In search of an Objective Measure
for the Perceptual Quality of Printed Images

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 23 maart 2004 om 15:30 uur

door

Judith DIJK

natuurkundig ingenieur
geboren te Düsseldorf, Duitsland.

Dit proefschrift is goedgekeurd door de promotoren:
Prof. dr. I.T. Young
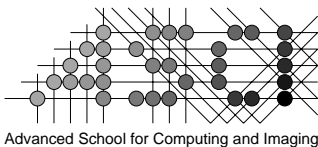Dr. P.W. Verbeek

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. I.T. Young,	Technische Universiteit Delft, promotor
Dr. P.W. Verbeek,	Technische Universiteit Delft, toegevoegd promotor
Dr. T.J.W.M. Janssen,	Océ Technologies B.V.
Prof. dr. H. de Ridder,	Technische Universiteit Delft
Prof. dr. ir. B.M. ter Haar Romeny,	Technische Universiteit Eindhoven
Prof. dr. ir. A.W.M. Smeulders,	Universiteit van Amsterdam
Dr. J. Walraven	Nederlandse Organisatie voor
	Toegepast Natuurwetenschappelijk Onderzoek/
	Technische Menskunde
Prof. dr. ir. L.J. van Vliet,	Technische Universiteit Delft, reservelid

Dr. J. Walraven heeft als begeleider in belangrijke mate aan de totstandkoming van het proefschrift bijgedragen.

Aan Sander en Onno
Aan mijn ouders

Contents

Chapter 1

Introduction

Most, if not all of our visual communication is based on images, whether "true" images of objects, or visual codes for conveying information, like alphanumeric symbols and graphics. Just like printed text can be of high or low quality, as measured by its legibility, the quality of printed natural images may also vary. However, how to measure this is far from simple. It is an issue addressed in many studies on image quality, e.g. [Webster et al., 1993, Davies and Rose, 1993, Marmolin, 1986, Engeldrum, 1999b, Daly, 1993b, Lubin, 1995, Barten, 1999, Farrell, 1999] [Ahumada, 1993, Kayargadde, 1995, Janssen and Blommaert, 2000a]. This study is another attempt. Studies on image quality are important because much industrial effort is dedicated to producing and reproducing images. This is particularly true for the design of electronic imaging and printing systems. If, for example, we change the inks in a printer, the printed image will change. Some changes will improve the image, some will degrade it. The question is, how can this be quantified?

One way to do so is by asking people to subjectively evaluate image quality. It has been shown that humans tend to respond more or less in the same way when using this technique [Roufs, 1989]. However, it is very time-consuming to do this kind of tests for every factor that may affect the imaging process. It would be really helpful if the quality could be assessed on the basis of a mathematical model. Such a model, using only a restricted set of quantifiable image properties, should predict the subjective quality response (rating) of the human observer. Unfortunately, but not unexpectedly, such a model does not exist as yet. In this thesis we attempt to obtain such a model for a small subset of all images, namely natural images that are produced by an ink jet color printer.

In the next section we first introduce the concept of image quality. Next, we focus on the main goal of this research. We end this chapter with an outline of this thesis.

1.1   Definitions and goals

As a starting point for defining image quality, we regard images as projections of the outside world [van der Schaaf, 1998]. Images can be treated as carriers of visual information [Janssen and Blommaert, 1997]. We mainly look at natural images, of which we manipulate various parameters that may be expected to affect image quality.

For our purposes, we used digitally recorded images. So, continuous scenes are sampled at discrete points (pixel positions), after which one or more values are stored. If the image is grey-valued, only one value per pixel is stored, whereas for a color image three or more values are stored for one pixel. Sometimes spectral color images are recorded, where a spectrum is stored for each pixel position.

Quality can be defined as the extent to which certain properties of the picture live up to someone's criteria. This definition has two important words: *extent* and *criteria*. If someone does not have very strict criteria, it is easier to produce "high quality" pictures.

If we combine both definitions, we may define image quality as: "the extent to which the image corresponds to the internal expectation of the human observing the image". Note that in this definition the human observer is still present, and not necessarily as passive observer, but also as goal oriented operator. The expectation depends upon the task for which the image is used.

If we have to see certain details in the image for a given task, the quality will have to meet other criteria than when we just want it for decorating a room. The first kind of image quality is called task-oriented image quality and the second kind is called perceptual image quality [Roufs, 1992, Kayargadde, 1995].

Roufs [Roufs, 1989] uses the term perceptual image quality instead of subjective image quality, and does so for two reasons. First, the term subjective would imply that quality differs widely across subjects. However, subjects are able to make consistent judgement of the quality of an image, and these judgements do not differ very much over subjects. Second, subjective image quality would also imply that the aesthetic features of images are important, which is not the case when we look at different versions of the same original image [Roufs, 1992, Kayargadde, 1995]. So, perceptual image quality is related to what we observe, not influenced by personally or culturally determined aesthetic values.

The quality of a reproduced image can be studied with respect to the original image. The "distance" (difference) between the reproduced image and the original can be measured. The image quality is redefined as image *fidelity*, or, in the case of image compression, as image *degradation*.

Image quality can also be studied without such a direct comparison. People are able to tell how good an image is, without seeing the original image, or the original *scene* the image is a reproduction of. We probably have some idea of how an image should look in our memory. So, to a certain extent, an image quality measure should also be able to work in an absolute sense, that is, without reference to a standard. Although, strictly speaking, the internal reference makes this "relatively" absolute.

It is a reasonable conclusion that the quality of the image has something to do with the *naturalness* of the image. If people in the image have purple faces, this probably will not be considered to be a good image. On the other hand, the image which is the most faithful reproduction (has the same color values) may not necessarily be perceived as the most natural image [Janssen and Blommaert, 2000a].

There is also a difference between image quality and visual comfort. Sagawa [Sagawa, 1999] reported experiments in which subjects scaled the subjective impression of comfort caused by visual stimuli, instead of the image quality. He used images in which the vividness of the colors (saturation) was changed. He found that images with lower saturation were more comfortable. However, others [Fedorovskaya et al., 1997, de Ridder et al., 1995] found that images with a slightly higher saturation are preferred. Sagawa concluded that observers use other criteria when scaling comfort than when scaling quality, because the functional aspect of the color usage is different.

Another problem is the importance of the image *content* as a determinant of the quality of the image. For example, we tend to judge an image of ourselves more critically than the images of unknown people. In this study, we are not interested in aesthetic aspects, like the contents and layout of the scene.

But even within an image some parts of the image are more important than others when it comes to judging image quality. If the background of an image is not in focus, this may not be a problem, but if the eyes of a person are a bit fuzzy, this may reduce image quality. How to deal with this variable is far from easy, since it may well be subject to individual differences.

Setting aside these "details", we define the major goal of this thesis as

finding an objective measure for the quality of a printed image that corresponds to perceptual image quality.

In addition to this major goal, a number of, closely related, secondary goals can be defined. Some of these secondary goals and corresponding research questions are:

Improving insight in the complex subject of perceptual quality When estimating the quality, does a subject evaluate the image as a whole or only sub-parts of the image? How do current ideas about the relation between naturalness and quality hold up for various sets of printed images?

Determining the relation between perceptual attributes and image quality
How does the quality depend on certain perceptual attributes of the image, such as sharpness and color distribution.

Developing methods for color image processing How can we optimize tools used for grey-value images, when these are applied to color images?

1.2   Outline

The outline of this thesis is as follows. In chapter 2, 3 and 4 we present the theoretical background and experimental approach for this study. In chapter 2 we present a short overview of how color can be quantified and what color variables are used. In chapter 3 we will discuss some of the current image quality models. In chapter 4 we will present the methods and materials used in our experiments. The topics discussed in that chapter include: how to characterize and calibrate a printer, what images can be used for this kind of experiments, and how to design experiments in which the measuring instrument is the human observer.

In chapters 5-8 we discuss the experimental part of this thesis. We have performed four groups of experiments with different parameters. In chapter 5 we report new measures for sharpening and smoothing. In chapter 6 we report a new measure for describing the sharpness of an image. In chapter 7 we report experiments on the effect of changing the luminance distribution of the image by a gamma manipulation. In chapter 8 we present results of experiments in which both the chroma was scaled and the luminance distribution was varied by a gamma manipulation.

In chapter 9 a generic method is described to use grey value image processing algorithms for color image processing. The grey value algorithm is applied to the lightness component of the color image is such a way that the colors of the image can still be rendered. This has been further explored for the case of a gamma manipulation as well as a sharpness manipulation. In chapter 10, an outline for a quality measure is suggested. This quality measure is useful to select the best image of a series of images that only differ due to a manipulation that affects the lightness value in the image. In the last chapter we look back at the results and discuss what we did and did not achieve.

Chapter 2

Preliminaries

2.1 Color basics

We have chosen to look at the quality of color (chromatic) images, although the difference with black-and-white (achromatic, grey-value) images is actually just a matter of dimensions. In terms of computer graphics, color is defined by three numbers (e.g. R, G, B), whereas a grey-value image requires manipulation of just one number (n = R = G = B). In this section we discuss some of the basics of color imaging and color measuring, that are relevant for this thesis, starting with the stimulus-response (input-output) aspect.

2.1.1 Introduction

The question "what is color" is a difficult question to answer. However, for the present purpose we can treat it as a psycho-physical phenomenon, in which a physical attribute of light correlates with a particular psychological response. The essential physical attribute is the spectral power distribution (SPD) of the light entering the eye, although spatio-temporal variables may also affect the color responses.

The SPD of a color depends on

- the SPD of the *light source*
- the spectral properties of some material that modify the SPD of the light source by selective absorption, reflection and/or transmission.

Color is much too often only identified by its physical characteristics, but that is only the *stimulus*, input to the color *response*, i.e. the color experience.

We can make a distinction between the color of a light, such as a green traffic light,

and a property of a surface, like the green of grass. That we perceive this difference is far from obvious, when considering that color always enters the eye as colored light [Walraven, 1992]. These two modes of color perception are most commonly called *aperture* and *object color* [Katz, 1911].

The psychological or visual response is determined by many factors, including the spatial and temporal distribution of the image projected on the retina. However, for the simple case of an isolated patch of light seen in isolation, we only have to consider the light absorption in the photo receptors of the eye and the consecutive transformation in a color response. How the color response is derived from the signals from the photo receptors is typically not a subject treated in studies of image quality, because the impossibility to measure all the steps in the visual chain from retina to visual cortex.

We can quantify colors in several ways, either physically, in terms of the SPD, or in so-called colorimetric units, which take into account the spectral (broad-band) filtering properties of the (standardized) human eye. But how are these colorimetric units related to the physical values that we measure?

We shall address the physics of the colors in section 2.1.2. In section 2.1.3 color according to the eye is discussed. In addition to the quantitative specification of color, one also has to consider the perceptual aspects. By that we mean the qualifications that we use to classify colors. Colors not only differ in the hue (like red or green), but also in luminance (bright or dim) and saturation (vivid or pale). There are also various other qualifications, which will be discussed in section 2.1.4.

2.1.2   Colorimetrics

Visible light consists of electromagnetic waves with wavelengths between about 380-780 nm. The distribution of the energy over the wavelengths, or spectral power distribution (SPD), is also called the spectrum of the light in question. Depending on how the light is produced, one can distinguish between emission, reflectance and absorption (or transmission) spectra. Different light sources emit different light spectra. In figure 2.1 three emission spectra of standard light sources are shown: the standard lamp light, and two different types of (simulated) daylight: D65 (blue sky) and D50 (clouded sky).

Light sources do not always emit a smooth spectrum. This is illustrated in figure 2.2, which shows the emission spectra of the RGB phosphors of an IIYAMA color monitor, with color temperature D65 [1]. Since 1969, CRT monitors have usu-

---

[1]Color temperature is the temperature of a Plankian radiator whose radiation has the same chromaticity as that of a given color stimulus. Light sources are described by their color temperature. D65 is a daylight simulator with a color temperature of 6500 K.
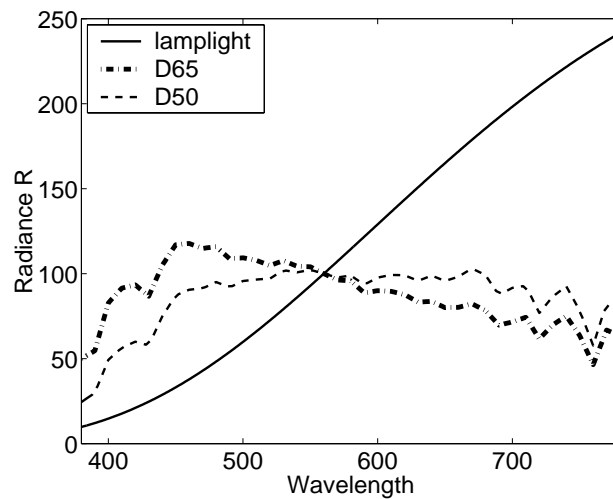
Figure 2.1: The emission spectra of three standard light sources: A (tungsten lamp light), and two daylight simulators: D65 (blue sky) and D50 (clouded sky).
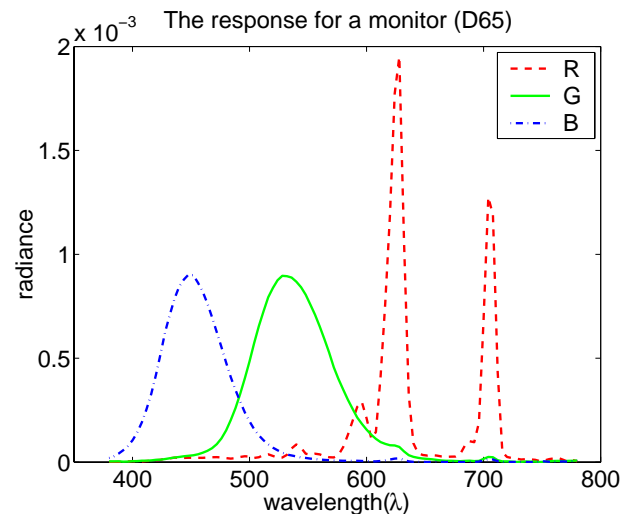
Figure 2.2: The emission spectrum of a color monitor.

ally the following phosphor set [Hunt, 1987]:

Red	europium yttrium vanadate	x = 0.675	y = 0.325
Green	zinc cadmium sulphide	x = 0.285	y = 0.595
Blue	zinc sulphide	x = 0.154	y = 0.068

It is of interest to note that the three peaked RGB spectra shown in figure 2.2 are capable of generating light mixtures that produce colors that are indistinguishable from the colors associated with the continuous spectra shown in figure 2.1. This shows that the visual system does not derive a color signature on the basis of a thorough spectral analysis. That is where the properties of the eye enter the picture.
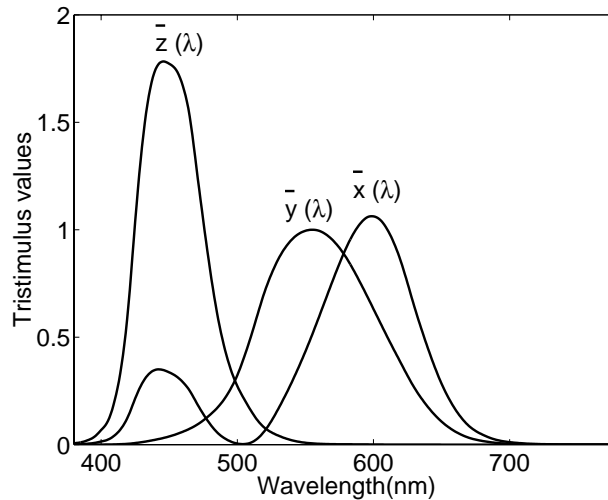
Figure 2.3: The color matching functions of the CIE 1931 standard observer.

2.1.3 Trivariance; color according to the eye

The visual system is equipped with 4 types of photo receptors, one for night (scotopic) vision, the rods (so called because of their cylindrical shape), and three different cone types (so called because of their tapered shape) for daylight (photopic) vision. The cones provide the sensors for color vision (at night we are color blind).

The three classes of cones absorb light over different, but largely overlapping, spectral ranges. These spectral ranges correspond roughly with the blue, green and red parts of the spectrum. The cones are referred to as short-wave (S), middle-wave (M) and long-wave (L) cones. In older literature one can also find the nomenclature blue, green and red cones.

It is possible to use entirely different spectra to produce colors that are indistinguishable for the eye. These so-called *metameric* colors appear if the absorptions in the three cone pigments are matched. If one looks at a natural scene on a television, all colors are metamers of the real colors of that scene.

To analyze color stimuli one is not obliged to use the LMS cone spectral sensitivities. As long as any three functions are linear combinations of the three "physiological filters" of the eye, they can be used to derive a three-dimensional metric that predicts the outcome of color mixtures. Such a metric is all that is needed for trichromatic color reproduction.

One of the standardized sets of three color filters is the CIE 1931 color-specification system [Wyszecki and Stiles, 1982]. One can use this system, for example, to explain how to produce a certain color on a television set.

In the XYZ color space, the responses (in terms of light absorption) of the filter are

given by

$$X = \int R(\lambda)\bar{x}(\lambda)d\lambda$$
$$Y = \int R(\lambda)\bar{y}(\lambda)d\lambda \tag{2.1}$$
$$Z = \int R(\lambda)\bar{z}(\lambda)d\lambda$$

in which $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ are the three filters, the so-called color matching functions. $R(\lambda)$ is the SPD, depending both on the light source $l(\lambda)$ and properties of the object $r(\lambda)$, such as reflection or transmission properties

$$R(\lambda) = l(\lambda)r(\lambda). \tag{2.2}$$

X, Y and Z are also called the tristimulus values of the color stimulus.

We often look at the relative contributions of X, Y and Z, because these values supply the information of the chromatic aspect of color (chromaticity). These relative values are defined as

$$x = \frac{X}{X + Y + Z}$$
$$y = \frac{Y}{X + Y + Z}. \tag{2.3}$$
$$z = \frac{Z}{X + Y + Z}$$

Because $x + y + z = 1$, it is sufficient to use only two of the three values. It is the convention to use x and y. The domain of x and y is limited, due to the fact that the range of differential cone responses is restricted, because of the considerable overlap of the cone spectral sensitivity functions. The use of equation 2.3 means that the information on the absolute intensity is lost. To restore this information, one can for instance use the luminance Y. x, y and Y provide the same information as X, Y and Z, and can also be used as a color space (see also figure 2.5 on page 24).

The boundary of the x, y space, or chromaticity diagram, is formed by the curve passing through x, y values of the spectral colors ("spectral locus") and the straight line ("purple line") connecting the spectral extremes (380 and 770 nm). This is visualized in figure 2.4. The location of some of the spectral colors is also given. All chromaticities within this boundary can only be made by mixing at least two different wavelengths, located at opposite sides of the chromaticity in question. All sources of colors can be used to mix colors, e.g. C_1 and C_2 in figure 2.4 can be used to produce all colors that lie on the line connecting them, including the white point. The chromaticity points of the RGB phosphors of a typical monitor (see figure 2.2) are also given. All colors within the triangle that connect these points

The CIE 1931 chromaticity diagram
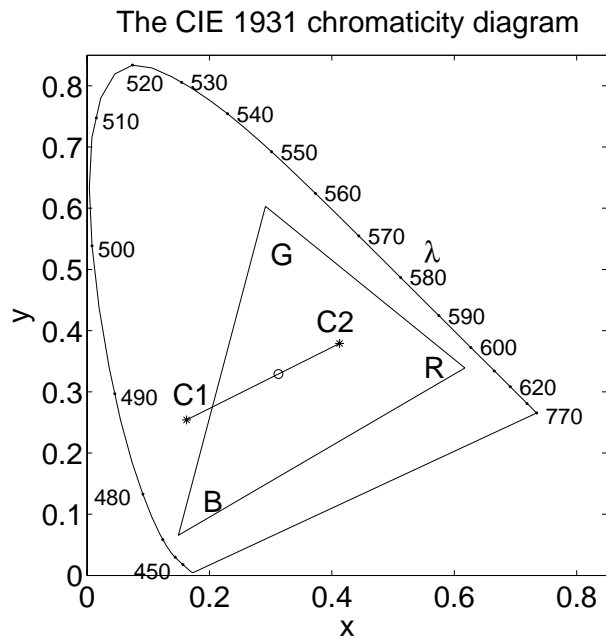
Figure 2.4: The CIE 1931 x, y chromaticity diagram. The colors that can be generated on a color CRT are bounded by the chromaticities of their phosphors (R, G, B). The circle denotes the D65 white point, C_1 and C_2 are two arbitrary colors that can be used to generate this white.

can be displayed on the monitor, all the points that are outside the triangle cannot. This difference leads to the definition of gamut, which is given as all colors that can be reproduced by a certain device. In this case the gamut is the area inside the triangle. More about gamuts and what to do with colors that are outside the gamut is given in chapter 4.

2.1.4 Perceptual color attributes

Although we experience color as one single attribute, it is possible to identify different basic properties or dimensions. This is a great help in bringing some order in the total scope of color sensations. Color can be ordered along three dimensions. The dimensions that are used depend upon how we perceive the color. If we perceive the color as the property of a *light*, for instance the red light from a traffic light, it is called an aperture or a self-luminous color. If we perceive the color as a property of a *surface*, such as a red apple, it is called an object color [Walraven, 1992].

For aperture colors, the dimensions are hue, brightness and saturation. The formal definitions (CIE 1987), with some minor extensions or clarifications (between parentheses [Walraven, 1992]) are

Hue attribute of a visual sensation according to which an area appears to be similar to one of the perceived colors red, yellow, green or blue (i.e. primary hues),

or to a combination of the two (i.e. intermediate hues).

Brightness attribute of a visual sensation according to which an area appears to
emit more or less light (either by itself or from incident light).

Saturation attribute of a visual sensation according to which the perceived color
of an area appears to be more or less chromatic, judged in proportion to
its brightness (the saturation being low at relatively low and high brightness
levels).

A few elementary, so-called primary hue elements can be defined. These are the
sensations we call blue, yellow, green and red. Each hue can then be described
as a mixture of two of these primary hues, such as yellow-red (orange) or blue-red
(purple). Not all combinations of the four primary hues make sense, though. One
cannot perceive a yellowish blue or a reddish green. This observation led to the so-
called opponent-color theory, first formulated by Hering [Hering, 1878] and further
developed and quantified by Hurvich and Jameson [Hurvich and Jameson, 1955].
The opponency can be described as a red-green and a blue-yellow channel. These
are the chromatic channels. There is also an achromatic channel.

The saturation of a color indicates the amount of balance between the chromatic
channel activity versus the achromatic channel activity.

We only deal will with object colors unless stated otherwise. For object colors, we
can use the same definitions for hue and saturation. The brightness depends both
on the incident light and the surface reflectance. The latter property is captured
in the term *lightness*. For a given illumination, a white object will always reflect
more light than a colored object, simply because it reflects at all wavelengths. Due
to this fact, we may use a white object as a reference for all other lightnesses in the
scene. We use the old CIE (1970) definition for lightness:

Lightness attribute of visual sensation according to which a body seems to reflect
(diffusely) or transmit a greater or smaller fraction of the incident light.

In the more recent CIE 1987 publication lightness is defined as relative brightness,
although there are good reasons to assign brightness and lightness to the two dif-
ferent domains of light and matter, respectively [Gilchrist et al., 1983]
[Walraven et al., 1990].

The attribute lightness requires, next to the response to the incident light, the
response to reference white. As long as the ratio between these two values is con-
stant, the lightness that is observed will be more or less constant, irrespective of the
overall light level [Jacobsen and Gilchrist, 1988]. This is called lightness constancy
[Walraven et al., 1990].

The difference between lightness and brightness is necessary to understand why we
see a difference between a white paper in a shadow (i.e. high lightness) and a black

paper in the sun (low lightness), even if the amount of light that the papers reflect is exactly the same.

2.1.5 Perceptual color spaces

There is a need for finding a relation between a color stimulus space (for instance XYZ) and a perceived color space. Unfortunately, such a space is not (yet) known. The next best thing is a color space in which small distances correlate to small perceptual differences. The CIE has recommended two color spaces for this purpose, that is, CIELUV and CIELAB [Hunt, 1977, Robertson, 1977]. These systems are intended for surface colors. CIELUV is mostly used for television and display systems, whereas CIELAB is the most used color space in the printing environment.

In the CIELAB color space five perceptual quantities are defined: the lightness L^*, the green-redness a^*, the yellow-blueness b^*, the chroma C^* (chromatic activity) and the hue h^*. These variables are defined as:

$$L^* = 116 f\left(\frac{Y}{Y_0}\right) - 16 \tag{2.4}$$

$$a^* = 500 \left[f\left(\frac{X}{X_0}\right) - f\left(\frac{Y}{Y_0}\right) \right] \tag{2.5}$$

$$b^* = 200 \left[f\left(\frac{Y}{Y_0}\right) - f\left(\frac{Z}{Z_0}\right) \right] \tag{2.6}$$

$$C^* = \sqrt{a^{*2} + b^{*2}} \tag{2.7}$$

$$h^* = \tan^{-1}(b^*/a^*) \tag{2.8}$$

in which $f(x)$ is defined as

$$f(x) = \begin{cases} x^{1/3} & \text{if } x > 0.008856 \\ 7.78x + \frac{16}{116} & \text{otherwise} \end{cases} \tag{2.9}$$

X_0, Y_0 and Z_0 are the tristimulus values of reference white.

CIELAB is based on typical office lighting conditions, i.e. a middle-gray background and average surround lighting [Berns, 1992]. If the surround luminance is very different, the tone reproduction changes. This is the reason that we can see much more contrast in transparency images than in printed images, because transparencies are watched in a dim environment, whereas printed images are looked at under normal lighting conditions. Hunt [Hunt, 1991] has shown that modifying the exponent of CIELAB can compensate for this surround effect. He suggested an exponent of 1/3.75 for a dim surround and an exponent of 1/4.5 for a dark surround.

CIELAB is far from perfect, therefore a number of other color spaces were defined: CIECAM97s [Luo and Hunt, 1998], CMC [Luo, 1998]. We use CIELAB because this is still commonly used in the printing world.

2.1.6   Device color spaces

We are not only looking at the world surrounding us, but are also reproducing it. To do this we record it (for instance with a camera) and display it (on a monitor, a printer, a TV set,...). Different devices can use different color spaces [Kang, 1997]. Each device typically has three or four primaries, colors used for generating its color space. Scanners and monitors have RGB primaries, the primaries of printers are the inks CMY or CMYK (see below). The goal is normally to display an image with the same colors as in the original scene. To do this, one should know the relation between the color space of the recorder and color space of the displaying device. One way of achieving this is to make a conversion of all devices to one standard perceptual color space (XYZ or CIELAB). To convert to another color space is then no more than converting to and from the standard color space. This idea is formalized by the ICC [ICC, 1998].

Most devices have RGB primaries, that is they use RGB values to display or record the colors. A normal assumption is that the RGB primaries are a linear combination of the XYZ values. The standard conversion for RGB to XYZ is given by

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = A \begin{pmatrix} R \\ G \\ B \end{pmatrix}, \tag{2.10}$$

in which A is a 3 x 3 matrix, for instance [Wyszecki and Stiles, 1982]

$$A = \begin{pmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.812 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{pmatrix} \tag{2.11}$$

Some nonlinearities in this relation can be modeled by an exponent relating linear RGB values to the monitor RGB values. This exponent is called gamma. If a monitor is used for exact color reproduction, the models given above are not accurate enough. To obtain a more accurate color reproduction, the monitor can be calibrated, for instance using the method described by Lucassen and Walraven [Lucassen and Walraven, 1990].

A printer has other primaries, that is the amount of the different inks that are used to print a color. The printing colors that are mostly used are Cyan, Magenta and Yellow, CMY. Black is normally added as ink to avoid printing too much ink and to enlarge the gamut (range of the printer). The letter denoting black is K (the last
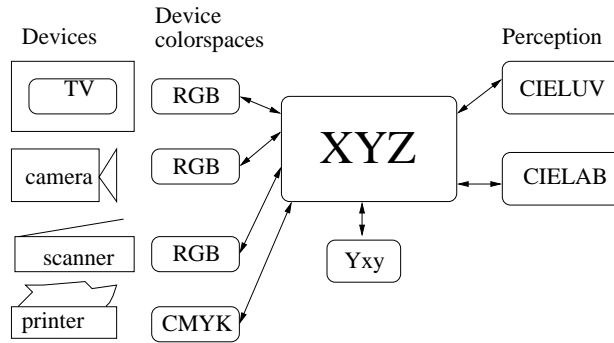
Figure 2.5: Relations between different color spaces.

letter of black) because B is already used for blue. The standard conversion from RGB to CMY is given by [Stone et al., 1988]

$$C = (255 - R)/255$$
$$M = (255 - G)/255 \tag{2.12}$$
$$Y = (255 - B)/255$$

The conversion from CMY to CMYK can be done in two ways

$$K = \min(C, M, Y)$$
$$C' = C - K$$
$$M' = M - K \tag{2.13}$$
$$Y' = Y - K$$

or [Bourgin, 1998]

$$K = min(C, M, Y)$$
$$C' = (C - K)/(1 - K)$$
$$M' = (M - K)/(1 - K) \tag{2.14}$$
$$Y' = (Y - K)/(1 - K)$$

These conversions are based on the assumption of a linear relation between RGB and CMY. Because the printing process is normally non-linear, this assumption often does not hold. Therefore, the characterization and calibration of a printer is fairly complicated. Relations between color spaces are given in figure 2.5.

More details concerning characterization and calibration are given in chapter 4.

Chapter 3

Studies on image quality

In this chapter we present a review of research that has been done in the field of image quality. We start with a general classification of image quality models, followed by a discussion on models from the literature. However, since there are so many models, we shall not try to be exhaustive. Moreover, many image quality models are just simplified versions of other models [Ahumada, 1993].

3.1   Classification of image quality models

It is possible to classify image quality models in different ways. A first distinction can be made between modeling *image* quality and *imaging device* quality. This is the difference between searching for the best image on one device and searching for the best device on which an image can be displayed. In other words, which version of the image is best or which device should we use to display the image? Here we are concerned with image quality, i.e. we only use the device as a tool for generating the images.

A second distinction can be made with regard to whether the quality of an image is related to a reference or an ideal model, or that the quality is modeled directly, independent of the reference. The first approach is called impairment approach [Engeldrum, 1999b], or fidelity approach [Daly, 1993b]; the second is called quality approach [Engeldrum, 1999b]. Impairment models determine a difference between an image and a reference image. This is mainly used in compression, where a difference is always a degradation of the original image. Quality models are more complex. Humans recognize objects in images, even if they have never seen the real scene. But this memory reference is quite elusive. The problem is to determine what the properties are that make a good image.

A third division is based on how much knowledge (vision models, task models) is used in creating the models. Lubin [Lubin, 1993] defines four different classes:

Model-free data collection This type of modeling is mainly used for evaluating different display systems. An image is displayed on different display systems, and subjects are used to evaluate the quality of the image on each system. For model-free data collection, no special knowledge, such as a vision model, is used. The disadvantages are that it is time-consuming and costly to do these experiments, as they always involve many subjects. The display systems or at least prototypes or models of these have to be available. Another disadvantage is that the number of display parameters has to be small.

Task modeling If the task can be decomposed into simple component tasks, one can also measure the performance of subjects on these smaller tasks. A relation between the performances on the smaller tasks and the physical parameters that are varied can be established. The task can be decomposed into simple component tasks, by using knowledge of the typical artifacts that are present in the image database used or that can occur with the use of the displaying systems. The task decomposition can also be done by multidimensional scaling.

Performance modeling Performance modeling is fitting the output of the visual system to the parameters that are varied at the input. The entire visual system is regarded as a black box. The advantage of this modeling is that once a good performance model is available, subject experiments can be avoided. An example of a performance model is Barten's SQRI measure [Barten, 1999].

Mechanistic modeling A mechanistic vision model is based on knowledge of physiological mechanisms. The functional response of the visual pathway is modeled. In the current models the output of the filters are tuned to different frequencies, spatial positions and orientations. The results of these filters are passed through a sigmoid function and then summed to one single scalar value. An example of mechanistic modeling is Daly's visible differences predictor [Daly, 1993b].

Performance modeling and mechanistic modeling require a vision model. This model can be based upon one channel (single channel metrics), or on several different channels of the image (multiple channel metrics) [Farrell, 1999].

Engeldrum has described the *process* of modeling image quality [Engeldrum, 1995, Engeldrum, 1999a, Engeldrum, 1999b]. He identifies four different types of variables. These variables are related through three different types of models. All variables and models together are called the Image Quality Circle, which is shown in figure 3.1.

The first type of variable is the technological variable, which is manipulated in the experiment. For instance, one can choose the type of paper an image is printed on. The second type of variable is the physical image parameter. This type can be mea-
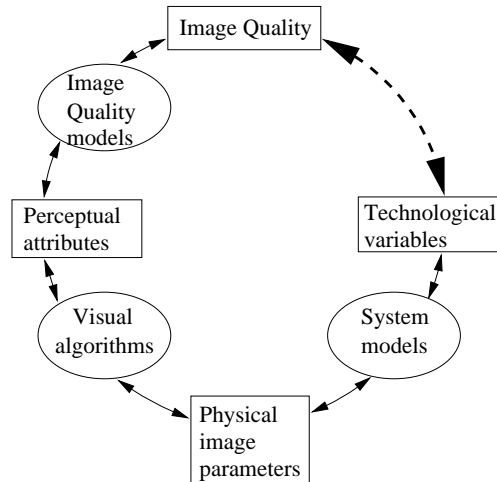
Figure 3.1: The image quality circle of Engeldrum

sured by physical instruments, for example, a scanner or a spectroradiometer. The third type of variable is called a perceptual attribute, such as brightness or darkness. The perceptual attributes are related to the image quality. (Engeldrum calls perceptual attributes customer perceptions). Although one usually only wants to know the relation between technological variables and image quality, all other parameters have to be estimated as well. The way to do so is to determine the relations between the different parameters, which is far from easy. This may explain why there are so may studies devoted to the topic of image quality. Still, a measure showing a high correlation with subjective scores is not yet available [van Dijk, 1997]. In recent work, both success and failure have been reported on mathematically predicting image quality from measured image features [Johnson and Fairchild, 2000, Engeldrum, 1995, Engeldrum, 1999b, Farrell, 1999, Ford, 1999]. One might conclude that as long as there is no simple way of describing an image, the mathematical prediction of its quality will always remain a problem.

3.2 Impairment models for black-and-white images

3.2.1 Root mean square error and its improvements

The easiest way of describing the difference between two images is the Root Mean Square Error (RMSE):

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3.1}$$

where x_i and y_i represent the grey level of pixel i in the original and the reproduced image, respectively. It is assumed that there is some correlation between the grey

value of the pixels and the lightness and luminance of the pixels. A more general version of the RMSE can be found using the Minkowski metric instead of the Euclidean

$$E = \frac{1}{n} \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \tag{3.2}$$

The RMSE or its generalized version does not often correspond with the perceived difference between the images (e.g. [Marmolin, 1986, Daly, 1993b]). Marmolin [Marmolin, 1986] derived other error measures by weighting the error in accordance with several assumed properties of the visual system. The general equation he used was

$$E = \frac{1}{n} \left(\sum_{i=1}^{n} |D_i|^p \right)^{1/p} \tag{3.3}$$

where D_i is a function of the original and reproduced pixel values. The D_i's that Marmolin tested were functions of one or more of the following variables: the mean value in a neighborhood around the pixel, the standard deviation in a neighborhood around the pixel and the gradient of the pixel. Marmolin found that the weighted measures corresponded better to the perceived similarity than the RMSE. However, none of these measures worked for all four images that were tested.

3.2.2   Impairment based on human vision models

The most complicated impairment models are based on models of the human visual system. We will describe two such models here: the visible difference predictor from Daly [Daly, 1993b] and Lubin's Sarnoff model [Lubin, 1995].

The Visible Differences Predictor

Daly [Daly, 1993b] made a model called the visible differences predictor (VDP). This model describes the difference between an original and a degraded image. The VDP uses a digital image processing approach. This approach can deal with nonlinearities in the system (in contrast with, for instance, the modulation transfer function (MTF)).

The input to the system is the original and the degraded image, provided in such a way that every pixel value in the gray-valued image represents the physical luminance value of the display device.

The output of the VDP is an image in which the value at each pixel represents the probability of detecting the differences between the input image and the reference image as a function of their location in the images.
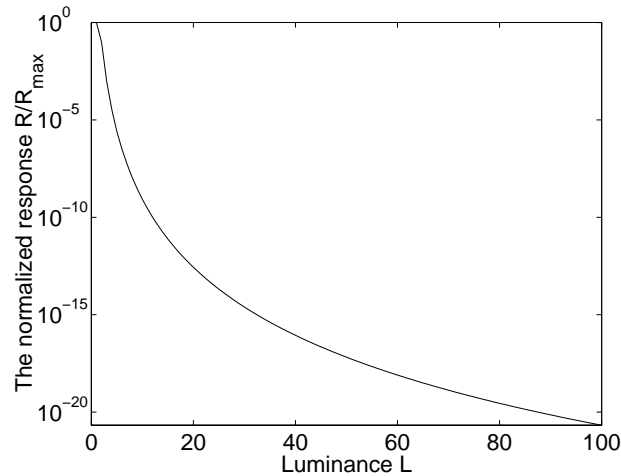
Figure 3.2: The normalized response function as used by Daly [Daly, 1993b] in the VDP.

The main part of the VDP algorithm consists of a model of the human visual system. All steps of the model have to be done for both the reference image and the degraded image, except for the probability, where the results for both images are combined into one image.

The model has three stages

Amplitude nonlinearity Daly models the visual sensitivity and perception of lightness with a simplified model. This model is shift-invariant, invertible, and implemented as simple point nonlinearities. The normalized response R/R_{max} is determined as

$$\frac{R(i,j)}{R_{max}} = \frac{L(i,j)}{L(i,j) + [c_1 L(i,j)]^b} \tag{3.4}$$

where i, j are pixel positions in x and y direction, $L(i,j)$ is the luminance and c_1 and b are constants (0.63 and 12.6). This response is shown in figure 3.2.

Contrast sensitivity function In the second stage, the variations in contrast sensitivity are modeled for different spatial frequencies (the CSF). The CSF varies due to the optics of the eye, the sampling aperture of the cone photo receptor, and both passive and active neural connections. The sensitivity S is modeled as a function of the radial spatial frequency ρ (c/deg), the orientation θ (deg), the light adaptation level l (cd/m^2), the image size i^2 (visual degrees), the viewing distance d (m) and the eccentricity (the angle of deviation from the direct line of sight) e (deg). The Daly relation, based on experimental data, is given by:

$$S(\rho,\theta,l,i^2,d,e) = P \min \left[S_1 \left(\frac{\rho}{r_a, r_e, r_\theta}, l, i^2 \right), S_1(\rho,l,i^2) \right] \tag{3.5}$$

where P is the absolute peak sensitivity of the CSF. The changes in resolution due to the accommodation level, the eccentricity and the orientation are given by r_a, r_e and r_θ, respectively:

$$r_a = 0.856 d^{0.14}$$
$$r_e = \frac{1}{1 + 0.24e}$$
$$r_\theta = \frac{1 - 0.78}{2} \cos(4\theta) + \frac{1 + 0.78}{2}$$

(3.6)

The effects of the image size and the light adaptation level are modeled in the following equations

$$S_1(\rho, l, i^2) = ((3.23(\rho^2 i^2)^{-0.3})^5 + 1)^{-1/5} A_l 0.9\rho e^{-B_l 0.9\rho} \sqrt{1 + 0.06 e^{B_l 0.9\rho}}$$
$$A_l = 0.801 \left(\frac{1 + 0.7}{l}\right)^{-0.2}$$
$$B_l = 0.3 \left(\frac{1 + 100}{l}\right)^{0.15}$$

(3.7)

Instead of a single viewing distance the CSF of the largest viewing distance and the closest viewing distance are determined and the envelope of these two CSF's is used in the evaluation. This assumes that the observer looks at *all* distances, so it is a conservative measure.

Detection mechanisms The final step in the human visual model is the modeling of detection mechanisms. Daly started this model by defining a set of filters for which the spatial and the orientational selectivity are modeled independently. Of each of these filtered the contrast $C_{G_{k,l}}$ in band k, l is given by

$$C_{G_{k,l}} = \frac{B_{k,l}(i, j)}{\bar{B_K}}.$$

(3.8)

with $B_{k,l}(i, j)$ the value of the filtered image at point i, j, and B_K is the mean of the image with the smallest low-pass filter.

Masking is the effect that signals are harder to discriminate in the presence of other signals. For instance, a distortion is easier seen against a uniform background then against a texture.

The masking function used in the VDP is

$$T_e^{k,l}[x, y] = (1 + (k_1(k_2 * m_n^{k,l}[x, y])^s)^b)^{1/b}$$

(3.9)

where T_e is called the threshold elevation image. s, and b are constants, s varies between 0.7 and 1.0 and b, k_1 and k_2 are set at 4, 0.0153 and 392.4980. $m_n[x, y]$ is the normalized mask contrast on pixel location $[x, y]$ in filter band

k, l. Because we are only interested in masking in both the reference image and the distorted image, the mutual masking is defined as

$$T_{em}^{k,l}[x, y] = \min[T_{e1}^{k,l}[x, y], T_{e2}^{k,l}[x, y]] \tag{3.10}$$

where T_{e1} is the masking in the reference image and T_{e2} is the masking in the distorted image. The contrast function is transformed to a probability that a signal will be detected using the psychometric function $P(c)$

$$P(c) = 1 - e^{(-c/\alpha)^\beta} \tag{3.11}$$

where c is the contrast. The constants α and β describe the threshold of the contrast, and the slope, respectively.

The probability $P_n[x, y]$ of detection in band k, l as a function of location is given as

$$P_n[x, y] = 1 - \exp(-(\Delta C_{k,l}[x, y]/(T_{em}[x, y]/CSF))^\beta) \tag{3.12}$$

where the contrast $\Delta C_{k,l}$ is given by

$$\Delta C_{k,l}[x, y] = C1_{k,l}[x, y] - C2_{k,l}[x, y] = \frac{B1_{k,l}[x, y]}{\bar{B}_K} - \frac{B2_{k,l}[x, y]}{\bar{B}_K} \tag{3.13}$$

where $B1_{k,l}[x, y]$ is the results of the filtering in band k, l for the reference image and $B2_{k,l}[x, y]$ for the distorted image. $\bar{B}_K$ is the mean of the baseband. The psychometric function is determined for each band in the filter set. Probability summation is used to find the total probability of detection

$$P_t(i, j) = 1 - \prod_{k,l}[1 - P_{k,l}(i, j)]. \tag{3.14}$$

This quality measure has been tested on different data sets, and it has been shown that it produces a reasonable fit to all sets. One disadvantage however, is that the peak sensitivity of the CSF, P, has to be set for every dataset [Daly, 1993a, Lubin, 1995].

The Sarnoff visual discrimination model

Lubin [Lubin, 1995] also describes a visual discrimination model (VDM), called the Sarnoff model (after the laboratory). As the Daly Visual Differences Predictor, it generates a Just Noticeable Differences (JND) map output for two input images, normally an original image and a reference image. Contrary to Daly, Lubin gives two measures out of this map that could be used: the mean JND, used for ratings, and

the maximum JND, used for thresholding differences. One JND unit in the output image corresponds to a 75% probability that the observer will see the difference.

Next to the two input images, the Sarnoff model needs additional parameters. For our application these are:

- The physical distance between sample points on the input image

- The distance of the (modeled) observer from the image plane

The input images are convolved with an approximation of the point spread function (PSF) of the eye (from Westheimer [Westheimer, 1986])

$$Q(\rho) = 0.952e^{-2.59|\rho|^{1.36}} + 0.048e^{-2.43|\rho|^{1.74}} \tag{3.15}$$

where ρ is the distance in minutes of arc from a point of light, and $Q(\rho)$ is the intensity of light at a distance ρ, relative to the maximum.

In the next step sampling by the retinal cone mosaic is simulated by convolution with a Gaussian, followed by point sampling to obtain an image on a grid. Martens and Meesters [Martens and Meesters, 1998] used a σ of 0.35 and a sampling distance of 1 or 2 arcmin.

$$I(x,y) = \text{Point sampling}(Gauss(I_{in} * Q(\rho), \sigma)) \tag{3.16}$$

The image is decomposed into contrast images at different scales (a contrast pyramid) using a technique that is similar to that of Peli [Peli, 1990]

$$c_k(x,y) = \frac{I(x,y) * (G(x,y,\sigma_k) - G(x,y,\sigma_{k+1}))}{I(x,y) * G(x,y,\sigma_{k+2})} \tag{3.17}$$

where $c_k(x,y)$ is the contrast at scale level k on point (x,y), and $(G(x,y,\sigma)$ is the Gaussian convolution kernel. σ_k is $\{32, 16, 8, 4, 2, 1, 0.5\}$ cycles per degree.

Four different orientations θ are used: $0°$, $45°$, $90°$ and $135°$. In each orientation the second derivative of a Gaussian o and its Hilbert transform h are quadrated and added

$$e_{k,\theta}(x,y) = \left(o_{k,\theta}\right)^2 + \left(h_{k,\theta}(x,y)\right)^2. \tag{3.18}$$

The energy measure $e_{k,\theta}$ is normalized by

$$\hat{e}_{k,\theta}(x,y) = \frac{e_{k,\theta}(x,y)}{(M_t(v_k, L_k(x,y)))^2} \tag{3.19}$$

where v_k is the peak frequency for the pyramid level k, and L is the local luminance value. M_t has to be adjusted using for example the Contrast Sensitivity function of Barten [Barten, 1999].

To reproduce the shape of Nachmias' contrast discrimination function [Nachmias and Sansbury, 1974], the energy at each level is put through a sigmoid function

$$T(\hat{e}_{k,\theta}(x,y)) = \frac{2\left(\hat{e}_{k,\theta}(x,y)\right)^{n/2}}{\left(\hat{e}_{k,\theta}(x,y)\right)^{(n-w)/2} + 1} \tag{3.20}$$

n is a value around 2; w is a value smaller than 1.

The results are convolved with a disc-shaped kernel of diameter five to account for the effect that the optimal visual sensitivity is not the same for the filter result (about 1 cycle per patch) as for the human visual system (5 cycles per patch).

The distance measure is determined as follows:

$$D(x,y) = \left\{\sum_{i=1}^{m}[P_{i,1}(x,y) - P_{i,2}(x,y)]^Q\right\}^{1/Q} \tag{3.21}$$

where m is all pyramid levels and all orientations, and $P_{i,1}$ the result of the adjustment at the right orientation and pyramid level for the first image. Q is set at 2.4.

If the model is adjusted correctly, a value of one will correspond to one JND.

Meesters and Martens have tested a slightly modified Sarnoff model for a certain setup [Martens and Meesters, 1998]. They do not find that this model performs better than simple root mean square error based measures (see subsection 3.2.1).

3.3   Impairment models for colored images

When looking at color images, another dimension is added to the quality models. For color images it is interesting to measure the perceptual difference of the colors. The most obvious measure is to take the difference in a color space that is perceptually uniform, i.e. in which distances are correlated to visual differences. This indicates the need for a perceptually uniform color space.

3.3.1   CIELAB color difference

One of the most frequently used color spaces to measure such differences is the CIELAB color space. The CIELAB color difference is defined as the Euclidean distance in CIELAB space, or , in formula [Hunt, 1977, Robertson, 1977]

$$\Delta E_{ab}^* = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \tag{3.22}$$

This would give an idea of the perceptual mismatch if distances in CIELAB correspond to perceptual differences. However, this is only approximately true. In 1994 the CIE proposed a new formula [CIE, 1995]:

$$\Delta E^*_{CH} = \sqrt{(\frac{\Delta L^*}{k_L S_L})^2 + (\frac{\Delta C^*}{k_C S_C})^2 + (\frac{\Delta h^*}{k_H S_H})^2} \qquad (3.23)$$

The difference between ΔE^*_{ab} and ΔE^*_{CH} is that for the latter the different color attributes (L^*, C^*, H^*) are weighted differently (which is depicted in S_L, S_C and S_H). The second difference is that specific experimental conditions, such as the amount of surrounding light, are taken into account into the values of k_L, k_C and k_H. This means that the weighting of the perceptual attributes is different for a setup in which there is much light than for a setup with less light.

In 1997 the CIE presented an interim color appearance model: CIECAM97s. This model should be adequate for most practical applications which require better estimation of color appearance than CIELAB. In CIECAM97s, mathematical scales are defined that correlate with various perceptual appearance attributes. The Cartesian color space, constructed with the dimensions J (lightness), a ($C\cos(h)$) and b ($C\sin(h)$) can be used as a uniform color space. The standard CIELAB ΔE is used to measure colors that have no spatial variation. Zhang and Wandell [Zhang and Wandell, 1996, Zhang et al., 1997]
[Zhang and Wandell, 1998] made an extension to this measure, named SCIELAB (spatial CIELAB). Before the ΔE calculation, suitable layers of the color image are filtered by an addition of Gaussian low pass filters. This makes it more suitable for measuring differences between colored textures. However, they mainly used it to measure differences between uniform colors and halftone patches, for which application it is obvious that some spatial integration of the halftone points should be applied.

3.4 Image quality models

3.4.1 Square root integral

Barten [Barten, 1999] has described an image quality model based on the square-root integral (SQRI). The rationale of this measure is that image quality is related to discriminability, and that the number of discriminable levels increases approximately linearly with the square root of the modulation of the spatial frequency components [Barten, 1999, Granger and Cupery, 1972].

The SQRI is given by

$$J = \frac{1}{\ln 2} \int_0^{u_{max}} \sqrt{\frac{M(u)}{M_t(u)}} \frac{du}{u}, \qquad (3.24)$$

where u is the angular spatial frequency at the eye of the observer, u_{max} is the maximum angular spatial frequency to be displayed. $M(u)$ is the MTF of the display and $M_t(u)$ is the modulation threshold function of the eye. J is the display quality in units of JND's. $M_t(u)$ is approximated by

$$1/M_t(u) = a(u)ue^{-bu}[1 + ce^{bu}]^{1/2} \tag{3.25}$$

with

$$
\begin{aligned}
a(u) &= \frac{540(1 + 0.7/l)^{-0.2}}{1 + \frac{12}{w(1+u/3)^2}} \\
b &= 0.3(1 + 100/L)^{0.15} \\
c &= 0.06,
\end{aligned}
\tag{3.26}
$$

where u is again the spatial frequency in cycles/degree, w is the angular display size in degrees, calculated from the square root of the picture area, and L is the effective display luminance in candelas per square meter. For several parameters, such as resolution, picture size and viewing distance, Barten showed a good relation between the calculated SQRI values and measured perceptual image quality [Barten, 1989]. He therefore concluded that the SQRI is a good measure for perceived image quality. However, the model is limited to achromatic images.

The SQRI model is spatially one-dimensional. Barten [Barten, 1999] made a two dimensional extension, which is equivalent to averaging the SQRI over different orientations. In practice, he used four orientations: $0°$, $90°$, $45°$ and $135°$.

3.4.2   Image quality by modeling perceptual attributes

A number of researchers determined image quality by modeling one or more underlying perceptual attributes. Most perceptual attributes are predominantly affected by only one parameter, which simplifies the estimation of the relations between the parameters and the image quality [Kayargadde, 1995].

Engeldrum [Engeldrum, 1995] used a nonlinear function to model image quality

$$\text{ImageQuality} = a_0 + (a_1 att_1^p + a_1 att_2^p + ... + a_n att_n^p)^{1/p} \tag{3.27}$$

where a_k and p are constants and att_i is a perceptual attribute. Both the attribute values and the image quality values were obtained using absolute scaling experiments. The attributes he investigated were sharpness, color accuracy, color uniformity, colorfulness and local impairments. He found that in this setting some tested attributes were not significant. However, some of these attributes may be significant if the range of the attributes would be larger, which would enlarge the

differences between the different images. Engeldrum did not attempt to predict the perceptual attributes.

Kayargadde [Kayargadde, 1995] tried to model perceptual attributes. He did so for two different attributes: blur (unsharpness) and noisiness (e.g. speckle).

For the blur estimation algorithm Kayargadde assumes that there exists at least some location in the original *scene* where the luminance distribution corresponds to an ideal step edge. An image is a result of imaging the scene, a process that inherently contains blurring. He models the blurring as a Gaussian blurring, which, in turn, is modeled by the σ of the Gauss.

The blur estimation algorithm consists of two steps. First, the regions of locally one-dimensional (1D) edges in the image are detected by determining the local maxima in the gradient magnitude. Two dimensional structures are removed by excluding points with a high two dimensional energy. Second, the edge parameters are determined at those locations by fitting a Gaussian shaped edge. The variables that are fitted are the amplitude, sigma, offset, displacement and angle of the edge. Both edge detection and estimation are carried out using polynomial transforms.

Kayargadde defined an objective measure for blur

$$S_b = 1 - \frac{1}{[1 + (\sigma_{bi}/\sigma_{b0})^2]^{1/4}} \tag{3.28}$$

in which σ_{bi} is the average spread of the image blurring kernel and σ_{b0} is the intrinsic blur in the early visual pathway. This measure correlates well with the perceived unsharpness in images. Kayargadde used a σ_{b0} of 0.65 arc min. Barten also used a Gaussian function to describe the MTF of the eye, with

$$\sigma_{b0} = \sqrt{\sigma_0^2 + (C_{ab}d)^2} \tag{3.29}$$

where σ_0 a constant (0.5 arc min), C_{ab} a constant (0.08 arc min/mm) and d the pupil size in mm.

In the same way as for the objective measure for perceived blur, an objective measure for perceived noise is determined. The main assumption here is that there are a sufficiently large number of locations in the scene, where the luminance distribution is locally zero-dimensional, i.e. homogeneous areas. The uniform locations are found by selecting points with a small gradient energy. The probability density function (PDF) of this gradient energy is used to obtain a good value for what is small. The standard deviation of the noise is obtained by fitting a model to the PDF. For images with no locations in which the luminance distribution is locally zero-dimensional, such as an image of sand, this measure will overestimate the amount of blur in the image.

