaging/software/index.html.

191. Triantaphillidou, S. and Jacobson, R. E. Measurements of the modulation transfer function of image displays, *Journal of Imaging Science and Technology*, Vol. 48, pp. 58-65, 2004.

192. Jin, E. W., Keelan, B. W. and Chen, J. Softcopy quality ruler method: Implementation and validation, *SPIE Proceeding: Image quality and system performance VI*, Vol. 7242, 2009.

193. Smoyer, E. M., Taplin, L. A. and Berns, R. S. Experimental Evaluation of Museum Case Study Digital Camera Systems, *Proceedings of the IS&T Archiving Conference*, pp. 85-90, 2005.

194. Karniyati. Evaluating a camera for archiving cultural heritage, Munsell Color Science Laboratory Technical Report, Rochester Institute of Technology, 2005.

195. Roehrig, H., Krupinski, E. A., Chawla, A. S., Fan, J., Gandhi, K., Furukawa, T. and Ohashi, M. Noise of LCD display system, *International Congress Series*, Vol. 1256, pp. 162-168, 2003.

196. Roehrig, H., Krupinski, E. A., Chawla, A. S., Fan, J. and Gandhi, K. Spatial noise and threshold contrasts in LCD displays, *SPIE Proceeding: Image Perception, Observer Performance, and Technology Assessment*, Vol. 5034, 2003.

197. Ray, S. Camera exposure determination, in: Jacobson, R. E. *et al* (editors), *The manual of photography* 9th, Oxford, Focal Press, 2000, pp. 310-335.

198. Elizabeth, A. Digital image workflow, in: Allen, E. and Triantaphillidou, S. (editors), *The manual of photography* 10th, Oxford, Focal Press, 2010, pp. 457 - 474.

199. Elizabeth, A. Digital image manipulation, in: Langford, M. and Bilissi, E. (editors), *Langford's advanced photography* 7th, Oxford, Focal Press, 2008, pp. 229-266.

200. Garrett, J. and Fairchild, M. Visual psychophysics and color appearance, in: Gaurav Sharma (editor), *Digital color imaging handbook*, London, CRC Press, 2003, pp. 115-171.

201. Engeldrum, P. G. Thresholds and Just Noticeable Differences, in: *Psychometric scaling*, Winchester, Imcotek Press, 2000, pp. 53-78.

202. Hunt, R. W. G. The effect of daylight and tungsten light adaptation on colour perception, *Journal of the Optical Society of America*, Vol. 40, pp. 362-371, 1950.

203. Engeldrum, P. G. Indirect interval scaling-Category scaling methods, in: *Psychometric scaling*, Winchester, Imcotek Press, 2000, pp. 123-138.

204. Gulliksen, H. A least squares solution for paired comparisons with incomplete data, *Psychometrika*, Vol. 21, pp. 125-134, 1956.

205. Morrissey, J. H. New method for the assignment of psychometric scale values from incomplete paired comparisons, *Journal of Optical Society America*, Vol. 45, pp. 373-378, 1955.

206. Rees, D. G. Summarizing data by numerical measures, in: *Essential statistics*, London, Chapman & Hall, 2001, pp. 31-46.

207. Choi, S. Y., Luo, M. R., Pointer, M. R. and Rhodes, P.A. Investigation of large display colour image appearance II. *Journal of Imaging Science and Technology*, Vol. 52, 2008.

208. Keelan, B. W. A detailed example of objective metric design, in: Keelan, B. W. (editor), *Handbook of image quality*, New York, Marcel Dekker, 2003, pp. 197-206.

209. Stroud, K.A. Probability, in: *Engineering mathematics: programmes and problems*, Basingstoke, Macmillan, 1995, pp. 953-998.

210. Cohen, E. and Yitzhaky, Y. No-reference assessment of blur and noise impacts on image quality, *Signal Image and Video Processing*, 2010, pp. 289-302.

211. MacLennan-Brown, K. Quantification of artefacts inherent within digital imaging chains, PhD Thesis, London, University of Westminster, 2001.

212. SPSS Inc, SPSS 14.0, 2005.

213. Mellers, B. J. and Cook, A. D. J. The role of task and context in preference measurement, *Psychological Science*. Vol. 7, pp. 76-82, 1996.

214. Jenkin, R. and Richardson, M. Comparison between the Effective Pictorial Information Capacities JPEG 6b and 2000, *Opto-Ireland 2005: Imaging and vision*, Vol. 13, 2005.

215. Dainty, J. C. and Shaw, R. Image noise analysis, in: *Image science: principles, analysis and evaluation of photographic-type imaging processes*, London, Academic Press, 1974, pp. 276-319.

216. ITU-R. BT 1543. 1280 × 720, 16 × 9 progressively-captured image format for production and international programme exchange in the 60 Hz environment, International Telecommunication Union - Radiocommunication Sector, 2001.

217. Sugawara, M., Mitani, K., Kanazawa, M., Okano, F. and Nishida, Y. Future prospects of HDTV, *SMPTE Journal*, Vol. 115, pp. 10-15, 2006.

218. Pappas, T. N. Safranek, R. J. and Chen, J. Perceptual criteria for image quality evaluation, in: Bovik, A. (editor), *Handbook of image & video processing* 2nd, London, Elsevier, 2005, pp. 939-959.

219. Bilissi, E. Films-types and technical data. in: Langford, M. and Bilissi, E. (editors), *Langford's advanced photography* 7th, Oxford, Focal Press, 2008, pp. 85-107.

220. Barten, P. G. J. Image quality measure, in: *contrast sensitivity of the human eye and its effects on image quality*, Bellingham, SPIE Press, 1999, pp. 153-174.

221. Jenkin, R., Jacobson, R.E. and Richardson, M. A. Use of the First Order Wiener Kernel Transform in the Evaluation of $SQRI_n$ and PIC Quality Metrics for JPEG Compression, *SPIE Proceeding: Image quality and system performance*, Vol. 5294, 2004.

222. Cohen, S. S. Computing aspects of statistics, in: *Practical statistics*, London, Edward Arnold, 1988, pp. 177-197.

223. Keelan, B. W. A general equation to fit quality loss functions, in: Keelan, B. W. (editor), *Handbook of image quality*, New York, Marcel Dekker, 2003, pp. 119-128.

224. Osberger, W. Perceptual Vision Models for Picture Quality Assessment and Compression Applications, PhD Thesis, Brisbane, Queensland University of Technology, 1999.

225. Topfer, K. Preference in color and tone reproduction, in: Keelan, B. W. (editor), *Handbook of image quality*, New York, Marcel Dekker, 2003, pp. 285-304.

226. Hunt, R. W. G. How to make pictures and please people, in: MacDonald, L. W. and Luo, M. R. (editors), *Colour image science*, Chichester, John Wiley & sons, 2002, pp. 421-437.

227. Fairchild, M. D., Wyble, D. R. and Johnson, G. M. Matching image color from different cameras, *SPIE Proceeding: Image quality and system performance V*, Vol. 6808, 2008.

228. Ilea, D. E. and Whelan, P. F. Image segmentation based on the integration of colour–texture descriptors, *Pattern Recognition*, Vol. 44, pp. 2479-2501, 2011.

229. Keelan, B. W. and Jin, E. W. Weighting of field heights for sharpness and noisiness, *SPIE Proceeding: Image quality and system performance VI*, Vol. 7242, 2009.