Kayargadde defined an objective measure for perceived noise with the same mathematical formula that he used for perceived blur

$$S_n = 1 - \frac{1}{[1 + (\sigma_{ni}/\sigma_{n0})^2]^{1/4}} \tag{3.30}$$

in which σ_{ni} is the standard deviation of the noise in the image and σ_{n0} is equivalent of this for the noise in the early visual pathway. The value of σ_{n0} is determined by fitting results of subject experiments to the algorithm. Kayargadde found that this measure is a good correlate of the noisiness of images. He tested two different types of noise: white noise and pink noise with a low correlation length.

Kayargadde determined the relation between perceptual space spanned by the perceptual unsharpness and noisiness and objective space spanned by the objective measures S_n and S_b. He found that the unsharpness and the noisiness are not totally independent. He showed that a combination of the unsharpness and the noisiness provides a good measure for the perceived image quality. It is obvious however, that images with smooth gradients in combination with speckled texture, like a clutch of speckled eggs, pose a problem for this kind of modeling, because the basic assumption that there is a sufficient number of locations in the scene, where the luminance distribution is locally zero-dimensional, is not satisfied.
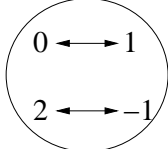
3.4.3   Image quality as a function of naturalness and usefulness

Janssen and Blommaert[Janssen and Blommaert, 1997] regard an image as the input for the visuo-cognitive system (eyes and brain). They treat this system as an information-processing system, which implies that the best image is the image that can be best interpreted and recognized. Janssen and Blommaert [Janssen and Blommaert, 2000b] define two different variables for interpretability and recognizability: "Usefulness" as the precision of the internal representation of the image and "Naturalness" as the degree of correspondence of the image with stored knowledge concerning what the image is supposed to represent. For evaluating the image quality both variables have to be evaluated at the same time [Janssen and Blommaert, 2000b]. The problem of determining image quality is divided over three tasks: determining the usefulness, determining the naturalness, and determining the relation between these two variables.

Usefulness   Janssen and Blommaert assume that visual metrics can vary in time. For instance, in the night we perceive other luminance contrasts than during the day. The scales of the visual metrics are optimized (adapted) in such a way that the ability to discriminate between items in the outside world is maximized. The two restrictions to the scale are that a) the upper and lower bounds are fixed, and that b) the accuracy is limited (there are always errors due to noise).

0 ⟷ 1	4	9	16	25		original attribute strength
2 ⟷ −1	6	7	18	23		attribute strength with error
0	5	10	15	20	25	attribute strength after scale function
2	3	12	13	22	23	new attribute strength with error

Figure 3.3: An example of topological errors. In the first row some errorless attribute strengths are assumed. In the second row noise is added. It can be seen (as indicated by the errors) that the first two attributes are inversed, giving a topological error. In row 3 the scale function s(x) = $5\sqrt{x}$ is applied. It can be seen that with the same noise, the topological error does not occur anymore. Note that the maximum and the minimum of the attribute strengths are the same.

Usefulness is defined as the extent to which different attributes can be distinguished: which is called discriminability. In their work, Janssen and Blommaert only investigated different color attributes. Discriminability is narrowly related to the total number of *topological* errors made in the mapping of attribute strength onto the values of the used visual metric scale. That is, errors that occur when the ordering of a set of items by their values on this scale differs from the ordering of this same set by their (errorless) attribute strength. An example of topological errors is given in figure 3.3.

For each pixel value the attribute strength x (for example luminance) is measured. The momentary[1] distribution of x, $\mu(x)$ is defined by $\int_x \mu(x)dx = 1$. The visual metric is represented by scale s. It is assumed that this scale is monotonic, because otherwise many topological errors will occur. $\eta(s)$ is defined as the momentary distribution of the scale function s for a give distribution $\mu(x)$ of the attribute strengths and a given scale function $s(x)$. The relation between $\eta(s)$ and $\mu(x)$ is given by

$$\eta(s) = \mu(x) \left(\frac{ds(x)}{dx} \right)^{-1} \tag{3.31}$$

Janssen and Blommaert show that the probability of a topological error as function of the scale value s, p_{err}, can be given by

$$p_{err}(s) = \frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{d}{2\sigma}\right) = \frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{1}{2N\eta(s)\sigma(s)}\right) \tag{3.32}$$

where d is the ideal noiseless scale value difference between any pair of two values and σ is the noise level.

---

[1]The scale varies in time due to changes in lighting.

The overall probability of a topological error P_{err} is found by integration of p_{err} over the entire scale

$$P_{err} = \int_S p_{err}\eta(s)ds = \int_x p_{err}\mu(x)dx \tag{3.33}$$

The discriminability D is proposed as

$$D = 1 - 2^n \prod_{i=1}^{n} P_{err,i} \tag{3.34}$$

where n is the number of dimensions in which the items can differ and $P_{err,i}$ is the topological error in this dimension. We should define the number of "items" in the image. Janssen and Blommaert made the choice to set this value to 100. The values of σ and S also have to be chosen. When ratios of D are used to compare different (versions of) images, Janssen and Blommaert found that the choice for N, S, and σ is quite robust to changes in d/σ.

Naturalness The naturalness Janssen and Blommaert proposed is a matching of several object colors to memory standards of these object colors. They identified three objects with clear memory standards (memory colors): these are skin, grass and sky. This indicates that with this method the naturalness of an image, which does not obtain any of these objects, cannot be determined. This is the case for some of the images we used in our tests. In theory all objects in the image should be matched to their memory counterparts. However, because most color changes are global, matching only these three colors (if they are present) will normally be enough [Yendrikhovskij, 1998]. The degree of matching $m(.)$ is given by a normalized correlation measure:

$$m(\eta_0^2(s), \eta_p^2(s)) = \frac{\int \eta_0(s)\eta_p(s)ds}{\sqrt{\int \eta_0^2(s)ds \int \eta_p^2(s)ds}} \tag{3.35}$$

where η_0 is the observed scale value distribution and η_p the memory standard distribution.

When there are three color dimensions, e.g. L^*, a^*, b^* these are combined as

$$m = m_{L^*}m_{a^*}m_{b^*} \tag{3.36}$$

Relation between naturalness and usefulness Janssen and Blommaert proposed a relation between the image quality IQ, usefulness measured in D and naturalness measured in m as the weighted sum of D and m. So

$$IQ = \lambda * D + (1 - \lambda) * m \tag{3.37}$$

where λ is a variable between 0 and 1 indicating how the usefulness and the naturalness are related.

Using these formulas, Janssen and Blommaert also defined optimal settings for the discriminability D and identifiability m, given the input image.

3.5   Image quality measures

In this chapter we gave a review of image quality measures. From this review, it can be seen that it is not easy to develop a good image quality measure. Most quality measures are only tested by the people that propose them. In the next chapters we will do some experiments on image quality and try to develop out own measures. We developed an impairment measure for relative sharpening (chapter 5) of black-and-white images. We also developed an image quality measure for the sharpness of black-and-white images (chapter 6). This measure is similar to the measure that Kayargadde proposed. Finally, we did experiments in which we changed the luminance and the chroma of color images. We tried to model the results of these experiments using the approach proposed by Janssen and Blommaert. These experiments are described in chapters 7 and 8. In chapter 10 we propose a quality measure that can be used to select the best image of a series of images that only differ due to a manipulation that affects the lightness value in the image.

Chapter 4

Methods and Materials

In this chapter we describe the methods and materials used for the experiments. Two major topics are discussed here:

- How can we make and reproduce the stimuli (sections 4.1, 4.2 and 4.3)

- How can we perform reproducible experiments with human subjects (section 4.4)

We start with the theoretical and practical applications of using a printing process in experiments. The printer or press used in the experiments is called a device. This topic can be refined using three questions: how can we characterize devices, how can we calibrate these devices, and what can we do with colors that cannot be displayed on the device that we want to use (gamut mapping)? We use the following definitions [Morovic, 1998]:

Characterization: the measurements and analysis to experimentally establish the relationship between the device-dependent color space and a device-independent color space (printer model).

Calibration: the adjustment of a device or process so that it generates repeatable data.

Gamut: the set of colors that can be produced using a device.

Gamut mapping: the handling of colors that cannot be displayed using the device.

Gamut boundary: the set of colors that are on the border between the in-gamut colors and the out-of-gamut colors.

One major problem while looking at color images is that these images are usually viewed in different circumstances, for instance with different illuminations or with different levels of adaptation of the human eye. Therefore, it is hard to establish

the same *appearance* of an image [Yendrikhovskij et al., 1999]. As a consequence, Hunt [Hunt, 1987] has distinguished several kinds of color reproduction:

Spectral: the spectral reflectance curves of the original and reproduced colors are identical.

Exact: the original and reproduction colors have the same CIE chromaticities, relative luminances and absolute luminances.

Colorimetric: the original and reproduction colors have the same CIE chromaticities and relative luminances.

Corresponding: the reproduction in which the chromaticities and relative luminances of the colors are such that, when seen in the picture-viewing conditions, they have the same appearance as the colors in the original would have if they had been illuminated to produce the same average absolute luminance level as that of the reproduction.

Equivalent: the chromaticities, relative luminances, and absolute luminances of the colors in the reproduction are such that, when seen in the picture-viewing conditions, they have the same appearance as the colors in the original scene. The illumination color, luminance level and ambient contrast may be different in the reproduction picture viewing conditions.

Preferred: the reproduction in which the colors depart from equality of appearance to those in the original, either absolutely or relative to white, to give a more pleasing result to the viewer.

Most calibration techniques are based on equivalent color reproduction.

4.1   Characterization

The goal of the characterization of a printing process is to find the relationship between the values sent to the printer ($CMY_{printer}$ or $CMYK_{printer}$ values) and the tristimulus values XYZ of the printed colors. For printers there is no simple relationship between the tristimulus values of single colors and the tristimulus values when these colors are printed over each other [Morovic, 1998]. The colors are also dependent on the reflections of the light on the surface and in the ink, and of scattering. These properties depend on the paper that is used.

One other problem when experimentally establishing a printer model is a lack of control over printer hardware [Wyble and Berns, 2000]. Further, with most modern printers internal control algorithms such as gray component replacement and dot gain correction cannot be bypassed [Wyble and Berns, 2000]. The printer can be

seen as a black box, where steering ink fractions are the input and a printed image the output.

We describe two different methods of characterization. The first method, searching for an analytic solution, tries to model the printing process by an analytic description. The resulting model has to be calibrated with a few measurements, usually of full ink coverage.

In the second method it is assumed that the printing process is *locally* linear, or at least dependent on neighboring measurements. The tristimulus values of points can be determined by only using calibration values that are in the neighborhood. We describe two methods of interpolation, linear interpolation and natural neighborhood interpolation.

4.1.1 Analytic solutions

Neugebauer was the first to consider halftone reproduction as an additive process [Xia et al., 1999]. His model is based upon the assumption that the color of a unit area is determined by the addition of the tristimulus values of the different combinations of colorants present in the area [Neugebauer, 1937, Morovic, 1998, Kang, 1997]. The main assumption is that the placement of the inks is uncorrelated, so that the fraction of the area that is printed by a certain combination of inks is given by the multiplication of the possibilities that the single inks occur on a certain position. For instance, if the dot percentage of the colors are given by c, m and y, the fraction of the area which is printed with cyan and yellow ink is $cy(1 - m)$. c is the possibility that cyan is printed on a certain position, $1 - m$ is the possibility that magenta is not printed on this position. The fractions of all eight possible combinations of inks are given by the Neugebauer equations:

$$
\begin{aligned}
f_1 &= (1 - c)(1 - m)(1 - y) \\
f_2 &= c(1 - m)(1 - y) \\
f_3 &= m(1 - c)(1 - y) \\
f_4 &= y(1 - c)(1 - m) \\
f_5 &= my(1 - c) \\
f_6 &= cy(1 - m) \\
f_7 &= cm(1 - y) \\
f_8 &= cmy
\end{aligned}
\tag{4.1}
$$

where c, m and y are the percentage dot areas of the three colorants, f_1 is the fraction of the area that remains white, and f_8 is the fraction of the area that is hit by all three inks. The generalization to a four color system like CMYK is obvious.

The tristimulus values of the dot areas can be obtained by measuring the colorant combination it represents:

$$X = \sum_{i=1}^{8} f_i \, X_i$$
$$Y = \sum_{i=1}^{8} f_i \, Y_i \qquad (4.2)$$
$$Z = \sum_{i=1}^{8} f_i \, Z_i$$

where X_i, Y_i and Z_i are the tristimulus values of combination i.

The Neugebauer equations can be used to determine the XYZ values when CMY(K) patches are given. We are also interested in the CMY(K) values for certain XYZ values. Unfortunately, the Neugebauer equations cannot be inverted analytically, but numerical solutions have been reported [Kang, 1997].

A refinement is to replace all values by spectral values [Viggiano, 1990] [Balasubramanian, 1995]. Instead of the tristimulus values, the spectra of the combinations of colors are added according to their weight

$$r(\lambda) = \sum_{i=1}^{8} f_i \, r_i(\lambda) \qquad (4.3)$$

where $r_i(\lambda)$ is the reflectance spectrum of the overlay i of inks and $r(\lambda)$ is the resulting reflectance spectrum of the printed patch. Using the color matching functions (equation 2.1) we can determine the tristimulus values of this spectrum. If we assume that the transmissions of the inks are multiplicative ($T_g = T_c T_y$), these tristimulus values are the same as in equation 4.2. However, it can still be an advantage to work with the spectra, for instance when one changes the viewing illuminant.

4.1.2   Analytic model using spectra

1D model: transmission

To convert from CMYK to XYZ we used our own analytical model. This is based on the fact that the reflection of inks can be modeled by multiplying their absorption curves. Ink printed on paper can be seen as three layers of absorption layers (inks) on top of a reflector (paper), as visualized in figure 4.1. The inks are characterized by their transmission spectra T_C, T_M, T_Y and T_K. The paper is characterized by its reflection spectra R_{paper}. This model holds if the transmission process (ink) is spatially separated from the reflection process (paper). If this is not the case, the
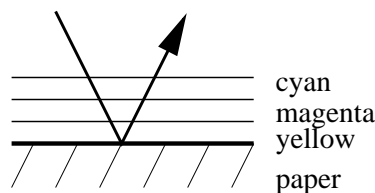
Figure 4.1: Ink printed on paper can be seen as three absorption layers and one reflector.
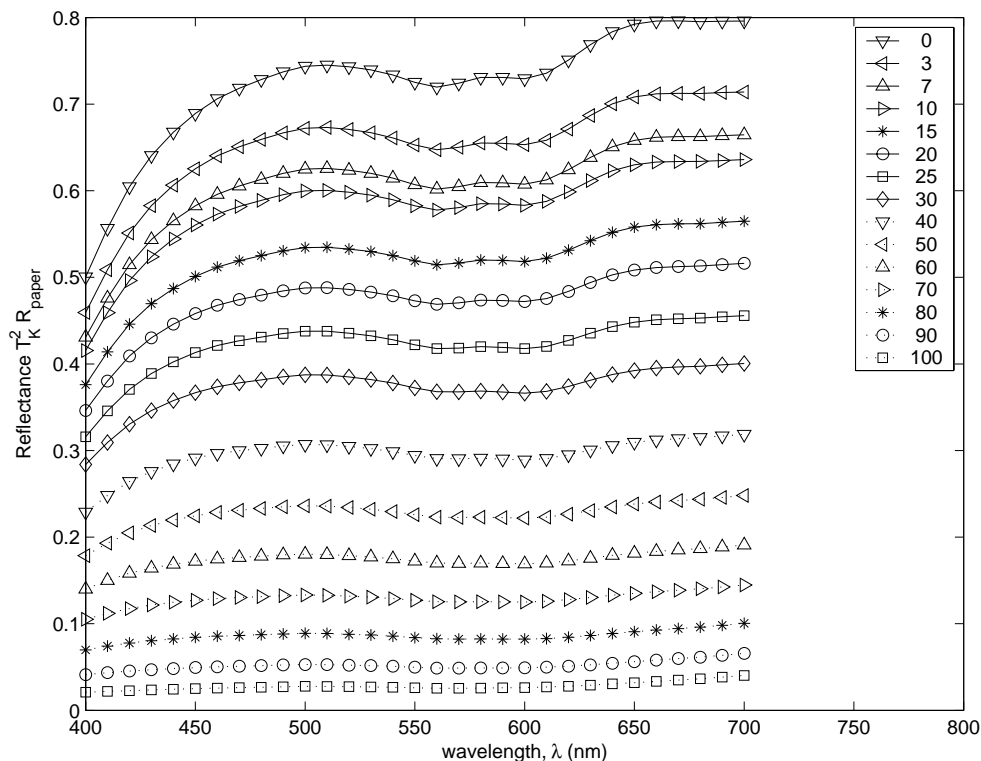
Figure 4.2: The reflectance spectrum for different amounts of black ink. The ink percentages are given in the legend. Note that the reflectance spectra change in a non-linear way. This is the effect of dot gain (see text for explanation).

Kubelka-Munk theory [Kubelka, 1948, Judd and Wyszecki, 1975, Allen, 1980] can be used to model the printing process. The assumption is that within a specific color, that is cyan, magenta, yellow or black, all of the dots are of the same thickness. Color differences are made by increasing or decreasing the size of the printed dots.

Edge effects: dot gain

The reflection spectrum is not only defined by the transmission of the solid inks. The edges of the dots also play a role. To explain this effect one should consider that the relation between the reflectance and the amount of ink is not linear (see figure 4.2 for black ink). This phenomena is called dot gain. Dot gain is caused by reflectances at the border of a dot, where light can be reflected through the ink
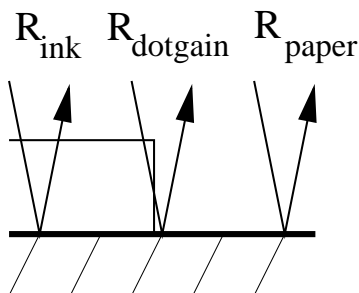
Figure 4.3: A visualization of the effect of dot gain. Some light is reflected by the paper without passing any ink (R_{paper}), some light is reflected by the paper and passes the ink twice (R_{ink}). Some ink passes the ink, but is reflected by the paper outside the area where the ink is printed ($R_{dotgain}$). This causes dot gain.
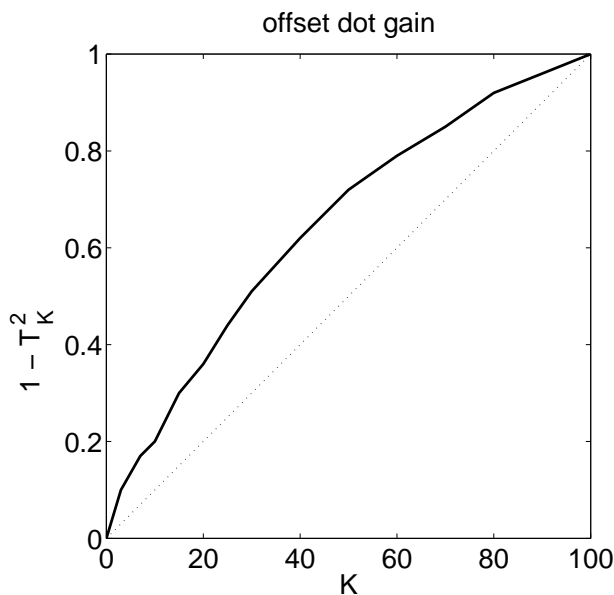
Figure 4.4: The dot gain for the black ink of the ISO printer. The data can be found in [ANSI, 1995]

after the reflection on the plain paper (see figure 4.3). In addition, dots often can be larger than specified. If we know the dot gain of a printer, we can determine the reflectance spectrum R_X of any fraction of ink X using the reflectance of a solid ink R_{ink} and the reflectance of the paper R_{paper}:

$$R_X = (1 - \text{dotgain\_corrected}(X))R_{paper} + \text{dotgain\_corrected}(X)R_{ink} \qquad (4.4)$$

The dot gain for the black ink for the ISO data is given in figure 4.4. Two points are of particular interest in this graph. The point $(0,0)$, indicating that if no ink is wanted no ink is printed (zero means zero) and $(1,1)$, indicating that one cannot print more than 100% ink. For the points in between the transmission is smaller than would be expected, which explains the name dot gain. In the model we assume that the dot gain for one ink is independent of the other inks. This would not work

if many edges of the dots of one ink are printed on top of the edges of dots of another ink. For an ink jet printer this hardly occurs because the dots are placed at random. In offset printing systems this effect is also negligible because the dots are printed on a fixed grid that is rotated for each ink. Note that we only correct for the dotgain of pure inks, and not for overprinted colors.

Since we are interested in combining inks, we need to combine the reflectances of the cyan, magenta, yellow and black inks. The reflectance spectrum of a combination of inks is given by the multiplication of the reflectance spectrum of the paper and the squared transmission spectra of the inks:

$$
\begin{aligned}
R_C &= T_C^2 R_{paper} \\
R_M &= T_M^2 R_{paper} \\
R_{CM} &= T_C^2 T_M^2 R_{paper} \\
&= \frac{R_C}{R_{paper}} \frac{R_M}{R_{paper}} R_{paper} \\
&= R_C R_M / R_{paper} \\
R_{CMYK} &= T_C^2 T_M^2 T_Y^2 T_K^2 R_{paper} \\
&= R_C R_M R_Y R_K / R_{paper}^3
\end{aligned}
\tag{4.5}
$$

An example of this is shown in figure 4.5. The reflectance spectra can be used to determine the tristimulus values of the colors using formula 2.1. To evaluate this method, we determined for all colors in the ISO set the XYZ values and determined the perceptual error ΔE_{Lab}. A unit in this error space ($\Delta E_{Lab} = 1$) corresponds roughly with a JND (just noticeable difference). The errors are given in figure 4.6. It shows that the errors are visible, but small in the higher L^* regions. The method does not seem to work in the lower L^* region, probably due to the fact that we measure inks that are printed as one layer, whereas for small L^* most colors are made by multi-layered inks.

4.1.3   Interpolation

Full characterization

The assumption underlying full characterization is that by producing a set of representative colors of the gamut, one can estimate how the colors in the neighborhood of a point in color space are reproduced. This set of reference points is a sparse sampling of the gamut. By assuming that the reproduction function is locally linear, the color of an arbitrary point can be determined by weighting the tristimulus values of the closest reference points [Morovic, 1998]. If the lightness L is significantly nonlinear with respect to the colorant, it can be useful to determine this

(a) 20 % Cyan spectrum

(b) 20 % Magenta spectrum

(c) 70 % Yellow spectrum

(d) 20 % Black spectrum

(e) Paper spectrum

(f) Combined spectrum
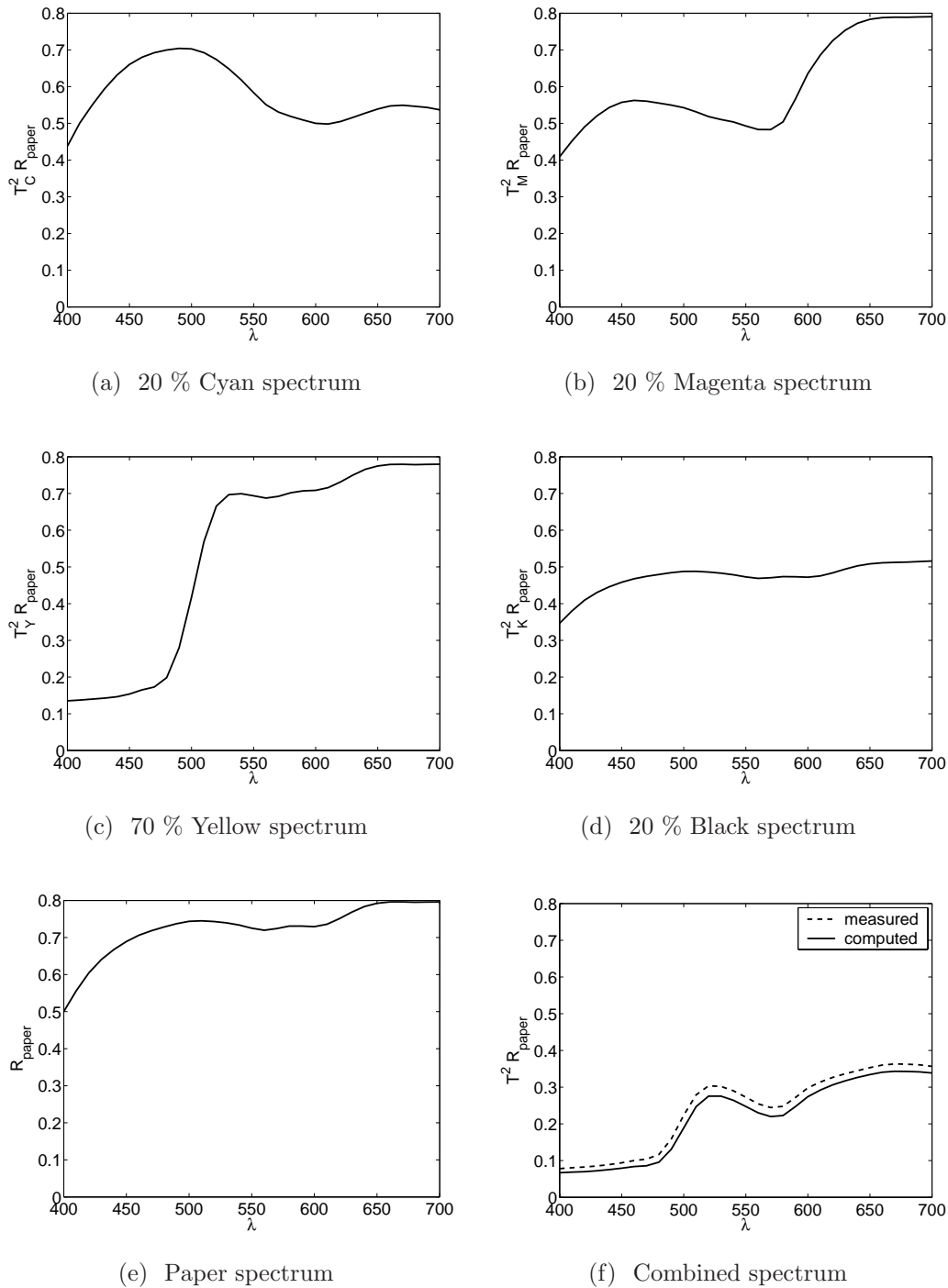
Figure 4.5: The steps in combining inks to produce a reflectance spectrum of combined inks. The reflectance spectra of the four single inks in the desired percentages are given in a-d. In e the reflectance spectrum of paper white is given, and in f the computed combined spectrum, along with the measured spectrum is given. The data can be found in [ANSI, 1995]
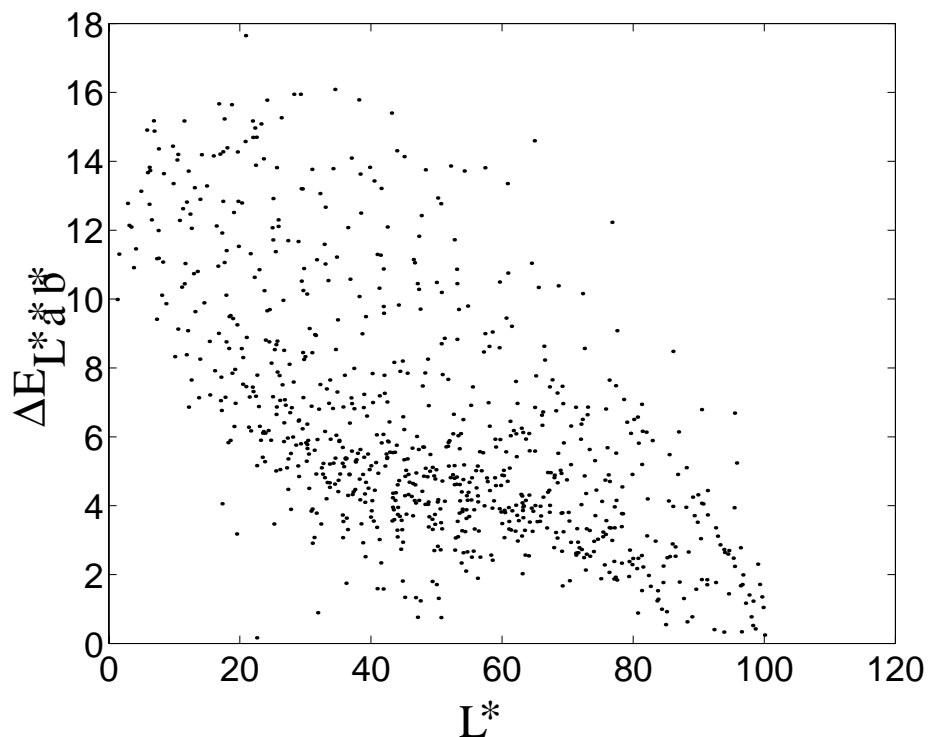
Figure 4.6: The perceptual errors when using the spectrum approximation for specifying the colors. A unit in this error space corresponds roughly with a JND (just noticeable difference).

nonlinear function first and vary the colorant in steps of equal lightness. Interpolation between the color spaces CMY and XYZ or CIELAB can be done both in the forward (CMY to XYZ) and in the inverse (XYZ to CMY) direction. However, for CMYK the forward interpolation can be done, but the reverse interpolation is ambiguous because the CMYK color space is four dimensional and the XYZ color space is three dimensional. One way to solve this problem is to interpolate between the CMY and XYZ color space, and to use a black printer algorithm to convert between CMY and CMYK. Black printer algorithms are discussed in section 4.1.4.

Linear interpolation

Linear interpolation starts from some regular structure, in most cases a cubic lattice of lattice points (in other words a stack of cubes) in color space. The color values of a new point are the weighted mean of the color values of the points on the cube that encloses the point. An example in two dimensions is given in figure 4.7.
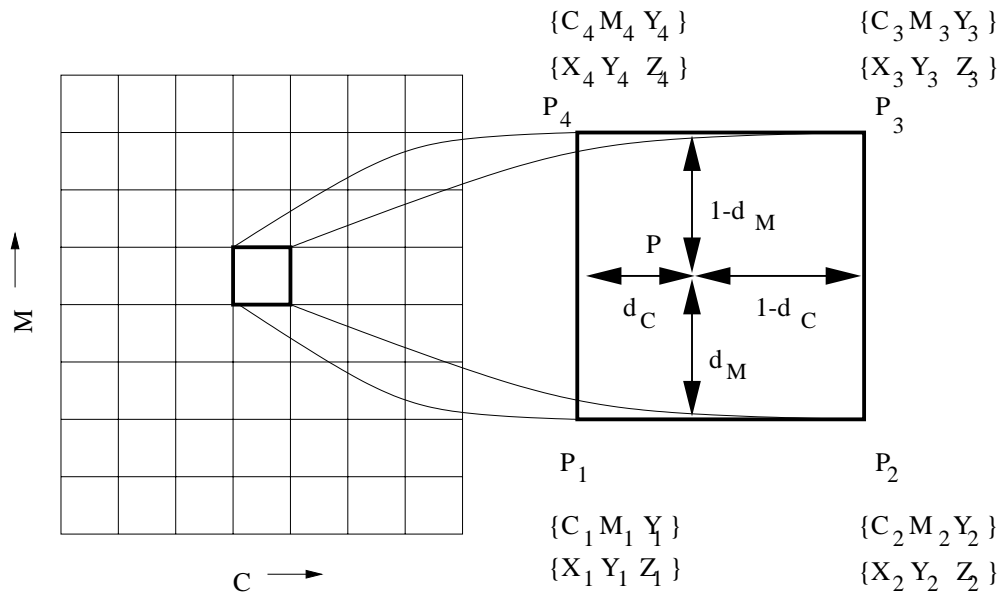
Figure 4.7: An example of trilinear interpolation in two dimensions. From the full C, M scale the block that includes point P is selected. The corner points of this block are the reference points. The tristimulus values of P are determined using the relative distances to the reference points d_C and d_M, and by the tristimulus values of the reference points.

The trilinear interpolation is done with the following formula

$$
\begin{aligned}
N_P = {} & (1 - d_C)(1 - d_M)(1 - d_Y)N_1 + d_C(1 - d_M)(1 - d_Y)N_2 + \\
& (1 - d_C)d_M(1 - d_Y)N_3 + d_C d_M(1 - d_Y)N_4 + \\
& (1 - d_C)(1 - d_M)d_Y N_5 + d_C(1 - d_M)d_Y N_6 + \\
& (1 - d_C)d_M d_Y N_7 + d_C d_M d_Y N_8 \qquad (4.6)
\end{aligned}
$$

with $N \in X, Y, Z$ and d_C, d_M, and d_Y are the relative CMY distances from the origin of the selected cube of reference points [Morovic, 1998]. Here one may replace XYZ also by $L^*a^*b^*$ or $L^*u^*v^*$. Note that the Neugebauer equations are the same as this one, but for only one cube, spanned by the full inks.

Above a certain value of n (the number of steps along each axis) the accuracy of the method will not improve significantly anymore. Morovic [Morovic, 1998] states that for certain ink jet printers, this number is 9. Full characterization gives the best characterization in most cases. The drawback is that the method needs a large number of measurements ($> 9^3$). This makes the method very time consuming[1], especially when the color reproduction system has a large drift in time. Another drawback is that a similar conversion from XYZ to CMY cannot be done successfully because the points are not on a regular grid in XYZ. A solution to this problem is

---

[1]This method is time consuming because for each measurement a color patch has to be printed. All colors have to be measured by a spectroradiometer or spectrophotometer, before the interpolation matrix can be estimated.

to determine the XYZ of a certain CMY, and adjust this CMY so that the XYZ converges to the XYZ that is needed (backwards estimation).

Instead of a regular block, other interpolation methods can be used such as triangulation or PRISM interpolation [Kang, 1997].

Natural neighborhood interpolation

In natural neighborhood interpolation [Sibson, 1981] the value of a point is determined by the values of neighboring points. The difference with full characterization is that the reference points do not have to be on a regular grid.

We describe this interpolation for the transformation from CMY to XYZ, but it can also be used in the reverse direction.

The CMY color space is divided into small volumes in which all points in the volume are closest to one of the reference points (see figure 4.8). This technique is Voronoi Tessellation [Okabe et al., 1992].

The point P that has to be interpolated has coordinates (C_P, M_P, Y_P). When P is introduced, a new tessellation with the reference points and point P is made (cf. figure 4.8). This tessellation differs only locally from the original tessellation. Some volumes are divided into two. The natural neighbors of point P are the reference points whose volumes are divided in two.

The natural neighborhood weight λ_i of reference point i is defined as the size of the volume, which belongs to i in the tessellation without P, and belongs to P if P is added. Theoretically, $\lambda_i = 0$ for points that are not natural neighbors of the point. The weight is normalized by the size of the area or volume belonging to the extra point. This gives that $\sum_i \lambda_i = 1$.

The interpolated value at of X, Y, and Z for point P can be determined with

$$
\begin{aligned}
X(P) &= \sum_i \lambda_i X(i) \\
Y(P) &= \sum_i \lambda_i Y(i) \\
Z(P) &= \sum_i \lambda_i Z(i)
\end{aligned}
\tag{4.7}
$$

where $(X(i), Y(i), Z(i))$ are the tristimulus values of the reference point i. So the interpolated value only depends on the values of the natural neighbors.

An example in 2D is given in figure 4.8. In figure 4.8(a) the reference points are seen, together with the Voronoi tessellation that divides this space into areas that are closest to one point. In figure 4.8(b) the point that we want to interpolate, point
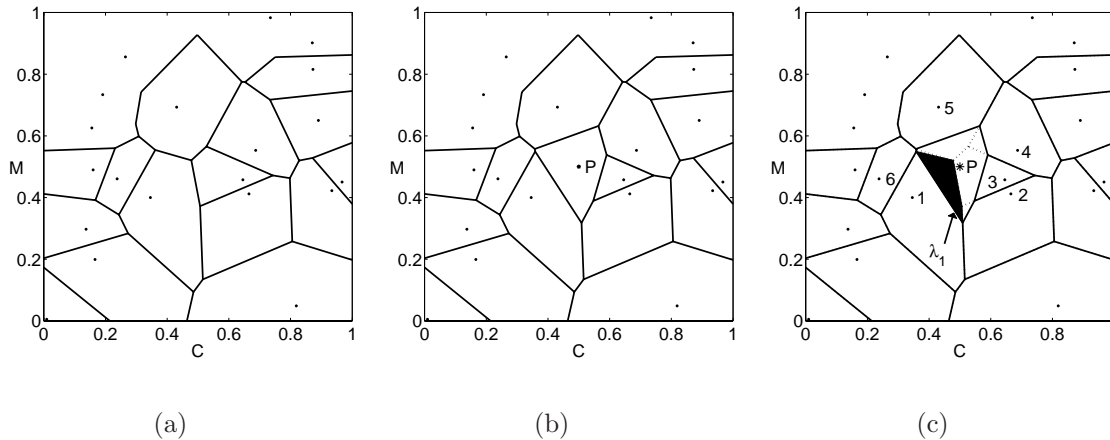
Figure 4.8: An example of the natural neighborhood interpolation in 2D. In (a) the Voronoi tessellation of the reference points is seen. In (b) point P that has to be interpolated is added, it can be seen that the tessellation changes locally. In (c) the natural neighbors of P are given the numbers 1-6. The natural neighborhood weight λ_i of reference point 1 is filled.

P is added. The Voronoi tessellation that now includes this point is only locally different from the tessellation in 4.8(a). In figure 4.8(c) the natural neighbors of point P are numbered 1-6. The natural neighborhood weight λ_1 of reference point 1 is filled. The λ_i's of 1, 3 and 5 are large, those of the other three points are small.

4.1.4   Black printer algorithm

Most printers have a fourth ink (black), in addition to cyan, magenta and yellow. The purpose of using the additional ink for images is twofold. On the one hand, the total amount of ink used for printing can be reduced for many colors, on the other hand the contrast is increased by increasing the absorption in the dark areas [Stone et al., 1988]. Text is usually printed in black.

Usually, different amounts of colored inks are used to produce a gray. The process of determining the mix of cyan, magenta and yellow to produce a gray is called gray balancing [Stone et al., 1988]. For most tristimulus values several CMYK combinations are possible. Normally one determines the CMYK values in two steps. First the relationship between XYZ and CMY is determined. Next, some amount of colored inks is replaced by a certain amount of black.

Because of the high non-linearity of printers, it is usually hard to predict how much black ink is needed to match the absorption of the removed colored ink [Stone et al., 1988]. The following equations are proposed by Bourgin [Bourgin, 1998]

without explanation (see also chapter 2):

$$C_3 = min(1, C_4 * (1 - K) + K)$$
$$M_3 = min(1, M_4 * (1 - K) + K) \quad\quad (4.8)$$
$$Y_3 = min(1, Y_4 * (1 - K) + K)$$

We use the following notation: C_3, M_3 and Y_3 are ink fractions that are printed without black (K), and C_4, M_4 and Y_4 are ink fractions that are printed with black. Because it is clear that black is only printed in the four ink scenario, we use K, where we could have used K_4.

The inverse relationships are:

$$K = min(C_3, M_3, Y_3)$$
$$C_4 = (C_3 - K)/(1 - K)$$
$$M_4 = (M_3 - K)/(1 - K) \quad\quad (4.9)$$
$$Y_4 = (Y_3 - K)/(1 - K)$$

We found that these relationships do not work for the ISO data set (an offset printer), and since we could not find other black printer algorithms, we decided to use a procedure based on the analytic spectral model.

Since we want to use the additive properties of the XYZ color space, we would want the hue to remain constant for a plane in the XYZ space. This does not hold for the hue angle h^*_{ab}, but it does so for h^*_{uv}. Proof for this is given in appendix A.

The basic idea behind the new interpolation is that a point in XYZ color space can be transformed into CMYK values using black and only two other inks. Depending on the color of the points these two inks can be cyan and magenta, cyan and yellow or magenta and yellow. An example of the gamut of the EPSON stylus 1520 printer is given in figure 4.9. If we look closely at the gamut boundary, we observe that the upper part of the boundary is made up of three areas, which represent the combination of two inks in different amounts. To print the color on this boundary, the combination of CMYK is uniquely defined by the two inks and white. If we move into the gamut towards black, we can make these colors by adding a certain amount of black. This idea can be modeled with the analytic model using spectra (section 4.1.2), if the spectrum of black can be considered as being constant. Note that the spectrum of the ISO data set, which is given in figure 4.5 (d), is indeed almost constant.

So, if we have a point P inside the XYZ gamut we can find the CMYK values if we know a) which boundary point Q to use and b) how much black should be added. The most simple way to find a boundary point is to take the intersection of the boundary and the line through P and the XYZ of black. However, because
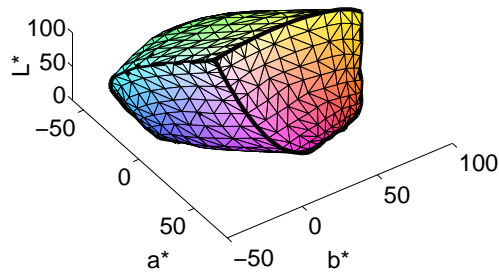
Figure 4.9: The gamut boundary of the EPSON printer.

the gamut is not black convex[2], one can end up outside the gamut. Therefore, we stretch the gamut in such a way, that it reduces to straight planes and becomes black convex.

In the stretched gamut we draw a line through P and black (K), and determine the intersection Q between this line and the gamut boundary. This is the situation depicted in figure 4.10.

Next we have to determine how much black should be added. The boundary conditions are no added black at the gamut boundary Q, and 100% at the XYZ black point. The straightforward solution is to make the amount of added black proportional, i.e.

$$K \approx T_K^2 = \frac{b}{a+b} \tag{4.10}$$

where a is the distance $\overline{KP}$ and b is the distance $\overline{PQ}$. As described in section 4.1.2 the spectral model must be refined by the dot gain correction. In our procedure we correct for this dot gain.

To evaluate this method, we used the data from the ISO report [ANSI, 1995], which includes a table with reference CMYK and corresponding XYZ values. If our interpolation method is correct, we should be able to convert the XYZ values to CMYK values that would return the original XYZ when printed. It is complicated to compare the CMYK values directly, because different CMYK values can give the same XYZ. Because of this, we first convert these values back to XYZ using the spectral method described in section 4.1.2. The results are shown in figure 4.11. It can be seen that the errors for the black printer algorithm are larger than if we use the original CMYK, as is to be expected. It can also be seen that the errors are visible, a certain amount are much larger than 1, but they are few and not very high. This implies that this method can be used, but produces some errors.

---

[2]We need a weak form of convex behavior. The normal definition is that if A, $B \in$ gamut, then the line $AB \in$ gamut. The definition we need is that if A is in the gamut and K is the black of the gamut, then the line $AK \in$ gamut. We call this form of convex behavior black convex.
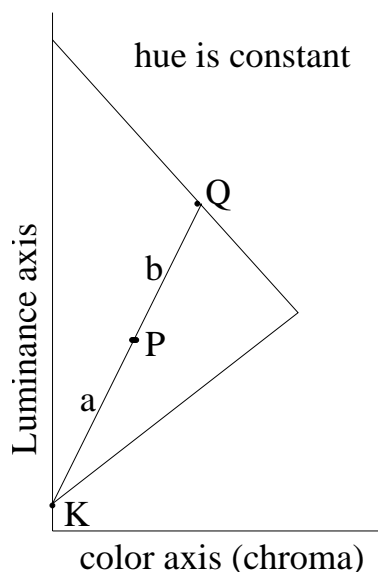
Figure 4.10: Illustration of the black-printer algorithm. P is the point to be interpolated. The triangular shape is the bisection of the gamut with the constant hue plane, which has the same hue value as the hue of point P. K is the point with minimal luminance or lightness, the black point. Q is the intersection of the line through P and K and the gamut boundary.

4.1.5   Characterization in practice

The images that we used are standard ISO images (ISO 12640:1997) that are used for evaluating the quality of printing. The images are specified in CMYK format. To indicate the original set of images that is connected to these values we use the notation $CMYK_{ISO}$. An ANSI report [ANSI, 1995] relates CMYK printing values to actually measured color values. The printer used in that proofing is an offset printer, and therefore, not relevant for our inkjet printer.

In the experiments we want to change variables in a perceptual color space, so we will have to convert the colors of these images to XYZ, in order to be able to transform to CIELAB or CIELUV.

The ideal procedure to convert the images from $CMYK_{ISO}$ to XYZ is to use the Analytic Model using Spectra 4.1.2. However, we developed this model *after* we had already performed the experiments. So, we actually used a sub-optimal procedure to convert the images from $CMYK_{ISO}$ to XYZ. On the other hand, the experimental pictures do not show any trace of irregularities or unnatural colors.

As discussed before, we used a black printer algorithm (equation 4.8) to transform the $CMYK_{ISO}$ pixel values to $CMY_{ISO\,based}$. Then we converted the pixels to XYZ with interpolation techniques. The ANSI report (1995) includes a table with reference $CMYK_{ISO}$ and corresponding XYZ values. The main problem with this table is that the tabulated reference data in $CMY_{ISO\,based}$ are not on a regular
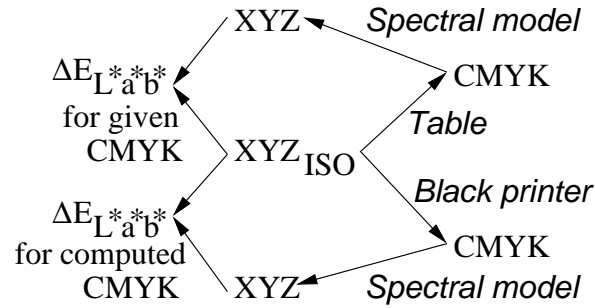
(a)
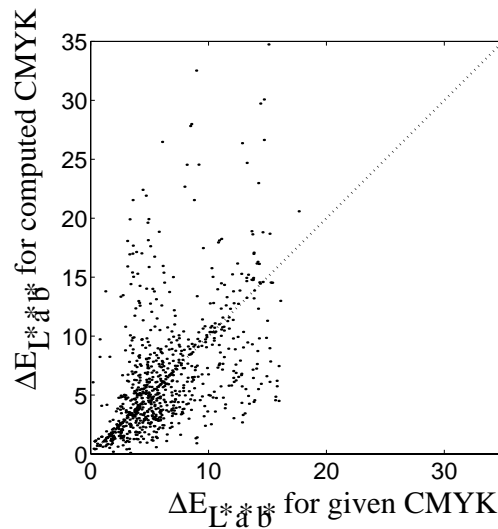
(b)

Figure 4.11: The results for the black printer algorithm are given. How the two different errors are determined is given in (a). For each XYZ color in the ISO set the CMYK values are determined using the algorithm. With these CMYK values the XYZ values are determined using the analytic solution using spectra. To see what the effect of the conversion from CMYK to XYZ is, the errors made with this analytic solution using spectra, using the original CMYK values, are also determined. In (b) the results are shown. It can be seen that the errors that are made using the black printer algorithm are somewhat larger, but comparable to the errors that are made using the given CMYK.

grid, so we used natural neighborhood interpolation (based on Voronoi). However, because this interpolation is very slow[3], we used two successive steps to convert

---

[3]This is partly due to the fact that we used a compiled program that is obtained from Dave Watson
`http://members.iinet.net.au/ watson/software.html`
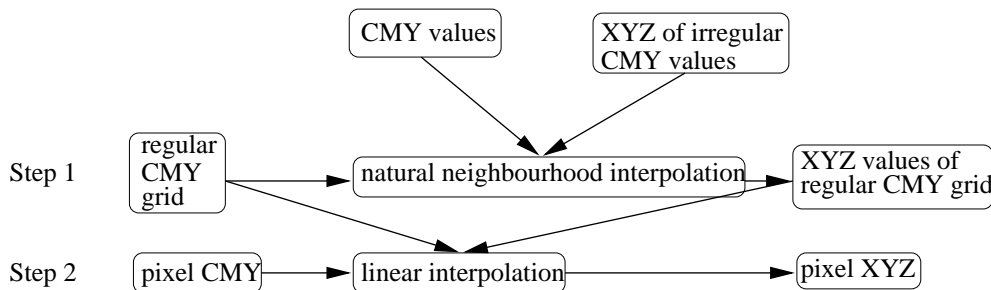
Figure 4.12: The relation between the interpolation and the reference sets. As regular grid a cubic lattice is used.

the pixels to XYZ. Figure 4.12 shows a schema of the procedure. The first step was converting uniform grids in $CMY_{ISO\ based}$ to XYZ using the slow non-linear interpolation method. The second step was converting the *pixels* in the image using the faster linear interpolation based on the regular $CMY_{ISO\ based}$ grid and its calculated XYZ values. Note that these images in XYZ are used as unambiguous defined test images. These images are derived, however, from the original ISO test images.

The reference white in the images is now the paper white of the ISO data $(X_{ISO\ based}, Y_{ISO\ based}, Z_{ISO\ based})$. In our experiments we use an EPSON Stylus color 1520, and plain EPSON photo quality paper with its own (paper) white $(X_{EPSON}, Y_{EPSON}, Z_{EPSON})$. We converted the white point of the image using [Hardeberg et al., 1999]

$$X_{out} = X_{in} * X_{EPSON}/X_{ISO\ based}$$
$$Y_{out} = Y_{in} * Y_{EPSON}/Y_{ISO\ based} \qquad (4.11)$$
$$Z_{out} = Z_{in} * Z_{EPSON}/Z_{ISO\ based}$$

in which (X_{in}, Y_{in}, Z_{in}) and $(X_{out}, Y_{out}, Z_{out})$ are the tristimulus values of the point before and after the transformation, respectively.

Intuitively this formula can be understood if one realizes that, unlike in the XYZ space, CIELAB "automatically" corrects for changes in the color of the "white" paper. So, any change due to changes in the color of white are not apparent as such. More precisely, to convert to CIELAB all colors are divided by the tristimulus values of paper white. So in order to convert from one white to the other, the tristimulus values have to be divided by the tristimulus values of the old paper white and multiplied by the tristimulus values of the new paper white. This formula fits in the spectral model if $r(\lambda) \equiv R_{EPSON}(\lambda)/R_{ISO}(\lambda)$ changes slowly with λ, then $R_{EPSON}(\lambda)\bar{x}(\lambda) \sim R_{ISO}(\lambda)\bar{x}(\lambda)$.

Just before printing the images were converted from XYZ_{EPSON} to CMY_{EPSON}. This conversion was done in the same way as the conversion from CMY_{ANSI} to
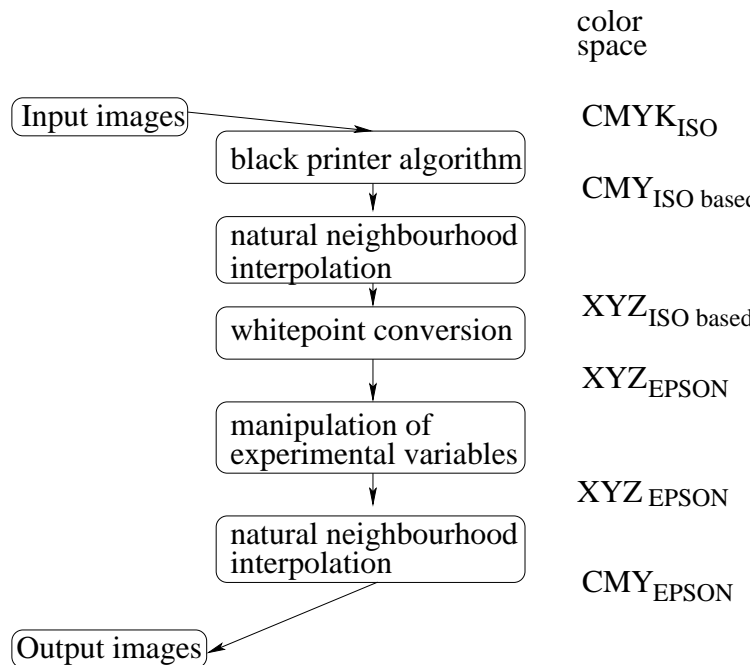
Figure 4.13: An overview of all the conversions steps used in the experiments.

XYZ. That is, a regular grid was constructed in XYZ and these colors were transformed to CMY_{EPSON}. To transform the image points we used these values and linear interpolation. We obtained a table with reference points using the techniques explained in section 4.2. All the conversion steps are given in figure 4.13. Note that the result can be seen as an empirically determined driver algorithm.

4.2   Calibration

For the interpolation from XYZ to the CMY_{EPSON} of the printer a reference table has to be used. To obtain these values, we measured the XYZ values of 1650 known CMY_{EPSON} colors, printed on the paper used in the experiment. The colors were measured using a spectroradiometer in a $0°/45°$ geometry [Sangwine and Horne, 1998]. This means that the light falls perpendicular on the paper, and the angle between the incident light and the spectroradiometer is $45°$, see figure 4.14. The relation between image width B, viewing angle α and viewing distance d, shown in figure 4.15, is given by

$$\alpha = 2\arctan(B/2, d) \tag{4.12}$$

The relation between the image width and pixel width b is given by

$$B = N_{pixels}b \tag{4.13}$$
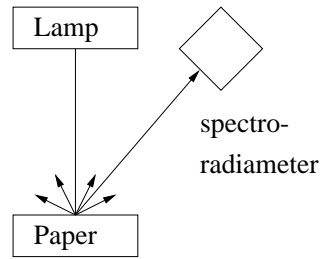
Figure 4.14: A 0°/45° geometry.
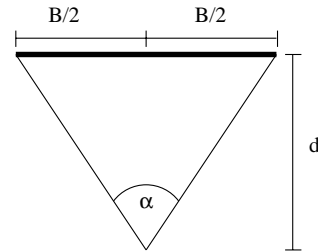
Figure 4.15: The relation between image width B, viewing angle α and viewing distance d.

where N_{pixels} is the number of pixels.

The chromaticity values of the inks are given in figure 4.16. The values are interpolated using the natural neighborhood interpolation.
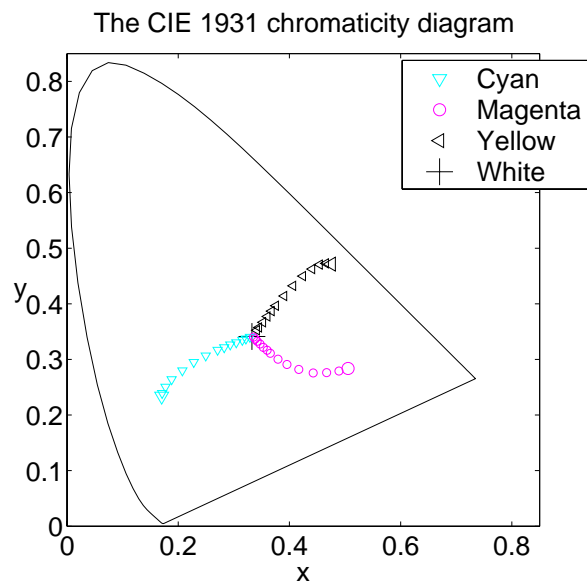
Figure 4.16: The location of the inks of the EPSON stylus 1520 printer in the xy diagram. 100% coverage of the ink is denoted by a larger symbol. For smaller amounts of ink, the location in the xy diagram is closer to the white point, this is the chromaticity of the paper used.
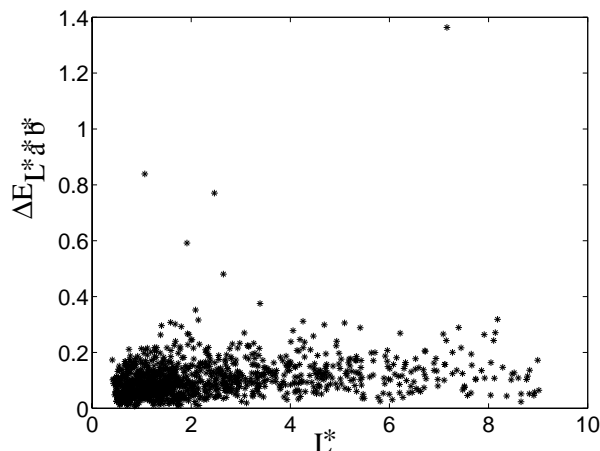
Figure 4.17: The results of the leave-one-out error procedure for one calibration set.

We measured the effect of the calibration by a leave-one-out method. Each point in the reference set is interpolated using all the other points of the reference set. The difference between the measured values and the computed (interpolated) values turns out to be $0.89 \pm 1.48 \ \Delta E_{L^*a^*b^*}$. The $\Delta E_{L^*a^*b^*}$ values for each color patch are given in figure 4.17. A $\Delta E_{L^*a^*b^*}$ of approximately 1 can just be seen by the human eye. It is clear that the calibration used here is quite adequate.

In figure 4.18 perceptual differences for different times and for different measurement setups are given. Figure 4.18(a) shows the perceptual error between measurements that were done with the same setup but with the reference data printed on different days. The setup used here was a spectrophotometer. The error is 2.1 ± 1.9 $\Delta E_{L^*a^*b^*}$. In figure 4.18(b) the measured printed patches were the same, but the measurement setup was in one case a spectrophotometer and in one case a spectroradiometer as described above. The error is $3.5 \pm 3.4 \ \Delta E_{L^*a^*b^*}$. As expected, the errors are somewhat larger than in 4.18(a), but still small enough.

The black printer algorithm (equation 4.8) was evaluated for the ISO data and the EPSON data using the following procedure. We selected colors with the *same* $\text{CMY}_{EPSON/ISO}$ and *different* $\text{CMYK}_{EPSON/ISO}$ values, and determined the $\Delta E_{L^*a^*b^*}$ value of these colors. The results are given in figure 4.19, which shows that this conversion does not work well for the ISO data set, the mean $\Delta E_{L^*a^*b^*}$ being too high (8.53). For the EPSON data set, however, the mean $\Delta E_{L^*a^*b^*}$ is only 1.75, so this conversion can be used for this printer. In this respect one should also realize, that there is no reason to expect that slight differences between the original and the reproduced ISO images would have an impact on the experimental results. Such a difference in the EPSON "original" would be present in all its derived test images, and it is unlikely, therefore, that this would effect the rank ordering of these images.

(a)   (b)
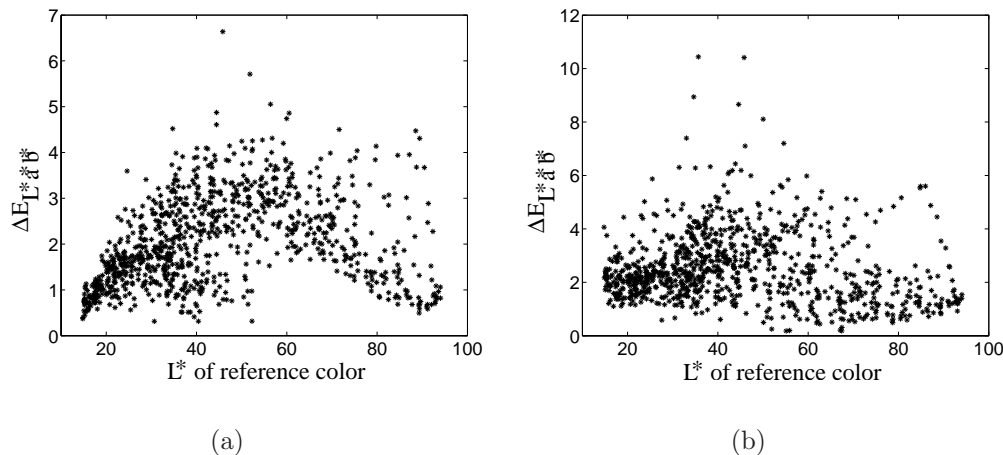
Figure 4.18: These graphs show the perceptual difference of measured calibration values for different circumstances. In (a) the measurement patches are printed on different days (half a year apart). The measurements are done with a spectrophotometer. It can be seen that the $\Delta E_{L^*a^*b^*}$ differences are small, typically < 4. In (b) the measurement patches that are measured are the same, but one set is measured with a spectrophotometer and one set is measured with a spectroradiometer. It can be seen that the differences are somewhat larger than in (a), but still small (mostly < 6).
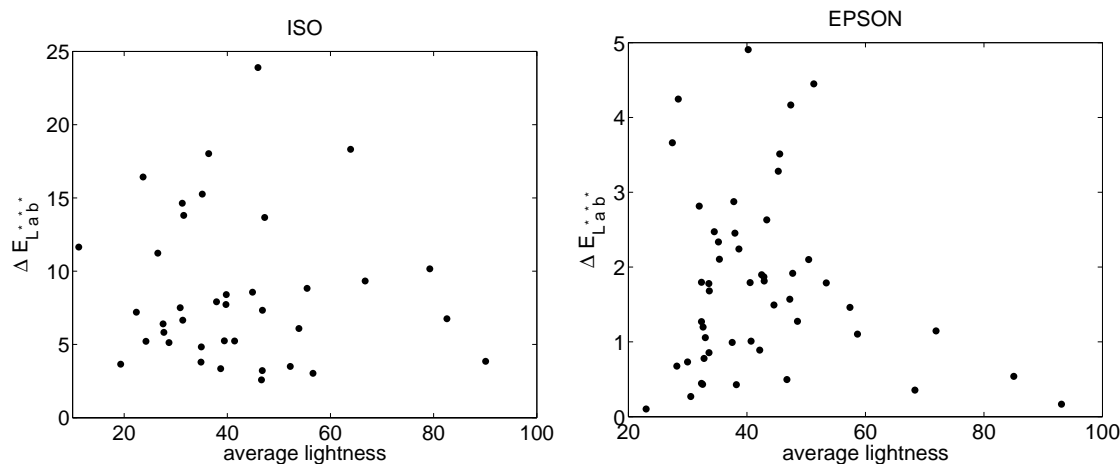
Figure 4.19: Comparison of the errors for points with the same CMY, but different CMYK values, using either the ISO data (a) or the EPSON data (b). Note the difference in scale.

4.3 Gamut mapping

The group of all colors that can be reproduced by any device can be seen as a solid in a color space [Morovic, 1998]. This solid is called the gamut of the device. For a printer, e.g., the gamut consists of all colors that can be printed. People often encounter the problem of different gamuts. An image that looks nice on a monitor, can be very poor when printed. This can be due to the fact that the gamut of the

printer is very different, and mostly smaller, than the gamut of the monitor. Not all monitor colors can be reproduced. The problem of gamut mapping is how to handle images in which some colors cannot be reproduced by the output device.

Gamuts of different devices are not equal in size and shape. The gamut of additive systems in CIEXYZ (or any other linear tristimulus space) is a convex polyhedron. Since most subtractive and hybrid systems are nonlinear, their gamuts have irregular shape [Sharma and Trussell, 1997].

The difference in the shape of the gamut is very large when comparing a normal CRT monitor and a hardcopy printer [Katoh et al., 1999]. In this case, problematic colors are saturated green and blue and light colors, because the monitors gamut is wider in these regions. The result of this is that not all colors that are produced with one device (e.g. a CCD camera) can be reproduced with another device (e.g. a printer). The goal of gamut mapping is to transform the colors in images in such a way, that the images can be optimally reproduced considering that it is not possible to always reproduce exactly on the output device.

The main goal is focused on the *overall appearance* of a given image rather than the exact appearance of the each colors.

The International Color Consortium (ICC) developed profiles to convert images from device color spaces to a standard perceptual color space and vice versa [ICC, 1998] (see also section 2.1.6). In these profiles four rendering intents for gamut mapping are defined: perceptual, saturation, colorimetric and absolute. The perceptual intent is used for reproducing pictorial or photographic images, and preserves the appearance of the image. The saturation intent is mainly used for images containing objects such as charts (business graphics). It preserves the vividness of the colors. With the colorimetric and absolute intent the relationships between the in-gamut colors are preserved. The difference between the two is that in the colorimetric intent the white point of the gamut is converted to that of the output device, whereas for the absolute intent nothing is changed. The absolute intent is best used for spot colors and when one is simulating one device on another. The colorimetric intent is appropriate for images which are in the output gamut after white point correction.

Since we are interested in the overall *appearance* of the input image, gamut mapping has to be applied to *perceptual* attributes such as hue or chroma. Usually perceptual uniform color spaces are used such as CIELAB and CIELUV [Katoh et al., 1999]. The second reason to use these color spaces is that colorimetric distances should be related to perceptual differences [Morovic, 1998].

One can try to map the total gamut of the input device onto the gamut of the output device (image independent mapping). However, usually better results are

obtained when mapping only the gamut of the *image*[4], because the latter will often better fit into the gamut of the output device and thus requires less distortion [Morovic, 1998]. The most obvious example for this statement is that if an image has only colors that are in the gamut of the output device, no gamut mapping should be done, whatever the input gamut may have been.

Morovic [Morovic, 1998] gives the following gamut mapping principles:

1. Make changes to as few colors as possible and make changes as small as possible.

2. Use a perceptually uniform environment.

3. Maintain perceived hue, rather than the hue as defined in CIELAB or CIELUV (see chapter 2).

4. Allow for different compression in different parts of the color space.

5. Maintain as much chroma as possible without sacrificing a significant amount of detail. Morovic states: "Psychophysical experiments suggest the significance of chroma in gamut mapping. This has often been ignored in papers on gamut mapping which consider lightness to be the most important attribute. However the best results will be achieved when a compromise between lightness and chroma is found. To do this, the simultaneous gamut mapping algorithms seem the most suitable."

Most gamut mapping methods are based on moving a point along a line in color space. There are two features to be considered, the mapping method and the mapping direction.

Mapping method: There are three common mapping methods [Katoh et al., 1999]:

- clipping

- linear compression

- soft clipping (non-linear clipping)

The choice of the mapping method depends upon the amount of out-of-gamut colors. If this amount is small, the preferred method is clipping, whereas if this amount is large, the preferred method is compression. A soft clipping method has the advantage of reproducing most of the common gamut colors accurately, while reducing the loss of detail that is due to clipping. These functions are illustrated in figure 4.20.

Mapping direction: The second feature of gamut mapping is the mapping direction, that is, the direction along which the colors are mapped in the used

---

[4]The gamut of the image is defined as the (limited) set of colors that occur in the image
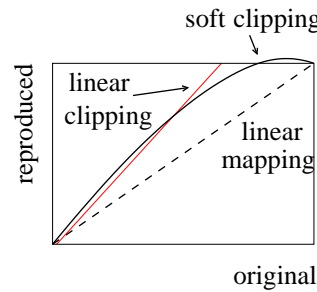
Figure 4.20: Linear clipping vs soft clipping

color space. Note that some mappings have to be performed in a specified sequence to obtain an image that is inside the gamut. Some often-used mapping directions for gamut compression are

Hue compression: Hue is the perceptual attribute that we can distinguish with the best precision. This is the reason that most authors propose to maintain the hue. A problem is that the hue as defined in the CIELAB and CIELUV space is still an imperfect representation of the *perceived* hue (see the definitions in chapter 2).

Lightness compression: The first attribute that is usually scaled is the lightness. This compression is used to map the maximum and minimum lightness of the gamuts onto each other. According to most authors the lightness should be linearly compressed; an even better way is to use a soft clipping function [Morovic, 1998]. An example of this compression is given in figure 4.21(a).

Chroma compression: Chroma compression is a simple linear compression in planes of constant hue and along lines of constant lightness, optionally with a soft clipping algorithm. A chroma compression is normally performed after a lightness compression. The compression can be done along each line independently (figure 4.21(b)), or for the whole image uniformly (figure 4.21(c)).

Simultaneous lightness and chroma compression: It is more successful to compress the lightness and the chroma simultaneously. These algorithms are more likely to preserve more of the colorfulness of the image [Morovic, 1998]. The most prominent of these is to compress the chroma and lightness towards the point where lightness is 50 and chroma is 0. As with the chroma compression the rate of compression can be determined along each line or uniform for the whole image. These two compressions are given in figure 4.21(d) and (e). With this compression, it may still be convenient to use a lightness compression to obtain gamuts with an equal minimum and maximum lightness, before the simultaneous light-

ness and chroma compression. If the simultaneous lightness and chroma compression is used directly, the loss of chroma may be very high.

The perceptual effects of these (technical) operations were studied by Katoh [Katoh et al., 1999]. He found that lightness compression reduces the contrast of the image, whereas chroma compression reduces the vividness of the image.

Instead of compression for all points one can also clip the points that are outside the gamut back onto the gamut boundary. The main disadvantage is that different colors close to the gamut boundary are clipped onto the same color. The advantage is that the colors of most points do not change. As said before, this advantage is usually smaller than the disadvantage.

Common used clipping directions are shown in figure 4.22 [Morovic, 1998]:

- orthogonal clipping; clipping along the line normal to the gamut boundary.

- chord clipping; clipping along a line towards a point on the L axis (for instance the point on the luminance axis with the same lightness as the cusp point[5]).

4.3.1 Gamut mapping in practice

The gamut of the EPSON printer is given in figure 4.23. It can be seen that the gamut is more regular for the lighter colors than for the darker colors. This effect is partly due to measurement noise. The gamut of the ISO data set is given in figure 4.24. As will be discussed below, the visualization of this gamut is more difficult than the visualization of the EPSON gamut. This difficulty explains also why the gamut is visualized with less points.

To perform the gamut mapping we need three things: a mapping method, a mapping direction, and a representation of the gamut boundary. As long as we refrain from mapping the hue, this representation can be given in a 2D plane of constant hue.

Finding the gamut boundary through Delaunay tetrahedration

Determining the boundary of a gamut can be done in various ways. We investigated two different methods.

---

[5]The cusp point is defined as the point with the maximal chroma for a certain hue.
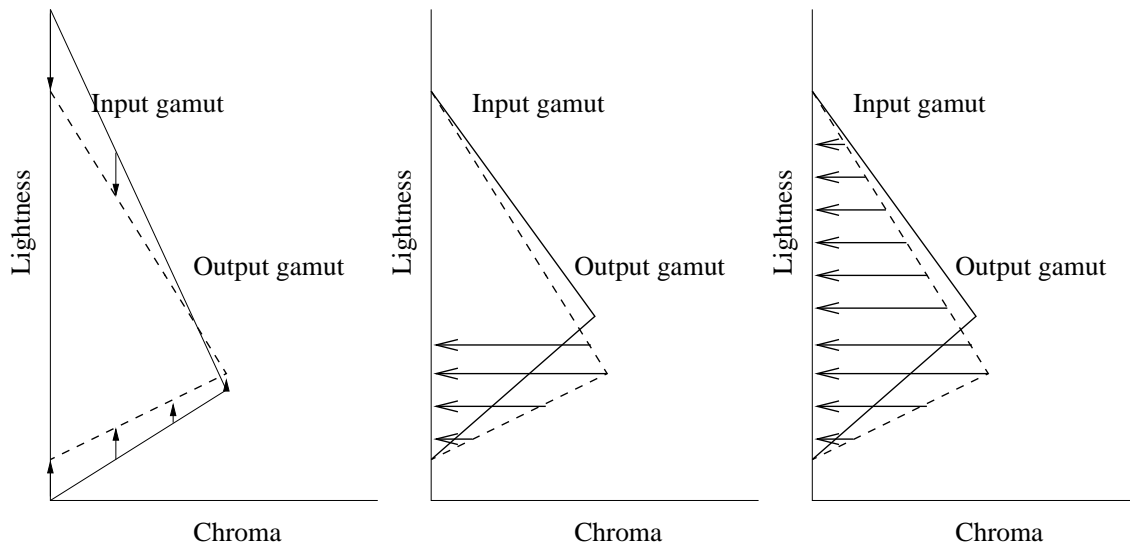
(a) lightness compression

(b) chroma compression for a limited number of light-ness lines

(c) uniform chroma com-pression
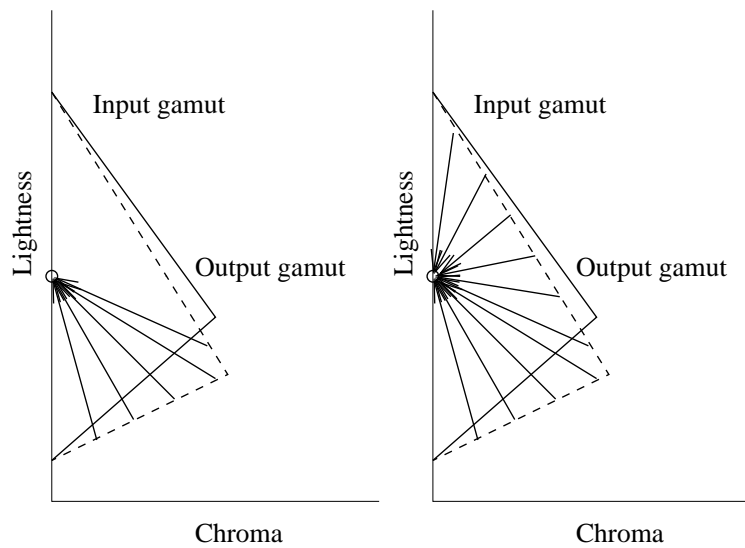
(d) simultaneous lightness and chroma compression

(e) uniform simultaneous lightness and chroma com-pression

Figure 4.21: Five different kinds of gamut compression. Note that the different gamut mapping strategies can be used after each other. In the figure, the output gamut obtained with lightness compression is used as input gamut for the other gamut compression strate-gies. The triangular shape of the gamut is schematic.

(a) orthogonal clipping   (b) chord clipping
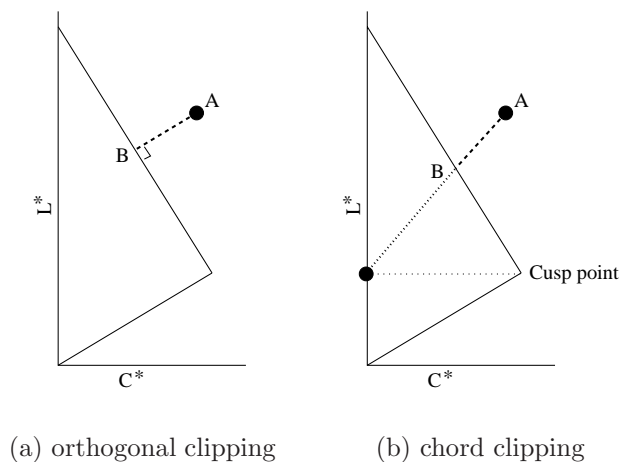
Figure 4.22: Two different kinds of gamut clipping

Method 1

To determine which points in the measured color space are on the gamut border, we first did a Delaunay tetrahedration of the reference points located near or on the gamut border. Because the measured color space is always 3 dimensional, this results in a number of tetrahedrons. If the gamut would be completely convex, the outer triangles of some tetrahedrons would be a good approximation of the gamut border. However, the gamut of a printer is normally (partially) concave.

Since we do not know a priori which or how many points are located on the boundary of the gamut, we made a Delaunay tetrahedration of all the reference points, and determined the intersection of all the faces of the tetrahedrons (triangles) with a constant hue plane. Sides of the triangles that have large sides were discarded, because this could be a concave part, where large sides connect the wrong points. This introduces a magic number, the threshold length. The result of this can be seen in figure 4.25 (a).

In this intersection, it is plainly seen that the number of points is much too large. Therefore, we select the outer line pieces by starting at the top or the bottom and selecting the line piece that is connected to the point and has the outer slope. The set of these line pieces is the intersection of the gamut at the constant hue plane. This set is given in figure 4.25 (b).

Method 2

Another way of determining the correct triangles is to use the known relations in the input color space[6]. This second method is also used by Hardeberg

---

[6]For this method it is assumed that the boundary in the input color space is transformed into the boundary in the output color space. More formally, the subset of boundary points of the gamut in the input color space are mapped onto a subset of the gamut in the output color space, and this subset is again the boundary.
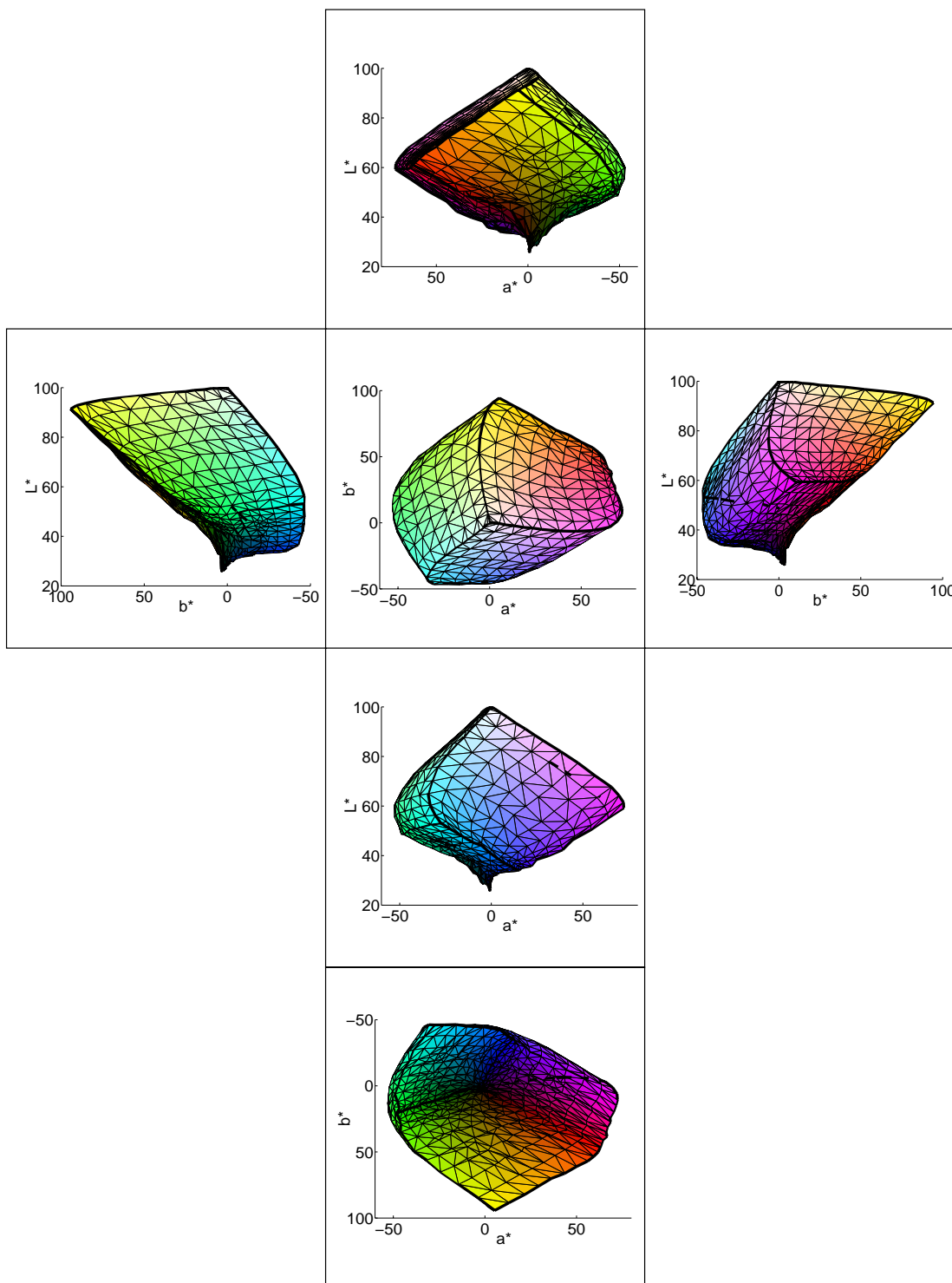
Figure 4.23: The gamut of the EPSON printer plotted in the CIELAB space for 6 different viewing points, corresponding to the 6 different faces of a cube. These images are made using the relations in the CMY_{EPSON} space. The solid lines indicate that two of the three inks are 0 or 100%.
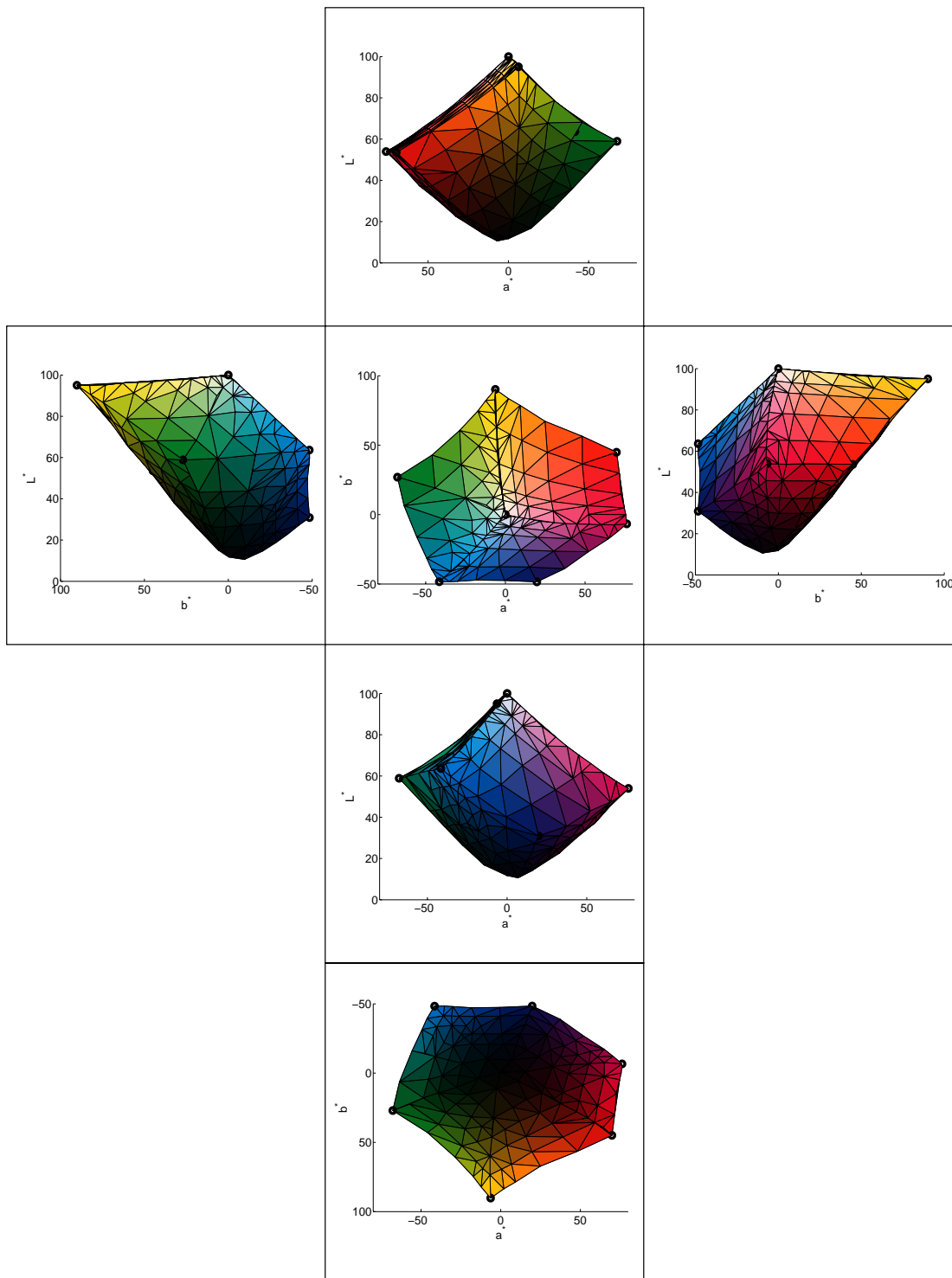
Figure 4.24: The gamut of the ISO data plotted in the CIELAB space for 6 different viewing points, corresponding to the 6 different faces of a cube. These images are made using the relations in the $CIELAB_{ISO}$ space (method 1). The circles indicate that C, M, and Y are either 0 or 100%. For all these points K is 0.
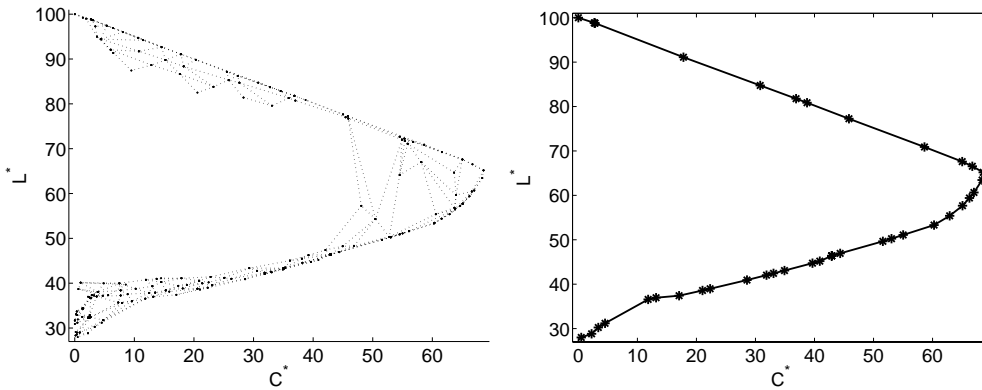
(a) Intersecting triangles in the 2D
constant hue plane

(b) Gamut boundary derived from (a)

Figure 4.25: The boundary of the gamut in a constant hue plane, using the first method. This implies first determining in the 2D constant hue plane before selecting the correct gamut boundary.

[Hardeberg and Schmitt, 1998]. If this color space is three dimensional, such as CMY or RGB, we can use the known ordering in the input color space. The boundary is divided into six 2-dimensional subspaces, in which one of the primaries is maximal or minimal. We can determine the Delaunay triangulation in these 2D subspaces, and then transform the triangles to the 3D measured color space. The results of this Delaunay triangulation is that we obtain the calibration points that are on the boundary of the gamut and the relations between these points.

To find the gamut boundary, we intersect all gamut triangles with the constant hue plane. The set of found line pieces is again the gamut boundary. This is visualized in figure 4.26.

If we use chord clipping (see figure 4.22 (b)) the mapping direction is known. To find the new position of point A, we have to find the intersection between the line in the gamut direction through point A and the line pieces. Note that there should be only one intersection. This is visualized in figure 4.27 (a). If we use orthogonal clipping (see figure 4.22 (a)), the mapping direction depends on the gamut boundary. We now determine the smallest distance to all line pieces. The point with the smallest distance is the new position B of point A. This means that for some points the mapping direction is not normal to the line piece. This mapping is given in figure 4.27 (b). This method is also used by Morovic [Morovic and Luo, 2000], who named the method "the flexible sequential line gamut boundary".

(a) intersection triangles / constant hue plane in 3D

(b) Gamut boundary derived from (a)
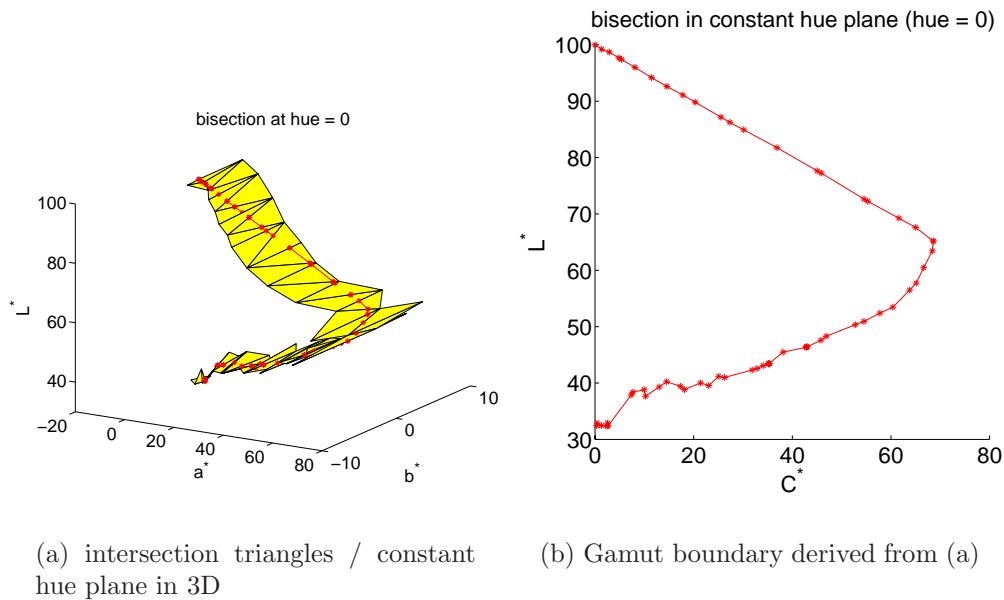
Figure 4.26: The boundary of the gamut in a constant hue plane using the second method. This implies determining the boundary triangles that intersect the constant hue plane in 3D. The gamut boundary is the intersection of the triangles and the constant hue plane.

(a) orthogonal clipping

(b) chord clipping
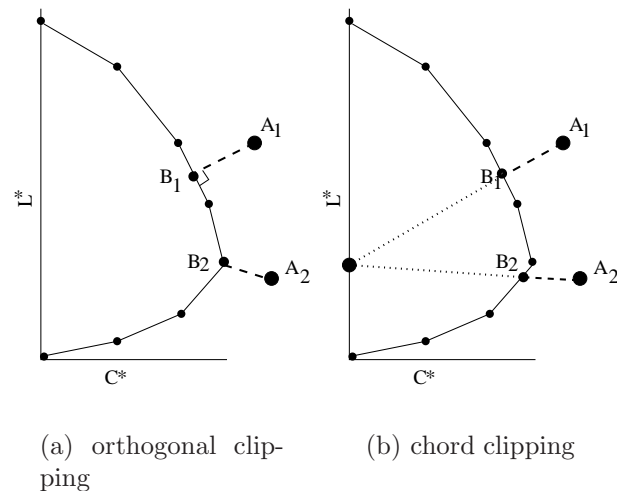
Figure 4.27: The line piece method for the two different kinds of gamut clipping. Note that for A_1, the mapping is orthogonal with respect to the line piece, which is not the case for A_2.

4.3.2   Testing the boundary found: leave-one-out method

In this section we measure how well the method described works. We used two data sets to determine the fit. The first data set is the standard ANSI CGATS TR 001-1995 [ANSI, 1995] set used for the calibration of offset printers. We measured the second data set ourselves on a EPSON 1520 ink jet printer, using EPSON photo
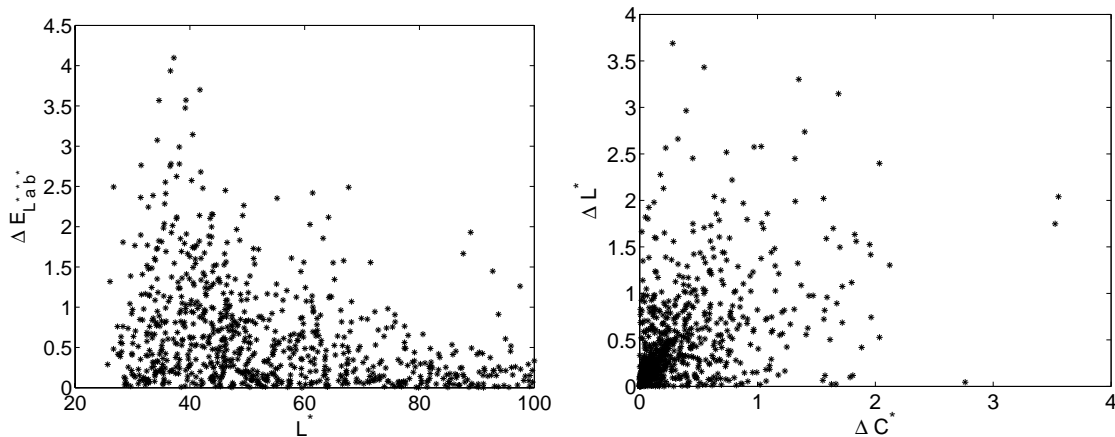
Figure 4.28: Perceptual errors (determined with a leave-one-out method) for the gamut of the EPSON printer using method 2. Note that the error in the lightness direction is independent of the error in the chroma direction.

quality paper. The input $CMYK_{EPSON}$ values were the same as given for the ISO set.

For the EPSON data we saw in section 4.2 that we can convert the data from $CMYK_{EPSON}$ to CMY_{EPSON} using equation 4.8. This means that we can use both method 1 and method 2. All data were plotted in CIELAB space.

We determined the error in the found boundary using a "leave-one-out" method, starting from the boundary points extracted from the data set. For each of these boundary points, the error is determined in the following way. First a new data set is selected consisting of all points except the boundary point under investigation. This data set is used to construct a new boundary point. The error is the difference between the original boundary point and the closest point in the constructed boundary. The calculation of the perceptual error, ΔE_{a*b*}, simplifies to $\sqrt{(\Delta L^*)^2 + (\Delta C^*)^2}$, because the hue is constant.

The perceptual error of the EPSON data set, using method 2, is plotted in figure 4.28. It can be seen that the error is typically below the visual threshold, indicating that this is a good way of determining the gamut.

Method 1 needs a magic number, that is the largest connecting line that is considered being a part of the gamut. This variable prevents very large line pieces on concave parts of the gamut. Instead of taking a fixed number, we made this number dependent on the lightness value of the two points. The maximum value of the line piece that is accepted, M_{maxval} is $rel\_maxval * (L_1 + L_2)$. The influence of this magic number can be seen in figure 4.29. It can be seen that this method works for the given relative maximum sizes, except for $rel\_maxval = 0.1$, where too many line pieces are rejected (figure 4.29 d).

(a) maximum size = 0.3

(b) maximum size = 0.2

(c) maximum size = 0.15

(d) maximum size = 0.1
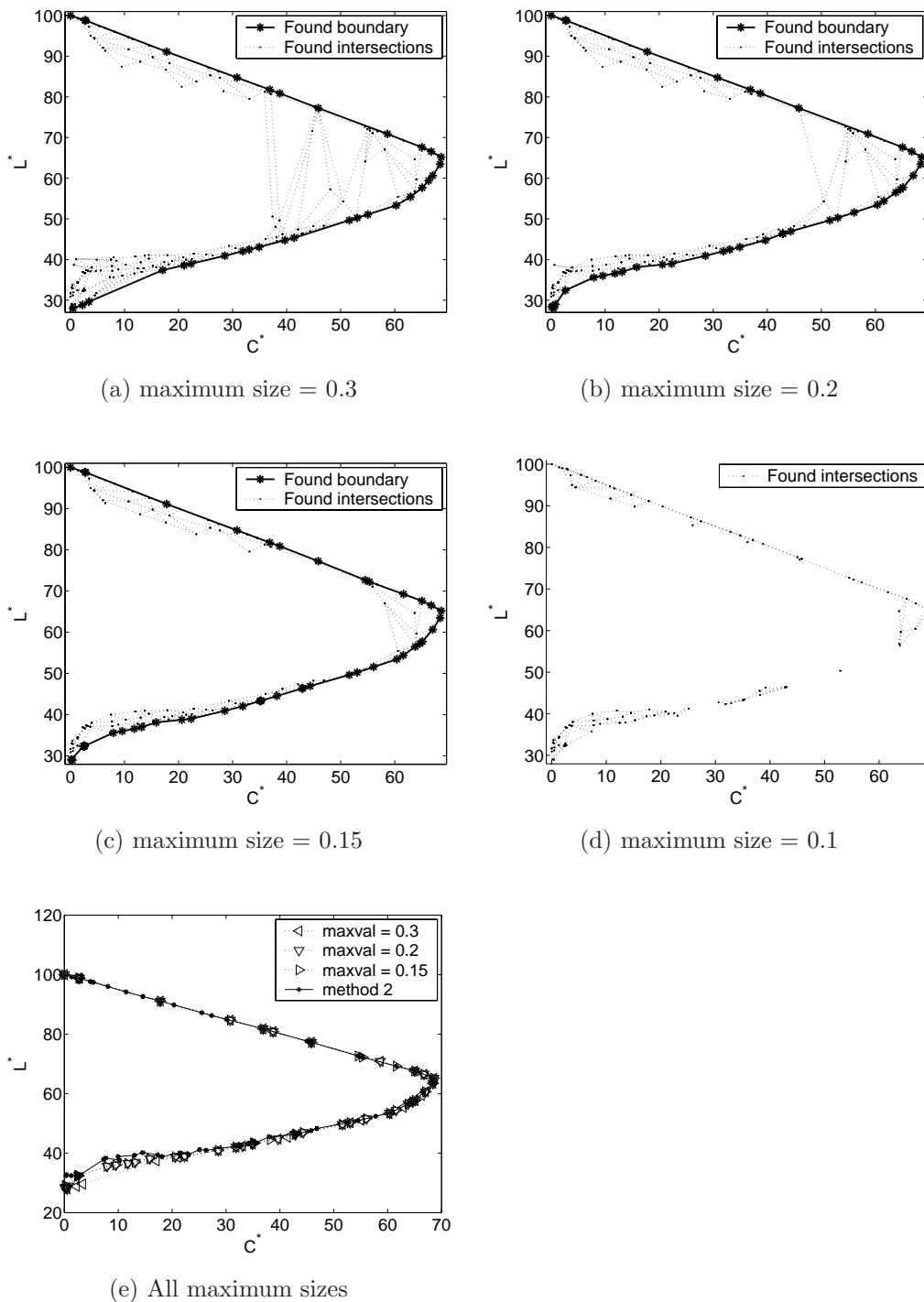
(e) All maximum sizes

Figure 4.29: The gamut boundary for the EPSON data set using different values for the relative maximum size (*rel_maxval*) of the connecting line piece using method 1. Boundaries could always be completed, except for *rel_maxval* = 0.1. The hue of the intersection is 0.
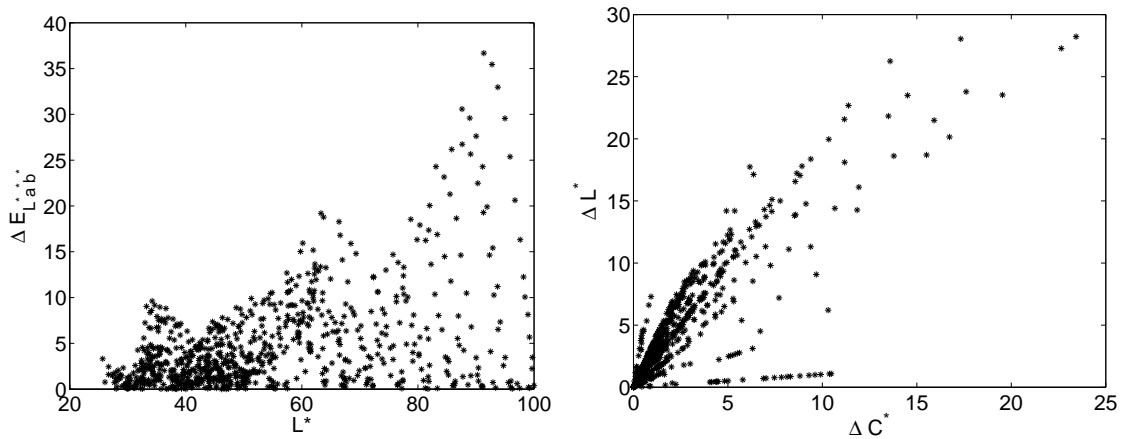
Figure 4.30: Perceptual errors (determined with a leave-one-out method) for the gamut of the EPSON printer using method 1. Note the dependence of the error in the lightness direction on the error in the chroma direction.

The perceptual error of the EPSON set, using the Delaunay tetrahedration in the measured color space, is plotted in figure 4.30. It can be seen that the errors are much larger than when we used the Delaunay triangulation in the input color space (see figure 4.28). The conclusion is that this method should not be used if we have information on the input color space. However, if this information is not available, and if the gamut boundary does not have to be determined very precisely, we may use this method.

For the ISO set we cannot determine the perceptual error, because it is not known which points are exactly on the gamut boundary. However, as is shown in figure 4.31 that the method using the Delaunay tetrahedration in the measured color space still gives reasonable gamut boundaries. However, it can be seen that the maximum size for which a good gamut is found is much larger than with the EPSON data set.

4.4   Methodology

After these sections on stimuli production we will focus in the rest of this chapter on performing and evaluating experiments with subjects. In the commonly used scaling methods it is assumed that the underlying variable is unidimensional. It is obvious, however, that for a variable such as image quality, this is not the case. Still, as Sjöberg [Sjöberg, 1987] has stated, if the differences between the stimuli are large enough, people can usually make consistent unidimensional ratings, even if the variables appear to be extremely complex. This consistency is obtained in the sense of rank ordering, but it may not be so easy to find more advanced quantitative properties. If the stimuli differ only marginally, classical scaling methods such as Thurstone's law of comparative judgement are the most appropriate, whereas for
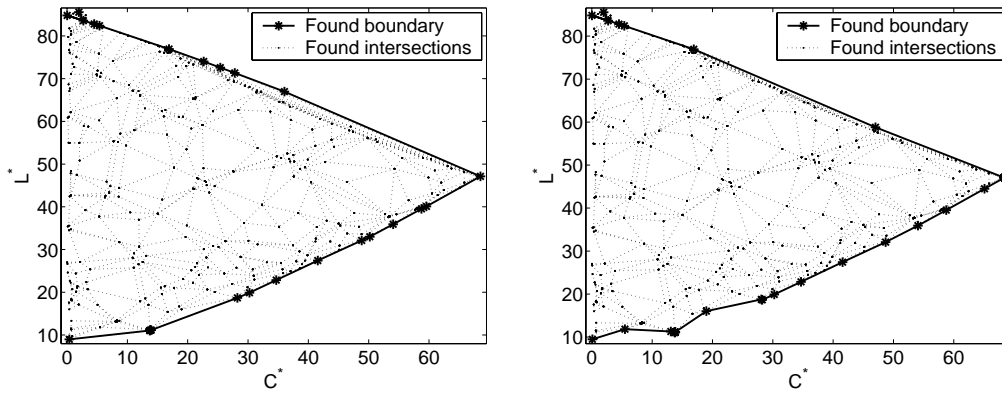
(a) $rel\_maxval = 0.6$

(b) $rel\_maxval = 0.5$
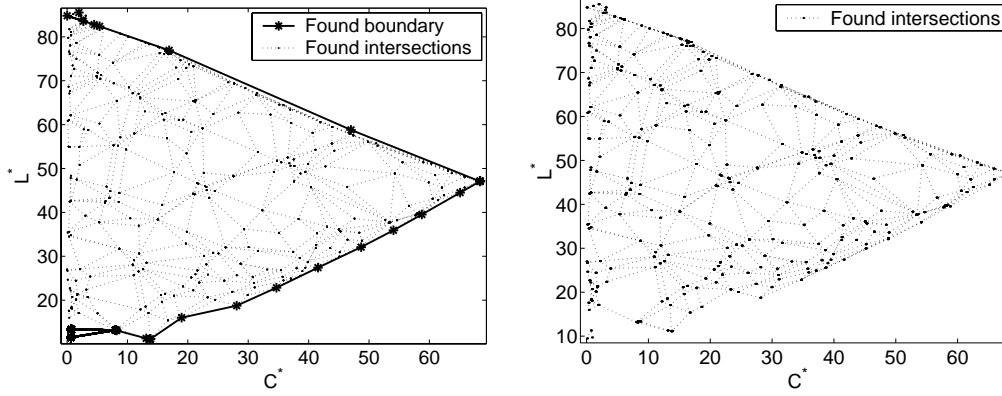
(c) $rel\_maxval = 0.4$

(d) $rel\_maxval = 0.3$
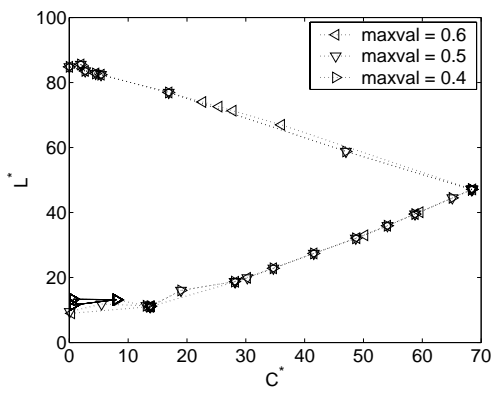
(e) All rel_maxval 's

Figure 4.31: The gamut boundary for the ISO data set using different values for the relative maximum size (rel_maxval) of the connecting line piece using method 1. Boundaries could always be completed, except for $rel\_maxval = 0.3$. The hue of the intersection is 0.

larger differences, absolute scaling is a better measurement method.

In this section we describe the methods we used for analyzing the results of subject experiments.

4.4.1   Rank ordering

With the rank ordering method the subjects are given a range of stimuli (in our case prints) which should be ordered according to a certain attribute (perceptual quality). The correlation between the subjective ordering and some measure of the variable in question can be calculated with the Spearman rank-order correlation coefficient r_s [Siegel and Castellan, 1988]. With this value, the null hypothesis H_0, i.e. the subjective ordering is not associated with the investigated measure (i.e. the subjective ordering and the measure under investigation are independent) can be tested against the hypothesis H_1, i.e. there is an association. The Spearman rank-order coefficient is defined as

$$r_s = 1 - \frac{6 \sum_{i=1}^{N} d_i{}^2}{N^3 - N} \tag{4.14}$$

in which N is the number of prints in the range and d_i is the difference in rank for the computational measure and the subjective ordering for each image in the range. If this coefficient is above a certain critical value (see table 4.1), the null hypothesis can be rejected. In our experiments, we use this also to see if subjects can see differences between the images. If this difference cannot be found, the results of the experiments cannot be used.

Table 4.1: The critical values for the Spearman rank-order correlation coefficient. These values hold for a two-tailed hypothesis with $\alpha = 0.05$, or a one-tailed hypothesis with $\alpha = 0.1$

N	5	6	7	8	9	10	12	14	16
critical value	1.000	0.886	0.786	0.738	0.700	0.648	0.587	0.538	0.503

4.4.2   Friedman two-way analysis of variance by ranks and Kendall Coefficient of Concordance

The Friedman two-way analysis of variance by ranks [Siegel and Castellan, 1988, StatSoft, 2000] is used to test the null hypothesis that the k samples, that are ranked, have been drawn from the same population. In our experiments, all N subjects scale the same k stimuli. The null hypothesis is that the distribution of the ranks over the stimuli is a matter of chance, that is, the means of the distributions

of the stimuli are the same. If the null hypothesis is true, the expected value of the sum of the ranks over the stimuli R_j would be $N(k+1)/2$. The Friedman test determines whether the found sum R_j is significantly different from the expected value. The statistic used is

$$F_r = \left[\frac{12}{Nk(k+1)} \sum_{j=1}^{k} R_j^2 \right] - 3N(k+1) \tag{4.15}$$

The probabilities of several values of F_r are tabulated for different numbers of subjects and stimuli. If k or N is large, it can be shown that F_r is distributed approximately as χ^2 with $k-1$ degrees of freedom.

If the obtained value of F_r is significant, we may reject the null hypothesis, that is we know that at least one of the stimuli is drawn from a population with a different mean. However, we do not know how much, and which stimuli are different. To test if stimulus u is different from stimulus v the following equation can be used

$$|R_u - R_v| \geq z_{\alpha/k(k-1)} \sqrt{\frac{Nk(k+1)}{6}} \tag{4.16}$$

where $z_{\alpha/k(k-1)}$ is the abscissa value from the unit normal distribution above which lies $\alpha/k(k-1)$ percent of the distribution, or in formula,

$$\alpha/k(k-1) = \frac{1}{2} + \frac{\sqrt{\pi}}{2} \mathrm{erf}(z_{\alpha/k(k-1)}). \tag{4.17}$$

Whereas the Friedman two-way analysis of variance by ranks is used to test the correlations between the stimuli, the Kendall Coefficient of Concordance W can used to test the correlations between the *subjects* [Siegel and Castellan, 1988, StatSoft, 2000]. W expresses the degree of agreement among the subjects and is defined as

$$W = \frac{\sum_{i=1}^{k}(\bar{R}_i - \bar{R})^2}{k(k^2-1)/12} \tag{4.18}$$

with $\bar{R}_i$ the average of the ranks assigned to the ith stimulus, and $\bar{R}$ the average of the ranks assigned across all stimuli. The denominator is the maximum possible sum of the squared deviations. The numerator would have this value if the subjects had complete agreement, in which case the average rankings would be 1, 2, ... k.

Instead of using Kendalls Coefficient of Concordance, one can also use a Spearman Rank Order Correlation between every two subjects. The number of rank-order correlations that has to be determined is $\binom{k}{2}$, which makes the procedure a bit

tedious for a large number of subjects. The relation between the average rank-order coefficient $\text{ave}(r_s)$ and the Coefficient of Concordance (W) is

$$\text{ave}(r_s) = \frac{kW - 1}{k - 1} \tag{4.19}$$

For small sample sizes the significance of W is determined and tabulated. For large sample sizes, it is found that the quantity

$$X^2 = N(k - 1)W \tag{4.20}$$

is approximately distributed as chi square with $k - 1$ degrees of freedom.

4.4.3   Analysis of Variance (ANOVA)

An Analysis of Variance (ANOVA) [StatSoft, 2000, Winer, 1970, Lane, 2003] is one of the methods that can be used to test if the differences that are found between different stimuli are statistically significant. The main assumption is that the probability function of each stimulus is normal. The null hypothesis is that the means of these distributions do not differ. The reason to use an ANOVA instead of other tests, for instance a t-test for independent samples, is that if we have more independent variables, we can test the significance of these variables with ANOVA and with less observations than if we would have performed several t-tests. The second advantage is that we can determine the significance of the interaction between variables. So, ANOVA is more powerful and gives more results. We shall give a short description of the ANOVA method.

We start with explaining the simple case of one variable. The following entities are determined: X_{ij} the response of subject i to stimulus j. X_{ij} can be summed over the stimuli: $P_i = \sum_{j=1}^{k} X_{ij}$ and over the subjects $T_j = \sum_{i=1}^{n} X_{ij}$. The sum over all observations $\sum_{i=1}^{n} \sum_{j=1}^{k} X_{ij}$ is called G. For the sums the average is also determined: $\bar{P}_i = P_i/k$, $\bar{T}_j = T_j/n$ and $\bar{G} = G/(nk)$.

The method is based on the idea that the variance in the entire experiment can be split into smaller parts. For instance, part of the variance is due to differences between subjects, part of the variance comes from differences between stimuli, and so on.

The total variance in the experiment is given by the sum of squares

$$\text{SS}_{total} = \sum_{i=1}^{n} \sum_{j=1}^{k} (X_{ij} - \bar{G})^2 \tag{4.21}$$

This variable has $kn - 1$ degrees of freedom, because there are n subjects and k stimuli, and only one linear relationship is assumed.

The variance in the experiment that is due to differences between the means of the subjects is given by

$$SS_{between\ subjects} = k \sum_{i=1}^{n} (\bar{P}_i - \bar{G})^2 \tag{4.22}$$

with $n - 1$ degrees of freedom. Note that if we use ranked data, this value will always be zero[7].

The variance within subject i is given by

$$SS_{within\ subject\ i} = \sum_{j=1}^{k} (X_{ij} - \bar{P}_i)^2 \tag{4.23}$$

This source of variance has $k - 1$ degrees of freedom. If this is pooled over all subjects, the total variance within the subjects is found as

$$SS_{w.subjects} = \sum_{i=1}^{n} \sum_{j=1}^{k} (X_{ij} - \bar{P}_i)^2 \tag{4.24}$$

with $n(k - 1)$ degrees of freedom.

It can be shown that $SS_{w.subjects}$ and $SS_{between subjects}$ are statistically independent and that

$$SS_{total} = SS_{w.subjects} + SS_{between subjects} \tag{4.25}$$

The variance within the subjects depends partly on differences in the stimuli and partly on uncontrolled or residual sources of variance. These variances are defined as

$$SS_{stimuli} = n \sum_{j=1}^{k} (\bar{T}_j - \bar{G})^2 \tag{4.26}$$

with $k - 1$ degrees of freedom, and

$$SS_{res} = \sum_{i=1}^{n} \sum_{j=1}^{k} \left[(X_{ij} - \bar{G}) - (P_i - \bar{G}) - (T_j - \bar{G}) \right]^2 \tag{4.27}$$

with $(k - 1)(n - 1)$ degrees of freedom. The relation between these variances is

$$SS_{w.subjects} = SS_{stimuli} + SS_{res} \tag{4.28}$$

where $SS_{stimuli}$ and SS_{res} are statistically independent.

A variance can also be given as a Mean Square. This is

$$MS = \frac{\text{variance}}{\text{degrees of freedom}} = \frac{SS}{df} \tag{4.29}$$

---

[7]Because all subjects give ranks from 1 to k, P_i is $(1 + k)/2$ for all subjects.

The F ratio

$$F = \frac{\text{MS}_{stimuli}}{\text{MS}_{res}} \tag{4.30}$$

can be used to test the null hypothesis, that is there is no effect for the different stimuli[8]. This variable has a sampling distribution which is approximated by an F distribution, which has $k - 1$ degrees of freedom for the numerator and $kn - k$ degrees of freedom for the denominator. If there is no difference between the stimuli, F is close to 1. The larger F is, the more probable the effect is. F can be compared to tabulated values of F to obtain the probability that the means are different.

If the data is ranked data, instead of the F ratio the χ^2_{ranked} statistic is usually used:

$$\chi^2_{ranked} = \frac{\text{SS}_{stimuli}}{(\text{SS}_{stimuli} + \text{SS}_{res})/n(k-1)} = \frac{n(k-1)\text{SS}_{stimuli}}{\text{SS}_{w.subjects}} \tag{4.31}$$

When no ties are permitted

$$\text{SS}_{w.subjects} = \frac{nk(k^2 - 1)}{12} \tag{4.32}$$

with this the χ^2 statistic becomes

$$\chi^2_{ranked} = \frac{12}{nk(k+1)} \sum T_j^2 - 3n(k+1) \tag{4.33}$$

which is the same as the Friedman Rank order coefficient.

So far we have studied one variable problems, given in the references [Winer, 1970, StatSoft, 2000, Lane, 2003]. In experiments in which there are two or more independent variables, we would want to know if the effects of the variables are statistically significant, and also if there is an interaction between the different variables. Here we derive the relationship for two variables with their interaction. For more variables the principle is the same.

The following entities are used $X_{ij_1j_2}$ is the response of subject i to the stimulus with independent variables j_1 and j_2. The number of subjects is n, the number of different values of variable 1 is k_1, the number of different values of variable 2 is k_2. $X_{ij_1j_2}$ can be summed over the subjects: $Y_{j_1j_2} = \sum_{i=1}^{n} X_{ij_1j_2}$. If we also sum over the variable k_2, the only remaining variance is due to differences in variable 1, and vice versa: $T_{j_1} = \sum_{j_2=1}^{k_2} Y_{j_1j_2}$ and $T_{j_2} = \sum_{j_1=1}^{k_1} Y_{j_1j_2}$. The sum over all variables except the subjects is given as $P_i = \sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} X_{ij_1j_2}$. The sum over one variable and the subjects is given as $Q_{ij_1} = \sum_{j_2=1}^{k_2} X_{j_1j_2}$ and $Q_{ij_2} = \sum_{j_1=1}^{k_1} X_{j_1j_2}$. The total sum of the responses is given as $G = \sum_{i=1}^{n} \sum_{j_2=1}^{k_2} \sum_{j_1=1}^{k_1} X_{ij_1j_2}$ The average of these sums are $\bar{Y}_{j_1j_2} = Y_{j_1j_2}/n$, $\bar{T}_{j_1} = T_{j_1}/(nk_2)$, $\bar{T}_{j_2} = T_{j_2}/(nk_1)$, $\bar{P}_i = P_i/(k_1k_2)$, $\bar{Q}_{j_1} = Q_{j_1}/k_2$, $\bar{Q}_{j_2} = Q_{j_2}/k_1$ and $\bar{G} = G/(nk_1k_2)$.

---

[8]Note the similarity with the Signal to Noise Ratio.

The total variance in the experiment is given by

$$SS_{total} = \sum_{i=1}^{n} \sum_{j_2=1}^{k_2} \sum_{j_1=1}^{k_1} (X_{ij_1j_2} - \bar{G})^2 \tag{4.34}$$

This equation reduces to 4.21 if $k_2 = 1$. SS_{total} has $nk_1k_2 - 1$ degrees of freedom.

The variance due to differences between subjects is

$$SS_{between\ subjects} = k_1k_2 \sum_{i=1}^{n} (\bar{P}_i - \bar{G})^2 \tag{4.35}$$

with n degrees of freedom.

The variance due to the first variable is given by

$$SS_{variable\ 1} = nk_2 \sum_{j_1=1}^{k_1} (\bar{T}_{j_1} - \bar{G})^2 \tag{4.36}$$

with $k_1 - 1$ degrees of freedom. This equation reduces to 4.26 if $k_2 = 1$. In the same way the variance of the second variable is given by

$$SS_{variable\ 2} = nk_1 \sum_{j_2=1}^{k_2} (\bar{T}_{j_2} - \bar{G})^2 \tag{4.37}$$

with $k_2 - 1$ degrees of freedom.

In our case, we have ranked data. The data is ranked with one variable, the image, constant. Assuming that the image is variable 2, $SS_{variable\ 2}$ is zero.

The new term is the variance due to the interaction of the two variables.

$$SS_{interaction} = n \sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} (\bar{Y}_{j_1j_2} - \bar{G})^2 - nk_1 \sum_{j_1=1}^{k_1} (\bar{T}_{j_1} - \bar{G})^2 - nk_2 \sum_{j_2=1}^{k_2} (\bar{T}_{j_2} - \bar{G})^2 \tag{4.38}$$

$(k_1 - 1)(k_2 - 1)$ degrees of freedom. Note that $SS_{interaction}$ is zero if k_1 or k_2 is 1. The total variance is given by

$$SS_{total} = SS_{total\ variable\ 1} + SS_{interaction} + SS_{res} \tag{4.39}$$

The residual sum of squares can be determined using

$$SS_{total\ variable\ 1} = \sum_{i=1}^{n} \sum_{j_1=1}^{k_1} (\bar{Q}_{j_1} - \bar{G})^2$$

$$SS_{res,variable\ 1} = SS_{total\ variable\ 1} - SS_{variable\ 1} \tag{4.40}$$

$$SS_{res,interaction} = SS_{total} - SS_{variable\ 1} - SS_{res,variable\ 1} - SS_{interaction}$$

Again the Mean Squares can be determined by dividing the variances by the proper residual variance

$$
\begin{aligned}
\mathrm{MS}_{variable\ 1} &= \frac{\mathrm{MS}_{variable\ 1}}{k_1 - 1} \\
\mathrm{MS}_{res,variable\ 1} &= \frac{\mathrm{SS}_{res,variable\ 1}}{(n-1)(k_1 - 1)} \\
\mathrm{MS}_{interaction} &= \frac{\mathrm{SS}_{interaction}}{(k_1 - 1)(k_2 - 1)} \\
\mathrm{MS}_{res,interaction} &= \frac{\mathrm{SS}_{res,interaction}}{(n-1)(k_1 - 1)(k_2 - 1)}
\end{aligned}
\tag{4.41}
$$

Because we test for two effects (the main effect of variable 1 and a two-way interaction), we obtain two F ratio's. Note that the main effect of variable 2 is not tested because our test setup does not allow this.

$$
\begin{aligned}
F_{variable\ 1} &= \frac{\mathrm{MS}_{variable\ 1}}{\mathrm{MS}_{res,variable\ 1}} \\
F_{interaction} &= \frac{\mathrm{MS}_{interaction}}{\mathrm{MS}_{res,interaction}}
\end{aligned}
\tag{4.42}
$$

The generalization to three or more variables is obvious. For three variables, for instance equation 4.39 is

$$
SS_{total} = \mathrm{SS}_{total\ variable\ 1} + \mathrm{SS}_{interaction\ 1,2} + \mathrm{SS}_{interaction\ 1,3} + \mathrm{SS}_{interaction\ 1,2,3} + \mathrm{SS}_{res}
\tag{4.43}
$$

To test if different stimuli are significantly different several different tests an be used, from which the most common test the Newman-Keuls test is. The Newman-Keuls test uses the variable q_r, which is defined as

$$
q_r = \frac{T_i - T_j}{n \mathrm{MS}_{error}}
\tag{4.44}
$$

r is the number of steps which the stimuli i and j are apart on an ordered scale. q_r is significant if it is larger than $q_{1-\alpha}(r, f)$, with f the degrees of freedom of MS_{error}. $q_{1-\alpha}(r, f)$ is the $(1 - \alpha)$ point on the q distribution. To avoid inconsistencies, the order of testing the different pairs of stimuli is always the same. In this procedure the significance of all tests is equal to α. But if we consider the collection of tests as a single test, the significance level is considerably lower than α. The Duncan test is identical to the Newman-Keuls test, but the critical values are corrected by protection levels to obtain a significance of α for the entire test.

4.4.4   Bisection

The bisection method is used to obtain information about the stimulus-response function of the stimuli. The subject is presented a number of stimuli and has to decide which print is in the center of a range of prints, for a certain criterion. For instance, if the prints differ in grey value, the subject has to select the print with the grey value that is exactly in the middle of the maximum and minimum grey value. In this way one point of the response function can be found. By varying the range, an estimate of the total response function is obtained. To test if the found bisection points are different from the physical center of the range, we traditionally used the student-t statistic.

Chapter 5

Relative Sharpness and Smoothness

5.1 Introduction

Natural images contain lines, edges and line and edge-free areas. In between, of course, are "textures". Some of these lines and edges are sharp, and some of the areas are smooth. Due to several causes, such as blurring and noise, lines and edges can be blurred, and smoothed areas can become less smooth. In image processing there are two types of operations to restore this, i.e. sharpening and smoothing. Adaptive operations can perform both. An example of an adaptive operation is anisotropic diffusion [Perona et al., 1994].

Two perceptual attributes that determine image quality are the sharpness of the lines and edges in an image and the smoothness of the areas. These perceptual attributes will depend upon the results of the image processing operations. In this chapter we will discuss measures for sharpening and smoothing. Such measures typically are *relative* measures. In the next chapter, however, we will propose a measure for *absolute* sharpness, independent of a reference image.

5.2 Sharpening and smoothing

In image enhancement two operators are commonly used: sharpening of the structures (lines and edges) and smoothing of the (almost uniform) areas. Most of the pixels in an average printer image "belong" to an area, and only a small number to lines and edges [Nishikawa et al., 1965]. This indicates that a filter, which performs both sharpening and smoothing, cannot be evaluated well by a measure that treats every pixel the same.

This observation has led us to formulate of *two* new objective measures for estimating the smoothing and sharpening effect of a filter. Since these measures, developed by Dick the Ridder, have worked satisfactorily in a neural network approach

[de Ridder et al., 1998, de Ridder, 2001] (i.e. they seemed to do so according to subjective appreciation), the question arose as to what extent these really conform to human judgement. To answer this question, we asked subjects to indicate relative smoothing and/or sharpening on a number of filtered images. The filter used was an edge-preserving smoothing filter, an anisotropic diffusion filter.

5.3   Sharpening and smoothing measures

In order to devise an informative performance measure for both sharpening and/or smoothing induced by a filter, the two effects have to be separated. Thereto, a scattergram is plotted of the pixels of the gradient magnitude of the original image I versus those of the gradient magnitude of the filtered version $f(I)$. The gradient magnitude measures the maximum deviation in a local environment around the pixel, or the absolute slope [van Vliet, 1993]. The gradient magnitude is defined as [Young et al., 1998]

$$|\nabla I| = \sqrt{(\frac{dI}{dx})^2 + (\frac{dI}{dy})^2}$$ (5.1)

where $\frac{dI}{dx}$ and $\frac{dI}{dy}$ are the derivative in the x and y direction, respectively. We used Gaussian derivative filters.

Figure 5.1 (a) show four examples. The filters used are described in section 5.4. Horn and Bachman [Horn and Bachman, 1978] have used a scattergram approach to determine the similarity of two images, in their case a real and a synthetic image. In contradiction to our approach, they constructed a scattergram of the pixels of the two images. However, it is hard to relate the cloud of points that are found to image similarity (see also [Katsulai and Arimizu, 1981]). By taking the gradient magnitude of the pixels instead of the grey value, we can relate pixel positions in the scattergram to the effect of the filter used.

Pixels are classified as either being sharpened or smoothed. In the first case, the gradient will be steeper: i.e. the pixel is plotted above the line $y = x$ in the scattergram. Pixels which are smoothed will end up below this line. The main assumption is that pixels on lines and edges are sharpened, and that the other pixels are smoothed. However, this assumption is not used to group the pixels. All sharpened and smoothed pixels are grouped into sets $\mathcal{A}$ and $\mathcal{B}$, where $\mathcal{A}$ are the pixels for which the filtered gradient is smaller than the original gradient (below the line $y = x$), and $\mathcal{B}$ are the pixels for which the filtered gradient is larger than the original gradient (above the line $y = x$). Note that in general $|\mathcal{B}| << |\mathcal{A}|$, since fewer pixels lie on edges than in smooth regions. Because the points in $\mathcal{A}$ are more cluttered (near the origin) than the points in $\mathcal{B}$, $|\mathcal{B}|$ *appears* larger. Lines $y = ax + b$ can be fitted through both sets using a robust estimation technique (medfit),

(a) Unsharp masking

(b) Gaussian smoothing

(c) Anisotropic diffusion

(d) Kuwahara
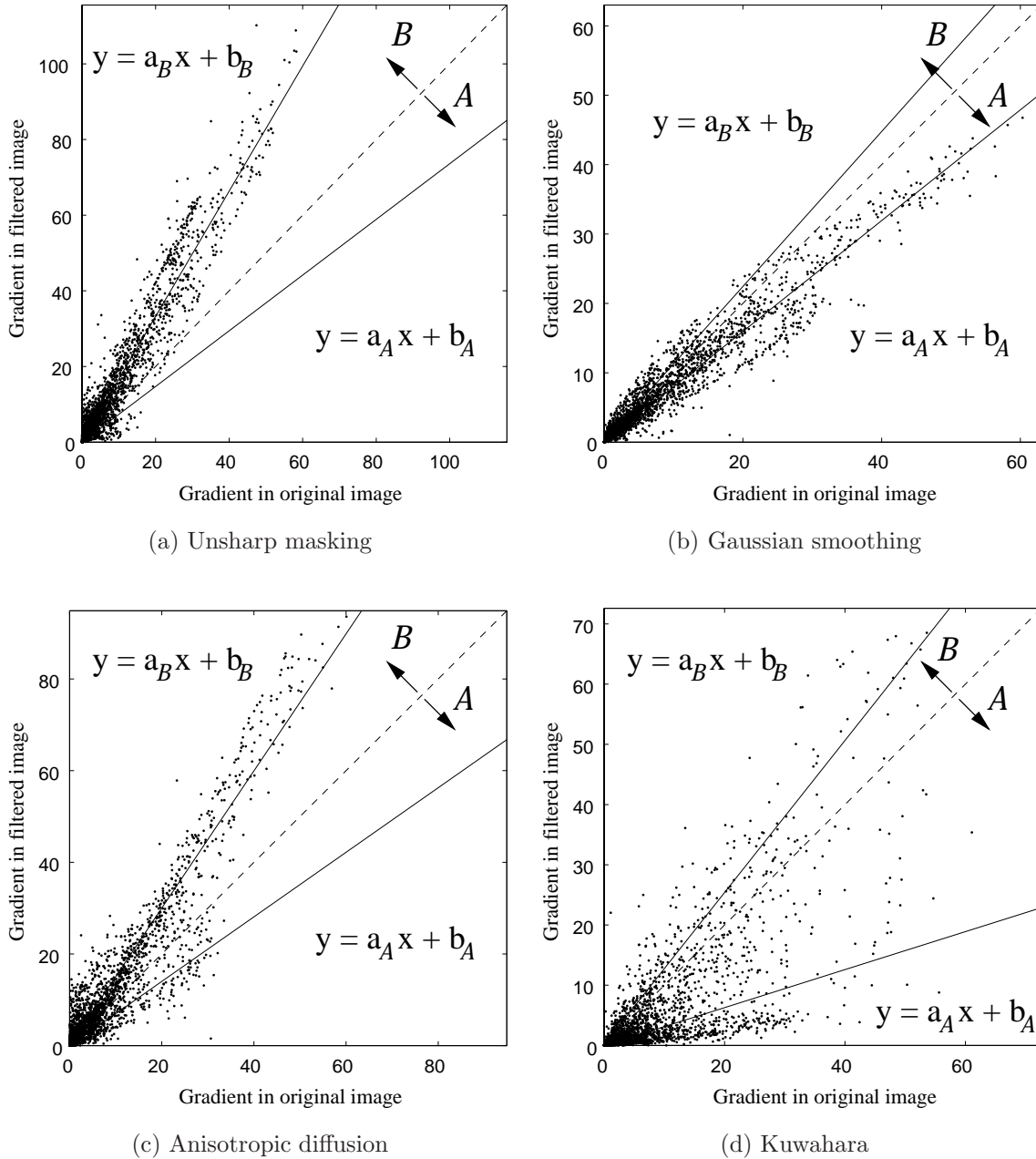
Figure 5.1: Scattergrams of gradient magnitude of an original image (x-axis) versus the gradient magnitude of the filtered version (y-axis). The filters are discussed in section 5.4. Only the pixel values of 3000 random points are plotted. See text for explanation.

which minimizes the absolute deviation [Press et al., 1988], to get an estimate for smoothing and sharpening, respectively, that is:

$$(a_{\mathcal{A}}, b_{\mathcal{A}}) \;=\; \arg\min_{(a,b)} \sum_{(x,y)\in\mathcal{A}} |y - (ax + b)| \tag{5.2}$$

$$(a_{\mathcal{B}}, b_{\mathcal{B}}) \;=\; \arg\min_{(a,b)} \sum_{(x,y)\in\mathcal{B}} |y - (ax + b)| \tag{5.3}$$

The estimation procedure is not the normal least square error method, but a method which is less sensitive to outliers. The slope of the lower line found, $a_\mathcal{A}$, will give an indication of the smoothing induced by the filter f. Likewise, $a_\mathcal{B}$ gives an indication of the sharpening effect of the filter. The offsets $b_\mathcal{A}$ and $b_\mathcal{B}$ are discarded, although it is necessary to estimate them to avoid biasing the estimate of $a_\mathcal{A}$ and $a_\mathcal{B}$. If these values would not be estimated, then we would see a sharpening and smoothing for all images[1]. Note that a demand is that $a_\mathcal{A} \leq 1$ and $a_\mathcal{B} \geq 1$, so the values are clipped at 1 if necessary[2].

To account for the number of pixels actually used to estimate these values, the slopes found are weighted with the relative number of points used for the estimate. Therefore, the numbers

$$Smoothing(f, I) \;=\; (a'_\mathcal{A} - 1)\frac{|\mathcal{A}|}{|\mathcal{A}| + |\mathcal{B}|} \tag{5.4}$$

$$Sharpening(f, I) \;=\; (a_\mathcal{B} - 1)\frac{|\mathcal{B}|}{|\mathcal{A}| + |\mathcal{B}|} \tag{5.5}$$

are used, where $a'_\mathcal{A} = \frac{1}{a_\mathcal{A}}$ was substituted to obtain numbers in the same range $[0, \infty\rangle$. These two values can be considered to be an amplification factor of edges and an attenuation factor of flat regions, respectively. Note that these measures depend on:

- image content;

- the filter used;

- any intermediate or post processing such as scaling or contrast stretching.

Given a certain image and using no further processing, these numbers can therefore be used to compare filter operation. Note that the procedure fails in the presence of excessive noise [3].

5.4   Edge-preserving smoothing

To judge the correspondence between the measures proposed in section 5.3 and human judgement, one could do experiments with, in principle, any kind of edge-

---

[1]In the model we use the gradient after filtering g_2 is linearly related to the gradient before filtering g_1, so $g_2 = cg_1$. This implied that all estimated lines should have no offset. However, this model is not exactly true.

[2]$a_\mathcal{A}$ can be larger than one if there is no smoothing and if a small offset b_A is estimated.

[3]For example, if we add Gaussian noise with standard deviation 10 to $\mathtt{trui}$, the sharpening found is -0.01 and the found smoothing is 0.01. The signal-to-noise ratio is 13.9, using $\mathrm{SNR} = 20\log_{10}\frac{s_a}{s_n}dB$ with s_a the standard deviation of the signal and s_n the standard deviation of the noise [Young et al., 1998]. For a Gaussian noise with standard deviation 20, the sharpening found is -0.04 and the found smoothing is 0.02. The SNR is 7.87.

preserving smoothing filter. However, one obvious demand is that the algorithm used has parameters which allow small differences in sharpening and smoothing to be created. This is necessary to be able to create a large number of images which span the sharpening-smoothing space. For instance, the Kuwahara filter [Kuwahara et al., 1976, Young et al., 1998] is not applicable, since the only parameter used is the window width (5, 9, 13, ...) resulting in too coarse a spacing.

The filters used in the experiment are:

Gaussian (unconditional smoothing): a purely smoothing Gaussian

$$f_G(I, \sigma) = G_\sigma * I \tag{5.6}$$

with

$$G_\sigma \equiv \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{5.7}$$

and $\sigma = 0.0, 0.1, ..., 2.0$.

Gaussian unsharp masking (unconditional sharpening): An edge enhancement technique that is also used in photography is unsharp masking [Yule, 1944, Young et al., 1998]. Edge enhancement can be separated in three steps: first the edges in the image are isolated, then these edges are amplified, and the last step is adding the amplified edges back into the image. The Laplace of Gaussian can be used for isolating the edges, giving

$$f_U(I, k) = I - kG_\sigma * \nabla \cdot \nabla I \tag{5.8}$$

where the amplifying term k is larger than zero. In the experiments k was varied in the range $0.0, 0.1, ..., 2.0$. The parameter σ was fixed at 1.0.

Anisotropic diffusion with unsharp masking (conditional smoothing): The diffusion equation proposed by Perona and Malik [Perona et al., 1994], which models intensity as pressure, is given by

$$I_{t+1} = I_t + \frac{1}{2}\nabla \cdot (C_t(x, y)\nabla I) \, \Delta t \tag{5.9}$$

where ∇I is an intensity gradient, C_t is a conduction function, $C\nabla I$ is an intensity current and $\Delta C\nabla I$ is accumulated intensity. For anisotropic diffusion, Perona and Malik take for C_t a function indicating the absence of an edge. The image is smoothed in places where no edges are present ($C_t = 1$) but not changed near the edges. Since the location of the edges is not exactly known, some monotonously decreasing function of the gradient magnitude is used, usually

$$C_t(x, y) = T(|\nabla I|, \tau)$$
$$T(m, \tau) = \exp\left(-\frac{m^2}{\tau}\right) \tag{5.10}$$

(a) `bicycle`

(b) `portrait`
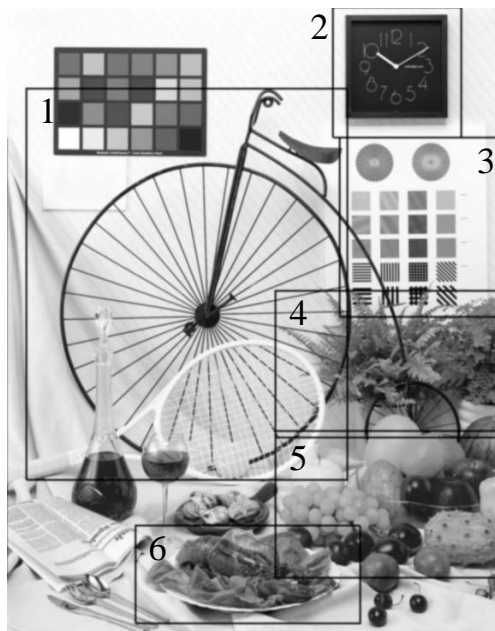
(c) `bicycle` (parts)
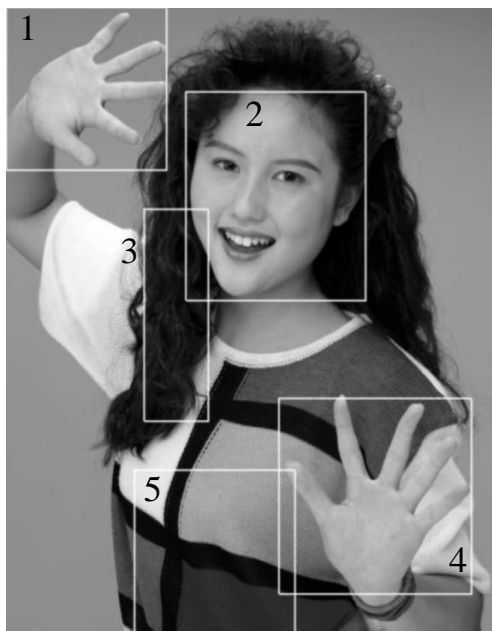
(d) `portrait` (parts)

Figure 5.2: The two images (a,b) and parts thereof (c,d) used in the experiments. In (c) the parts are called (1) `bike` (the part with only the bicycle), (2) `clock`, (3) `test pattern`, (4) `plant`, (5) `fruit` and (6) `lobster`. In (d) the parts are called (1) `right hand`, (2) `face`, (3) `hair`, (4) `left hand` and (5) `sweater`.

which falls to zero for large values of the gradient. The parameter σ_1 determines how large a gradient has to be in order to be considered an edge.

In practice, the flux (i.e. current through a certain area) between two pixels a and b is approximated by

$$\frac{1}{2}\Delta t C_t(x,y)\nabla I \equiv \phi_t(p,n) \quad = \tfrac{1}{2}\Delta t \tfrac{1}{2}(C_t(x_p,y_p)+C_t(x_n,y_n)) \cdot$$
$$(I_t(x_p,y_p)-I_t(x_n,y_n)) \qquad (5.11)$$

so that the update rule (one iteration) for one pixel p becomes

$$I_{t+1}(x_p,y_p) = I_t(x_p,y_p) + \sum_{n\in n_4} \phi_t(p,n) \qquad (5.12)$$

where n_4 denotes the 4-connected neighborhood of pixel p. The time step Δt in equation 5.11 is fixed to 0.25, giving updates as large as possible but keeping the scheme numerically stable [Niessen et al., 1994]. The number of iterations is a parameter, N.

In the experiments described below, a modification proposed by Catté [Catté et al., 1992] was used. That is, the gradient magnitude in equation 5.10 is calculated with a Gaussian derivative:

$$J_t \quad = G_{\sigma_2} * I_t \qquad (5.13)$$
$$C_t(x,y) \quad = T(|\nabla J|,\sigma_1) \qquad (5.14)$$

This introduces a second parameter σ_2, which can be used to suppress noise.

The diffusion operation described here has a purely smoothing effect[4]. To make the filter both sharpening and smoothing, images were pre-filtered with the unsharp masking filter $f_U(I,k)$ described above. This introduces another parameter, k. The total filter therefore is $f_A(f_U(I,k),\sigma_1,\sigma_2,N)$. In the experiments, the parameters were varied as follows: $k = 0.0, 1.0, ..., 5.0$; $\sigma_1 = 1.0, 2.0, ..., 5.0$; $\sigma_2 = 0.0, 0.25, ..., 2.0$ and $N = 5, 7, 10, 14, 20, 28, 40, 56, 80$.

We can use the diffusion equation 5.9 to relate the three filters to each other. If $C(x,y)$ is 1 for all x, y, the result of equation 5.9 reduces to a Gaussian with a σ of $\sqrt{t}$, or in formula

$$I(t) = G(\sigma = \sqrt{t}) * I(0) \qquad (5.15)$$

The actual smoothing is done by *adding* a Laplacian. This suppresses edge information. If we use one step backwards diffusion (that is, $t < 0$), we obtain

$$I(\Delta t) = I(0) + \frac{1}{2}\nabla^2 I \Delta t \stackrel{\Delta t = -2k}{=} I(0) - k\nabla^2 I \qquad (5.16)$$

---

[4]It can sharpen edges due to the smoothing of the regions that meet at an edge, but does so without overshoot. Incidentally, this is what the Kuwahara filter does. Unfortunately, the Kuwahara effect is so "heavy-handed" that it is not possible to find subtle distinctions over images over an allowed range of window sizes (5, 9, 13, ...)
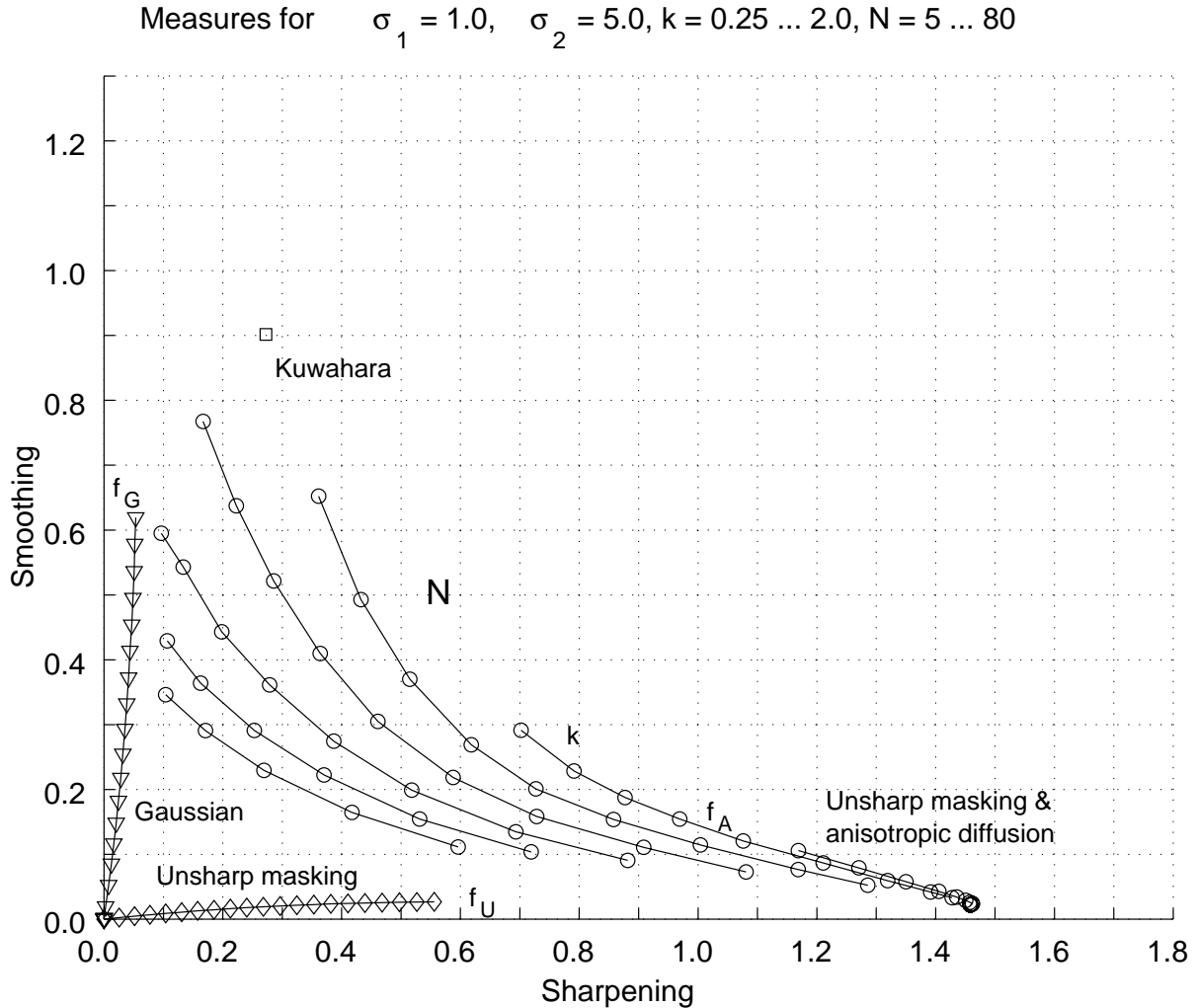
Figure 5.3: Sharpening/smoothing values for a number of different filters on the `portrait` image. Definitions of the Gaussian filter f_G, the unsharp masking filter f_U and the anisotropic diffusion filter f_A are given in the text. The Kuwahara filter is only given for N is 5. Note that for the anisotropic diffusion filter only a small subset of the filters is given, because σ_1 and σ_2 are constant. The points with the same k setting are connected by a line.

which is the same as equation 5.8 without G_σ. To be able to compare the different filters, we should use stepsizes in which $\sigma^2/2 \sim -k$.

Figure 5.3 shows an example of how various filtered versions of an image end up in the sharpening-smoothing space. In appendix B transfer functions of most of these filters are given.

5.5 Experiments

5.5.1 Experimental Setup

The experimental images were made from the two images shown in figure 5.2. They are ISO standard images taken from the CD-ROM 12640:1997 and are originally 300 dpi, in CMYK format. The images were first converted to RGB, using Adobe PhotoShop. Next, the images were converted to 32-bit grey scale floating point images. Following ITU [Bourgin, 1998], the luminance Y was defined as: $Y = 0.222R + 0.707G + 0.071B$. To reduce the amount of computation time needed, the images were reduced to $\frac{3}{8}$ of their original size in both the x and y direction, by pre-smoothing with a Gaussian ($\sigma = 2.4$) and linearly interpolating.

For both images, the sharpened and/or smoothed versions (f_G, f_U and f_A) were calculated with the parameter settings described in section 5.4, resulting in a large number of image points in sharpening-smoothing space. Finally, the images were printed on a 600 dpi HP LaserJet 4000N. Print size was 12.8×16 cm; dithering was done by the printer.

A room was used containing a special, somewhat tilted table on which the subject could sort the prints. The light source was a studio lamp, which provided homogeneous, indirect lightning of the prints. The luminance on the table was approximately 600 lux. The prints were put in plastic covers to prevent them from becoming dirty. A window was cut in the center to see the image on the print directly and not through plastic. Four subjects participated in the experiments, all having some experience in the field of image analysis.

In the instruction to the subjects, sharpening was explained as the sharpening of *edges* and smoothing as the smoothing of *regions*. Although this might introduce a bias in the outcome of the experiments, it was deemed necessary since some of the subjects had no clear concept of sharpening or smoothing.

5.5.2 Experiment A: stimulus response

The goal of this experiment was to see whether subjects could discriminate levels of sharpening and smoothing as defined in the previous sections. The subjects were given a range of prints and asked to rank order them by perceived sharpening or smoothing.

From the two-dimensional feature space, one-dimensional ranges were drawn containing images with constant smoothing or sharpening. Four different sharpenings and smoothings were used: 0.0, 0.1, 0.3 and 0.5. These values were used for constructing series Sm_x (Sh_y), consisting of – at most – 8 images with fixed smoothing
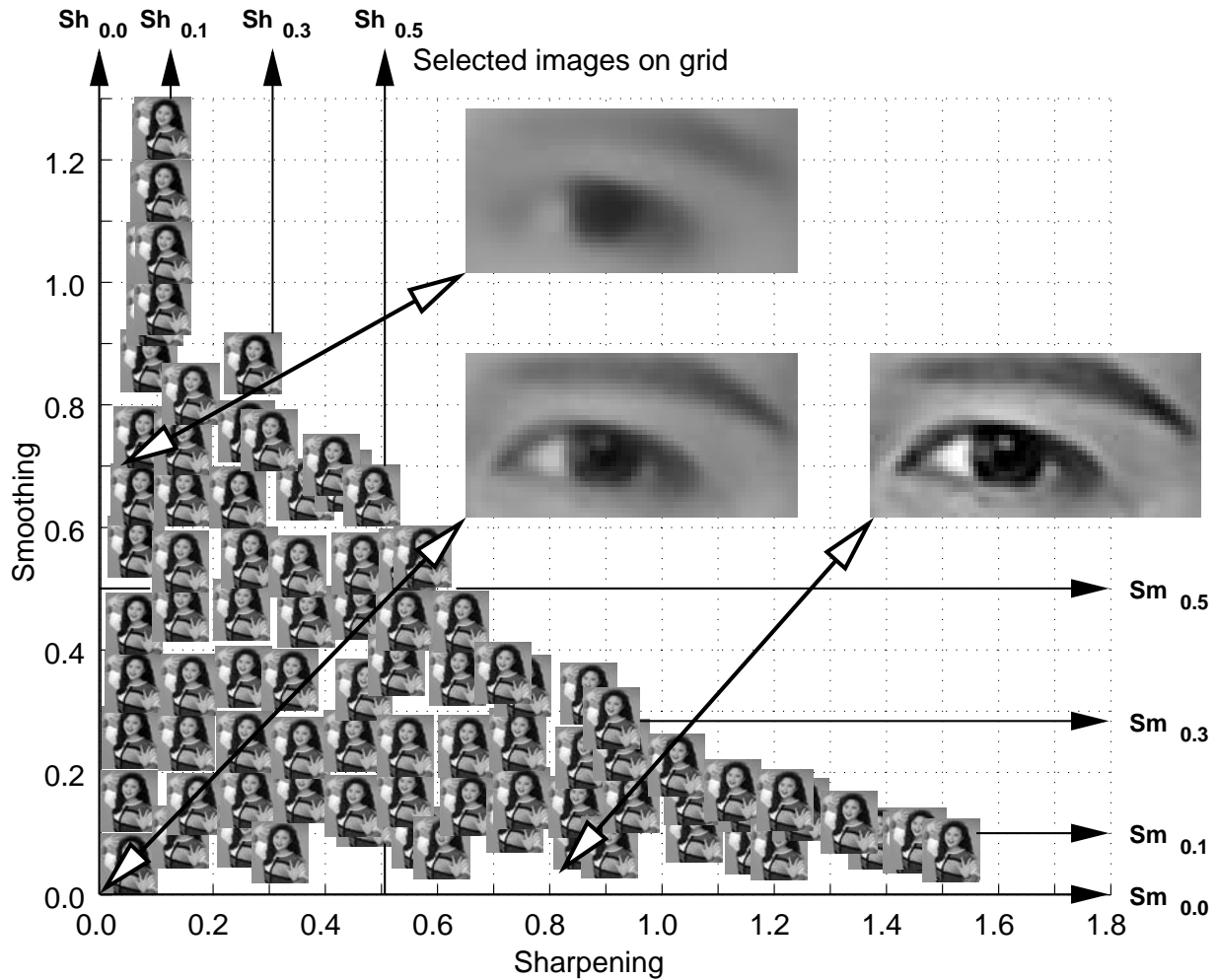
Figure 5.4: Images selected along a grid: along each of the horizontal and vertical lines indicated by the arrows, a range of images was chosen.

x (sharpening y). Sharpening (smoothing) was varied in steps of 0.1. The ranges used are shown in figure 5.4.

Since the various grades in smoothing and sharpening had to be selected from a limited sample (see figure 5.3) the specified values of 0.0 to 0.7 in general could only be approximated. This was done by selecting the nearest value to the desired value. As can be seen in table 5.1, however, the differences were always less than 0.05. Some ranges consisted of less than 8 images, so for some desired grid locations no nearby images could be found. Figure 5.4 shows an example of the resulting tiling.

5.5.3   Experiment B: preference

In the second experiment, subjective preference for a particular sharpening or smoothing value was tested. The subject was given a range of prints, asked to select three prints that she/he considered best and to order these three by quality.

Table 5.1: The grid location and actual sharpening (Sh) and smoothing (Sm) values of the image ranges used in experiment A. In the cases for which no image could be found near the grid location, this is indicated by "-". Ranges are indicated by what is constant (Sh/Sm) and at what value.

Range(s)	Grid		portrait		bicycle	
	Sh	Sm	Sh	Sm	Sh	Sm
$Sh_{0.0}$, $Sm_{0.0}$	0.00	0.00	0.00	0.00	0.00	0.00
$Sh_{0.0}$, $Sm_{0.1}$	0.00	0.10	0.00	0.10	0.00	0.10
$Sh_{0.0}$,	0.00	0.20	0.00	0.21	0.01	0.21
$Sh_{0.0}$, $Sm_{0.3}$	0.00	0.30	0.01	0.30	0.01	0.31
$Sh_{0.0}$,	0.00	0.40	0.01	0.39	0.01	0.38
$Sh_{0.0}$, $Sm_{0.5}$	0.00	0.50	0.01	0.52	0.02	0.51
$Sh_{0.0}$,	0.00	0.60	0.02	0.60	0.01	0.62
$Sh_{0.0}$,	0.00	0.70	0.02	0.70	0.04	0.69
$Sh_{0.1}$, $Sm_{0.0}$	0.10	0.00	0.09	0.04	0.08	0.04
$Sh_{0.1}$, $Sm_{0.1}$	0.10	0.10	0.12	0.10	0.10	0.09
$Sh_{0.1}$,	0.10	0.20	0.10	0.20	0.10	0.20
$Sh_{0.1}$, $Sm_{0.3}$	0.10	0.30	0.11	0.29	0.10	0.30
$Sh_{0.1}$,	0.10	0.40	0.09	0.41	0.10	0.39
$Sh_{0.1}$, $Sm_{0.5}$	0.10	0.50	0.09	0.49	0.09	0.50
$Sh_{0.1}$,	0.10	0.60	0.10	0.60	0.09	0.61
$Sh_{0.1}$,	0.10	0.70	0.10	0.69	0.10	0.69
$Sm_{0.0}$	0.20	0.00	0.21	0.05	0.23	0.04
$Sm_{0.1}$	0.20	0.10	0.21	0.11	0.20	0.10
$Sm_{0.3}$	0.20	0.30	0.20	0.31	0.20	0.30
$Sm_{0.5}$	0.20	0.50	0.21	0.50	0.19	0.50
$Sh_{0.3}$, $Sm_{0.0}$	0.30	0.00	0.27	0.02	0.30	0.04
$Sh_{0.3}$, $Sm_{0.1}$	0.30	0.10	0.31	0.12	0.30	0.10
$Sh_{0.3}$,	0.30	0.20	0.30	0.19	0.30	0.18
$Sh_{0.3}$, $Sm_{0.1}$	0.30	0.30	0.29	0.30	0.29	0.30
$Sh_{0.3}$,	0.30	0.40	0.31	0.40	0.28	0.41
$Sh_{0.3}$, $Sm_{0.1}$	0.30	0.50	0.30	0.49	0.28	0.51
$Sh_{0.3}$,	0.30	0.60	0.31	0.60	0.34	0.59
$Sh_{0.3}$,	0.30	0.70	0.25	0.69	0.33	0.73
$Sm_{0.0}$	0.40	0.00	-	-	0.39	0.05
$Sm_{0.1}$	0.40	0.10	0.40	0.10	0.40	0.09
$Sm_{0.3}$	0.40	0.30	0.42	0.29	0.41	0.31
$Sm_{0.5}$	0.40	0.50	0.41	0.49	0.40	0.50
$Sh_{0.5}$, $Sm_{0.0}$	0.50	0.00	0.52	0.04	0.5	0.04
$Sh_{0.5}$, $Sm_{0.1}$	0.50	0.10	0.49	0.11	0.50	0.10
$Sh_{0.5}$,	0.50	0.20	0.49	0.19	0.49	0.19
$Sh_{0.5}$, $Sm_{0.3}$	0.50	0.30	0.48	0.33	0.50	0.29
$Sh_{0.5}$,	0.50	0.40	0.49	0.40	0.50	0.41
$Sh_{0.5}$, $Sm_{0.5}$	0.50	0.50	0.49	0.40	0.48	0.48
$Sh_{0.5}$,	0.50	0.70	-	-	0.50	0.68
$Sm_{0.0}$	0.60	0.00	0.56	0.03	0.59	0.03
$Sm_{0.1}$	0.60	0.10	0.61	0.10	0.60	0.09
$Sm_{0.3}$	0.60	0.30	0.58	0.31	0.61	0.30
$Sm_{0.5}$	0.60	0.50	0.52	0.50	-	-
$Sm_{0.0}$	0.70	0.00	-	-	0.71	0.03
$Sm_{0.1}$	0.70	0.10	0.70	0.11	0.71	0.10
$Sm_{0.3}$	0.70	0.30	0.70	0.29	0.69	0.27
$Sm_{0.0}$	0.80	0.00	0.81	0.04	-	-
$Sm_{0.0}$	0.90	0.00	0.86	0.03	-	-

All ranges used in experiment A were also used in this experiment. In addition, an extra range was used in which both sharpening and smoothing were varied between 0.0 and 0.2, in steps of 0.1. Except for this last two-dimensional range, the prints were already ordered by either sharpening or smoothing in order to not unnecessarily complicate the preference experiment for the subjects.

5.6   Results of sharpening and smoothing experiments

5.6.1   Experiment A: stimulus response

The results for the 16 ranges used in experiment A are given in figure 5.5 and 5.6. For some ranges it can be seen that subjects are quite capable of ordering the prints, for instance in the `portrait`, $Sm_{0.5}$ range only 3 wrong orderings (orderings that are not consistent with the smoothing parameter) are made. For other ranges the subjects are much less capable of ordering the prints; for instance in the `bicycle`, $Sm_{0.5}$ range (too) many wrong orderings are made.

The correlation between the defined sharpening and smoothing measures on the one hand and the perceived sharpening and smoothing on the other, is measured with the Spearman rank-order correlation coefficient r_s [Siegel and Castellan, 1988]. The H_0 hypothesis tested is that there is no association between the perceived sharpening/smoothing and the sharpening/smoothing measure (see section 4.4 for more details). If H_0 can be rejected, we find the proposed measures to be perceptually relevant for sharpening and smoothing.

For $N_{stimuli} = 8$, with an error of 5%, the critical value above which the two-tailed H_0 hypothesis can be rejected is 0.736 (for $N = 7_{stimuli}$ this value is 0.786; for $N_{stimuli} = 6$, 0.886).

The rank-order coefficients for each subject are given in table 5.2.

It can be seen that for most subjects and ranges the null hypothesis can be rejected. So, the correlation reflects a "true" relationship between perceived and physical rank order.

The ranges for which the null hypothesis cannot be rejected are:

- **Subject 3 and 4, `portrait`, $Sm_{0.1}$**: we suspect that this is simply due to chance.

- **Subject 1, `bicycle`, $Sh_{0.5}$**: since `bicycle` is an artificial mix of many different images, with different scales of detail, we suspected subjects to base their judgements on different parts of the image.

  To verify this, sharpening and smoothing values of parts of the two images (shown in figure 5.2 (c) and (d)) were calculated. The parts were selected manually, in such a way that all parts of the images subjects claimed to have looked at were represented. When correlating the sharpening and smoothing values of the parts with the results for subject 1, it is clear that for the parts `fruit` and `lobster` H_0 can be rejected. After the experiments, the subject stated that he had looked mostly at the (`fruit`), `lobster` and `clock` image parts. The conclusion is that the large variations in the sharpening and
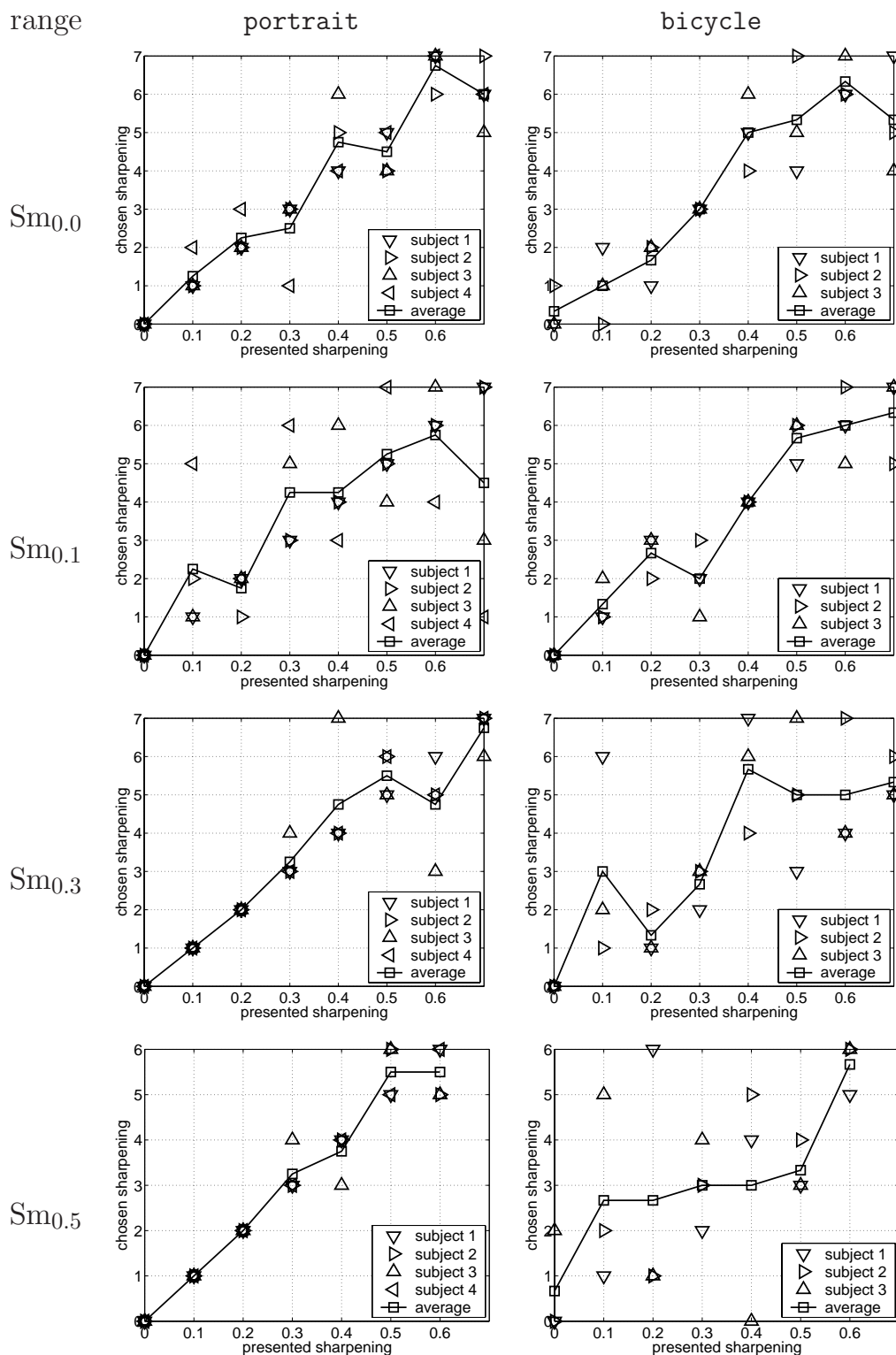
Figure 5.5: The results of the ordering for different smoothing levels (constant smoothing, varying sharpening). It can be seen that, for the $Sm_{0.5}$ range, ordering the images in the same order as given by the sharpening measure is less demanding for the `portrait` (left) than for the `bicycle` (right) image.
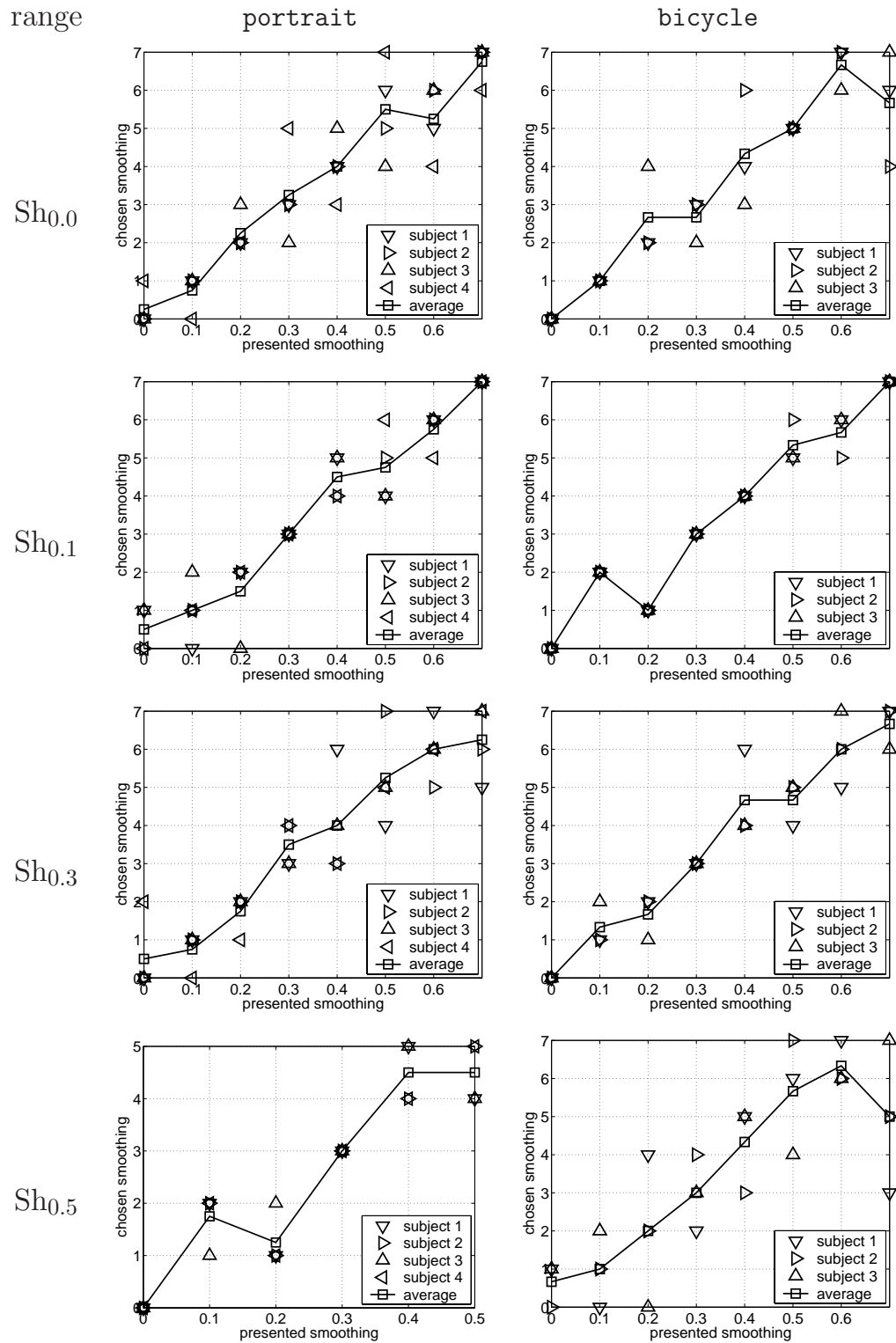
Figure 5.6: The results of the ordering for different sharpening levels (constant sharpening, varying smoothing).

smoothing values of these parts of the images are the reason for the results of subject 1 in this range.

Table 5.2: The Spearman rank-order coefficients for the ranges used in experiment A, per subject. For the values printed in italic, the null hypothesis cannot be rejected; that is, in these cases perceived sharpening and smoothing are uncorrelated with the physical measures. CV stands for critical value. Note that subject 4 did not participate in the `bicycle` experiment.

	Subject						
Range	1	2	3	4	Avg.	$N_{stimuli}$	CV
$Sh_{0.0}$	0.98	1.00	0.95	0.81	0.98	8	0.74
$Sh_{0.1}$	0.96	1.00	0.91	0.98	1.00	8	0.74
$Sh_{0.3}$	0.88	0.91	1.00	0.91	1.00	8	0.74
$Sh_{0.5}$	0.89	0.94	0.94	0.94	0.93	6	0.89
$Sm_{0.0}$	0.98	0.98	0.88	0.91	0.95	8	0.74
$Sm_{0.1}$	1.00	0.98	*0.69*	*0.17*	0.90	8	0.74
$Sm_{0.3}$	1.00	0.98	0.76	0.98	0.96	8	0.74
$Sm_{0.5}$	1.00	0.97	0.92	1.00	0.99	7	0.79

(a) `portrait`

	Subject					
Range	1	2	3	Avg.	$N_{stimuli}$	CV
$Sh_{0.0}$	0.98	0.83	0.93	0.97	8	0.74
$Sh_{0.1}$	0.98	0.95	0.98	0.98	8	0.74
$Sh_{0.3}$	0.93	1.00	0.95	0.99	8	0.74
$Sh_{0.5}$	*0.69*	0.88	0.91	0.93	8	0.74
$Sm_{0.0}$	0.95	0.88	0.83	0.96	8	0.74
$Sm_{0.1}$	0.98	0.93	0.91	0.98	8	0.74
$Sm_{0.3}$	*0.43*	0.98	0.79	0.78	8	0.74
$Sm_{0.5}$	*0.61*	0.93	*0.25*	0.98	7	0.79

(b) `bicycle`

- **Subjects 1 and 3, `bicycle`, $Sm_{0.3}$ and $Sm_{0.5}$:** in highly smoothed images, the effect of the sharpening operation is only preserved for large-scale, high edges. Small details and less prominent edges are smoothed away. While subjects tend to place emphasis on this loss of detail, the proposed sharpening measure is not heavily influenced by it. This would also explain why the `portrait` ranges do not show these deviations: there is far less detail present in this image, all important edges present are on more or less the same scale.

As shown in figure 5.5, subjects had much more difficulty in ordering the `bicycle` image than in performing the same task on the `portrait` image for some ranges. As was discussed above, this is likely due to the diverse content of the former image. A

second general conclusion is that different levels of smoothing seem to be more easily discriminated than levels of sharpening, because less ordering errors are made.

5.6.2   Experiment B: preference

In this second experiment, subjects were asked to give a first, second and third preference per range. These preferences were averaged with a certain weight: the first preference had a weight of 4, the second of 2 and the third of 1. The results are shown in figure 5.7.

The two left images (figures 5.7 (a) and (c)) show that, for smoothed images, subjects tend to prefer high sharpening to compensate for the smoothing away of the edges. The effect of smoothing away of the smaller edges can be seen in figure 5.8 and figure 5.9. In these figures the effect of different sharpening and smoothing values on small parts of the `bicycle` image is shown. The original subpart is shown in the bottom left of the figure. Both subparts contain lines and edges of different sizes. It can be seen that for high smoothing, the texture of the pineapple is smoothed. The dark dots and the contour of the pineapple are still sharp for a high sharpening. The same effect can be seen in the `plant` subpart. The leaves are sharp for all sharpening and smoothing values, but the texture on the leaves is not visible anymore for higher smoothing values. Clearly, edges play an important role in subject appreciation of an image. For `portrait`, however, the leftmost value, indicating a preference for high sharpening at low smoothing, is hard to explain.

The ranges in which subjects were asked for smoothing preference, figures 5.7 (b) and (d), show that subjects prefer sharpening rather than smoothing. For highly sharpened images, some smoothing is preferred to reduce the artifacts introduced by the sharpening operation.

These conclusions are corroborated by the results for the two-dimensional range, given in table 5.3.

Subjects seem to prefer a little bit of smoothing and quite a bit of sharpening. Apparently, the original is not the best image.

5.7   Conclusions

The two new measures for the amount of sharpening and smoothing introduced here, seem to correlate reasonably well with human perception.

Problems arise for images in which parts of the image require different sharpening and/or smoothing values, as is the case for the complex `bicycle` image. Subjects tend to look at different parts and combine their judgement into an overall decision. One possible future approach is to find these different image parts (possibly by using the measures themselves) and treat them separately.
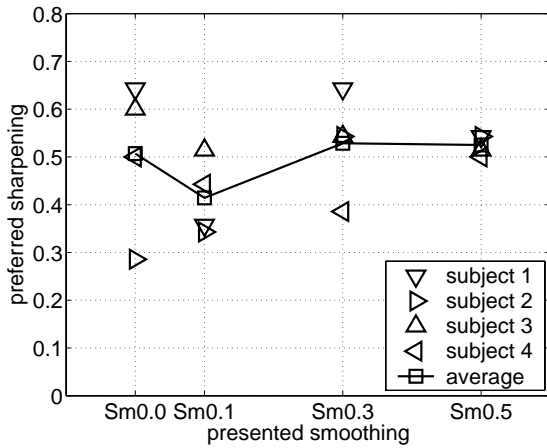
(a) `portrait` sharpening preference
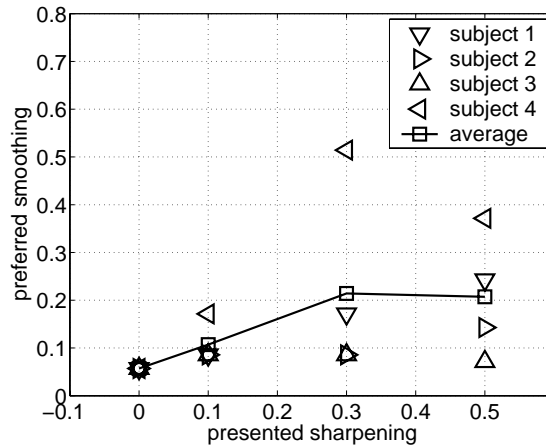
(b) `portrait` smoothing preference
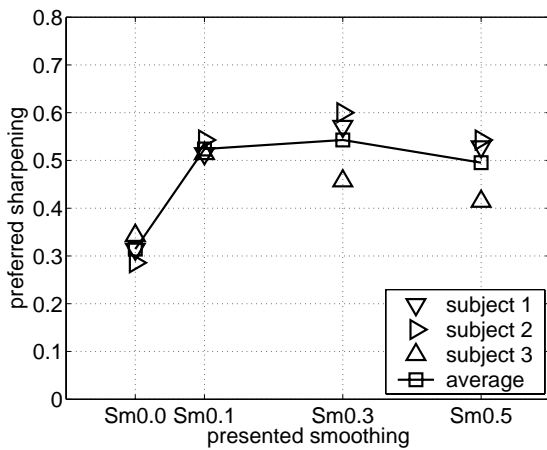
(c) `bicycle` sharpening preference
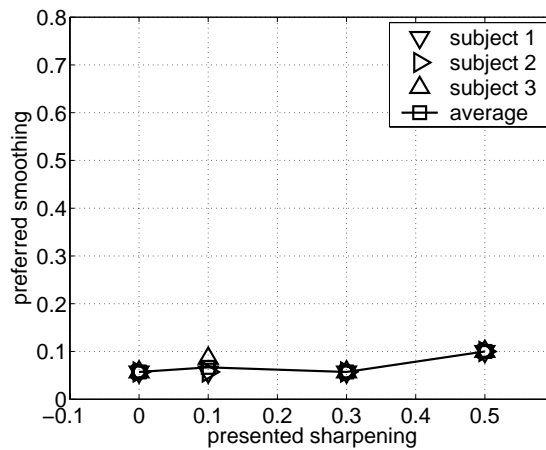
(d) `bicycle` smoothing preference

Figure 5.7: Results of the preference experiment B. For this set of images it appears that subjects prefer quite a bit of sharpening.

Our experiments have also showed that subjects tend to have much less problems in discerning various levels of smoothing than they have with levels of sharpening. This indicates that the two measures proposed are not equivalently spaced: the just noticeable difference of the smoothing measure is smaller than that of the sharpening measure.

The results of the preference experiment look promising. One can say that subjects prefer images in which the smoothing is low and the sharpening is high. However, this may not be the case for noisy images, in which a certain amount of smoothing will likely be appreciated.
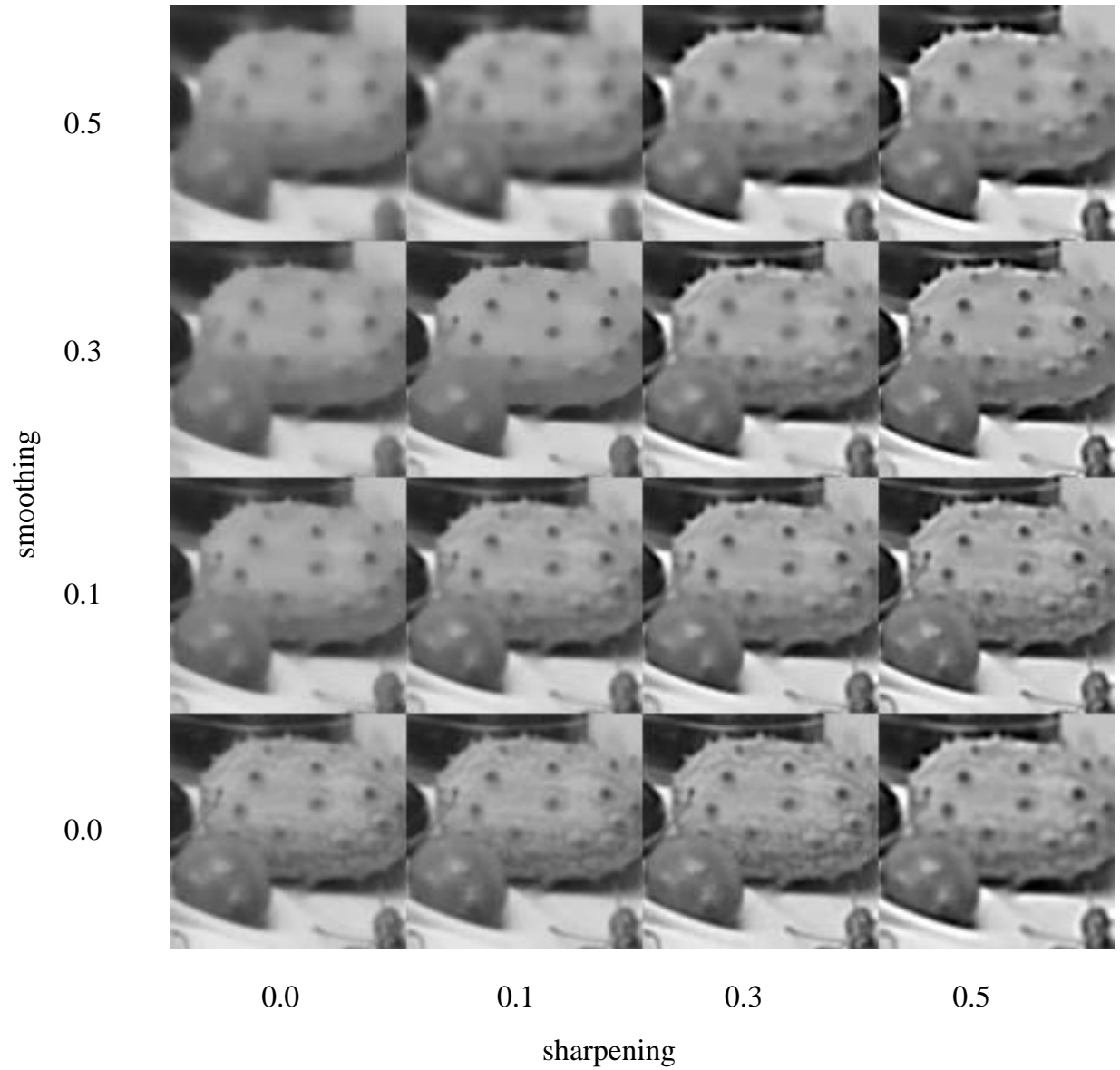
Figure 5.8: The effect of different sharpening and smoothing values for `pineapple`. This image is a part of the `bicycle` image.
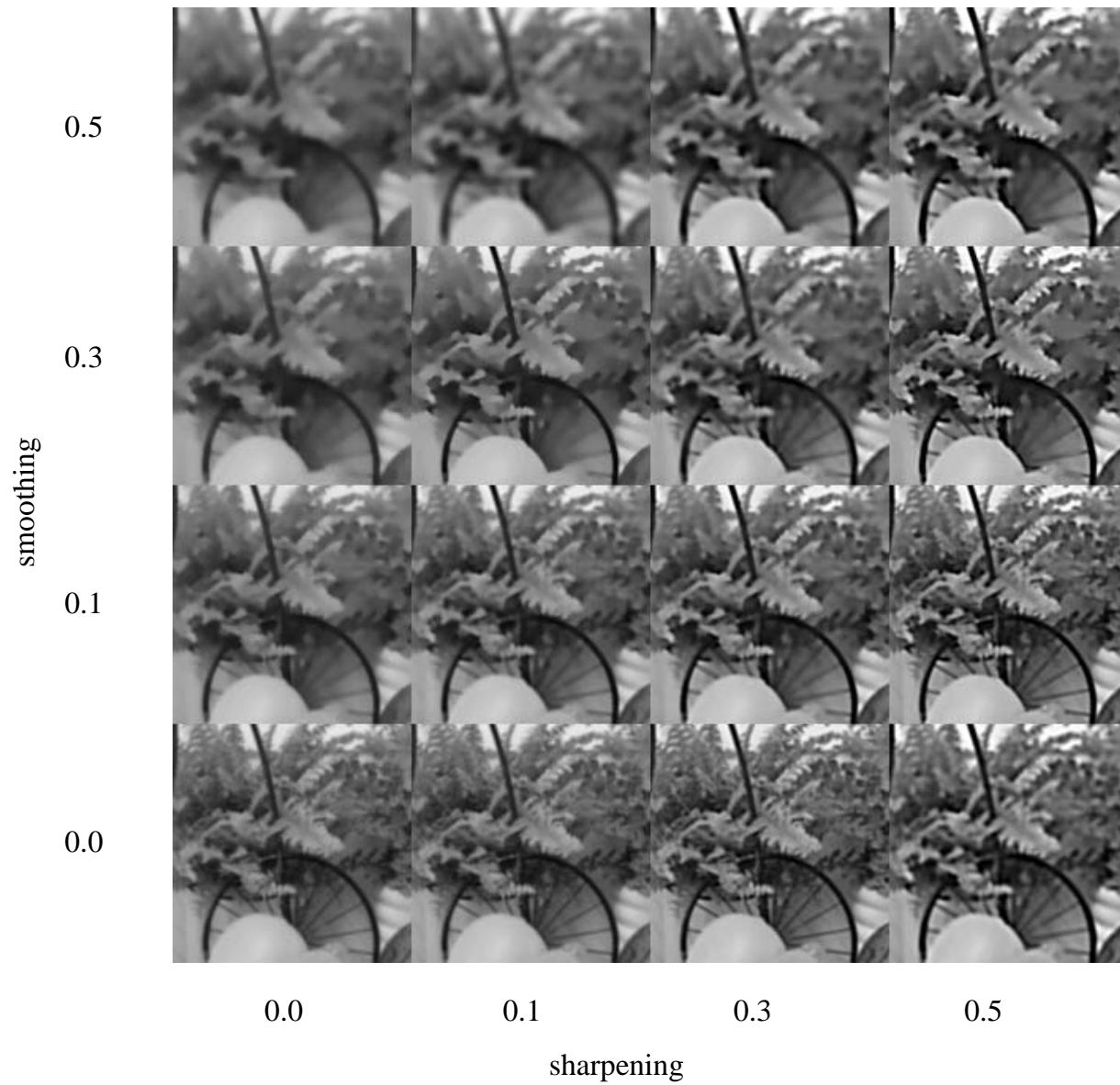
Figure 5.9: The effect of different sharpening and smoothing values for `plant`. This image is a part of the `bicycle` image.

Table 5.3: Results of sharpening (Sh) and smoothing (Sm) of the preference experiment B for the two- dimensional range. Note that subject 4 did not participate in the `bicycle` experiment.

Pref.	Subject								Avg.	
	1		2		3		4			
	Sh	Sm	Sh	Sm	Sh	Sm	Sh	Sm	Sh	Sm
1st	0.2	0.1	0.2	0.0	0.2	0.0	0.2	0.2	0.20	0.08
2nd	0.2	0.0	0.2	0.1	0.2	0.1	0.2	0.0	0.20	0.05
3rd	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.10	0.13

(a) `portrait`

Pref.	Subject						Avg.	
	1		2		3			
	Sh	Sm	Sh	Sm	Sh	Sm	Sh	Sm
1st	0.1	0.0	0.2	0.0	0.2	0.0	0.17	0.00
2nd	0.2	0.0	0.1	0.0	0.1	0.0	0.13	0.00
3rd	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00

(b) `bicycle`

Chapter 6

An objective measure for absolute sharpness

In the previous chapter we have discussed the effect of sharpening of an image. This measure is relative to the original image. But humans can tell whether one single image is sharp or not, indicating that sharpness is a property of one single image.

In this chapter, we will propose and discuss a new measure for the sharpness of natural (complex) images.

In the sharpening experiment it was already shown that images can be seen as a collection of areas that are more or less uniform, separated by lines and edges. We assume that perceptual sharpness is correlated to the sharpness of (some of) these lines and edges. In between these are textures.

To determine the sharpness we need to determine the location and orientation of the lines and edges (together called transients) so that we can perform measurements on their profile. To describe the sharpness of a profile we need a model. We use a Gaussian profile for the line and an integrated Gaussian (error function) for the edge, which can be characterized by its width (σ).

We computed Gaussian derivatives at several scales to obtain a response function or signature. At a transient, the response function can be predicted given the width and the amplitude of the line or edge. Conversely, we can estimate the width and amplitude of the line or edge from the measured response function.

The measured widths of all points have to be combined to provide one or more measures of sharpness. An obvious measure is the median width in the image. Because some estimated widths can be very large, we do not use the mean width. Another measure that can be used is the fraction of pixels for which the width is smaller than a certain value. This gives some insight into the number of sharp lines and edges in the image.
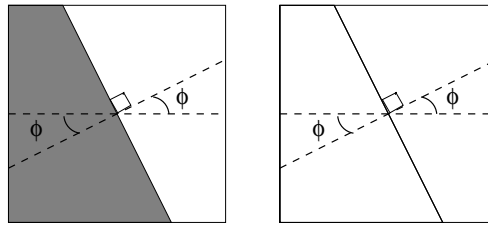
Figure 6.1: The definition of orientation ϕ for an edge (left) and a line (right)

Kayargadde [Kayargadde, 1995] proposed a similar measure for the perceptual sharpness of images. He used a polynomial transform (Hermite transform), to detect and estimate edge parameters, such as position, orientation, amplitude, mean value and width. Our methods is different from his on three points. The first is that we detect lines and edges and determine the width of both of them. The second difference is that we perform a numerical estimation of the amplitude and width, whereas Kayargadde derived an analytical relationship. And the last difference is that we estimate the orientation of the structures in the localization phase, where Kayargadde determines the orientation in the estimation phase.

6.1   Line and edge detection

Before we can determine the sharpness of individual transients, we must extract them from the image. For each transient we must also establish whether it is a line or an edge, and its orientation ϕ, as defined in figure 6.1. Since we must deal with both lines and edges, it is logical to use a filter bank based on quadrature filters (e.g. [Knutsson and Granlund, 1983]). A quadrature filter is a linear, complex-valued filter. The real and imaginary part act as line and edge filters, respectively. The magnitude of the response is phase-invariant, i.e. insensitive to whether the transient is a line or an edge. This is important when we discuss the suppression of spurious responses below.

The quadrature filter is sensitive to edges and lines under a limited range of orientations. To obtain the response under an arbitrary angle we use a steerable quadrature filter [Freeman and Adelson, 1991]: the response can be computed from the filter response under a finite set of angles. The details of the quadrature filter we use can be found in [van Ginkel, 2002]. The filter's characteristic frequency (f_c), the range of frequencies it is sensitive to (b_f), and the orientation selectivity (s_a) can be independently tuned. Our supposition is that perceptual sharpness relates to the sharpest line or edge in the image. This supposition only holds for natural images. We have tuned the filter in such a way that it will detect small-scale lines and edges $(f_c, b_f) = (0.16, 0.16)$. The orientation selectivity $s_a = 0.185$ [van Ginkel, 2002] was chosen as a trade-off between orientation selectivity, signal-to-noise ratio and

(a) a line

(b) an edge

(c) responses to the line

(d) responses to the edge
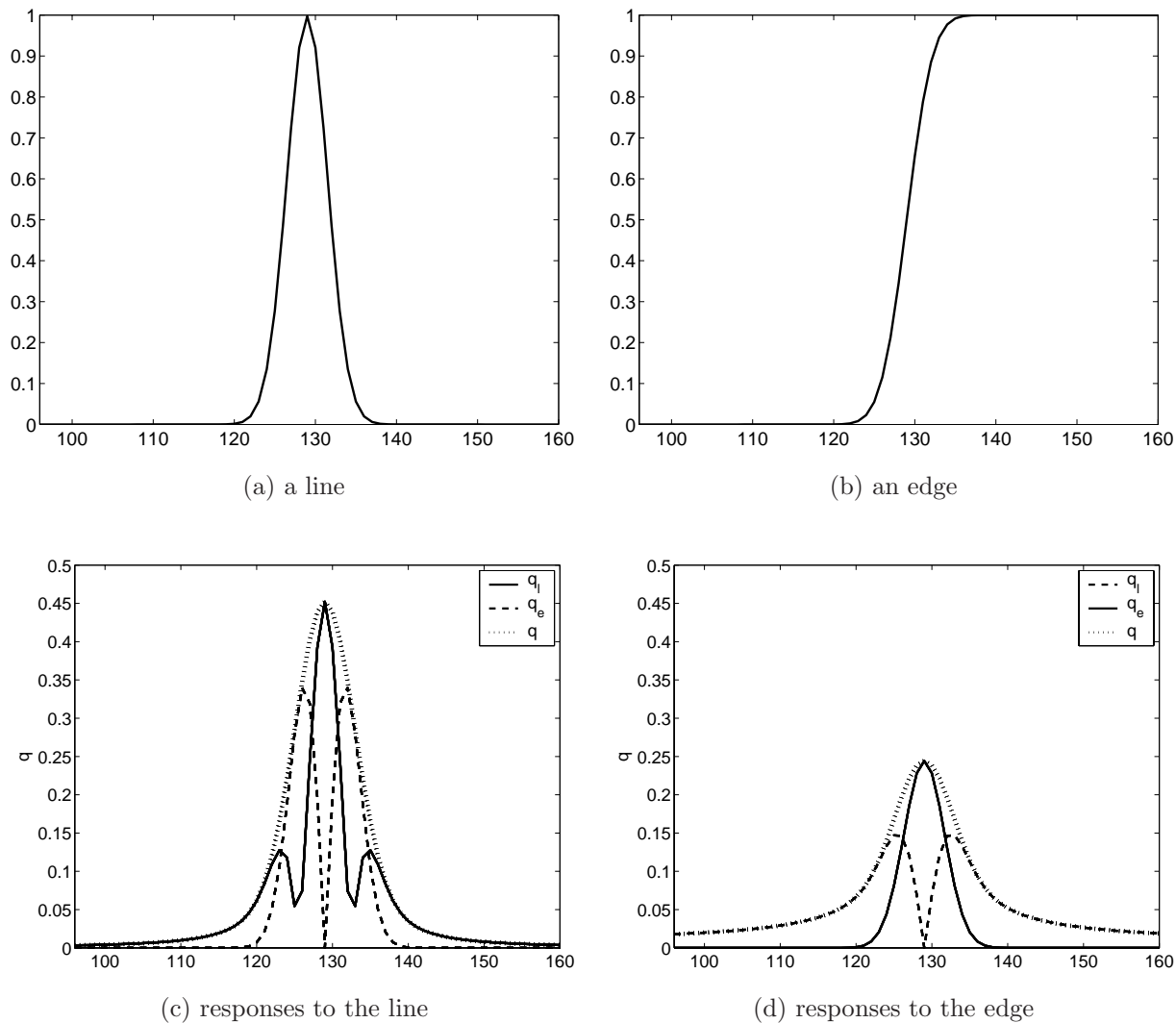
Figure 6.2: The total quadrature filter response q, the quadrature filter response for the line component q_l and the quadrature filter response for the edge component q_e. The input is a line (left) and an edge (right).

localization of the filter response. To properly span the entire range of orientations, given this selectivity, 17 filters are required. Angular differences, with respect to a FWHM criterion, up to $3.33s_a$ can be distinguished, which in this case corresponds to about 35°.

The last step is to determine whether a detected transient is an edge or a line, and to suppress spurious responses. To see why these occur, imagine a line. The line detector responds as it should, but the edge detector will also respond. It responds to the flanks of the line, although less strongly than the line detector. This is illustrated in figure 6.2. We resolve this problem by suppressing (inhibiting) the

secondary responses. The final line response l_{line} is given by

$$l_{\text{line}}(x, y) = \begin{cases} 0 & \text{if } q_l(x, y) < \max_{N(x,y)} q_e(x, y) \\ q_l(x, y) & \text{elsewhere} \end{cases}, \tag{6.1}$$

where q_l and q_e are the quadrature filter result for the line and edge component, respectively. The total quadrature filter result, q is given by $q = \sqrt{q_l^2 + q_e^2}$. $N(x, y)$ is a neighbourhood around (x, y). The size of the neighbourhood must be roughly equal to the width of the response lobes. By swapping the roles of q_l and q_e the same technique can be used to obtain the final edge response. Points for which the quadrature filter is larger in a different orientation are discarded. The result of the different line and edge detection steps for the input image trui are given in figure 6.3.

6.2 Line and edge characterization

In this section we explain how we can determine the amplitude and width of Gaussian lines in images. We assume that the noise level is low: there is no point in measuring a subtle feature like sharpness if the image is heavily distorted by noise. The method we use is a variant of Mallat's approach [Mallat and Hwang, 1992], using Gaussian derivatives rather than wavelets [van Asselt, 1997]. The idea is to compute the response of the Gaussian derivative operator, applied across the transient, while varying the scale of the Gaussian. The response depends on both the scale of the Gaussian and that of the transient. Since we know the former, we may estimate the latter.

It is convenient to adapt a local coordinate system (v, w) at each point (x_0, y_0) that is aligned with the orientation ϕ at that point:

$$\begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}. \tag{6.2}$$

The singularity functions of an infinitely long line and edge are given by

$$\begin{aligned} h_{\text{line}} &= A\,\delta(v) = A\,\delta\big((x - x_0)\cos(\phi) + (y - y_0)\sin(\phi)\big) \\ h_{\text{edge}} &= A\,u(v) = A\,u\big((x - x_0)\cos(\phi) + (y - y_0)\sin(\phi)\big) \end{aligned} \tag{6.3}$$

where A is the amplitude of the transient, ϕ the angle of the transient with respect to the x-axis, $\delta(x)$ the Dirac delta function and $u(x)$ the Heaviside step function. Real transients have a finite width. We model this by convolving these functions with a Gaussian with $\sigma_{l/e}$. The width $\sigma_{1/e}$ reflects the sharpness of the transient.

To find an estimate for the width $\sigma_{l/e}$ and amplitude A of the lines and edges, we construct a response function in the following way: we convolve the input

(a) the input image

(b) quadrature

(c) uncorrected line response

(d) uncorrected edge response

(e) lines and edges

(f) found lines

(g) found edges
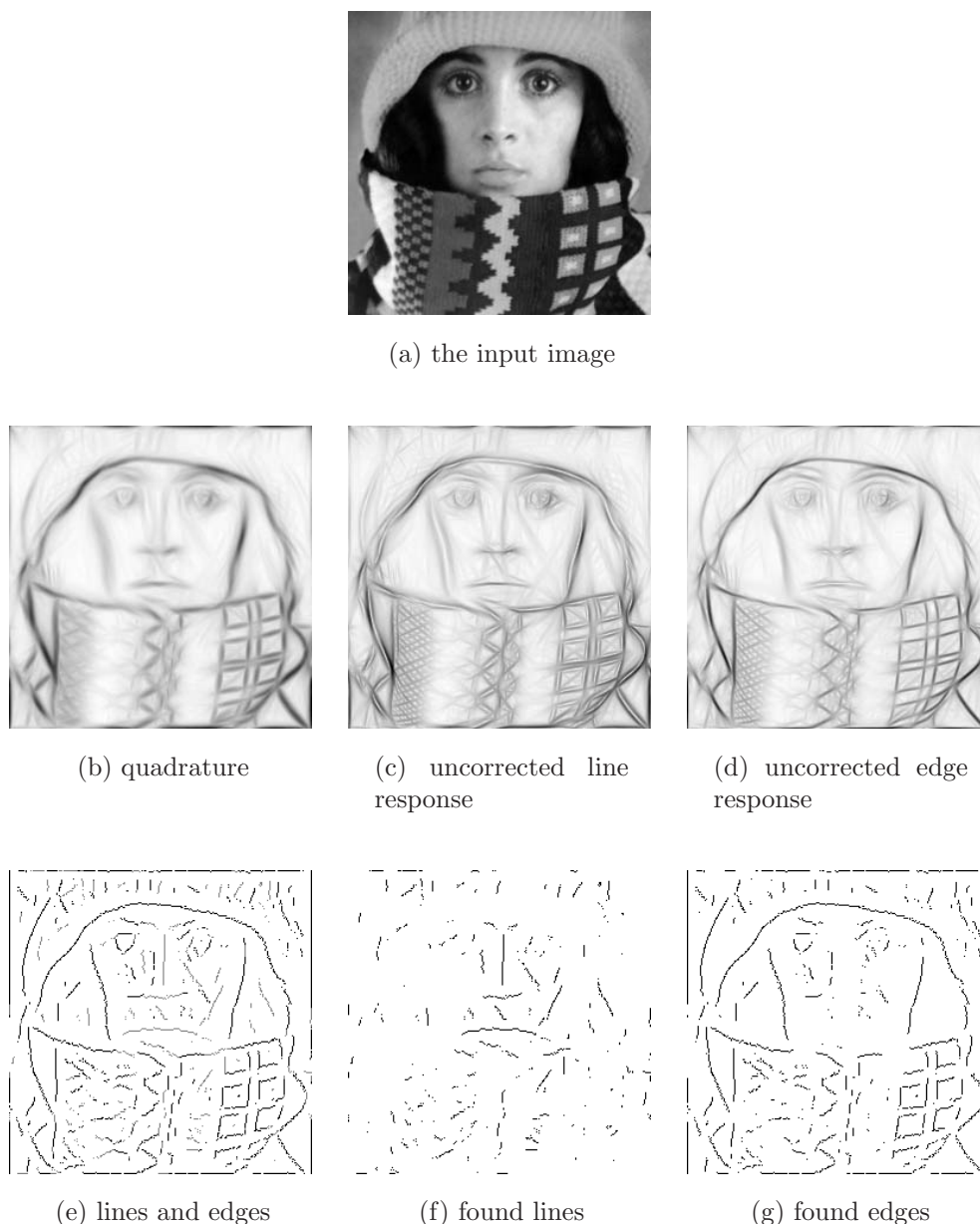
Figure 6.3: The line and edge detection steps. In (a) the original image is given, in (b) the response of the quadrature filter, in (c) and (d) the responses of the line and edge component filters, respectively. In (e) the final line and edge responses are given. Line responses are visualized in grey, edge responses in black. In (f) and (g) the final line and edge responses are given in separate images.

image with directional Gaussian derivatives along ϕ, sampling the scale axis exponentially: $\sigma = b^i$ ($b > 1$) with integer i. The Gaussian regularization has, in general, the effect that the response decreases as a function of scale, as noted by Lindeberg [Lindeberg, 1993]. We follow [Lindeberg, 1993] in using normalized, or scale-independent, Gaussian derivatives. This results in more pronounced response curves. The normalization consists of multiplying the response with σ.

The response curve at (x_0, y_0) is described by

$$r(\sigma) = \sigma \frac{\partial g_\sigma(v, w)}{\partial v} * \left(h(v, w) * g_{\sigma_{l/e}}(v, w)\right). \tag{6.4}$$

Using the commutativity of the convolution operator we obtain

$$r(\sigma) = \sigma \frac{\partial}{\partial v}\left(h(v, w) * g_s(v, w)\right) \tag{6.5}$$

with $s = \sqrt{\sigma_{l/e}^2 + \sigma^2}$. In what follows we consider the modulus of the response $M(\sigma) = |r(\sigma)|$. The expression for M for a line and edge respectively is given by the following two equations:

$$
\begin{aligned}
M_{\text{line}}(\sigma) &= |A\sigma \frac{\partial}{\partial v} g_s(v)| = \frac{|A|\sigma}{\sqrt{2\pi s^3}}|v|\exp(-\frac{v^2}{2s^2}) \\
M_{\text{edge}}(\sigma) &= \frac{|A|\sigma}{\sqrt{2\pi s}}\exp(-\frac{v^2}{2s^2})
\end{aligned}
\tag{6.6}
$$

The modulus maxima per scale are given by

$$
\begin{aligned}
\max M_{\text{line}}(\sigma) &= \frac{|A|\sigma}{\sqrt{2\pi e s^2}} \qquad \text{at } v = s \\
\max M_{\text{edge}}(\sigma) &= \frac{|A|\sigma}{\sqrt{2\pi s}} \qquad \text{at } v = 0.
\end{aligned}
\tag{6.7}
$$

The modulus maxima for lines and edges with different widths are given in figure 6.4. These responses are the theoretical responses of lines and edges. We find the widths of the lines and edges in the image by fitting the measured responses to the theoretical responses.

The selection of the modulus maximum for an edge is straightforward: the position of the maximum is the same as the position of the point itself. For the lines this is different, the maxima are shifted over s. To find these maxima we search for a maximum in an appropriately sized neighbourhood.

The minimalization can be done by some numerical minimization method; the Levenberg-Marquardt method. We start with the maximum for $\sigma = b^0$ and use 8 scales. b is chosen $2^{1/3}$, i.e. three samples per octave.

6.3   Sharpness and sampling matters

Natural images as used in the experiment are recorded by an optical system. The best optical system is a diffraction-limited optical system. In such a system the
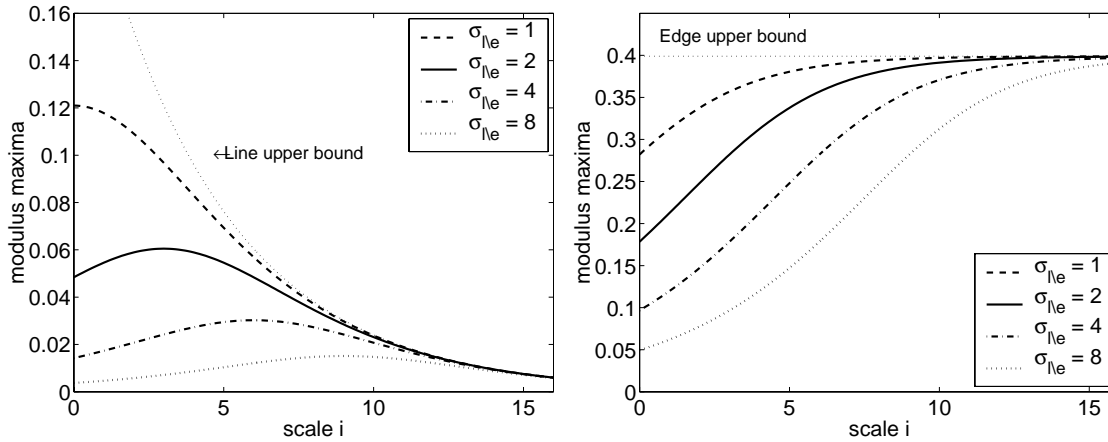
Figure 6.4: The modulus maxima of the response for a line (a) and an edge (b) for different scales i and different values of the width of the transient $\sigma_{l/e}$. The amplitude of the input transient is 1. The width of the Gaussian derivative is $2^{i/3}$ ($b = 2^{1/3}$).

amplitudes are low-pass filtered in the Fourier domain (pill box). Because intensities are measured instead of amplitudes, the optical transfer function (OTF) is a pill box convolved with a pill box, or, in formula

$$\mathcal{G}(f) = u(f_c^2 - f^2) * u(f_c^2 - f^2) \tag{6.8}$$

where f_c is the cut-off frequency. This formula can be rewritten as

$$\mathcal{G}(f) = \frac{2}{\pi} \left[\arccos\left(\frac{f}{f_c}\right) - \frac{f}{f_c}\sqrt{1 - \left(\frac{f}{f_c}\right)^2} \right] \tag{6.9}$$

An image is properly sampled if the Nyquist criterion is met, that is the sample frequency is twice the smallest frequency found in the image. A good approximation for the OTF, if the image is sampled according to the Nyquist criterion, is the Fourier transform of a Gaussian with width 0.9 [van Vliet, 1993]. In our sharpness estimation, ideal lines and edges recorded with such a system should have a width of 0.9. If the lines are blurred, the widths found should be even larger. However, in all tested original (not filtered) images we found widths smaller than 0.9. These smaller widths are found for real lines and edges, as shown in figure 6.5.

We tested two different hypotheses for these smaller widths:

- The smaller widths are found because approximation of the Gaussian for the OTF is not correct, especially for high frequencies.

- The smaller widths are found because the images are not properly sampled.

The first possible explanation is tested by making OTF-limited lines and edges, instead of Gaussian lines and edges. The width and amplitude of these OTF-limited lines and edges are estimated with the Gaussian derivative method also
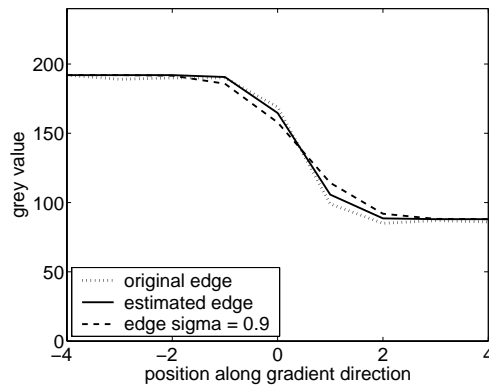
Figure 6.5: Example of an edge with width smaller than 0.9. The original edge is taken from `trui`. The width of the fitted edge is 0.63. It can be seen that this edge is a better fit to the original edge than the edge with width 0.9.
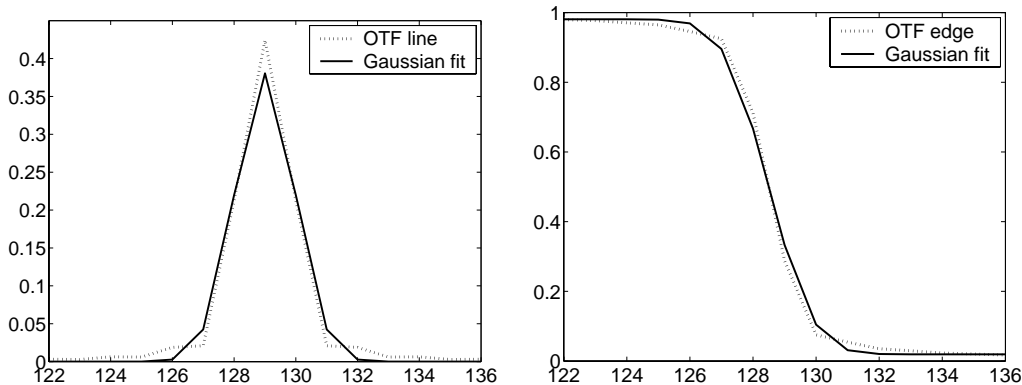
Figure 6.6: the fitted results for a line (left) and an edge (right) for OTF limited lines and edges. The input line and edge are a delta and step function that are convolved with the PSF of an optical system. The fitted lines are Gaussian fits, for the line the amplitude and width are 0.91 and 0.95, respectively, for the edge the amplitude and width are 0.97, 1.01, respectively.

used before. The results are given in figure 6.6. The found width for the line and edge are 1.11 and 0.95, respectively. From this simulation we can conclude that the fact that the Gaussian is not the exact OTF is not the reason that these smaller widths are measured.

The second possible explanation is that the image is not properly sampled according to Nyquist. This is tested by estimating the amplitude and width of undersampled lines and edges. These lines and edges are constructed by resampling a correctly sampled line or edge with width and amplitude 1. The undersampling is done with bilinear interpolation. The results are shown in figure 6.7. It can be seen that widths smaller than 0.9 are possible with undersampling. Note the different behaviour of the amplitude for the line and the edge. From these images we can conclude that the smaller widths can occur due to undersampling.
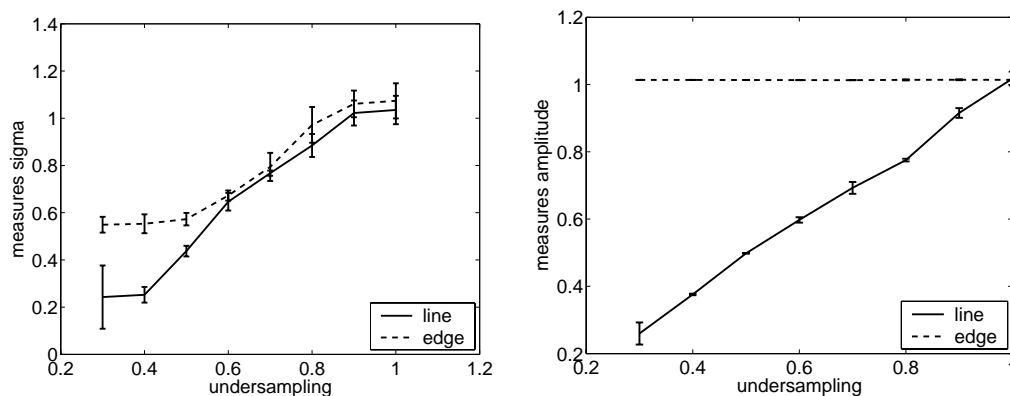
Figure 6.7: Estimation of the widths (left) and amplitudes (right) of undersampled lines and edges. The input line and edge have an amplitude and width of 1. It can be seen that for undersampled images, the estimated widths can be smaller than 0.9.
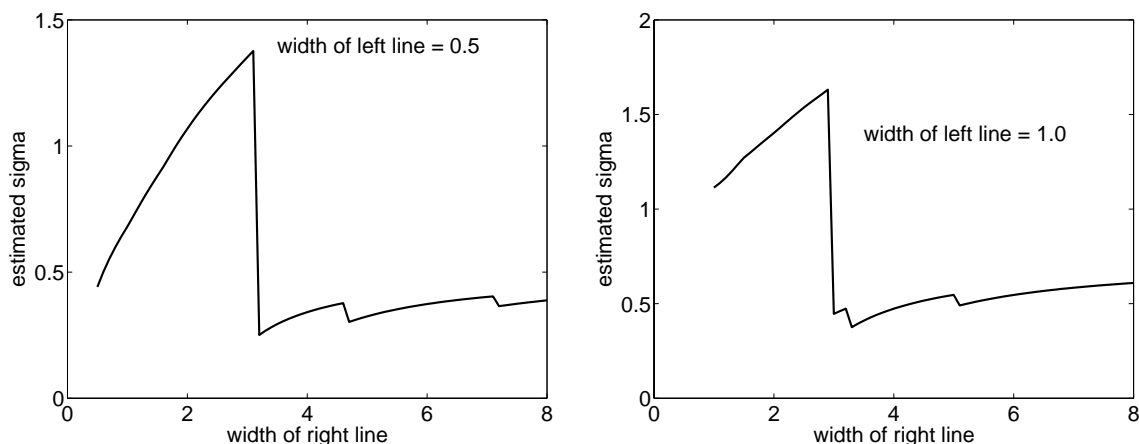
Figure 6.8: The results for estimating the widths of asymmetric lines. The left part of the line is a normal line width a fixed width. This width is 0.5 for the left image and 1.0 for the right image. The right part of the line is a normal line with an increasing width.

6.4   Tests on asymmetric lines

We tested how the sharpness measures react on asymmetric lines, that are lines which have a certain width on one side and another width on the other. The test images are constructed from two normal lines. The estimated widths are shown in figure 6.8. It is shown that the estimated width increases, as could be expected, and then suddenly falls back to about 0.5. This behaviour can be explained by examining the results somewhat closer. For lines which have a small width on one side and a large width on the other, the transient that is found in the localization procedure is not a line but an edge. In figure 6.9 some examples are shown for fitted lines and edges. It can be seen that the fit is rather good for lines that are quite different from the model used. Transient type switches from a line to an edge if the line response is smaller than the edge response. However, this evaluation is
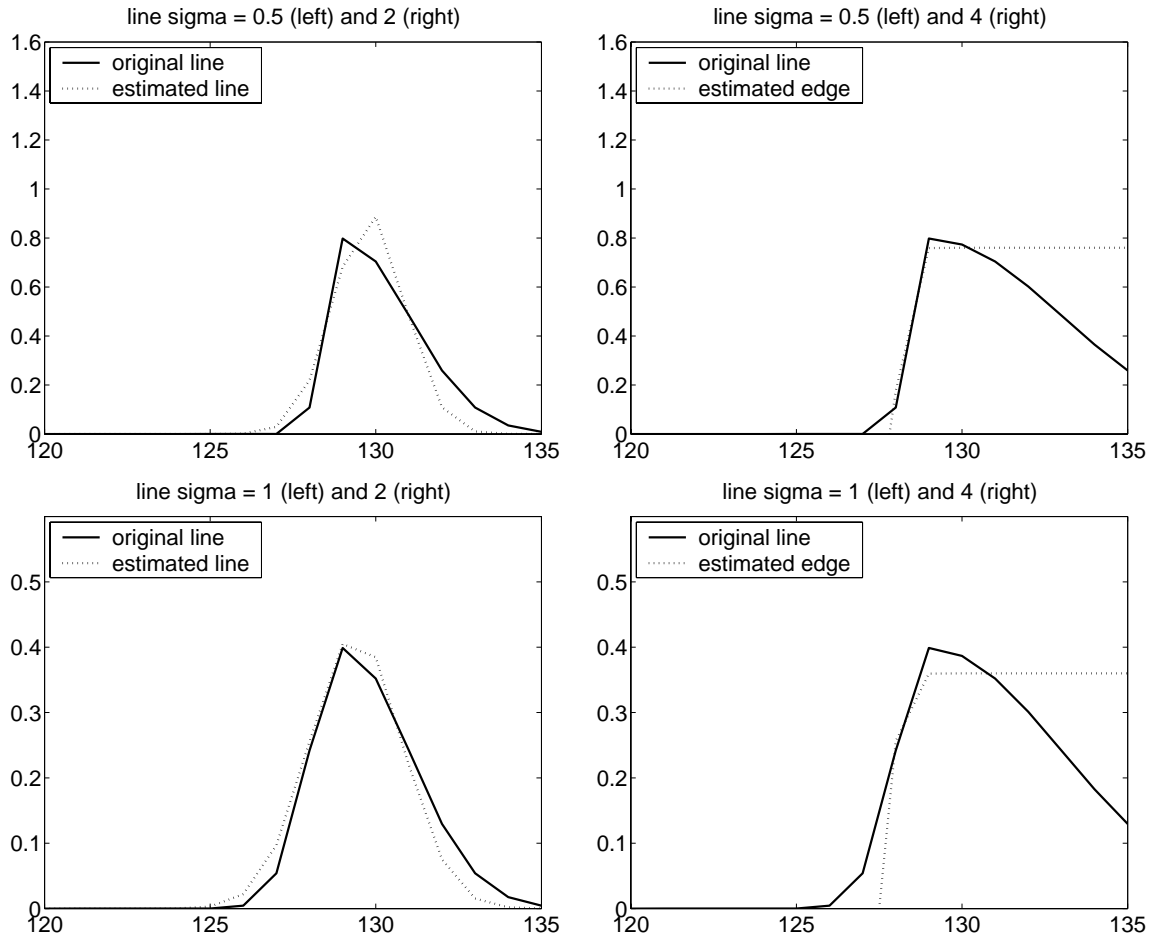
Figure 6.9: Examples of fitted asymmetric lines.

done at the pixel position with the largest total response. For the small line, the total response is largest for the transient position. This is shown in figure 6.10 (b). For widths larger than approximately three, the edge response at the transient position is larger than the line response, so the transient type becomes an edge. The response found is shown in figure 6.10 (a).

For the large line, the position of the largest response shifts from the gradient position to its neighbor, shown in figure 6.10 (d). At this neighboring position, the edge response is larger than the line response (shown in 6.10 (c)), which causes that the transient type to be interpreted is an edge.

6.5   Sharpness measures

The final step in defining a sharpness criterion is to obtain one or a few sharpness measures using the widths. We define and evaluate two sharpness measures. The first measure we looked into is the fraction of transient points with a width smaller
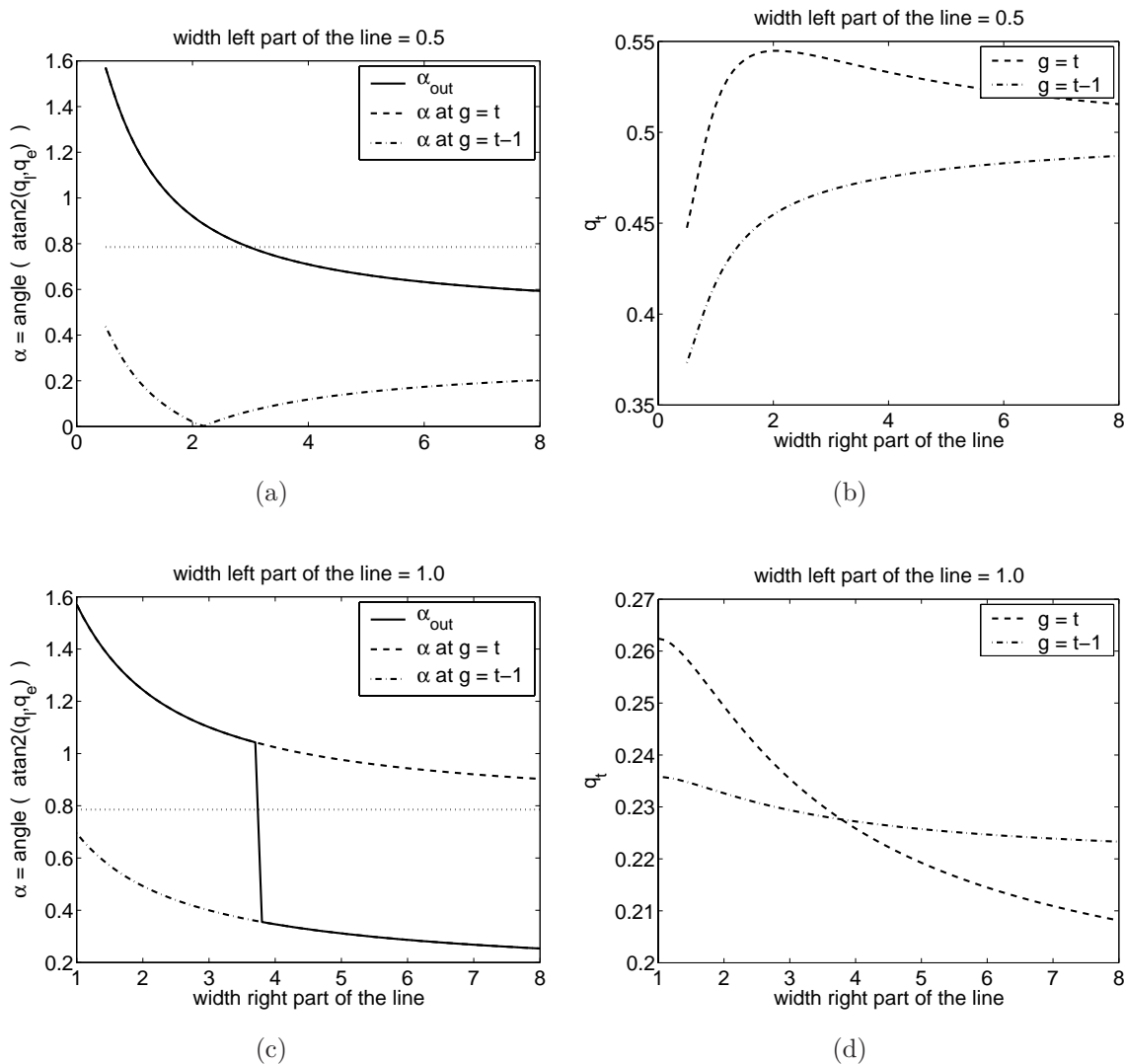
Figure 6.10: The quantities from which the transient type is determined. The line is localized at gradient position t. In (a) and (c) the arctan of the edge response and the line response is shown. If this value is larger than $\pi/4$, the transient type is an line. If it is smaller, the transient type is an edge. In (b) and (d) the total responses for the line position t and the neighboring position $t-1$ are shown. For the small line (b) the response for the line position is always larger. For the large line (d) the response for t is smaller than the response for $t-1$ for large widths.

than 2. This is a very intuitive measure, just the number of points with a small width. We call this measure A. The second measure is the median of the widths. Ideally, we would want to use the smallest width, or a small percentile. However, due to sharpening this value is often close to zero, so that there cannot be any difference between sharpened images. Therefore, we use median (the 50th percentile) of the width as an estimate for the average width. We call this measure B. The mean is not used because this measure is sensitive to highly overestimated widths.

(a) `trui`   (b) `erika`   (c) `cameron`   (d) `lenna`
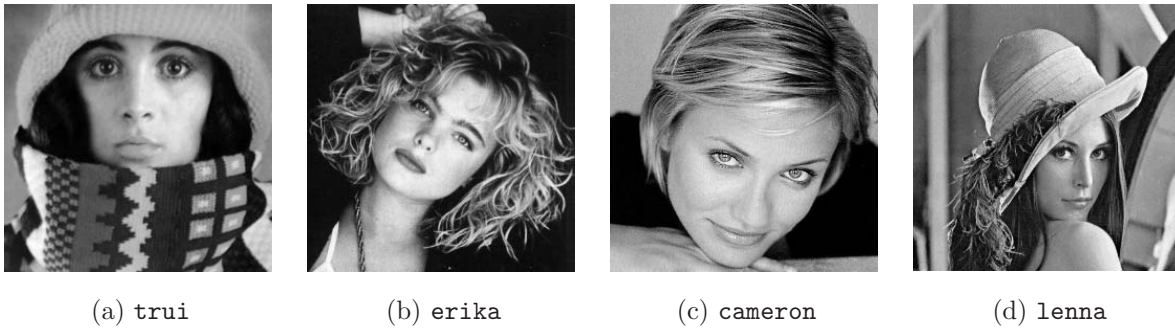
Figure 6.11: The test images.

(a)   (b)
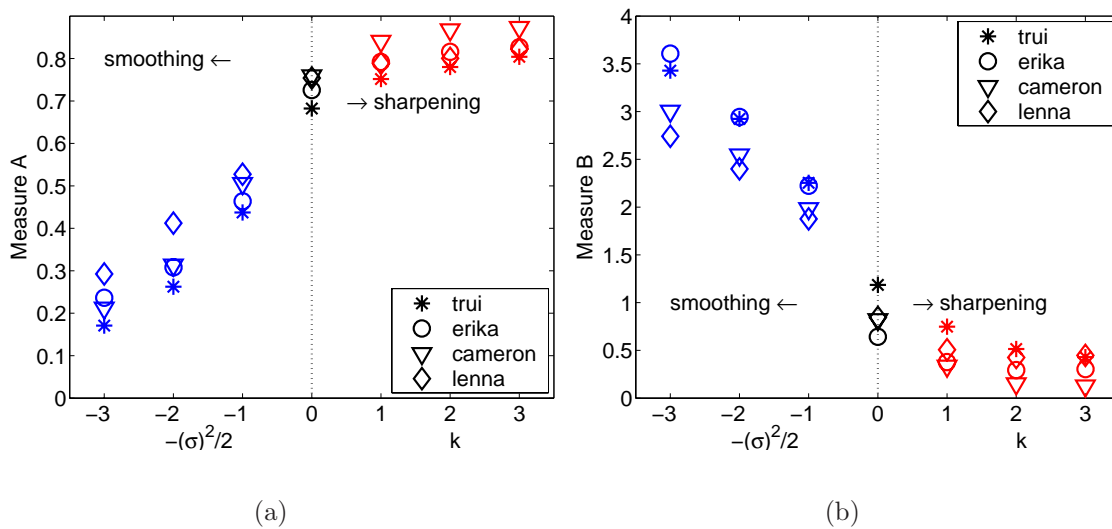
Figure 6.12: Computed sharpness measures, respectively based on measure A: the fraction of transient points with a width smaller than 2 (a) and measure B: the median of the width (b). In (a) the values are positively correlated with the sharpness of the image, in (b) the values are negatively correlated with the sharpness of the image.

6.6 Tests of the sharpness measures

We tested the two proposed sharpness measures for four different test images. These images are given in figure 6.11. The images were manipulated in two ways. The first manipulation is a Gaussian blurring with $\sigma^2/2 = \{0.5, 1.0, 1.5...3.0\}$. The second is unsharp masking with $\sigma = 1.0$ and k is $\{0.5, 1.0, 1.5...3.0\}$. These filters are discussed in section 5.4 on page 89. In the results we plot these two manipulations in one plot by putting the original in the middle (denoted by the dotted line), with the blurring to the left and the sharpening to the right. The spacing to the left is $-\sigma^2/2$, the spacing to the right is k.

The results for the two sharpness measures are given in figure 6.12. We expect

the result of measure A (the fraction of pixels with a width smaller than two) to decrease for a larger blurring, and the result of measure B (the median) to increase for a larger blurring. It can be seen that this is indeed the case. This means that both measures can be used as a measure for the difference in sharpness between images. However, the question is whether this also applies to the *perceived* sharpness of the images. We investigated this by performing a small pairwise-comparison experiment, comparing the sharpness of the original images with respect to each other. The images were printed on plain paper. The size of the images was 6 cm by 6 cm. Six subjects compared all pairs, six subjects only the pairs without `lenna`. It was found that `cameron` is rated as the sharpest image, closely followed by `lenna`. The least sharp image is `trui`. Note that this result may depend on the size of the image.

If the measures are *absolute* sharpness measures, this order should also be found for the computed sharpness measures. According to both measures, `trui` is indeed the most unsharp image. The ordering of the other three images is not the same as that found for the pairwise comparison experiment, but these values are not significantly different in the first place. We conclude that both measures may possibly be used for both relative and absolute measures of sharpness, that is, with and without some reference image.

6.7   Results of perceptual experiments

In the sharpening experiments, discussed in chapter 5, we asked subjects to order a range of images that were sharpened with unsharp masking, and subsequently smoothed with anisotropic diffusion [Perona et al., 1994, Catté et al., 1992].

Two standard ISO images (CD-ROM 12640:1997) were used in the experiment, `bicycle` and `portrait` (figure 6.13).

The correlation between the perceived sharpness on the one hand and the computed sharpness measures on the other, is measured with the Spearman rank-order correlation coefficient r_s [Siegel and Castellan, 1988]. The null hypothesis is that there is no relation between the two rankings. The rank-order coefficients for the different subjects and ranges are given in table 6.1.

Table 6.1 shows that the null hypothesis can be rejected for many ranges (see bold numbers). The null hypothesis can be rejected for all ranges with no smoothing. So, there is a correlation between computed and perceptual sharpening. It can also be seen that the two different computational measures do not perform equally well: measure B results in more cases where the correlation does hold. The ranges for which the sharpening measure in the previous chapter is not significant are also not significant for the new sharpness measure. We argued that one range was not

(a) `portrait`   (b) `bicycle`
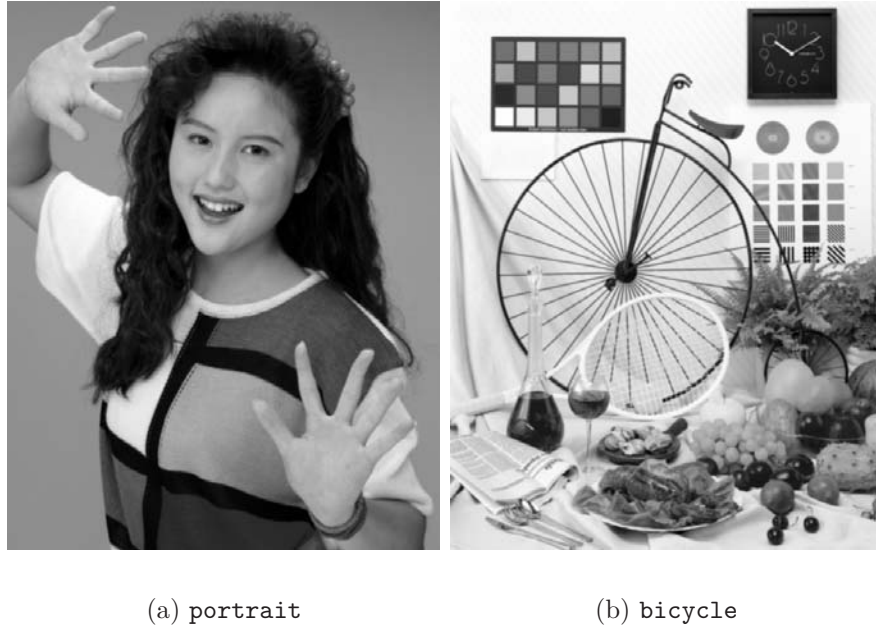
Figure 6.13: The test images for the perceptual experiments..

significant due to coincidence, the other two would not be significant due to the effect of the large smoothing. If this is the case, the supposition that perceptual sharpness depends on the sharpest line or edge in the image is not true for images that are filtered with an anisotropic diffusion filter. Maybe the sharpness of the visual system does not depend on the sharpest line or edge in the image, but on the sharpness of certain structures from which we know that they are sharp, such as eye lashes and hair. As eye lashes have a different orientation at a smaller scale than the main orientation of the eye, they may be smoothed by the anisotropic diffusion filter, introducing an unsharpness. This can also be the explanation for the other not-significant ranges. Another explanation is that the assumption for Gaussian lines and edges cannot be used for images filtered with anisotropic filters. By applying a small Gaussian smoothing the profile of the lines and edges will become more Gaussian.

For measure A (fraction of points) also many ranges are significant, indicating this measure may be used as an estimate for sharpening. However, compared to measure B, more ranges are not significant.

The conclusion is that both objective measures are in agreement with perceptual relative sharpness for images that are not heavily smoothed, but measure B perforsm slightly better than measure A.

Table 6.1: The Spearman rank-order coefficients per subject. For the values printed in italic, the null hypothesis cannot be rejected; that is, in these cases the perceived sharpness is independent of the measures. CV stands for critical value. N is the number of images in the range. Note that subject 4 did not participate in the `bicycle` experiment.

Range	Subject 1	2	3	4	N	CV
$Sm_{0.0}$	**0.88**	**0.88**	**0.79**	**0.81**	8	0.74
$Sm_{0.1}$	**0.83**	**0.79**	*0.48*	*0.12*	8	0.74
$Sm_{0.3}$	*0.74*	**0.76**	*0.64*	**0.76**	8	0.74
$Sm_{0.5}$	**0.88**	**0.84**	**0.80**	**0.88**	7	0.79

(a) `portrait` measure A (fraction of transient points)

Range	Subject 1	2	3	N	CV
$Sm_{0.0}$	**0.79**	**0.79**	**0.79**	8	0.74
$Sm_{0.1}$	*0.69*	**0.86**	*0.64*	8	0.74
$Sm_{0.3}$	*0.09*	**0.76**	*0.35*	8	0.74
$Sm_{0.5}$	*0.45*	*0.73*	*0.02*	7	0.79

(b) `bicycle` measure A (fraction of transient points)

Range	Subject 1	2	3	4	N	CV
$Sm_{0.0}$	**0.95**	**0.95**	**0.86**	**0.88**	8	0.74
$Sm_{0.1}$	**0.95**	**0.90**	*0.71*	*-0.14*	8	0.74
$Sm_{0.3}$	**0.83**	**0.81**	*0.40*	**0.81**	8	0.74
$Sm_{0.5}$	**1.00**	**0.96**	**0.93**	**1.00**	7	0.79

(c) `portrait` measure B (median)

Range	Subject 1	2	3	N	CV
$Sm_{0.0}$	**0.98**	**0.81**	**0.86**	8	0.74
$Sm_{0.1}$	**0.81**	*0.69*	**0.86**	8	0.74
$Sm_{0.3}$	*0.36*	**0.78**	*0.62*	8	0.74
$Sm_{0.5}$	*0.54*	*0.86*	*0.14*	7	0.79

(d) `bicycle` measure B (median)

6.8 Conclusions and discussion

We have found that we can measure the sharpness of simple line and edge images. We first locate these lines and edges in the image. Then we determine the sharpness of these lines and edges by fitting a Gaussian line or edge profile to the Gaussian derivative signature.

We have defined two measures: measure A is the fraction of transient points with a width smaller than 2. This is a very intuitive measure, just the number of points with a small width. Measure B is the median of the widths of the transient points. We found that both measures correlate to perceptual sharpness.

The two measures were tested in perceptual experiments. It was found that measure B yields results that correspond slightly better to perceptual sharpening than measure A. But both measures can be used as a measure for sharpening for images that are not heavily smoothed. Given the fact that the ranges with high smoothing do not correspond very well to the measure, we will have to reconsider the supposition that an image is as sharp as the sharpest line or edge. In the previous chapter we suggested that the problem of these ranges could be that the smaller edges are removed from the image due to the anisotropic diffusion, and this indicates that

more lines and edges than only the sharpest are important for the evaluation of an image.

A topic for further research is to study the distribution of the widths and amplitudes to see if other measures that correlate to the sharpness of images can be defined. Another topic is the relation between the estimated width and lines and edges that are not Gaussian lines and edges.

Chapter 7

Luminance gamma manipulation

7.1 Introduction

An important variable of natural images is contrast. For printed images, the maximum contrast range is determined by the white of the paper used and the darkest color that can be printed. Not only the absolute range of contrast is important, but also the number and discriminability of contrast steps within the range. These variables can be altered by gamma manipulation, as is used in television and photography[1]. This manipulation is given by

$$Y_{out} = k_1 Y_{in}^{\gamma} - k_2, \tag{7.1}$$

where Y_{in} and Y_{out} are the original and reproduced luminances, respectively. The constants k_1 and k_2 are used for keeping the output range constant. If the minimum luminance is zero and the maximum luminance is one, k_1 is 1 and $k_2 = 0$. The exponent γ, gamma, which determines the compression or expansion of luminance steps, is the parameter addressed in this chapter. This equation can be rewritten as

$$Y_{out} = \min(Y_{in}) + (\max(Y_{in}) - \min(Y_{in})) \left(\frac{Y_{in} - \min(Y_{in})}{\max(Y_{in}) - \min(Y_{in})} \right)^{\gamma} \tag{7.2}$$

The relation between the input and output luminance is given in figure 7.1. In 7.1 a, the simple case when the minimum luminance is zero and the maximum luminance is one (no constants k_1 and k_2) is given. In 7.1 b the minimum and maximum luminances are free.

Roufs [Roufs, 1989, Roufs et al., 1994, Roufs, 1992] investigated the effect of gamma on perceptual image quality. The displaying devices were monitors [Roufs, 1989]

---

[1]In photography the gamma of the image can be changed by using different paper. "Hard" paper has a high gamma and "soft" paper has a low gamma.

(a)   (b)
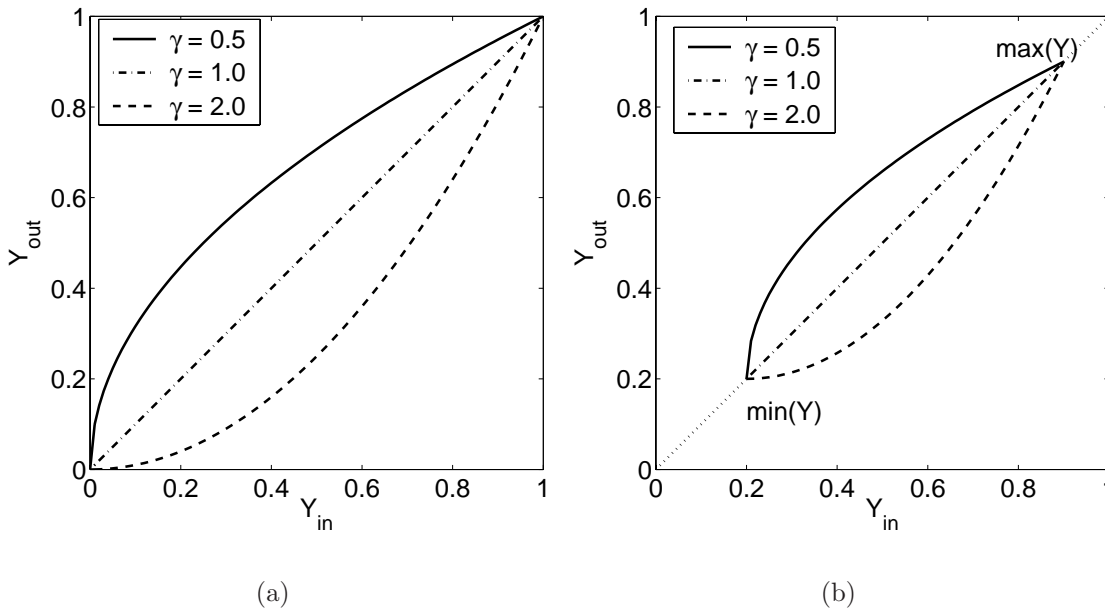
Figure 7.1: Two versions of equation 7.2. In a) the maximum luminance is 1 and the minimum luminance is 0. In b) the maximum luminance is 0.9 and the minimum luminance is 0.2.

or slide projections[Roufs et al., 1994, Roufs, 1992]. Important conclusions from these experiments were that differences between individual subjects are of minor importance, but differences between test scenes may have a major influence on the perceived image quality[Roufs, 1992]. This dependency on image content can also be observed in the results of a study on the effect of surround (dark or light) on perceived lightness contrast [Daniels et al., 1997].

In the experiments described in this chapter we investigated how gamma affects the perception of *color* images. Thereto we varied the gamma of the luminance distribution in the image, while keeping the color coordinates x and y of each point constant. We also looked at differences between printed color images and black-and-white images. These experiments are extensions of trial experiments exploring black-and-white images [Dijk et al., 1999], and two color images [Dijk et al., 2000].

7.2   Method

7.2.1   Test material

The stimulus set consisted of five different images (figure 7.2). These images are ISO standard images (from the cdrom ISO 12640:1997), `portrait`, `cafe`, `basket`, `bicycle` and `musicians`. The 300 dpi versions of these images were used. Since we

(a) portrait   (b) cafe   (c) bicycle

(d) basket   (e) musicians

Figure 7.2: The five stimuli.

were interested in the role played by the image content, we chose test images that were quite different in this respect.

These images are given in CMYK format. Although this is a device dependent color space, we chose to use these images because they are used as standards in the field of printing.

The images were converted from CMYK to XYZ using the methods described in subsection 4.1.5. The white point of the resulting images was the white of the paper used (EPSON photo quality ink jet paper).

Next we performed a chroma mapping (see section 4.3) over the whole image to prevent too many colors from being out-of-gamut after the γ transformation. This step has the effect of slightly desaturating the colors; that is, the chroma C^* (in CIELAB) of each point was replaced by $0.85 * C$. The effects of this saturation with respect to the quality will be discussed in chapter 8.

The luminance values corresponding with the various gamma values were calculated with equation 7.1. The maximum luminance was the luminance of the paper (approximately 80% of the light was reflected) and the minimum luminance was the luminance of the darkest black that could be printed (approximately 6-7% of the light was reflected). This range control was achieved by using suitable values of k_1 and k_2 in equation 7.1.

Just before the transformation from XYZ to CMY, an orthogonal gamut clipping (see section 4.3) was done in which the few remaining colors that were out-of-gamut (despite the earlier applied gamut mapping) were mapped onto the gamut. Note that with this gamut mapping the hue of the point remains the same.

The images were converted from XYZ to CMY, again using the methods described in subsection 4.1.5. The color images were sent to the printer (an EPSON Stylus color 1520) as a postscript file with CMY values.

In addition to the color images we also made black-and-white images by transforming Y_{out} directly to K. The effect of the gamma manipulation is illustrated in figure D.1. The input image is a part of the `bicycle` image. The images were printed on EPSON photoquality paper. The size of the paper is A4 and the size of the images 12.8 x 16 cm.

7.2.2   Experimental setup

Three experiments were performed, respectively on the stimulus response relation, perceptual quality and perceptual naturalness:

Experiment A: Stimulus response relation The goal of the first experiment was to determine the relation between the gamma of the stimulus and the perceptual response of the subject. This was done with two different tasks: rank ordering and bisection.

The subject first had to rank order images according to gamma, which, for this type of stimuli, inherently implies ordering according to average lightness (or luminance). Words to this effect were also used in the subject instruction. The rank ordering provided information as to how well subjects are able to discriminate between prints with different gamma settings.

For the bisection task the subject had to decide which print was reproduced with a gamma exactly in between the "lightest" print and the "darkest" print. Only one print could be chosen.

Five different ranges were used, see figure 7.3. Three ranges were presented, in which gamma respectively varied from 0.6 to 1.7 (full range), 0.6 to 1.3 (low range), 1.0 to 1.7 (high range) with equal steps of 0.1. We also presented two
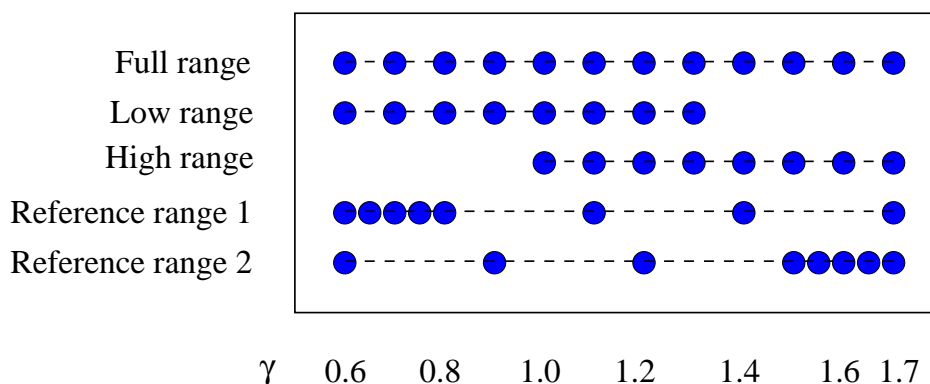
Figure 7.3: The different ranges used in the experiment. Reference range 1 was used for the `portrait`, `cafe` and `bicycle` image, reference range 2 was used for the `basket` and `musicians` image.

reference ranges using unequal steps in gamma, to bias the average gamma towards low values (reference range 1) or high values (reference range 2). Reference range 1 was used for `portrait`, `cafe` and `bicycle`. This range has small steps in the low gamma values ($\gamma \in \{0.6, 0.65, 0.7, 0.75, 0.8, 1.1, 1.4, 1.7\}$). Reference range 2 was used for `basket` and `musicians`. This range has small steps in the high gamma values ($\gamma \in \{0.6, 0.9, 1.2, 1.5, 1.55, 1.6, 1.65, 1.7\}$). These reference ranges were used to test if people select the *perceptual* or numerical center of the range. The ranges are visualized in figure 7.3.

Experiment B: Quality In this experiment, using the same ranges, the perceptual quality for a particular gamma was tested. The pictures were arranged according to the subject's own rank ordering in the stimulus response experiment. The subjects had to order the prints according to quality. They were instructed to select the print they would use to decorate the wall of their living room, and then the second best, and so on.

Experiment C: Naturalness In the third experiment the subjective naturalness of a print for a particular gamma was tested. The setup of the experiment was the same as for the quality experiment, but now the subject was asked to order the prints according to naturalness. They were instructed to select the print they thought looked the most like the original *scene*. Note that the subjects could not know how the original scene would actually look.

7.2.3 Viewing conditions

In the experiments the prints were viewed under homogeneous lighting in a Macbeth SpectraLight II light booth. The luminance reflected from the print was maximally 450 cd/m^2 (white paper). The color temperature of the illuminant was approximately 6430 K. The prints were put into transparent plastic covers to prevent

smudges. A window was cut in the center of these covers, to see the image on the print directly and not through the plastic cover.

7.2.4 Subjects

Eight subjects took part in the experiments, all completely naive with respect to image analysis. The subjects were undergraduate students, paid to do these experiments. Three of the subjects were male and five female. They were 17 to 23 years old. The subjects had normal color vision and normal or corrected-to-normal visual acuity.

7.3 Results

7.3.1 Experiment A: stimulus response relation

Rank ordering

In experiment A the rank ordering of the prints by the subjects was compared to the ordering in physical gamma. The (average) ordering as function of the gamma of the image is shown in figure 7.4 (for black-and-white pictures) and 7.5 (for color pictures).

The results indicate that the subjects are quite capable of ranking the images according to the physical γ values. In the full, low and high ranges some errors (order reversals) were made (figure 7.5 a, b, and c and figure 7.4 a, b, and c). Some of these errors reflect the limitations of the discriminability of the gamma steps. We believe, however, that some of the errors are simply due to loss of attention by the subjects, a problem also encountered in previous studies [Dijk et al., 1999, Dijk et al., 2000], and a general problem in this type of experiment.

In the reference ranges (figure 7.5 d and e and figure 7.4 d and e), more errors were made in the images in the part of the range where the step sizes were smaller, probably because the differences between two consecutive images were smaller there as well. Note that there is no indication that the biasing of the gamma range towards low or high has not affected the rank ordering.

Bisection

The bisection results, averaged over subjects, are plotted for the black-and-white images in figure 7.6, and for the color images in figure 7.7.

(a) full range   (b) low range   (c) high range

(d) reference range 1   (e) reference range 2
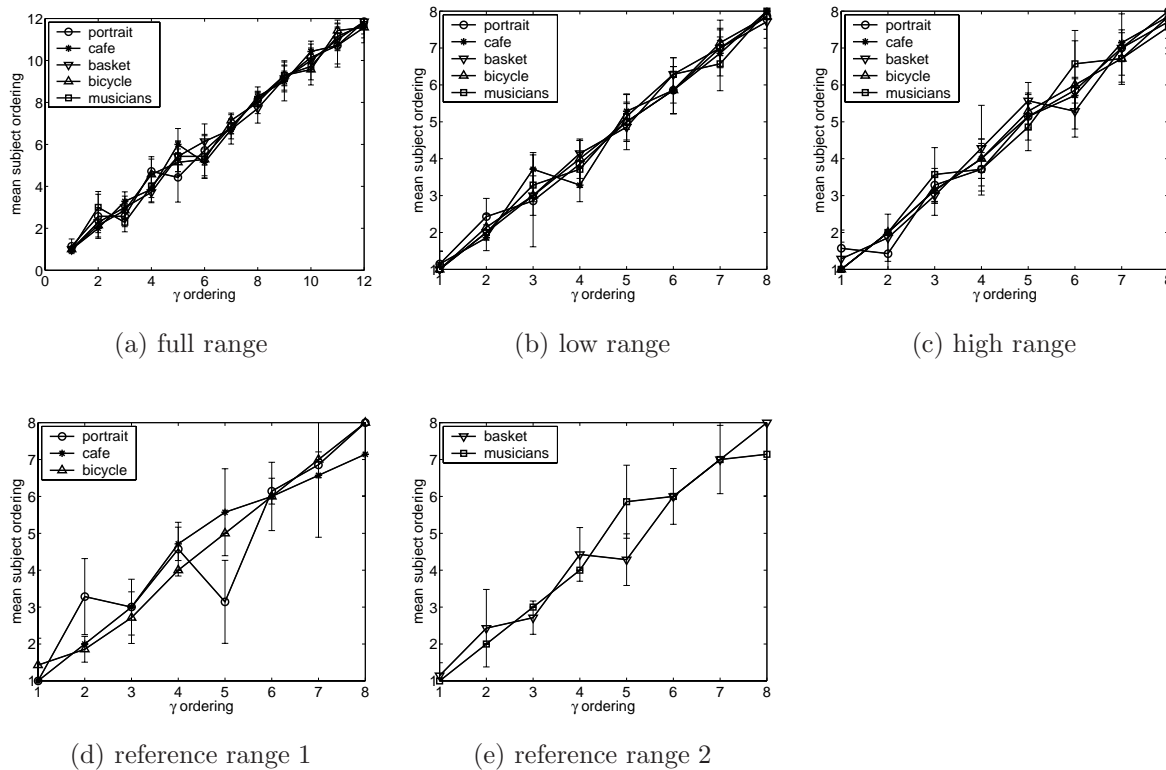
Figure 7.4: The results of the mean rank ordering for the black-and-white stimuli. The ordering averaged over the subjects is given as function of the ordering of gamma (from low to high) of the image. Note that for the `portrait`, `cafe`, and `bicycle` image the reference range (given in figure d) is a different selection of γ values steps than for the `basket` and `musicians` image (figure e).

The reference range (either 1 or 2, depending on the image) was used to see if subjects showed a consistent response for the *perceptual* center of the range, that is, if the image that was in the center of the range depended only on the "lightest" print and the darkest print, and not on the steps in gamma in the presented range of images. One subject did not show this behaviour, for eight of the ten reference ranges he chose the numerical center. Because this indicates that he was not able to understand this experiment well, he was excluded from the results.

Figure 7.7 and 7.6 show that the results of most ranges are shifted towards the left of the range. This would indicate that the stimulus response function is steeper on the left side than on the right side of the range.

We tested this by looking at the bias, i.e. the deviation from the numerical center of the range, and the error in this bias. The significance was tested with the Student's T test. The values are given in table 7.1. For some values, this shift is indeed significant. Starting with the full range, the shift is significant for the color images but not for the black-and-white images. The shift in the low range is not significant

(a) full range   (b) low range   (c) high range

(d) reference range 1   (e) reference range 2
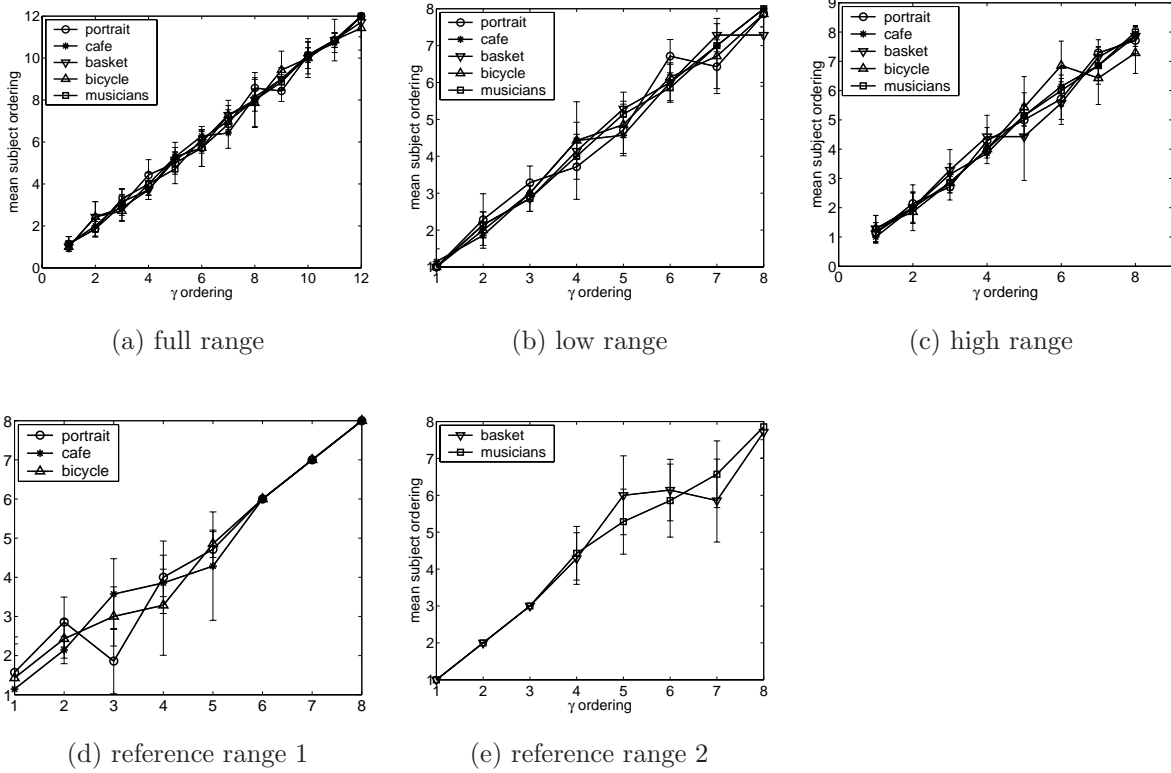
Figure 7.5: The same experimental results as shown in figure 7.4, but now for the color stimuli.
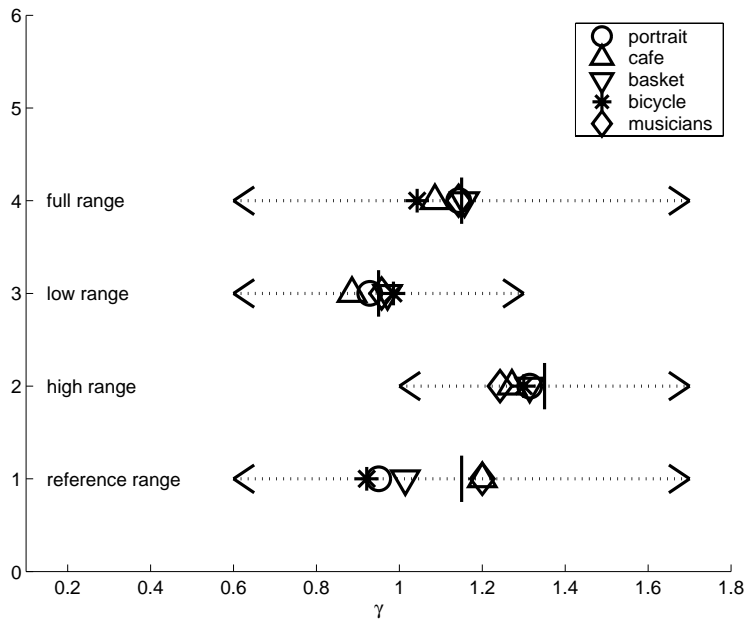
Figure 7.6: The results of the bisection experiment (A), averaged over subjects for the black-and-white images. The horizontal arrows delineate the range of γ that is presented.
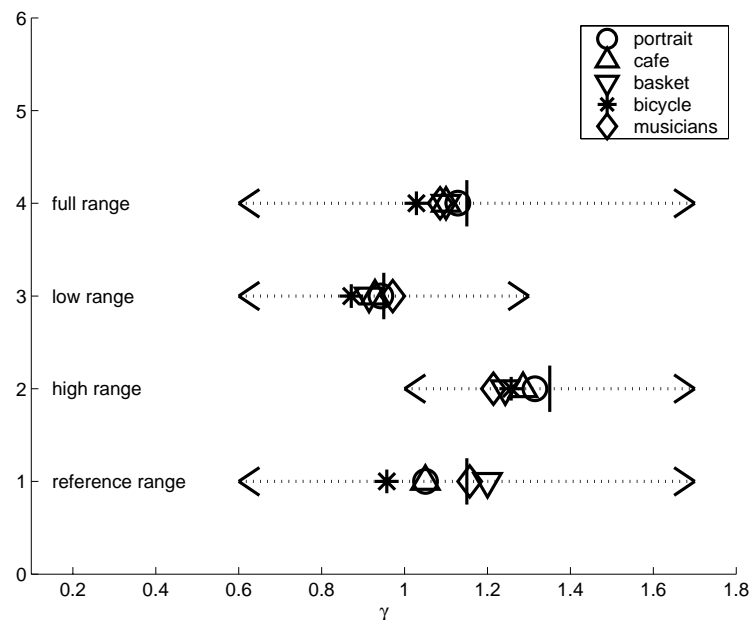
Figure 7.7: The results of the bisection experiment (A), averaged over subjects for the color images. The horizontal arrows delineate the range of γ that is presented.

for both the black-and-white and the color images, the shift in the high range is significant for both types of images.

The shifts in the reference range are not significant, except for the ranges for which the full range is also shifted. This means that the irregularity of the ranges has no significant effect on the results.

In conclusion, this result is consistent with a stimulus-response function that is more or less linear at the lower gamma range, and has a decreasing slope towards the higher gamma values. Significant shifts tend to occur for ranges including the high range. This is consistent with the non-linearity of the response function (compression at high levels). Most range effects can be shown for the `bicycle` image individually. Because this is not the case for the other images, one might suspect that the range effects are only due to `bicycle`. However, if the overall range effects are determined without `bicycle`, the results are not different except for the reference range, indicated that the range effects do not mainly depend on this image.

In an earlier experiment [Dijk et al., 2000], we found that the stimulus response function was a stimulus-response function, consistent with a slightly steeper slope at the low gamma values than at the high gamma values. This result was based on only two images and four subjects. Clearly, the non-linearity of the stimulus response function can only be measured with more data than in the earlier experiment. Still, the trend of the data is nevertheless the same as in the present experiment.

Table 7.1: The bias, i.e. the deviation from the numerical center of the range, for the ranges in the bisection experiment. Zero means that the bisection point is exactly in the center of the gamma range, a value larger than zero means that the bisection point is shifted towards the higher gamma values and a value smaller than zero means that the bisection point is shifted towards the lower gamma values. The bold numbers are significant.

	bias			
	all images		all images except `bicycle`	
range	bw	color	bw	color
full	*-0.04 ± 0.04*	**-0.06 ± 0.04**	*-0.02 ± 0.04*	**-0.05 ± 0.04**
low	*0.00 ± 0.03*	*-0.02 ± 0.04*	*-0.01 ± 0.03*	*-0.01 ± 0.04*
high	**-0.06 ± 0.03**	**-0.09 ± 0.04**	**-0.06 ± 0.03**	**-0.09 ± 0.04**
reference	*0.05 ± 0.07*	**-0.07 ± 0.06**	*-0.06 ± 0.07*	*-0.04 ± 0.04*

		bias	
range	image	bw	color
full	`portrait`	*-0.01 ± 0.11*	*-0.02 ± 0.07*
	`cafe`	*-0.06 ± 0.08*	*-0.05 ± 0.12*
	`basket`	*0.01 ± 0.16*	*-0.05 ± 0.13*
	`bicycle`	**-0.11 ± 0.11**	**-0.12 ± 0.07**
	`musicians`	*-0.01 ± 0.05*	*-0.06 ± 0.08*
low	`portrait`	*-0.02 ± 0.07*	*-0.01 ± 0.15*
	`cafe`	*-0.06 ± 0.08*	*-0.02 ± 0.07*
	`basket`	*0.02 ± 0.10*	**-0.04 ± 0.04**
	`bicycle`	*0.04 ± 0.06*	**-0.08 ± 0.07**
	`musicians`	*0.01 ± 0.05*	*0.02 ± 0.14*
high	`portrait`	*-0.04 ± 0.06*	*-0.04 ± 0.11*
	`cafe`	**-0.08 ± 0.05**	**-0.06 ± 0.06**
	`basket`	*-0.04 ± 0.10*	*-0.11 ± 0.12*
	`bicycle`	*-0.05 ± 0.08*	**-0.09 ± 0.09**
	`musicians`	**-0.10 ± 0.08**	*-0.14 ± 0.10*
reference	`portrait`	*-0.20 ± 0.17*	*-0.10 ± 0.12*
	`cafe`	*0.05 ± 0.16*	*-0.10 ± 0.12*
	`basket`	*-0.14 ± 0.14*	*0.05 ± 0.16*
	`bicycle`	**-0.23 ± 0.16**	**-0.19 ± 0.17**
	`musicians`	*0.05 ± 0.00*	*0.01 ± 0.11*

7.3.2   Experiment B and C: quality and naturalness

The two smaller ranges (low and high) were used to test for the presence of any range effects, that is, a tendency for preferring the middle of a range. The results indicate, that the seven subjects preferred the same or nearly same image, irrespective of

the stimulus range in which it was presented. This was also found in a previous experiment [Dijk et al., 2000]. As expected, the quality and naturalness selection of the eighth subject, excluded already in experiment A, did shift with the range.

After each experiment the subjects were asked what criteria they had used. One of the subjects gave unusual criteria for the quality and naturalness experiments, that is, she preferred the images in the order from the darkest to the lightest. Because we wanted to have a homogeneous group of subjects, we chose to exclude this subject. We tested the homogeneity of the remaining group of subjects using the Friedman two-way analysis of variance and Kendall Coefficient of Concordance. The results of these tests, given in appendix F, show that for most experiments the ranks are not random, and thus that the subjects can be seen as one group. For some images this was not the case. This does not mean that the (average) responses cannot be used for a quality measure, but for certain subjects the deviation from the mean quality was quite high.

The quality and naturalness rank orderings (averaged over subjects) are given in figure 7.8.

We can see here that there are differences between the color and black-and-white responses, and between naturalness and quality responses. The image with the highest quality is for most images close to the original image (γ is 1). However, for the `bicycle` image the image with the highest quality is the image with gamma is 1.3 (color) and 1.2-1.3 (black-and-white).

The shift in preference can be explained by looking at the images and the remarks that the subjects made. It seems that variations in γ affect the balance between chromatic and achromatic aspects of the image, in the sense that at higher γ the blacks and whites become more enhanced, causing the colors to lose some of their colorfulness (cf. [Hunt, 1991]). Decreasing the γ has the opposite effect, so reducing the γ, and hence, luminance contrast, makes the colors stand out, even to an extent that it is no longer natural. Evans ([Evans, 1974]) already made related observations and reported how the color of a stimulus centered in an achromatic surround becomes more colorful with decreasing luminance contrast. The `bicycle` image contains many achromatic colors. It could benefit therefore from more luminance contrast, and that is what the subjects indicate by using γ's greater than 1. Indeed they do the same in the black-and-white version. The `cafe` image needs much color, as it is an outdoor image of a sunny day. As we reduced the chroma of the image, this image should benefit from more colorfulness, as is the case for a γ slightly lower than 1. The `musicians` image is a balanced image in color and luminance. Most subjects indicate that they look mostly at the face of the blond and the dark women. If $\gamma > 1$, the contrast in the dark face becomes lower, whereas for $\gamma < 1$, the face of the blond woman loses contrast. The `portrait` image is balanced in color and luminance contrast, and cannot gain from a gamma increase or
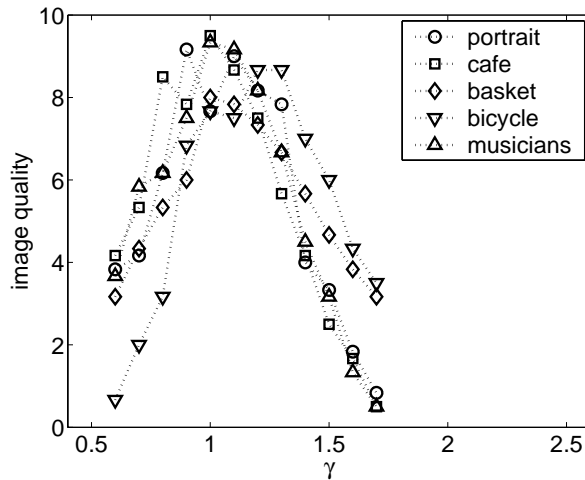
(a) black-and-white quality
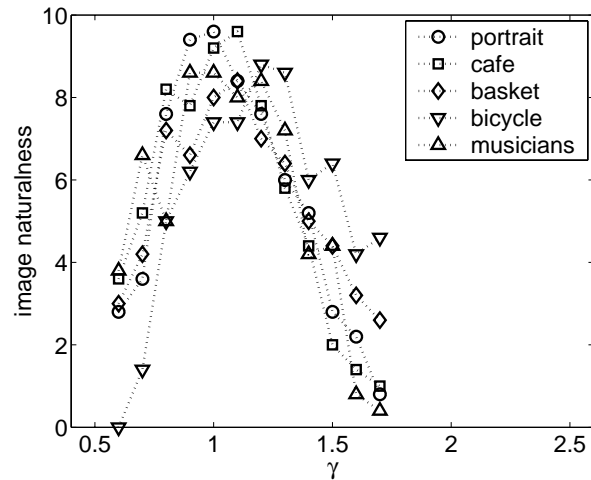
(b) black-and-white naturalness
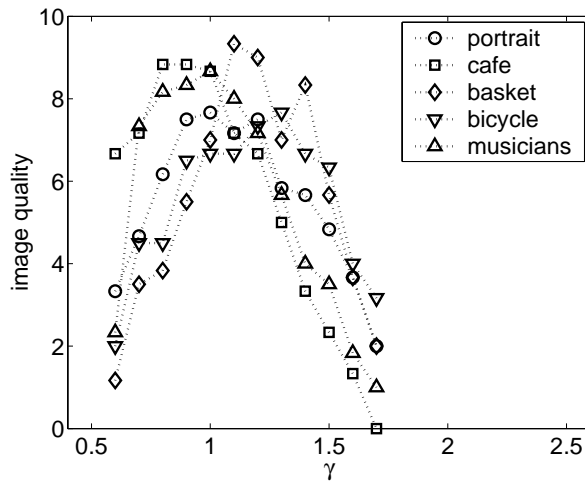
(c) color quality
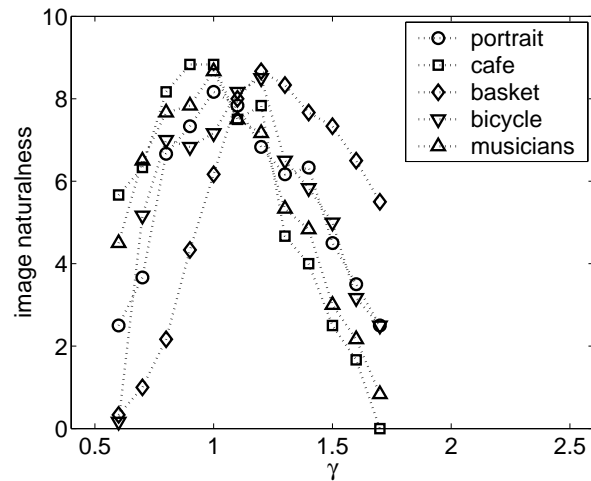
(d) color naturalness

Figure 7.8: The subject responses for the quality and naturalness experiments. There is a statistically significant interaction between gamma and chroma for all groups.

decrease. The `basket` image is a colorful image, that can only slightly be improved by a $\gamma > 1$.

In an earlier experiment [Dijk et al., 2000] we investigated the quality of the `musicians` and `bicycle` image. For the `musicians` image we found in the earlier experiment that the preferred gamma was 1.0, both for the color and the black-and-white versions. This is the same as we found in the experiments that are presented here. For the `bicycle` image, we found in the earlier experiment that the preferred gamma for the color image was 1.2, and for the black-and-white image 1.4. In the experiments presented here we found that the preferred gamma for the color image is 1.3 and

for the black-and-white image 1.2-1.3. This is slightly different than found before, but this is probably due to differences between the subjects. The trend that the preferred gamma for the `bicycle` image is higher than 1 is found again. However, in the experiments presented here, we did not find that the preferred gamma for the black-and-white image is higher than for the color image.

In the following part of this chapter we will first investigate which settings of the experiment had a statistically significant effect on the responses of the subjects. After that, we will show how the usefulness-naturalness model
[Janssen and Blommaert, 2000b] may be used to fit the quality responses.

Statistical Analysis

To investigate if the different settings had a significant effect on the responses of the subjects, the Analysis of Variance (ANOVA) [StatSoft, 2000, Winer, 1970] method was used. This method is described in subsection 4.4.3.

We tested the responses of the subjects for the following variables:

Image: the input image the subject ranked

Gamma: the gamma of the manipulated image

Q/N: whether quality (Q) or naturalness (N) was the attribute the subject ranked

Color/BW: whether the manipulated image was given in color or in black-and-white.

The subjects ranked images that only differed in gamma, therefore we could not test the main effects for the image, for quality/naturalness and for color/bw (see also page 82). However, we can find interactions between these variables and gamma.

To obtain more insight, we grouped the responses into different groups in which one or more variables were constant. All results are given in table 7.2.

The four main groups of interest are color quality, black-and-white quality, color naturalness and black-and-white naturalness. The subject responses for these four groups are shown in figure 7.8. For all these groups we found that there is a main effect for gamma. That is, the quality/naturalness results depend on the gamma setting, regardless of the image. There is also a two-way interaction between image and gamma, which indicates that the way the results depend on the gamma setting, is influenced by the image. These effects can be seen in all groups that were evaluated.

Next, we evaluated the results grouped into color images and black-and-white images. For both groups, no interaction between quality/naturalness and gamma was found, which indicates that there is no reason to assume that the way the results

(a) `bicycle` color

(b) `musicians` color

(c) quality color

(d) naturalness color
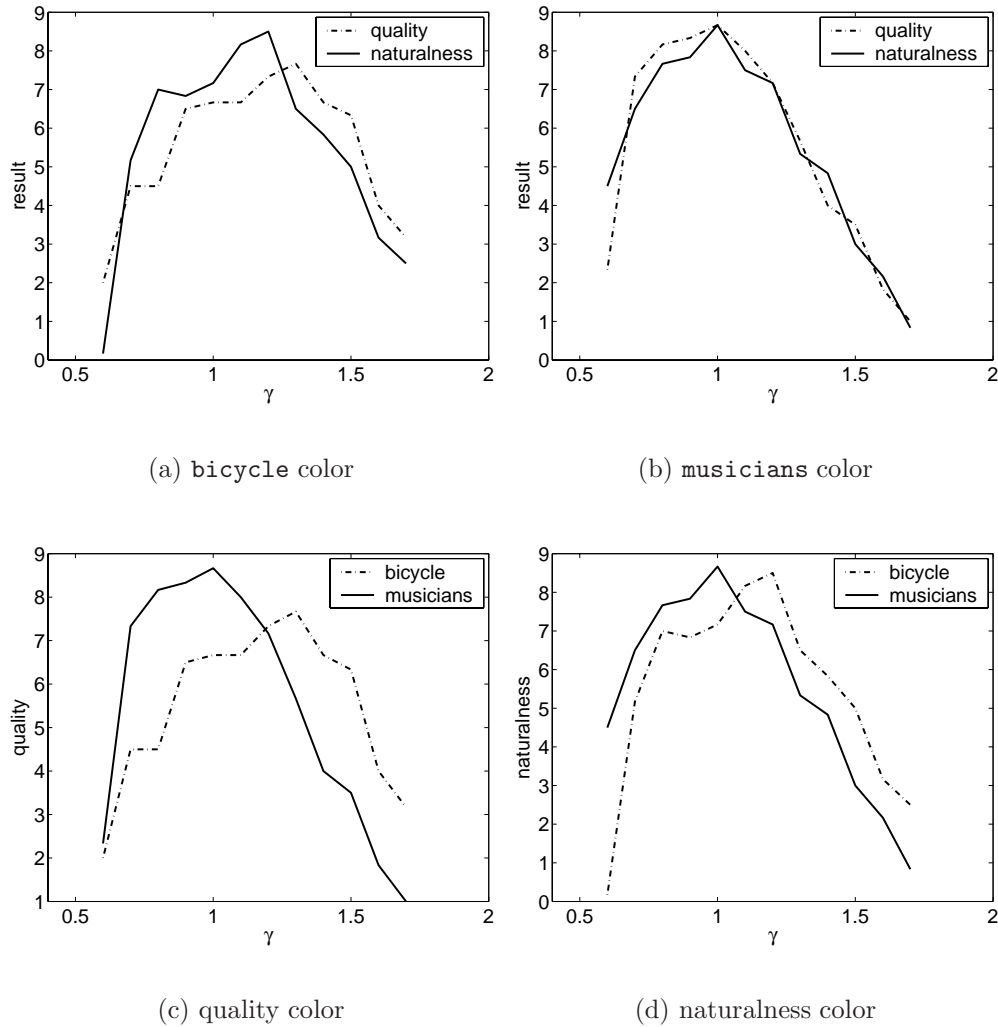
Figure 7.9: The quality and naturalness responses for the `bicycle` (a) and `musicians` (b) color images. For the `musicians` image the curves are almost the same, whereas for the `bicycle` image the curves are shifted with respect to each other. This effect visualizes the three-way interaction between gamma, image and quality/naturalness.

depend on the gamma setting, is influenced by whether the subject scaled quality or naturalness. However, for the color stimuli, there is a three-way interaction between quality/naturalness, the image and the gamma. This indicates that the influence of the image on the gamma dependence is different, depending on whether quality or naturalness is scaled. This is not the case for the black-and-white stimuli. The idea of the three-way interaction can be observed in figure 7.9. For the `musicians` image the naturalness and quality scaling is almost the same, whereas for the `bicycle` image the peak for the quality response is at a higher gamma value (1.3) than the peak for the naturalness response (1.2).

Summarizing, when comparing the quality and naturalness results, we conclude that for both groups there is no interaction between color/black-and-white and

Table 7.2: The results of ANOVA for different groups of responses. F is the F-ratio, p the probability. Effects that are *not* significant are given *in italics*.

	effect	F	p
Color quality			
	Gamma	**7.98**	**0.000**
	Image * Gamma	**2.47**	**0.000**
BW quality			
	Gamma	**19.25**	**0.000**
	Image * Gamma	**1.65**	**0.011**
Color naturalness			
	Gamma	**11.47**	**0.000**
	Image * Gamma	**3.03**	**0.000**
BW naturalness			
	Gamma	**9.67**	**0.000**
	Image * Gamma	**1.44**	**0.046**
Color			
	Gamma	**11.33**	**0.000**
	Q/N * Gamma	*0.53*	*0.873*
	Image * Gamma	**3.01**	**0.000**
	Q/N * Image * Gamma	**1.73**	**0.006**
BW			
	Gamma	**16.64**	**0.000**
	Q/N * Gamma	*1.04*	*0.424*
	Image * Gamma	**1.86**	**0.002**
	Q/N * Image * Gamma	*0.60*	*0.978*
Quality			
	Gamma	**19.65**	**0.000**
	Col/bw * Gamma	*0.57*	*0.844*
	Image * Gamma	**2.99**	**0.000**
	Col/bw * Image * Gamma	*0.82*	*0.778*
Naturalness			
	Gamma	**14.52**	**0.000**
	Col/bw * Gamma	*1.24*	*0.285*
	Image * Gamma	**2.67**	**0.000**
	Col/bw * Image * Gamma	**1.81**	**0.003**

gamma. However, for the naturalness results, there is a three-way interaction between color/black-and-white, the image and gamma. This indicates that influence of the image on the gamma dependence, is different for color or black-and-white images. This is not found for the quality results.

Table 7.3: The results of linear regression for the color images and the black-and-white images.

image	variables	explained variance	naturalness coefficient	D_L^* coefficient
color images	naturalness & D_L^*	0.85	0.85	0.22
	naturalness	0.80	0.90	
	D_L^*	0.15		0.40
black-and-white images	naturalness & D_L^*	0.90	1.04	-0.19
	naturalness	0.87	0.94	
	D_L^*	0.15		0.41

Searching for a relation between quality, naturalness and physical measures

In the rest of this chapter we try to fit a quality model to the results we found using linear regression. This should answer the important question whether there is a relation between quality, naturalness and physical measures in the image. The quality model we use is the usefulness-naturalness model [Janssen and Blommaert, 2000b] (see subsection 3.4.3 on page 37), in which two objective measures, discriminability and naturalness, are presented. Since we cannot use the naturalness measure Janssen and Blommaert developed (this is based on grass, skin and sky, which are not present in some of our images), we use the subjective measured naturalness. The discriminability measure is defined in equation 3.34. This measure can be calculated for different color spaces, as indicated by the subscripts.

To test the relation between quality on one hand and naturalness and discriminability on the other hand, we used linear regression. We found that, consistent with the usefulness-naturalness model, that quality depends for a large part on the naturalness of the image. The explained variance was 80% for the color experiment and 87% for the black-and-white experiment.

The first discriminability measure we tested was D_{L^*}, the discriminability in the lightness dimension. The relation between gamma and D_{L^*}, and results for the linear regression are shown in figure 7.10. With only this measure only 15% of the variance of both the color and black-and-white experiment was explained. However, the discriminability had a significant effect on the image quality.

Combining the discriminability and the naturalness data, 85% of the variance in the color experiment and 90% in the black-and-white experiment can be explained. The predicted quality is plotted against the observed quality in figure 7.11.

Instead of using D_L^*, we could also use a discriminability measure that takes one or two color dimensions into account. We evaluated the following discriminability
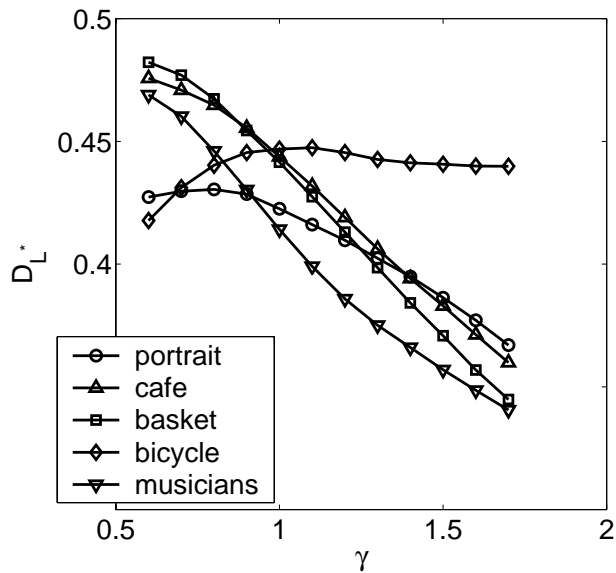
(a)
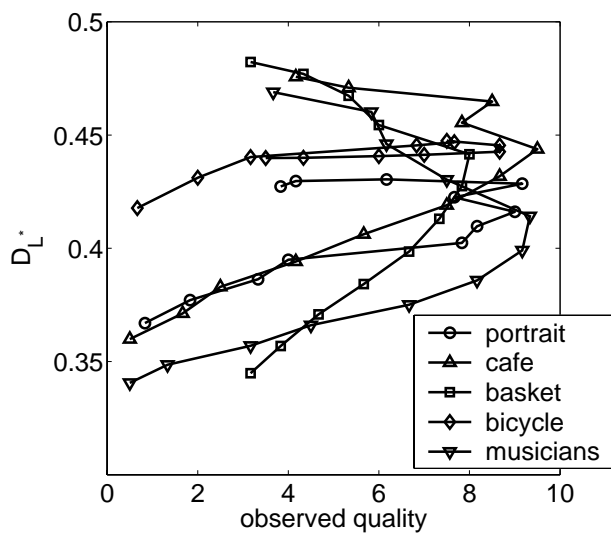
(b) black-and-white
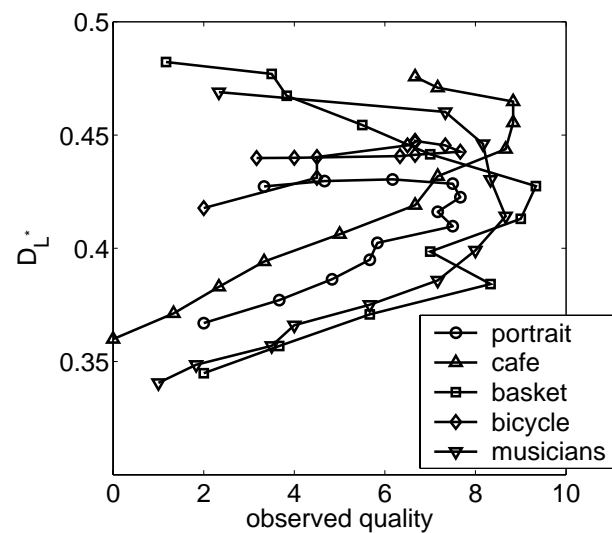
(c) color

Figure 7.10: The relation between gamma and D_{L^*} is given in (a). The results (scatter plots) of the linear regression analysis, using only D_{L^*} is given in (b) for the black-and-white images and in (c) for the color images.

measures: $D_{L^*a^*b^*}$, $D_{L^*u^*v^*}$, $D_{L^*u'v'}$, $D_{L^*S^*}$ and $D_{L^*C^*}$. These discriminability measures performed somewhat poorer than D_L^*. This might be expected, because we only changed the achromatic attribute in the image, so another color axis would only give more noise.

On the basis of these results, and the usefulness-naturalness model, we can reasonably well predict the effect of gamma on the quality of an image, using only

measures of naturalness and the discriminability of the image. However, we moved
the problem from measuring image quality to image naturalness. We would like to
use the objective measure Janssen and Blommaert developed for naturalness, but
this measure assumes that all images contain either skin, sky or grass, which is not
the case in two of our five images.

7.4   Conclusions and discussion

In this chapter we presented the results of experiments in which we manipulated
color images in the XYZ color space by applying a gamma transformation on the
luminance (Y), while keeping the chromaticity (x,y) constant. In order to evaluate
the influence of the chromatic components, we also printed the same images in
black-and-white.

The aim of the first experiment was to determine the stimulus-response relation
between the physical gamma (the stimulus) and perceptual response ("perceived
gamma"). It was shown that the subjects are quite capable of ranking the im-
ages according to the physical gamma values. These results are consistent with a
stimulus-response function which is more or less linear at the lower gamma range,
and has a decreasing slope towards the higher gamma values.

The gamma manipulation changes the mean luminance level and the local contrast
of the image. For a gamma smaller than one the mean luminance level is increased
and the contrast in darker parts of the image is increased at the cost of the contrast
in the lighter parts of the image. For a gamma larger than one these effects are
inversed. The relation between the mean luminance and perceived mean lightness is
(approximately) described by a third root power function (see equation 2.4). So, if
the subjects evaluate the mean luminance level of the image, the stimulus response
function should be increasing towards the higher gamma values. Assuming that the
increase in contrast in terms of luminance is equal to the decrease in contrast in
the lighter parts of the image, and vice versa, we would expect a stimulus response
function that is steeper at the lower gamma values than at the higher gamma
values. Because this is the case, we hypothesize that the subjects evaluate the local
contrast rather than the mean luminance level in the bisection task. We have not
tested this hypothesis. This is one of the questions that could be answered through
a continuation of this research.

The aim of the second and third experiment was to measure the subjective quality of
the images and the subjective naturalness of the images, respectively. We found that
for all images there was a maximum quality, and that decreasing the quality from
the preferred gamma has the same effect on the quality judgement as increasing the
gamma from the preferred gamma. The differences in quality estimations over the
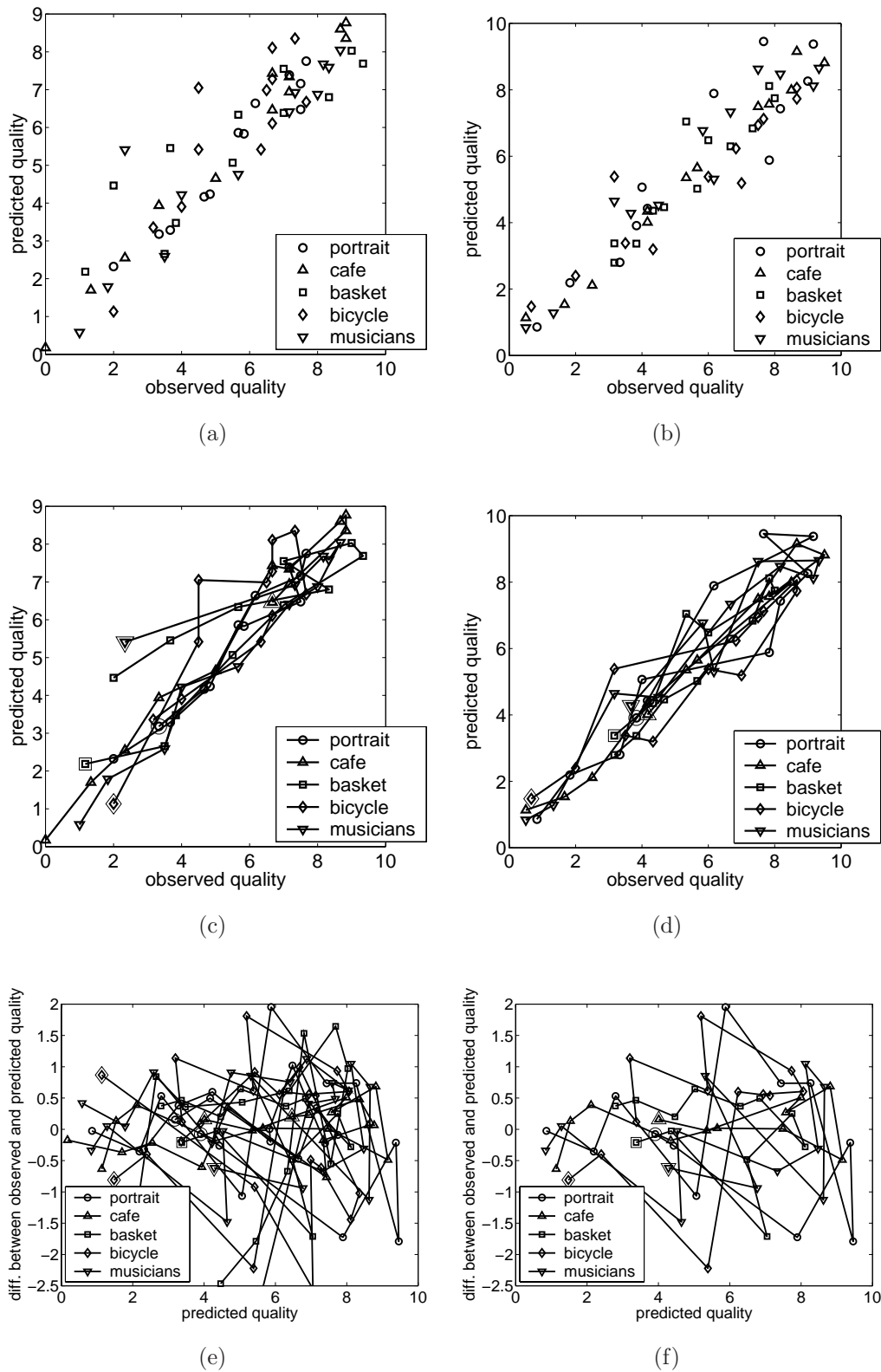
Figure 7.11: The result of the linear regression. In (a) and (b) the results are given as a scatter plot, in (c) and (d) the same results are connected by lines. These lines indicate the relation with gamma. The points with $\gamma = 0.6$ are plotted with double contours.

subjects was not negligible. For some images we did not find agreement between the different subjects. This indicates that if one finds a quality measure for the average subject, the individual subject responses can still show large differences.

The gamma manipulation affects both the contrast in the achromatic colors and the colorfulness of the colored parts. These effects may explain the differences between the quality responses for the different images. A suggestion for future work is to manipulate the colors locally, so that the colored parts remain colored and the more achromatic parts gain contrast. This is similar to contrast stretching, a technique that is used in photography where the colors in the shade are manipulated differently than the colors in full light.

Using the statistical analysis method ANOVA, we found that for all groups of experiments there was a main effect for gamma and an interaction between the image and gamma. The dependency on gamma is not the same for all images, but is influenced by the image that is evaluated. Because of the image differences and the effects that a gamma manipulation has on the image, this could also be expected. Some images benefit from more luminance contrast (for instance images that do not contain much color), other images may benefit from more colorfulness. For the color stimuli a three-way interaction between gamma, image and quality/naturalness was found. This was not the case for the black-and-white stimuli. So, the way the image influences the subject's response on colored images manipulated with a gamma manipulation is different depending on whether the subject was expected to scale quality or naturalness.

Using linear regression we found that we could predict the quality of the image using the discriminability measure of the usefulness-naturalness model [Janssen and Blommaert, 2000b] and the subjective naturalness. This is a confirmation of the usefulness-naturalness model, which predicts quality using the discriminability and a more objective naturalness measure. The measure that was developed by Janssen and Blommaert to estimate the naturalness of an image, could not be used because we used images that do not contain skin, grass or sky. So we shifted the problem of estimating the quality of an image to estimating the naturalness of that image.

Chapter 8

Luminance gamma manipulation combined with chroma scaling

8.1 Introduction

In the previous chapter, we studied the effect of changing the achromatic contrast on the quality of images. The achromatic contrast was changed by means of a gamma manipulation. In this chapter, we study the combined effect of changing both the achromatic contrast and the chroma of the image.

Fedorovskaya et al. [Fedorovskaya et al., 1997] and de Ridder et al.
[de Ridder et al., 1995] did experiments in which the chroma was changed in the CIELUV space. Fedorovskaya et al. tested two different ways of changing the chroma: 1) through the addition or subtraction of the same amount of chroma to or from the chroma value of each pixel; 2) through multiplication of the chroma value of each pixel by a constant. They found that the so-called colorfulness [Hunt, 1977] of an image is the main perceptual attribute underlying image quality and naturalness when chroma varies. Colorfulness could be modeled as a function of the average chroma and the variability of chroma as derived from the color point distribution in the CIELUV color space. Both Fedorovskaya and de Ridder found that the highest quality image is not the most natural one. The subjects preferred more colorful images, although they realized that these images looked somewhat unnatural. This was also found by Janssen and Blommaert [Janssen and Blommaert, 2000b], who modeled quality as a function of naturalness and discriminability (see subsection 3.4.3). The discriminability increases for images with higher chroma.

In the experiments described in this chapter, we changed the images using two consecutive steps. First, the chroma was changed using a variable scaling factor. Then, the luminance distribution in the image was changed using a gamma manipulation,

while keeping the color coordinates x and y of each point constant[1]. In the previous chapter, the chroma scaling factor was a constant (0.85). This scaling factor was used to reduce the number of out-of-gamut colors. The gamma manipulation was the same manipulation that was used in the previous chapter.

8.2   The relation between gamma and chroma

In the experiment we first scaled the chroma C^* in the CIELAB space. In this space we also have the variables a^*, b^*, L^* and h^*. After this we applied a gamma transformation (an exponent relating the input to the output) to the luminance (Y) in XYZ space, while keeping the chromaticity values (x, y) constant.

In this section, we consider the theoretical relation between γ and chroma. The change in hue h^* of the CIELAB space is also studied.

First, we study how the chroma changes from C_1^* to C_2^* due to a change in Y from Y_1 to Y_2. The chromaticity (x, y, z) is kept constant, so

$$x_1 = x_2 \text{ and } y_1 = y_2 \text{ and } z_1 = z_2 \tag{8.1}$$

where subscript 1 and 2 denote the values before and after the change, respectively.

Since $x = X/(X + Y + Z)$, $y = Y/(X + Y + Z)$ and $z = Z/(X + Y + Z)$, we can rewrite equation 8.1 as

$$\begin{aligned} \frac{X_1}{c_1} &= \frac{X_2}{c_2} \\ \frac{Y_1}{c_1} &= \frac{Y_2}{c_2} \\ \frac{Z_1}{c_1} &= \frac{Z_2}{c_2} \end{aligned} \tag{8.2}$$

with $c_1 = X_1 + Y_1 + Z_1$ and $c_2 = X_2 + Y_2 + Z_2$. Combining the equations in 8.2 gives

$$\begin{aligned} X_2 &= X_1 \frac{Y_2}{Y_1} \\ Z_2 &= Z_1 \frac{Y_2}{Y_1} \end{aligned} \tag{8.3}$$

---

[1]note that this means that the hue is constant in the CIELUV space, but not in the CIELAB space

The following relations are used:

$$X_{rel} = X_1/X_0$$
$$Y_{rel} = Y_1/Y_0$$
$$Z_{rel} = Z_1/Z_0 \tag{8.4}$$
$$Y_{ratio} = Y_2/Y_1$$

with (X_0, Y_0, Z_0) the tristimulus values of reference white. In approximation, a_1^* and a_2^* are given by

$$a_1^* = 500(X_{rel}^{1/3} - Y_{rel}^{1/3})$$
$$a_2^* = 500(X_{rel}^{1/3}Y_{ratio}^{1/3} - Y_{rel}^{1/3}Y_{ratio}^{1/3}) \tag{8.5}$$
$$a_2^* = Y_{ratio}^{1/3}a_1^*$$

and b_1^* and b_2^* by

$$b_1^* = 200(Y_{rel}^{1/3} - Z_{rel}^{1/3})$$
$$b_2^* = 200(Y_{rel}^{1/3}Y_{ratio}^{1/3} - Z_{rel}^{1/3}Y_{ratio}^{1/3}) \tag{8.6}$$
$$b_2^* = Y_{ratio}^{1/3}b_1^*$$

This approximation holds for tristimulus values that are sufficiently large relative to reference white (which is true for almost all points).

With these equations chroma (C^*) is given by

$$C_1^* = \sqrt{a_1^{*2} + b_1^{*2}}$$
$$C_2^* = \sqrt{a_2^{*2} + b_2^{*2}}$$
$$C_2^* = \sqrt{Y_{ratio}^{2/3}(a_1^{*2} + b_1^{*2})} \tag{8.7}$$
$$C_2^* = Y_{ratio}^{1/3}C_1^*,$$

and hue (h^*) by

$$h_1^* = \arctan(b_1^*/a_1^*)$$
$$h_2^* = \arctan(b_2^*/a_2^*)$$
$$h_2^* = \arctan(Y_{ratio}^{1/3}b_1^*/Y_{ratio}^{1/3}a_1^*) \tag{8.8}$$
$$h_2^* = h_1^*$$

$$\tag{8.9}$$

So, if $Y_2 > Y_1$, Y_{ratio} is larger than 1 and $C_2^* > C_1^*$.

The hue h^* of the point does not change for any change of Y.

Now we look at the effect a gamma transformation has when applied to the luminance Y. The transformation is given by

$$Y_2 = k_1 Y_1^\gamma - k_2 \tag{8.10}$$

in which k_1 and k_2 are constants which are used to keep the maximum and minimum luminance of the image constant.

From this follows that

$$Y_{ratio} = \frac{Y_2}{Y_1} = \frac{k_1 Y_1^\gamma - k_2}{Y_1} = k_1 Y_1^{\gamma-1} - k_2/Y_1$$

$$C_2^* = (k_1 Y_1^{(\gamma-1)} - k_2/Y_1)^{1/3} C_1^* \tag{8.11}$$

We interpret the effect of the gamma transformation on the chroma as a new scaling factor, $(k_1 Y_1^{(\gamma-1)} - k_2/Y_1)$. This scaling factor depends, besides on γ, also on the luminance of the original point Y_1. For gammas larger than 1, this scaling factor is smaller than 1 for all luminance values (this can be seen by the fact that Y_2 is always smaller than Y_1 for a gamma larger than 1). For gammas smaller than 1 this scaling factor is larger than 1 for all luminance values. So, the larger the gamma, the smaller the chroma of each point. Note that for $\gamma = 1$, $k_1 = 1$ and $k_2 = 0$, so the scaling factor is 1.

8.3   Method

8.3.1   Test material

The stimulus set consisted of four different images (figure 7.2 a-d). These images are ISO standard images (from the cdrom ISO 12640:1997), `bicycle`, `basket`, `cafe` and `portrait`. Because of practical limits to the maximum length of the experiments, we used four images instead of the five used in the previous chapter.

We started with the same conversions as in subsection 7.2.1, that is, the images were converted from CMYK to XYZ using the methods described in subsection 4.1.5. The white point of the resulting images was the white of the paper used.

Next, we changed the chroma C^* (in CIELAB) of each point by

$$C_{new}^* = C_{old}^* s_{chroma}, \tag{8.12}$$

where s_{chroma} is the chroma scaling factor. In the previous chapter we used a chroma scaling factor of 0.85 for all images.

The second step, the gamma transformation, was applied by using equation 8.10. Note that the final chroma value is

$$C_{final}^* = (k_1 Y_{in}^{\gamma-1} - k_2 Y_{in}^{-1}) s_{chroma} C_{old}^* \qquad (8.13)$$

Just before the last step, the transformation from XYZ to CMY, an orthogonal gamut clipping (see section 4.3) was done in which the out-of-gamut colors were mapped onto the gamut. For a high chroma scaling factor more points need to be clipped than for a low chroma scaling factor. Note that with this gamut mapping the hue of the point remains the same.

After the conversion from XYZ to CMY using the methods described in subsection 4.1.5, The color images were sent to the printer as a postscript file with CMY values. The dithering (dot percentages and placing of the dots), as well as the Grey Component Replacement, was done by the printer.

In total 16 different output images were made from each input image, with γ varying from 0.9 to 1.2, in steps of 0.1, and chroma varying from 0.7 to 1.15 in steps of 0.15. We show the effect of the combined gamma manipulation and chroma scaling in figure D.2 for a part of the `bicycle` image. This input image is taken because it contains small structures and because it contains memory colors, that is, colors that are recalled in association with familiar objects in long-term memory [Bartleson, 1968]. All images were printed on an EPSON Stylus color 1520 printer. The images were printed on plain paper (EPSON photo quality ink jet paper).

8.3.2   Experimental setup

The transformed images were used for two experiments, respectively dealing with quality and naturalness

Experiment A: Quality In the first experiment the perceptual quality for a particular gamma was tested. The set of test images consisted of 16 images. In this experiment two rank orderings of the images were determined in a number of consecutive steps, see figure 8.1.

Experiment B: Naturalness In the second experiment the subjective naturalness for a particular gamma was tested. The setup of the experiment was the same as the setup for the quality experiment.

8.3.3   Viewing Conditions

The experimental environment was the same as that used in the previous chapter. The prints were presented in a Macbeth SpectraLight II light booth, with homogeneous lighting. The luminance reflected from the print was at most 450 cd/m^2

16 input images

4 random groups

4 new groups:
the best images
the second best,
and so on

The two best images
and the two worst
are already set, the
other images are
again ranked

The first
final ordering

4 new groups
composed from
the first final
ordering

The two best images
and the two worst
are already set, the
other images are
again ranked

The second
final ordering
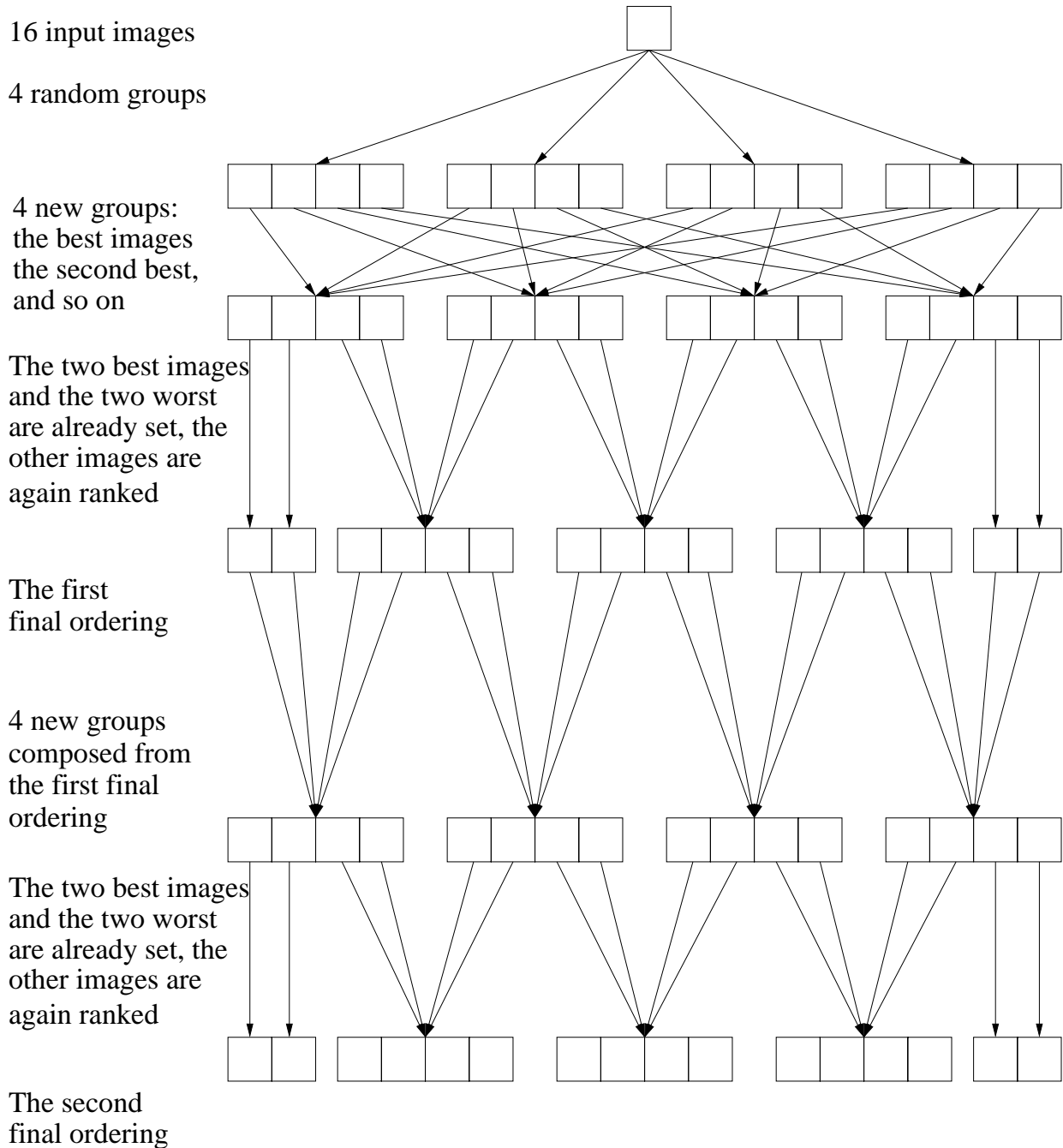
Figure 8.1: The consecutive steps to obtain two rank orderings. The different sets that have to be ranked are formed with images from sets above.

(white paper). The color temperature of the illuminant was approximately 6430 K. The prints were put into transparent plastic covers to prevent smudges. A window was cut in the center to see the image and a white border on the print directly and not through plastic. Because the covers were matte and absorbed some light, the reference white was provided by the paper used.

8.3.4 Subjects

Eight subjects, all completely naive with respect to image analysis, took part in the experiments. The subjects were undergraduate students, paid to do these experiments. Five of the subjects were male and three female. They were 17 to 23 years old. The subjects had normal color vision and normal or corrected-to-normal visual acuity.

8.4 Results

We tested the homogeneity of the group of subjects using the Friedman two-way analysis of variance and Kendall Coefficient of Concordance. The results of these tests are given in appendix G. It was found that for all experiments the rank ordering was not random, indicating that the subjects can be seen as one group.

The rank ordering of the quality and naturalness responses (averaged over subjects) is given in figure 8.2.

In this figure we can see two striking features. The first is that the quality and naturalness responses differ more for different gamma settings than for different chroma scaling factors. This can best seen in the `portrait` results, which are almost vertical. We can also see directly that the quality and naturalness responses vary much between the different images. For the `portrait` image, the best image is one with a high gamma, whereas for the `basket` image, the best image is found for gamma equals 1. Comparing the responses for naturalness and quality, we see that for the different images the quality response is different from the naturalness response. If we ignore the chroma effect, the most natural image can be found for gamma is 1.0 (`cafe`, `basket` and `bicycle`) or 1.1 (`portrait`). The most unnatural image within the range is either the image with the least color (chroma = 0.7, gamma = 1.2) or the image with the most color (chroma = 1.15, gamma = 0.7).

We can compare these results to the results from the previous chapter by comparing the line for which the chroma scaling factor is 0.85 to the results found earlier. These results are given in table 8.1. Most of these values are numerically different (up to 20% difference). Since the variance introduced by having different subject groups in both experiments cannot be estimated, it is not possible to test the statistical significance of these differences. Given the magnitude of the differences, it seems likely that these are caused by normal differences between subject groups.
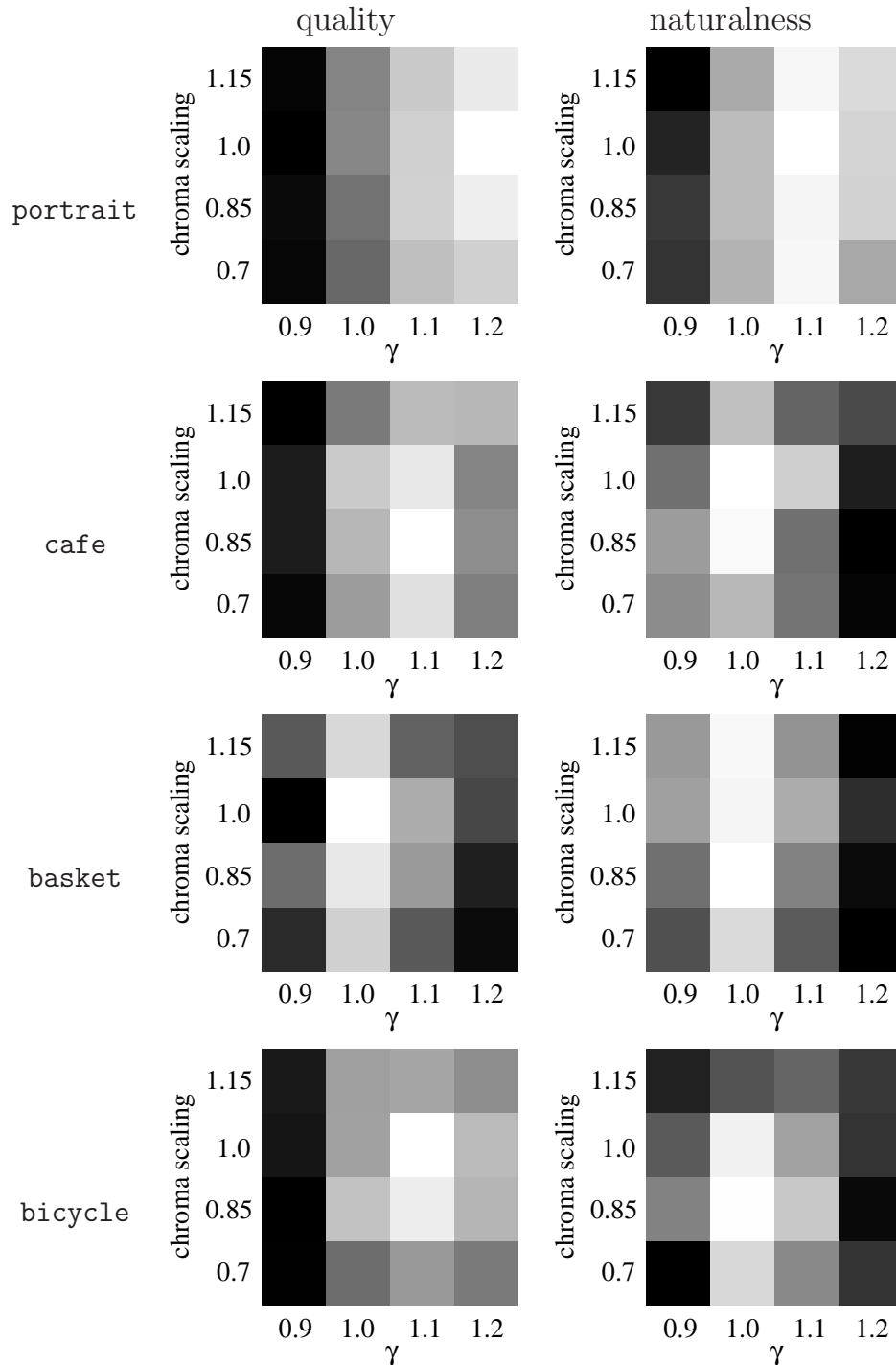
Figure 8.2: Representation of the subject responses for the quality and naturalness experiments. In these graphs the average quality or naturalness responses for 16 images are given against the gamma and chroma scaling factor. The whiter the point, the higher the preference is.

Table 8.1: Results from the gamma experiment compared with the results from the gamma chroma experiment. The values given are the images with highest quality or highest naturalness. For the gamma and chroma experiment only the images with a chroma scaling factor of 0.85 are included.

	gamma experiment		gamma-chroma experiment	
	quality	naturalness	quality	naturalness
portrait	1.0	1.0	1.2	1.1
cafe	0.8-1.0	0.9-1.0	1.1	1.0
basket	1.1-1.2	1.2	1.0	1.0
bicycle	1.3	1.2	1.1	1.0

8.4.1 Statistical analysis

Using an Analysis of Variance (cf. section 4.4.3), we investigated whether the different gamma and chroma settings had a significant effect on the responses of the subjects. We tested the responses of the subjects for the following variables:

Image: the input image

Gamma: the gamma of the manipulated image

Chroma scaling: the chroma scaling factor of the manipulated image

Q/N: whether quality (Q) or naturalness (N) was the attribute the subject ranked

Note the differences between the variables. Image, Gamma and Chroma scaling are physical variables of the stimuli. Q/N is a result of the interaction with the observer, who is asked to produce results reflecting either the quality or the naturalness of the image. The resulting ranking is a complex interaction between all these variables.

Since the subjects ranked images that only differed in gamma and/or chroma scaling, we could not test the main effects for the image or quality/naturalness (see also page 82). However, we *can* find interactions between these variables and gamma and/or chroma scaling. Table 8.2 shows the results of the ANOVA.

There is a main effect for gamma and chroma scaling for both the quality and naturalness responses. These main effects can also be seen in figure 8.3, which shows more or less pronounced optima for the gamma and chroma settings.

The interaction between gamma and chroma scaling is also significant, meaning that the way that different gamma settings influence the perceived quality or naturalness depends on the chroma scaling setting and vice versa. Because there is a colorimetric relation between gamma manipulation of Y and chroma scaling (see section 8.2) this was expected. One can imagine that a gamma manipulation of Y on an image with a low chroma scaling has a different effect on both quality and naturalness than a gamma manipulation on an image with a high chroma scaling. This effect

Table 8.2: The results of ANOVA for different groups of responses. F is the F-ratio, p the probability. Effects that are *not* significant are given *in italic.* The interaction between gamma and image is significant for both quality and naturalness, the interaction between chroma scaling and image is not significant for either quality and naturalness.

	effect	F	p
Quality			
	Gamma	**16.28**	**0.00**
	Chroma scaling	**4.19**	**0.01**
	Image * Gamma	**7.88**	**0.00**
	Image * Chroma scaling	*0.76*	*0.65*
	Gamma * Chroma scaling	**2.16**	**0.03**
	Image * Gamma * Chroma scaling	*1.38*	*0.10*
Naturalness			
	Gamma	**8.45**	**0.00**
	Chroma scaling	**3.71**	**0.02**
	Image * Gamma	**11.04**	**0.00**
	Image * Chroma scaling	*1.55*	*0.14*
	Gamma * Chroma scaling	**3.81**	**0.00**
	Image * Gamma * Chroma scaling	**2.42**	**0.00**
All			
	Gamma	**9.70**	**0.00**
	Chroma scaling	**4.49**	**0.01**
	Q/N * Gamma	**21.90**	**0.00**
	Q/N * Chroma scaling	*1.08*	*0.37*
	Image * Gamma	**10.19**	**0.00**
	Image * Chroma scaling	*1.18*	*0.31*
	Gamma * Chroma scaling	**3.65**	**0.00**
	Q/N * Image * Gamma	**3.86**	**0.00**
	Q/N * Image * Chroma scaling	*1.25*	*0.27*
	Q/N * Gamma * Chroma scaling	*1.75*	*0.08*
	Image * Gamma * Chroma scaling	*1.39*	*0.09*
	Q/N * Image * Gamma * Chroma scaling	**2.47**	**0.00**

can be seen best for the responses for the naturalness responses for the `cafe` image in figure 8.2d.

There is a significant interaction between image and gamma, but the interaction between the image and chroma scaling is not significant. This is also reflected in the results shown in figure 8.3. It can be seen that if only the gamma and the image differ, there are four distinct curves, whereas if only the chroma scaling and the image differ, the curves are almost flat and non-distinct. The fact that the interaction between image and chroma scaling is not significant can be explained

(a) gamma quality   (b) gamma naturalness

(c) chroma quality   (d) chroma naturalness
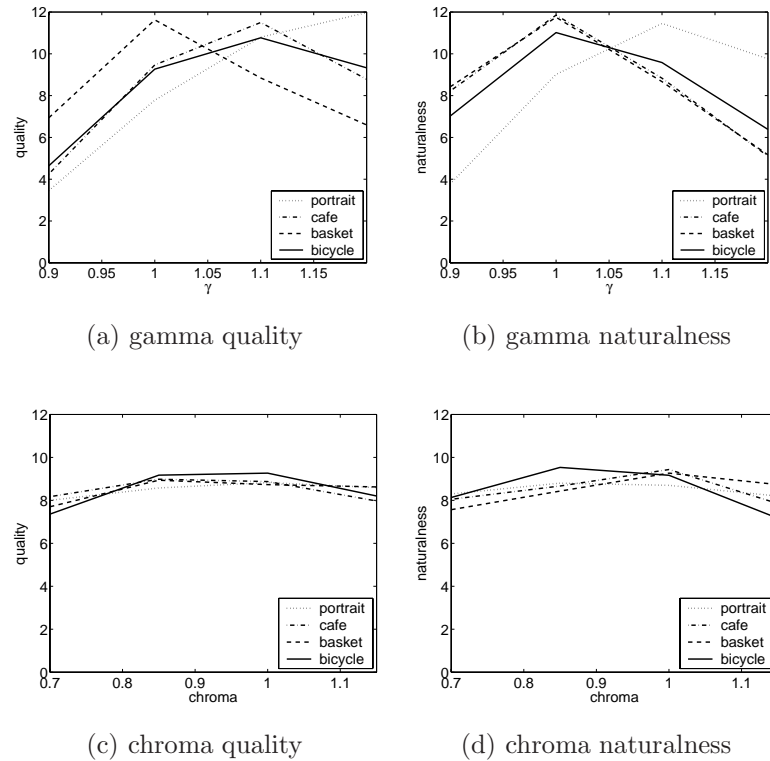
Figure 8.3: The subject responses for the quality and naturalness experiments, averaged over chroma scaling (upper row) and gamma (lower row). Note that the optima for quality and naturalness are more pronounced for gamma than for chroma.

by the fact that the responses for the different chroma scaling values are almost constant.

The fact that the effect of chroma scaling on the quality responses is so small (but significant!) can be explained in the following way. Differences between images with different chroma scaling factors may be relatively small, so that small deviations between image quality responses cannot be measured. However, the differences between the images could be easily seen. Another explanation could be that the stimulus transformation is inherently different for the gamma manipulation and the chroma scaling. The gamma manipulation of Y is non-linear with fixed minimum and maximum luminance. The chroma manipulation is linear, with only a fixed minimum. Whether these differences really explain the results can only be tested by experiments in which the chroma and luminance are treated the same. Still another explanation could be that in natural environments humans are used to evaluate images that differ (linearly?) in chroma, whereas the need for a global gamma manipulation does not occur in natural images.

The three-way interaction between image, gamma and chroma scaling is significant for the naturalness experiment, but not for the quality experiment. This interaction implies that the way that the naturalness response is influenced by the gamma and

the chroma scaling setting, is influenced by the image that is evaluated. This interaction is probably due to the `cafe` responses (see figure 8.2 d), which is very different from the other three curves. The other curves for quality and naturalness depend mainly on gamma, and only for a small part on the chroma scaling.

Evaluating the combined results of both experiments, the following can be noted with respect to the Q/N factor. As expected from the results for quality and naturalness separately, the main effects for gamma and chroma scaling, as well as the interactions between gamma and chroma scaling, and between gamma and image are significant. The interaction between chroma scaling and image is not significant, as was also the case when we looked at the separate results for quality and naturalness.

There is a significant interaction between quality/naturalness and gamma, but not between quality/naturalness and chroma scaling. This is again due to the small differences in the subject responses for images that only differ in chroma scaling. The three-way interaction between quality/naturalness, image and gamma is significant, the three-way interaction between quality/naturalness, image and chroma scaling is not significant.

The four-way interaction between gamma, chroma scaling, image and quality/naturalness is also significant. This four way interaction means that the way that the image influences the response of the subject on different stimuli, is influenced by the fact whether they scale quality or naturalness. According to Lane [Lane, 2003], four-way and higher interactions are "usually difficult to interpret and rarely meaningful".

Searching for a relation between quality, naturalness and physical measures

In the rest of this chapter we shall try, as we did in section 7.3.2, to fit a quality model to the results we found, using linear regression. Again we are searching for a relation between quality, naturalness and physical measures in the image.

The quality model we used is the usefulness-naturalness model (see subsection 3.4.3 on page 37). As in section 7.3.2 we used the subjectively measured naturalness and the discriminability measure as defined in equation 3.34. The color space used in the calculation of the discriminability is indicated by the subscript.

To test the relation between quality on the one hand and naturalness and discriminability on the other, we used linear regression. We found, confirming the usefulness-naturalness model, that the naturalness of the image affects its quality. The explained variance is 35%.

The first discriminability measure we tested was D_{L^*}, the discriminability in the

Table 8.3: The results of linear regression for the images individually.

image	variables	explained variance	naturalness coefficient	D^*_L coefficient
all	naturalness & D^*_L	0.39	0.646	0.224
	naturalness	0.36	0.603	
	D^*_L	*-0.01*		*0.103*
portrait	naturalness & D^*_L	0.94	0.649	0.444
	naturalness	0.81	0.905	
	D^*_L	0.64		0.817
cafe	naturalness & D^*_L	0.76	0.797	1.03
	naturalness	*-0.01*	*0.230*	
	D^*_L	0.30		0.592
basket	naturalness & D^*_L	0.75	1.09	0.445
	naturalness	0.63	0.812	
	D^*_L	*-0.01*		*-0.24*
bicycle	naturalness & D^*_L	0.63	*0.268*	-0.73
	naturalness	*0.10*	*0.402*	
	D^*_L	0.58		-0.78

lightness dimension. With only this measure none of the variance of the experiment was explained. We also evaluated the following discriminability measure: $D_{L^*a^*b^*}$, $D_{L^*u^*v^*}$, $D_{L^*u'v'}$, $D_{L^*S^*}$ and $D_{L^*C^*}$. None of these discriminability measures have a significant effect on the perceived quality. However, when combining D_{L^*} and naturalness, 44% of the variance is explained. Both D_{L^*} and the naturalness have a significant effect on the perceived image quality. Combining naturalness and one of the other discriminability measures does not improve the explained variance. In figure 8.4, the quality, predicted on the basis of the naturalness and D_{L^*}, is plotted as function of the observed quality. The conclusion is that the observed quality can be partly described by naturalness and lightness discriminability. However, a large part of the variance is not explained. Surprisingly, D_{L^*} outperforms the other discriminability measures, despite the lack of color information in this measure. This is consistent with results reported earlier in this chapter, which show that the influence of gamma on the results is larger than the influence of the chroma scaling factor. This also explains why the explained variance in this experiment is less than the explained variance in the gamma experiment (see chapter 7).

The results of linear regression for each image individually is given in table 8.3. For each image the explained variance is higher than the explained variance over all images. It can be seen that the coefficients for each image are different, explaining that the variance over all images is smaller. It is strange that for the `bicycle` image the coefficient for D^*_L is negative. We do not have an explanation for this as yet.
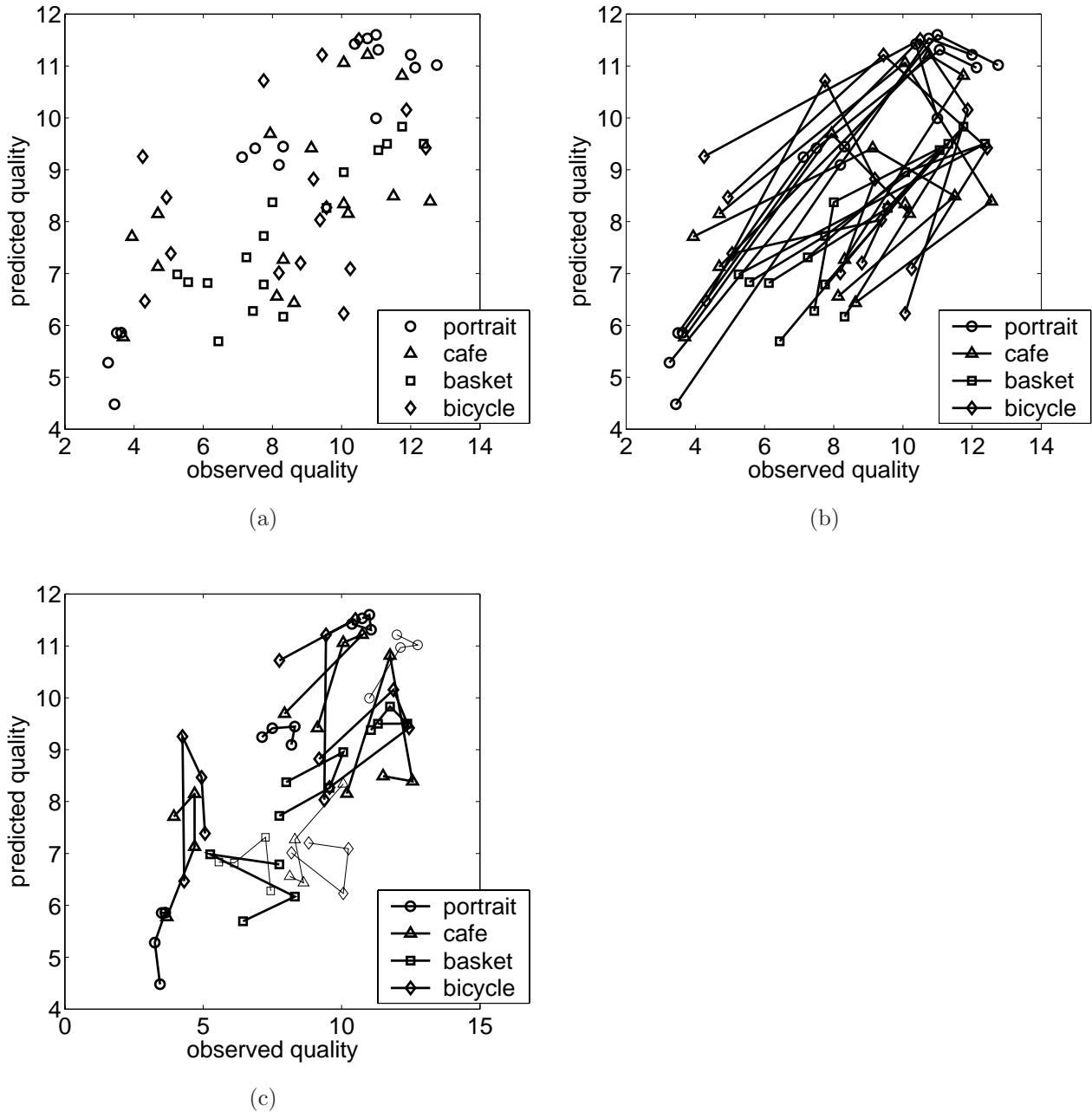
Figure 8.4: The quality predicted in the basis of observed naturalness and D_{L^*}. Predicted quality = 0.646 * observed naturalness + 0.224 * D_{L^*}. In a) no connection between points is made, in b) points with the same chroma scaling factor are connected and in c) points with the same gamma value are connected.

8.5 Conclusions and discussion

In these experiments we evaluated quality perception for images that were scaled in chroma and transformed by a gamma transformation. The statistics of subject responses indicate that the subjects can be seen as one group, similarly responsive

to the experimental stimuli. We found that both gamma and chroma scaling have an effect on the perceptual quality. However, the gamma setting has more influence on the perceptual quality than the chroma scaling.

An explanation for this may be found in intrinsic differences in the stimulus production. The gamma manipulation is non-linear with fixed minimum and maximum luminance, whereas the chroma manipulation is achieved by a linear scaling factor, in combination with a fixed minimum.

Another explanation may be found in the human visual system. Maybe in natural environments humans are used to evaluating images that differ (linearly?) in chroma, whereas gamma manipulation does not occur naturally.

It would be interesting to test if the results would be different if the chroma scaling and gamma were treated in the same way. This is achieved if the luminance and chroma are scaled linearly. A problem with this is that to scale the luminance linearly either the black point or the white point cannot remain the same. Both the chroma and the luminance can also be changed using the gamma manipulation. However, a gamma manipulation of chroma is not normally used in natural images.

A manipulation that does occur in natural images is local contrast stretching of the luminance. In future work it would be interesting to evaluate the influence of local contrast stretching on the quality of images.

With statistical analysis some significant interactions were found. These are an interaction between gamma and chroma scaling, an interaction between gamma and the image and an interaction between gamma and the attribute (quality or naturalness) that the subjects scaled. Higher interactions are also found, but these are hard to explain.

With the usefulness-naturalness model we can explain 44% of the variance in the quality responses, using the naturalness that the subjects scaled and a discriminability measure. In the previous chapter, we could explain 85% of the variance in the quality responses, also using the naturalness and a discriminability measure.

Chapter 9

Lightness filtering in color images with respect to the gamut[1]

9.1 Introduction

For a long time, image processing research has concentrated on grey-value images, but with the advance of digital color reproduction, image processing algorithms for color images are being developed. One difference with grey-value images is that the color of each pixel is defined by three values instead of one. Another difference is that the range of colors, that can be visualized or rendered, is not a simple scale between black and white, but a 3D body in color space, the gamut (see chapter 4).

When image processing operations are performed on color images, it is normal that the production of out-of-gamut pixels is not prevented. Colors that were inside the gamut of the displaying apparatus before the image processing may unwantedly be converted to colors outside this gamut. So, in order to fully render the image, some *gamut mapping* algorithm has to be used to bring out-of-gamut colors into the gamut. Several standard gamut mapping algorithms are discussed in section 4.3. In these algorithms neither the color before processing, nor the kind of processing are taken into account, and thus may reduce the effect of the image processing operation. In this chapter we propose a generic method that allows grey image processing without affecting color rendering, that is, without exceeding the limits of the gamut of the apparatus in question.

---

[1]This research is described in: P.W. Verbeek and J. Dijk, Nederlands octrooiaanvrage 1022258, Werkwijze voor het bewerken van een kleurenbeeld, ten name van Technische Universiteit Delft te Delft.

9.2   The generic method

Image processing of color images typically focuses on hue, chroma or saturation, and lightness. Apart from a global change of hue and chroma, known as white correction, most processing algorithms only deal with the achromatic variables (the grey image). We propose a generic method for converting *grey* image processing into lightness or luminance processing for *color* images. This method assumes that the colors of the original image are located within the gamut of the rendering device. The colors are defined here as points in the CIELAB color space. We stretch the gamut in such a way that it reduces to straight lines.

For a given color point an algorithm for grey-value image processing defines an "ideal" lightness correction. At the same time, however, the gamut defines, for the given hue, a limited "acceptability" area of lightness and chroma. The ideal correction may cause the point to exceed that area. The method proposed here provides a number of "recipes" to perform the lightness correction in such a way that 1) the colors stay inside the gamut (or "acceptability area") and 2) different colors before processing remain different after processing. In this way, the result of the image processing is kept automatically and inconspicuously within the gamut. The hue of the point is always kept constant, because this is the attribute that we can distinguish with the highest precision (the same reason why hue is kept constant in the gamut mapping procedures, see section 4.3, page 64). For some recipes the chroma is also kept constant; for others it is not.

The desired lightness correction causes the original color point to travel along a fixed path. The path is defined by the original point, the gamut boundary and the chosen recipe. Along the path the lightness changes monotonically with position. The intersections of the path with the upper and lower boundary of the gamut define the range between the maximum and minimum lightness correction. This is illustrated in figure 9.1, which shows how a point within a given plane of constant hue (note that the hue should be constant) can be moved within that plane, depending on the path.

Some grey value manipulations, such as sharpening, affect the high frequencies of the grey values but leave the low frequency information constant. For some lightness manipulations, we want to keep this feature. When using sharpening, for instance, we may assume that color changes can only be seen in the low frequencies. The goal of the manipulation is to sharpen the image while trying to keep the color information of the low frequencies the same. To achieve this with the proposed method for the lightness, we must add the constraint that not only the shift for a desired lightness correction will stay in the gamut, but also the shift for the same lightness correction in the opposite direction. For the total image this will lead to equal size lightness corrections in the positive and negative direction. If the overall color should remain the same, a similar constraint for the chroma is needed. Note
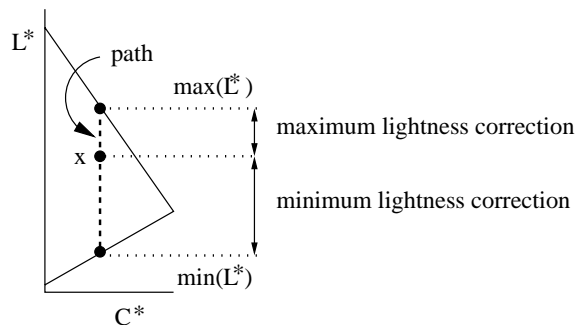
Figure 9.1: Illustration of a path for a given point x and a given recipe (*chroma = constant*) The maximum and minimum lightness correction are determined by the intersections with the gamut boundary.

that the hue of a point always remains the same, as dictated by the model.

Summarizing, we propose a generic method, that can be used to apply grey value algorithms on color images without violating the limits set by the color gamut. Different implementations of the method are given as different recipes.

9.3   Application fields

We identify the following application fields:

Digital photography Digital photography is becoming more and more standard, for both professional and amateur photographers. One of the advantages of digital photography is that the user is able to change the image before the photo is reproduced. This is done with an image processing program, such as PhotoShop or Paintshop Pro. The proposed method can be used to optimize the image quality of the photos. The quality of the result can be determined by the user or by an image quality measure (which is the main theme of this thesis). The method can only be used if the gamut is known (either analytically or numerically). Fortunately, this is often the case because the gamut information is also used for the printer's gamut mapping.

Inkjet or laser printer/copier Most copiers and some printers have an option to lighten or darken the outcome, usually by means of a gamma manipulation. The proposed method can automatically and inconspicuously adapt chroma to this - already standard - darker/lighter choice. The proposed method also introduces new options that allows sharpness and contrast to be used as new parameters, in addition to the darker/lighter option[1]. These variables may improve the performance of the printer/scanner. Because of the increased

---

[1] Note that these parameters were also used in earlier chapters of this thesis

complexity, however, it may be difficult to find the optimal solution. The optimum can be found by visual evaluation or by an image quality measure.

Offset print For the offset printer the proposed method can also automatically and inconspicuously adapt chroma to the - already standard - darker/lighter choice. The method also introduces sharpness and contrast as new parameters. By using these parameters, part of the traditional craftsmanship can be automated, although the optimization of the parameters may be hard.

Color monitor including TV In the field of digital television the method may also be used to optimize images using the parameters darker/lighter, sharpness and contrast. Digital television is a highly advanced and specialized market. Adapting to the gamut of the monitor, however, is relatively simple.

9.4   Application example 1: gamma manipulation

The proposed method can be used to apply a variety of grey value algorithms on color images. In this section we show the results for contrast improvement using gamma manipulation.

9.4.1   Gamma manipulation

The effect of a gamma manipulation is that the lightness values are distributed nonlinearly over the range that is used. This may increase the contrast in one or more regions of the lightness range, at the cost of decreasing the contrast in other regions. In figure 9.2 two different forms of gamma manipulations are shown, one operating in the same way over the *whole* lightness range, and one in which the operation is applied separately to the low and the high lightnesses of the image.

What is shown in figure 9.2 is mathematically detailed in the following. The gamma operation shown in figure 9.2 (a) is described by:

$$L^*_{out} = \min(L^*) + (\max(L^*) - \min(L^*)) * \left(\frac{L^* - \min(L^*)}{\max(L^*) - \min(L^*)} \right)^{\gamma}. \qquad (9.1)$$

This is the same gamma manipulation as used in the gamma experiment (chapter 7, equation 7.2), but now applied to lightness instead of luminance. When this manipulation is used for $\gamma > 1$, the higher (lighter) lightness range gains more contrast, at the expense of the contrast of the darker colors. At the same time the mean lightness is decreased (all new colors are darker than the original colors). When $\gamma < 1$, the opposite occurs (more contrast in the darker colors, less contrast in the lighter colors, and mean lightness increases).

(a) (b)
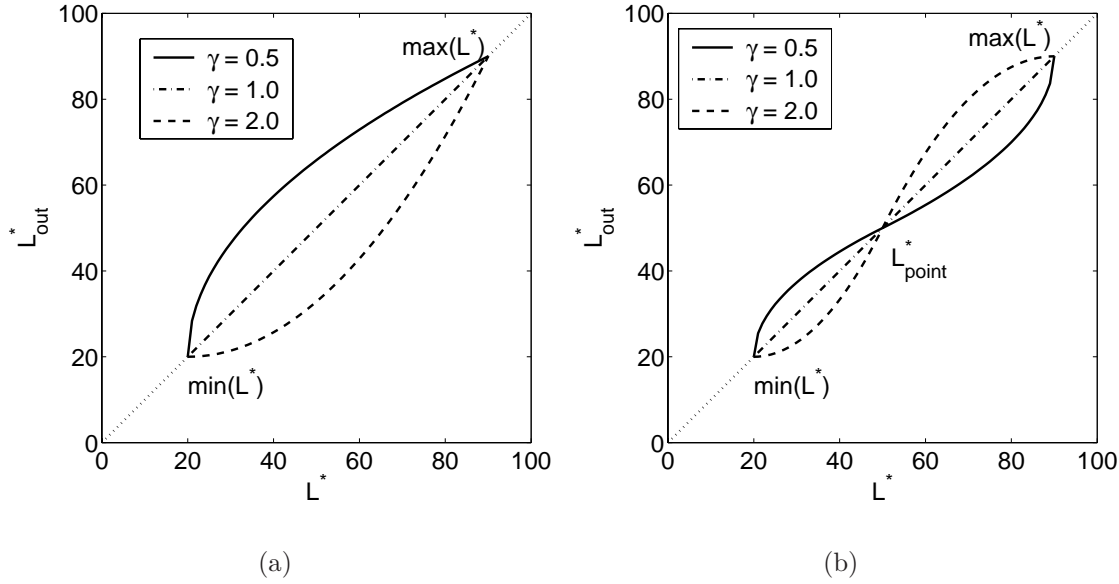
Figure 9.2: Two different forms of gamma manipulation. In a) the result of equation 9.1 is given for three different gamma values. It can be seen that for $\gamma < 1$ the lightness of the image (L^*_{out}) is always higher than for the original image (L^*_{in}), and that darker colors have more contrast. For $\gamma > 1$ the opposite holds true. In b) the result of equation 9.2 is shown for three different gamma values. If $L^*_{point} = L^*_{average}$, the average lightness remains approximately the same. However, the contrast of the colors does change. For $\gamma > 1$ the mid tones gain more contrast and the high and low colors lose contrast. For $\gamma < 1$ the opposite occurs.

The gamma operation shown in figure 9.2 (b) is not applied to the whole lightness range, but to two different parts: the high lightness values and the low lightness values. The gamma manipulation is then defined as:

$$
L^*_{out} = \min(L^*) + (L^*_{point} - \min(L^*)) * \left(\frac{L^* - \min(L^*)}{L^*_{point} - \min(L^*)} \right)^{\gamma}
$$
$$
\text{for } L^* \leq L^*_{point}
$$
$$
L^*_{out} = \max(L^*) + (L^*_{point} - \max(L^*)) * \left(\frac{L^* - \max(L^*)}{L^*_{point} - \max(L^*)} \right)^{\gamma}.
$$
$$
\text{for } L^* > L^*_{point}
$$

(9.2)

Note that for $\gamma < 1$ the high and the low lightness values gain more contrast at the cost of the contrast of the center lightness values (the high lightness values of the low range and the low lightness values of the high range). The mean lightness remains approximately the same, if the original mean lightness is chosen as L^*_{point}. Another interesting choice for L^*_{point} is the lightness of the cusp point for the hue in question (L^*_{cusp}). This is the point in the constant hue plane at which the chroma of the hue is at its maximum (see also figure 9.3 (d)).
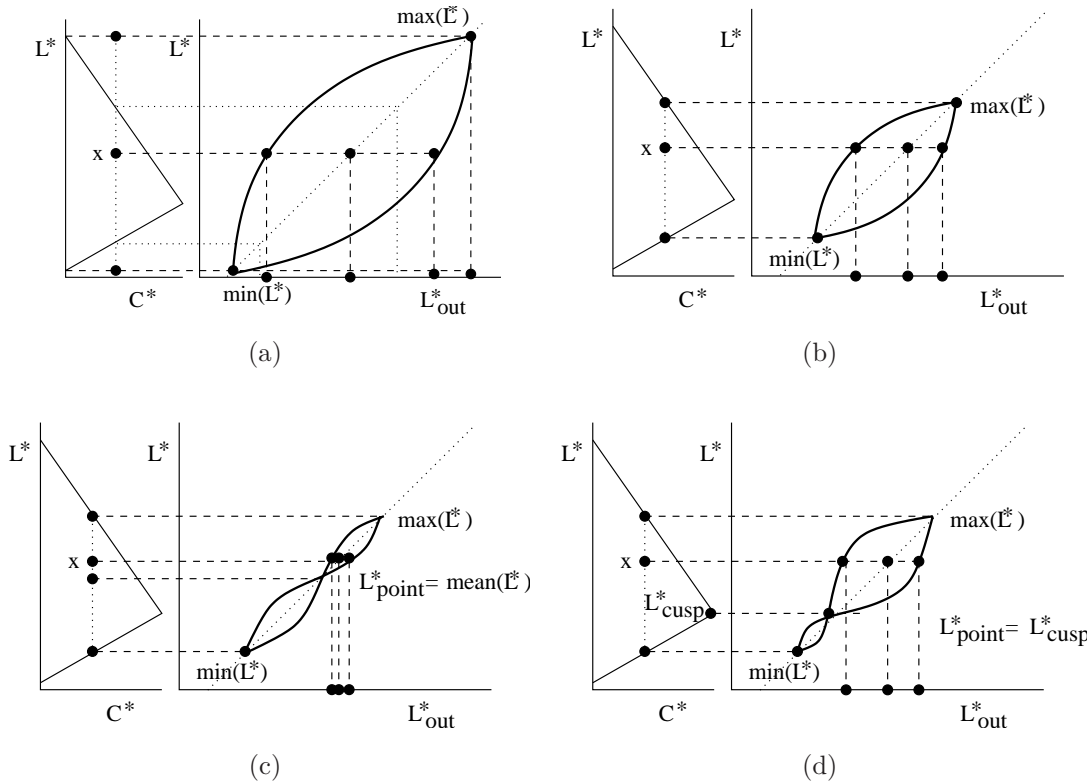
Figure 9.3: Different forms of gamma manipulation, while keeping chroma constant. The original color point is indicated by x. Note that axes in the graphs in the right part of the figure are unusual; the graphs are tilted versions of figure 9.2. In (a) the normal, not gamut-limited, gamma manipulation is shown. It can be seen that one of the resulting points is indeed out-of-gamut. In (b) the gamma manipulation as defined in 9.1 is shown. The intersections with the gamut for constant chroma are the maximum and minimum used in the gamma manipulation. In (c) and (d) the gamma manipulation as defined in 9.2 is shown. The intersections with the gamut for constant chroma are the maximum and minimum. In (c) L^*_{point} is the mean lightness, in (d) $L^*_{point} = L^*_{cusp}$

.

In figure 9.3 is shown how the original color point, x, can travel over the path indicated by the dotted line in the constant hue plane. This path has the property that chroma (C^*) is constant. Other paths are also possible, as will be discussed in the next section. In figure 9.3 (a) the normal, not gamut-limited, gamma manipulation is shown. It can be seen that points can indeed be out-of-gamut. In figure 9.3 (b) the gamma manipulation defined by equation 9.1 is used. In the right hand part of this figure the limits, $\max(L^*)$ and $\min(L^*)$ of the lightness range are shown. These are dictated by the available gamut, so whatever the value of L^*_{in}, it will always be such that the lightness of point x will stay within the available gamut. In figure 9.3 (c) and (d), the same principle is shown, but now for a gamma manipulation according to equation 9.2.

9.4.2   Recipes for gamut-limited gamma manipulation

The gamut-limited manipulation in figure 9.3 applies to the rather simple case of constant chroma. We also discuss other possibilities, all together resulting in

(a) $C^* =$ **constant** The lightness is manipulated while keeping the chroma C^* constant. The maximum and minimum value are the maximum and minimum lightness L^* for this particular C^* value. The effect of this recipe for the two different forms of gamma manipulation has already been shown in figure 9.3.

(b) $C^*/L^* =$ **constant** [2] The lightness is manipulated while keeping the ratio C^*/L^* constant. The minimum value is per definition 0, the maximum value is the lightness value for which the line C^*/L^* intersects the gamut boundary.

(c) Mapping towards black and white The lightness is manipulated in such a way that the point in the chroma/lightness space moves towards black for a lightness decrease and towards white for a lightness increase.

(d) Mapping away from black and white The lightness is manipulated in such a way that the point in the chroma/lightness space moves away from black for a lightness increase and away from white for a lightness decrease. The maximum and minimum lightness are given by the intersections of both lines with the gamut boundary. This recipe is the opposite of the previous recipe.

(e) Mapping over lines from black over the cusp point to white This recipe is a combination of the two previous recipes. For lightness values below the cusp point, the point in the chroma/lightness space moves to black (for a lightness decrease) and away from black (for a lightness increase). For lightness values larger than the cusp point, the point moves away from white for a lightness decrease and to white for a lightness increase.

The different paths in the constant hue plane that result when applying the above recipes are illustrated in figure 9.4.

On these paths the two positions of the color point are indicated that result for $\gamma = 2$ and $\gamma = 0.5$. We would not expect the recipes (d) and (e) to give nice results. *Mapping away from black and white* (d) adds chroma to almost all points, which may introduce colors that are too colorful for the luminance, as if they are self luminant. *Mapping over lines from black over the cusp point to white* (e) introduces a discontinuity at the cusp point. However, it can be instructive to see the results of such manipulations.

---

[2]Note that in the CIELUV space, the attribute saturation is defined as C^*/L^*. In CIELAB, however, this attribute is not defined.

(a) $C^* = $ constant

(b) $C^*/L^* = $ constant

(c) Mapping towards black and white

(d) Mapping away from black and white

(e) Mapping over lines from black over
the cusp point to white
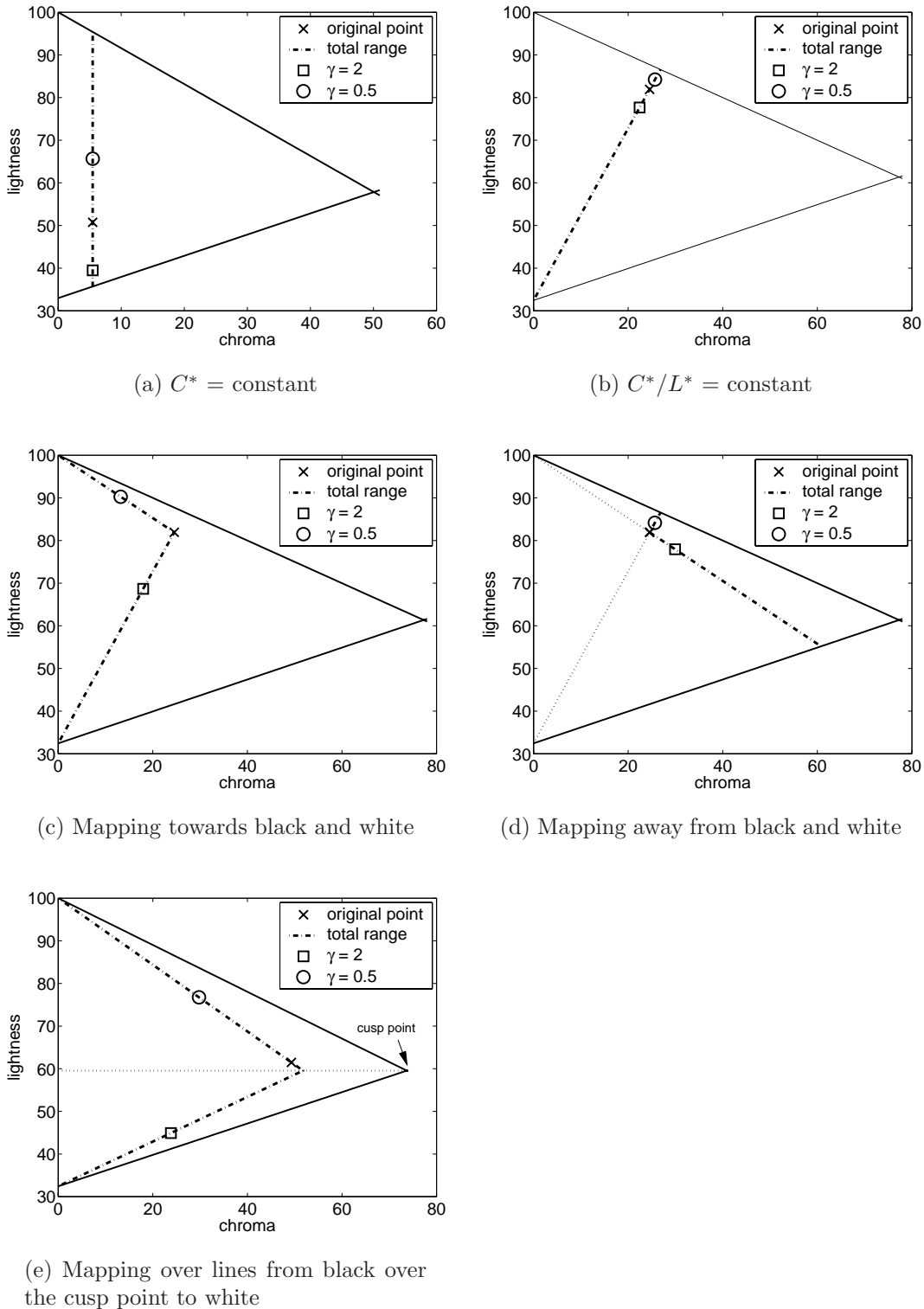
Figure 9.4: Paths along which a color point may move within the constant hue plane, when applying different recipes for gamma manipulations. For each recipe for an original point, x, the range for lightness changes is given, along with the result for $\gamma = 2$ and $\gamma = 0.5$. More details about the paths are given in the text.

9.4.3   Results for gamut-limited gamma manipulation

First gamma manipulation (equation 9.1)

The visual results for the different recipes using the first gamma manipulation are given in figure E.1 - E.6. In each image five different gamma settings are given. The image for which $\gamma = 1$ is for all recipes the same as the original image. The results are shown for a part of the `musicians` image, the blond girl. The input image was subsampled with a factor two because otherwise the results were too large to print on the EPSON stylus 1520 used.

In figure E.1 the results for the normal gamma manipulation are shown, where out-of-gamut pixels are not prevented. Out-of-gamut pixels are mapped onto the gamut with orthogonal clipping. This results, for low gamma values, in a skin color that is too yellow.

In figure E.2 the results for the $C^* = constant$ recipe are shown. It can be seen that the effect of the gamma manipulation is smaller, because the mean lightness change is smaller. The results are better in that here the skin tones are more natural than the results with the normal, not gamut-limited, gamma manipulation (shown in figure E.1).

In figure E.3 C^*/L^* is kept constant, instead of C^*. The effect is that images with $\gamma < 1$ gain chroma, whereas images with $\gamma > 1$ loose chroma. Therefore, the quality of these images is somewhat less than the images in figure E.2.

As could be expected, the results with the *mapping towards black and white* recipe, shown in figure E.4, cause a loss of chroma. All colors are mixed with either black (for $\gamma > 1$) or white (for $\gamma < 1$). The resulting images are less vivid, and therefore the quality is less than that of other recipes. The results for the *mapping away from black and white* recipe (shown in figure E.5), do not differ very much from each other. This is due to the fact that the size of the path in this recipe is much smaller than the paths in other recipes.

The results for the *mapping over the cusp point* are shown in figure E.6. The expected color problems indeed occur, some colors are too colorful. This effect can be best seen in the lips and the background for $\gamma = 0.6$, and for the skin tones for $\gamma = 1.4$.

Second gamma manipulation (equation 9.2)

The results for the different recipes using the second gamma manipulation are given in figure E.7 - E.10. In each image five different gamma settings are given. The image for which $\gamma = 1$ is for all recipes the same as the original image. In these

figures images in the top row are made with $L^*_{point} = L^*_{average}$, whereas the images in the bottom row are made with $L^*_{point} = L^*_{cusp}$. Because the *mapping towards black* and *mapping away from black* recipe did not give nice results for the first gamma manipulation, they are not shown for the second gamma manipulation.

The results with normal gamma manipulation are shown in figure E.7. As the cusp point is a feature of the gamut, which is normally not used, the results are only shown for $L^*_{point} = L^*_{average}$. It can be seen that the image gains contrast for $\gamma > 1$ and looses contrast for $\gamma < 1$.

The results with the $C^* = constant$ recipe are shown in figure E.8, and the results with the $C^*/L^* = constant$ recipe are shown in figure E.9. It can be seen that there are only small differences between these images. The skin tones of the $C^* = constant$ recipe seem to be somewhat more natural than the skin tones of the normal gamma manipulation, but this is only a small effect. The images in the top row and the bottom row are almost the same, indicating that the choice of L^*_{point} is not critical.

The results with the *mapping over the cusp point* recipe are shown in figure E.10. In the images in the top row, for which $L^*_{point} = L^*_{average}$, one can see that some colors are unnatural because they are too colorful. This effect can be seen best in the lips of the woman and the background of the top left image. In the images in the second row, however, these unnatural colors do not occur. This is due to the fact that in the bottom row, L^*_{point} is the lightness of the cusp point, so colors that had a lightness lower than the cusp point still have a lightness that is lower than the cusp point, and vice versa. However, the resulting images have lost some colorfulness compared to the normal gamma manipulation and the $C^* = constant$ recipe.

In conclusion, gamma manipulation with $C^* = constant$ yields the best results. In this case the image does not loose much color, as can be the case with the other recipes (especially *mapping towards black and white*). Keeping C^*/L^* constant also works rather nicely. The differences with gamma manipulation without taking the gamut into account are larger for the first gamma manipulation than for the second gamma manipulation. For the second gamma manipulation, the two choices of L^*_{point} give more or less the same results, except for the *mapping over the cusp point* recipe.

9.5 Application example 2: sharpening

9.5.1 Sharpening

In image processing the use of sharpening can change the lightness of a point rather drastically. If sharpening is applied to the lightness of color images, some of the resulting colors may be out-of-gamut, because the lightness exceeds the (vertical)

limits set by the lightness (and chroma) of the constant hue plane. To avoid this, the lightness correction should depend on the available "gamut space", just as in the case of gamma manipulation. Sharpening is usually achieved by adding a high frequency filtered version of the image to the original image. In these experiments we use real unsharp masking, where the high frequency filtered version is found by subtracting a Gaussian-smoothed image:

$$\tilde{I}_{high} = (1 - \tilde{G}(I, \sigma))\tilde{I} = \tilde{F}_{high}\tilde{I} \tag{9.3}$$

where ˜ denotes the Fourier transform. The real unsharp masking sharpening filter is defined as

$$\tilde{F}_R(\alpha)\tilde{I} = \tilde{I} + \alpha \cdot \tilde{I}_{high} \tag{9.4}$$

Another example of a sharpening filter is Gaussian unsharp masking, which is discussed in section 5.4. We also used inverse unsharp masking to blur the input image. This filter is defined as

$$\tilde{F}_b(\alpha_{inverse})\tilde{I} = \frac{\tilde{I}}{\tilde{I} + \alpha_{inverse} \cdot \tilde{I}_{high}} = \frac{1}{\tilde{F}_R(\alpha_{inverse})}\tilde{I} \tag{9.5}$$

The width of the edges that are amplified by real unsharp masking is determined by σ, and the amount of sharpening, that is, the size of the lightness change, is controlled by the parameter α. In some recipes α can be changed by the user, but in other recipes the amount of sharpening is determined by the maximum and minimum lightness difference and the distance of the point to the gamut. To avoid that the results are largely influenced by outliers, I_{high} is clipped between the 5th and 95th percentile. An undesired effect of unsharp masking is that not only the edges are enhanced, but that also the (high frequency) noise is amplified.

Since the sharpening should only work on high frequency parts of the image, we use the property of images [Nishikawa et al., 1965] that the pixels that make up lines and edges are outnumbered by the pixels in homogeneous areas. Further, we assume that the chroma changes are only seen in the low frequencies, so that we can change the chroma of the line and edge pixels without changing much of the color impression.

9.5.2   Recipes for gamut-limited sharpening

When using gamut-limited sharpening, there are a number of approaches to be considered. We tested the following recipes:

(a) **Sharpening within the gamut** For each point the maximum and minimum sharpening should be within the gamut. If this is not the case, the chroma C^* is reduced to obtain a larger lightness range[3].

---

[3]Note that this does not entirely comply to the described method. There is no fixed path for each point, but the path is moved along the x-axis until the lightness range between the maximum and minimum sharpening fits into the gamut.

(b) Gamut limited sharpening For each point the desired lightness difference is adjusted so that the maximum lightness difference (either up or down) just fits within the gamut, without changing the chroma. This is achieved by adjusting the value of α so that the maximum lightness change fits exactly in the gamut.

(c) Mapping towards black and white The lightness is manipulated in such a way that the point in the chroma/lightness space moves towards black for a lightness decrease and towards white for a lightness increase. In other words, the new point lies on the line through the original point and white ($\Delta L^* > 0$) or black ($\Delta L^* < 0$). Note that this has the effect that the chroma of all points is reduced.

(d) Mapping halfway towards black and white The points are again mapped over lines toward black and white. For each point the lightness difference is scaled in such a way that the maximum lightness difference is halfway to the minimum distance to black or white. If, for instance, the original lightness difference is 10, the distance from the original point to black and white is 40 and 60, respectively, and the maximum lightness difference is 30, then the effective lightness difference is (0.5 * 40/30) * 10. The chroma is adjusted so that the new point lies on the line through the original point and black or white. The size of α is optimized by the recipe.

(e) Mapping away from black and white The new point is located on the line through the original point and black (for $\Delta L^* > 0$) or white ($\Delta L^* < 0$). The maximum lightness difference is the distance to the upper and lower boundary, respectively. All points gain chroma. This is the opposite from the *mapping towards black and white* recipe.

The different paths in the constant hue plane that result when applying the above recipes are illustrated in figure 9.5. On these paths, the original point and the output point for a given lightness correction are indicated. For some recipes the total output range can be defined, for some recipes only the (positive and negative) lightness change for the given point is shown.

9.5.3   Results for gamut-limited sharpening

The results for the different recipes are shown in figure E.11 for the `blond` image. The original image is blurred with inverse unsharp masking to obtain a test image that can be improved. The two images on the top row are given for comparison. These are the test image, and the image which is filtered using "normal" sharpening, where the gamut is not taken into account. If the original image is located in the gamut and if α is equal to $\alpha_{inverse}$, the latter equals the original. Pixels that are out-of-gamut are clipped onto the gamut using orthogonal clipping.
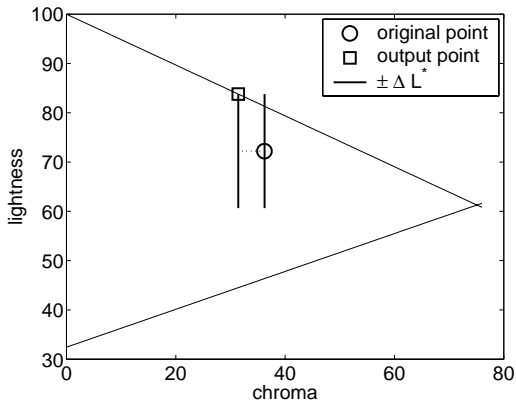
(a) sharpening within the gamut
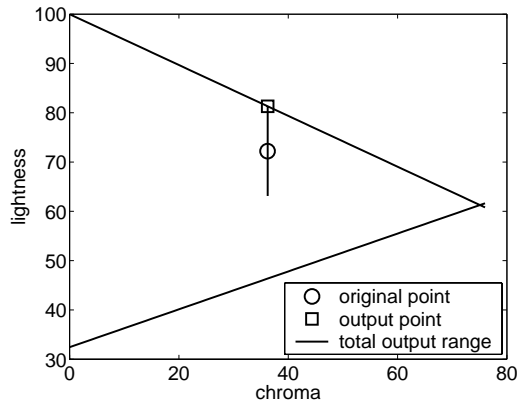
(b) gamut limited sharpening
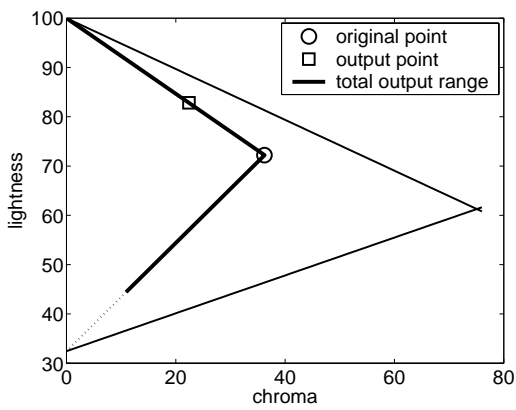
(c) mapping towards black and white
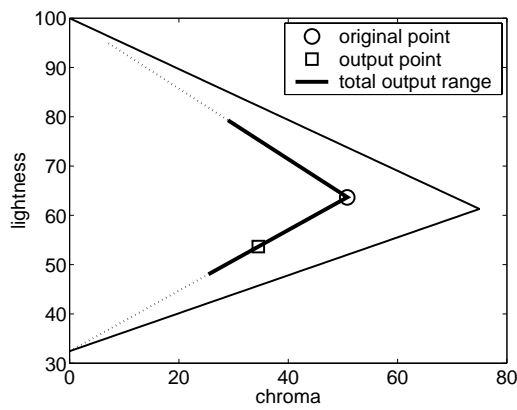
(d) mapping halfway towards black and white
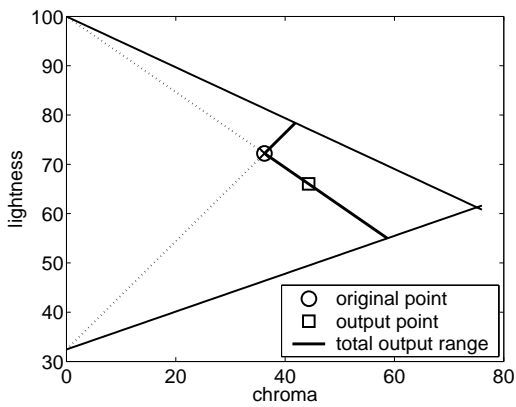
(e) mapping away from black and white

Figure 9.5: Different paths for sharpening manipulations. The triangles are the intersection of the gamut with a constant hue plane. More details about the paths are given in the text.

The rest of figure E.11 shows the images, that are changed using the new method for several recipes. For all recipes the values of σ and α are chosen 1.0 and 0.6, respectively. It can be seen that the result for *gamut limited sharpness* (b) and *mapping halfway towards black and white* (d) are the sharpest. This can for instance be seen on the location of the eyes and the mouth. This effect is probably due to the fact that for these two recipes α is optimized with respect to the gamut instead of set by the user. The difference between the two results is that the image produced by *gamut limited sharpening* (b) is more colorful than the image produced by *mapping halfway towards black and white* (d). This suggests that the chroma change in the low frequency for the *mapping halfway towards black and white* (d) recipe are quite visible, despite the fact that the chroma change is not large for most points.

The *mapping away from black and white* (e) recipe adds chroma on places where the lightness is changed. The colors in the resulting image will therefore be more colorful than the colors in the original image. This can be best seen in figure E.11 in the colors of the mouth of the woman.

The *sharpening within the gamut* (a) and *mapping towards black and white* (c) recipe are comparable to the normal sharpening procedure. Because for these recipes the size of α is adjusted by the user, we need to compare the results for different α's. This is done in figure E.13 for *sharpening within the gamut* (a) and in figure E.14 for *mapping towards black and white* (c). For both recipes the results tend to become somewhat "greyish" for $\alpha > 0.6$. For *sharpening within the gamut* (a) the chroma reduction is due to the fact that the points are shifted towards the achromatic axis until the sharpening fits within the gamut, and for larger α these shifts will be larger. In the *mapping towards black and white* (c) recipe all points are shifted towards black and white, also causing a chroma reduction. In the *mapping away from black and white* (e) recipe (shown in figure E.15 the some areas in the resulting image become to colorful, e.g. the hair becomes too yellow.

In figure E.13 the influence of sigma is shown. It is clear that a sigma of 0.5 is too small for the `blond` image, the main result of the filter is that noise is amplified. For a sigma of 4.0 the lines that are sharpened are too coarse, which can for instance be seen at the upper lip of the woman. The sigma's 1.0 and 2.0 give nice results.

We conclude that the results of the original ("normal") sharpening method can be improved upon by taking the gamut into account. *Gamut limited sharpening* (b) is the best recipe for the `blond` image, because the chroma is not reduced and because the value of α is optimized by the recipe. Other recipes, except may also give somewhat better results than normal sharpening.

9.6 Conclusions and discussion

In this chapter we have proposed a generic method that allows grey image processing for lightness processing on color images without affecting color rendering, that is, staying within the gamut of the apparatus in question.

In the proposed method the lightness correction is a function of

- the particular lightness and chroma of the original point in question.

- the desired lightness change, as implemented by the grey value image processing algorithm (like gamma manipulation or sharpening)

- the choice of recipe to perform the lightness change. For each point, a path is determined, which is specified by the recipe. The lightness and chroma changes depend on the maximum and minimum of this path.

The monotonic character of the lightness correction guarantees that the color relations within the image remain undisrupted. That is, all points that were different before the lightness correction, remain different.

The results for two applications of the method were shown: 1) sharpening and 2) contrast improvement using gamma manipulation.

For gamma manipulation, the lightness of all pixels changes. Therefore, *mapping towards black and white*, where the chroma reduction is proportional to the lightness difference, results in greyish images, whereas *mapping away from black and white*, where the chroma increase is proportional to the lightness difference, results in images that are too colorful. The differences with gamma manipulation without taking the gamut into account are larger for the first gamma manipulation (equation 9.1) than for the second gamma manipulation (equation 9.2). For the second gamma manipulation, the two choices of L^*_{point} give more or less the same results, except for the *mapping over the cusp point* recipe.

For the sharpening method, the lightness correction is small for most points, since most pixels do not belong to lines or edges. Therefore, it is to be expected that the chroma can be changed proportionally to the lightness. However, for all recipes where the chroma is changed, this "small" correction is perceptually quite visible. The *gamut limited method*, for which the chroma does not change, gives the best results for the image used here.

In general, we conclude that for these two applications the gamut-limiting process is only advisable when the limiting is applied to the lightness dimension, rather than the chroma dimension.

Several application fields for this method can be identified, such as digital printers and digital photography. A sharpening procedure that improves most images can

be implemented easily. It might be hard, however, to find automatically the optimal settings for each image.

Chapter 10

Searching for a quality measure for images that are manipulated in the achromatic domain

10.1 Exploring the constant hue plane

The primary goal of this study has been to find a quantitative measure for the perceptual quality of color images. This turned out to be a too ambitious goal, but we nevertheless have come to a point at which it is possible to start speculating on which direction to go. The results of our experiments on gamut-limited gamma manipulation and sharpness enhancement, indicate that it pays off to separately look at the chromatic domain (hue, chroma) and the achromatic variables (lightness, luminance). One conclusion that we reached on the basis of these experiments is that improvements in image quality are best done in such away that hue and chroma are not severely manipulated. Manipulations should be confined to the achromatic domain. This leaves lightness and luminance as the best choice for quality improvement.

In this chapter we propose a measure for the quality of images that are manipulated in lightness or luminance. Such a measure might be used to explain the results that are found in the gamma and gamma/chroma experiments.

The basic assumption is that an image has a higher quality if the colors are optimally distributed over the available gamut space, so as to make all lightness nuances visible. This is also implied in the discriminability measure of the usefulness-naturalness model [Janssen and Blommaert, 2000b]. According to this model, quality also depends on the naturalness of the colors. Maybe the naturalness does not change very much for images that are only manipulated in lightness or luminance, because the hue is always kept constant. We test such a model for the three input images shown in figure 10.1. These images are subparts of the `musicians` image.

(a) `blond`   (b) `oriental`   (c) `dark`

Figure 10.1: The images used to test the model

$h_{center} = 3\pi/5$ $h_{center} = 2\pi/5$

$h_{center} = 4\pi/5$ $h_{center} = \pi/5$

$h_{center} = \pi$ $h_{center} = 0$

$h_{center} = 6\pi/5$ $h_{center} = 9\pi/5$

$h_{center} = 7\pi/5$ $h_{center} = 8\pi/5$
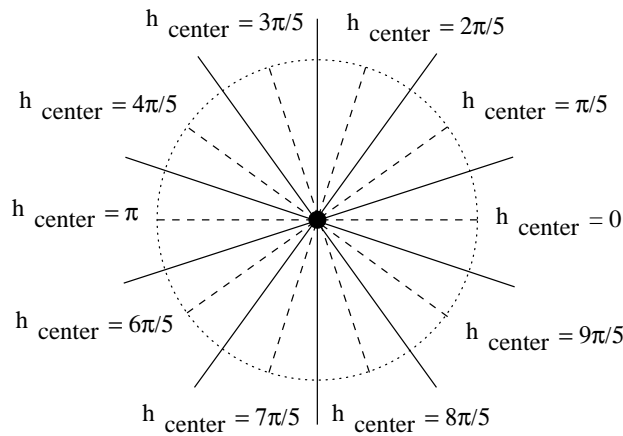
Figure 10.2: The division of the hue circle in ten sectors. For each sector h_{center} is given.

In chapters 4 and 9 we determined the boundary of the gamut in a constant hue plane. In this plane the gamut is a deformed triangle, which resembles a flag. In the remainder of this text we will call this intersection a flag.

We hypothesize that an image has a higher quality if the colors are optimally distributed in the flag (see later figure 10.3). But what is optimal? We proceed as follows in determining the optimal lightness distribution. We determine the location of the center of mass of the color points in the flag. In order to sample sufficient hue points per hue plane, we increase its thickness. We divide the hue circle in ten sectors, so each hue slice spans a hue angle of $\frac{2\pi}{10} = \frac{\pi}{5}$ rad. The central hue plane of the slice we call h_{center}. This is visualized in figure 10.2. Note that for the test images, only the hue slices containing skin and background have a large number of pixels (see also 10.3). The center of mass can be measured along the two axes of the constant hue plane, lightness and chroma. However, we have already decided that we will stick to the achromatic domain, so we only manipulate the position of the center in the vertical direction within the gamut. This position can be defined

as the ratio of the two distances that separate the center from the upper and lower boundary of the gamut, respectively. If we call the smaller one of these distances d_1 and the other d_2, we can define this ratio as

$$q = d_1/d_2 \qquad\qquad (10.1)$$

Note that for a point located in the (vertical) center of the gamut, $d_1 = d_2$. In that case, $q = 1$, which we consider as optimal because the lightness nuances surrounding that point can be stretched over equal ranges of lighter and darker tones. Note that this is also the idea behind contrast stretching.

10.2 Implementation

Here we go into further detail about the implementation of the method.

- First, we determine the distribution of pixels within each of the ten hue slices. Note that the h_{center}'s are spaced apart according to $n\pi/5$, with n an integer.

- The lightness/chroma values of the pixels that are located in a certain hue slice can be projected on the plane of the center, and as such are plotted in figure 10.3. We also plot the gamut intersection for h_{center}. Note that most pixels have a different hue, and that some of them may be located outside the h_{center} gamut. The latter pixels are the result of slight differences in the gamuts of the hue planes adjacent to h_{center}.

- The center of mass of the pixels in the chroma direction c_{mean} is determined for each group. Then the center of mass in the lightness direction, l_{mean}, is determined. To avoid that this value depends too much on achromatic colors, only the pixels that are in the column $\{c_{mean} - \Delta c, c_{mean} + \Delta c\}$ are used to determine this value. The center of mass of the pixels thus selected is determined. From this center the distances d_1 and d_2 are measured and used for calculating the quality measure q. The gamut boundary used is the gamut boundary for h_{center}.

- The images that comprise the pixels that are located within each hue slice can be visualized. An example is given in figure 10.4. With the settings used (10 slices, $h_{center} = n\pi/5$, the skin tones are mainly located in the second slice (h_{center} is 0.63). This holds for all three skin types (white, oriental, dark).

10.3 Preliminary results and evaluation

In a first evaluation of the results we address the values of q for the skin tones, which are all located in the second slice (h_{center} is 0.63, figure 10.4). The fractions q for this
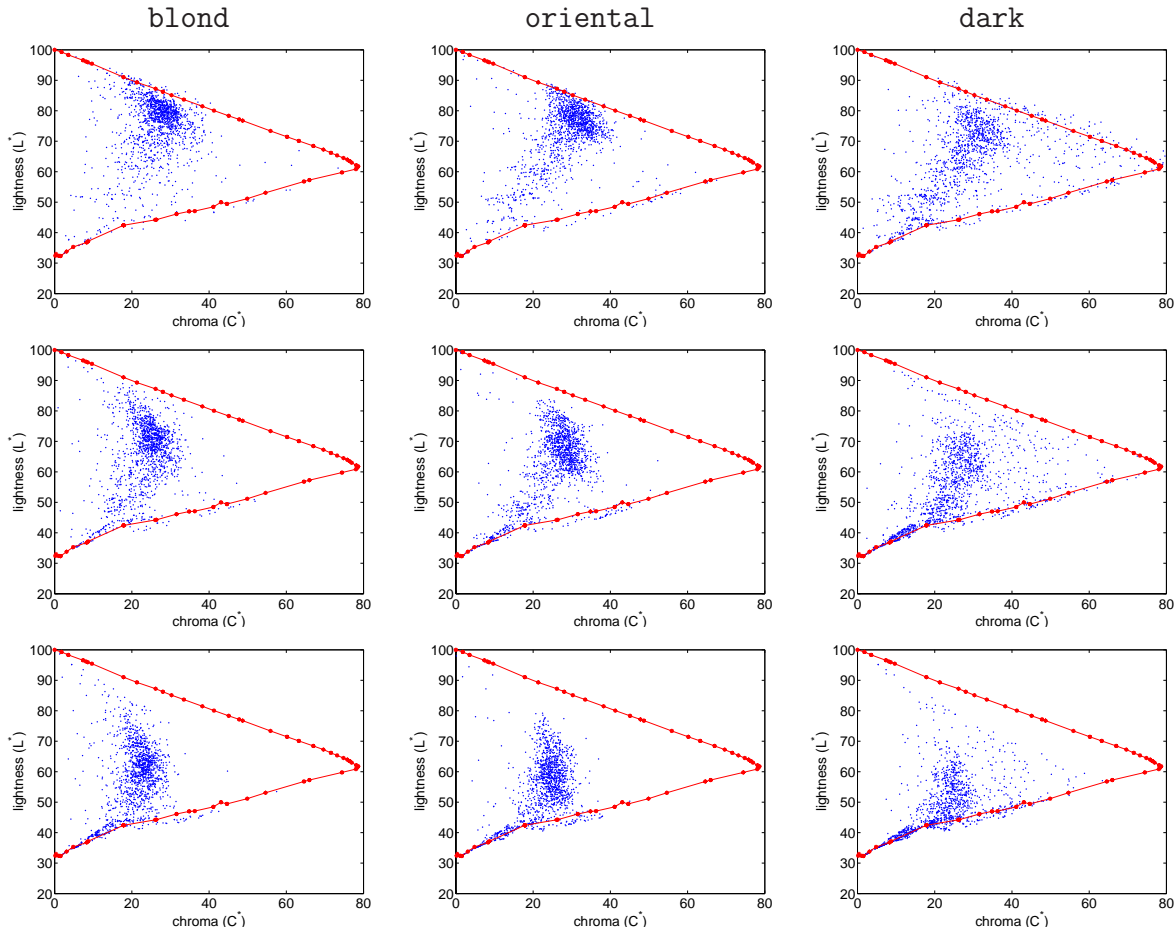
Figure 10.3: The pixels located in the slice with $h_{center} = 0.63$ for gamma manipulated images. The images are `blond` (left column), `oriental` (center column) and `dark` (right column). The gamma of the images is 0.6 (top row), 1.0 (center row) and 1.4 (bottom row).

slice are shown in table 10.1 for three different sizes of the column encompassing the cluster of color points within the flag of h_{center}.

From remarks that subjects gave after the preference ordering experiment in both the gamma and gamma/chroma experiment, we know that the blond and dark woman are important subparts for judging the quality of the total image. The blond woman is rather pale and can do with a little bit more contrast in the lighter parts of the image, as is accomplished with $\gamma > 1$. It is of interest that our quality measure q is consistent with that observation (see bold values of q in table 10.1). The opposite holds true for the dark woman, the face gains more nuances for $\gamma < 1$. Again, this is confirmed by the values of q in table 10.1. The contrast in the face of the oriental woman seems to be just right, so that $\gamma = 1$ would be appropriate. Here too, we found high values of q, when gamma is equal or close to one.

The data shown in table 10.1 are plotted in figure 10.5, which shows in more detail that the three woman portraits reach maximal quality for slightly different

Table 10.1: Values of q as a function of gamma for the three different skin tones located in the hue slice with $h_{center} = 0.63$. The data are taken from the gamma experiments discussed in chapter 7. The maximum values are given in bold.

	blond	oriental	dark	
0.6	0.30	0.29	0.64	
0.7	0.40	0.40	0.87	
0.8	0.51	0.52	**0.88**	
0.9	0.64	0.66	0.67	
1.0	0.80	0.83	0.50	
1.1	**0.97**	**0.95**	0.37	$\Delta c = 2$
1.2	0.86	0.77	0.27	
1.3	0.71	0.61	0.19	
1.4	0.60	0.48	0.15	
1.5	0.51	0.38	0.14	
1.6	0.43	0.30	0.13	
1.7	0.40	0.24	0.13	
	blond	oriental	dark	
0.6	0.32	0.32	0.68	
0.7	0.43	0.44	**0.90**	
0.8	0.55	0.58	0.84	
0.9	0.68	0.74	0.65	
1.0	0.84	**0.94**	0.49	
1.1	**0.98**	0.85	0.38	$\Delta c = 5$
1.2	0.82	0.69	0.30	
1.3	0.69	0.56	0.25	
1.4	0.59	0.46	0.20	
1.5	0.50	0.38	0.16	
1.6	0.43	0.31	0.12	
1.7	0.37	0.25	0.10	
	blond	oriental	dark	
0.6	0.36	0.39	0.81	
0.7	0.47	0.53	**0.94**	
0.8	0.60	0.69	0.73	
0.9	0.75	0.87	0.58	
1.0	**0.92**	**0.92**	0.46	
1.1	0.90	0.74	0.37	$\Delta c = 10$
1.2	0.75	0.60	0.29	
1.3	0.62	0.49	0.23	
1.4	0.52	0.40	0.18	
1.5	0.44	0.33	0.14	
1.6	0.37	0.27	0.11	
1.7	0.32	0.22	0.09	

$$h_{center}$$

$$0 \cdot \frac{2\pi}{10} = 0$$

$$1 \cdot \frac{2\pi}{10} = \frac{\pi}{5}$$

$$2 \cdot \frac{2\pi}{10} = \frac{2\pi}{5}$$

$$3 \cdot \frac{2\pi}{10} = \frac{3\pi}{5}$$

$$4 \cdot \frac{2\pi}{10} = \frac{4\pi}{5}$$

$$5 \cdot \frac{2\pi}{10} = \pi$$

$$6 \cdot \frac{2\pi}{10} = \frac{6\pi}{5}$$

$$7 \cdot \frac{2\pi}{10} = \frac{7\pi}{5}$$

$$8 \cdot \frac{2\pi}{10} = \frac{8\pi}{5}$$

$$9 \cdot \frac{2\pi}{10} = \frac{9\pi}{5}$$
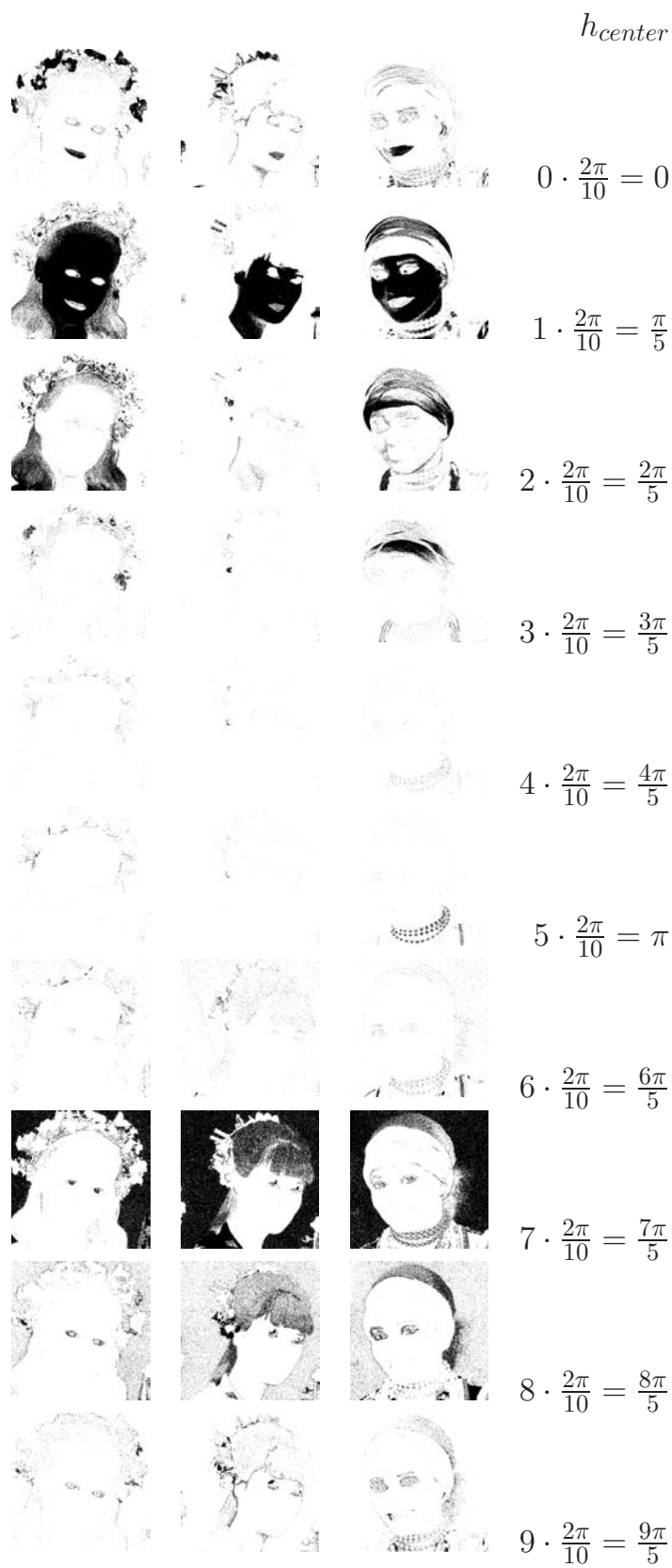
Figure 10.4: The pixels with hues contained within the different hue slices.

(a) $\Delta c = 2$ (b) $\Delta c = 5$ (c) $\Delta c = 10$
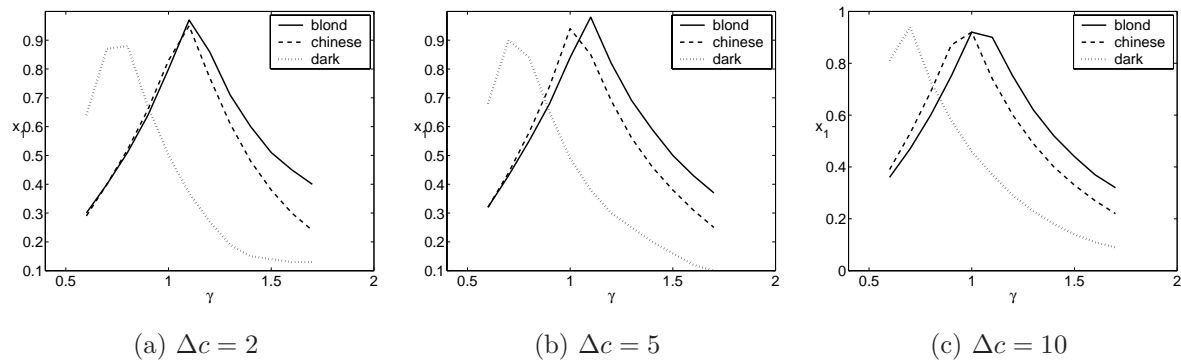
Figure 10.5: The quality measure q as a function of γ, for different sizes of the column width (Δc). The data are from table 10.1

(a) blond   (b) oriental   (c) dark
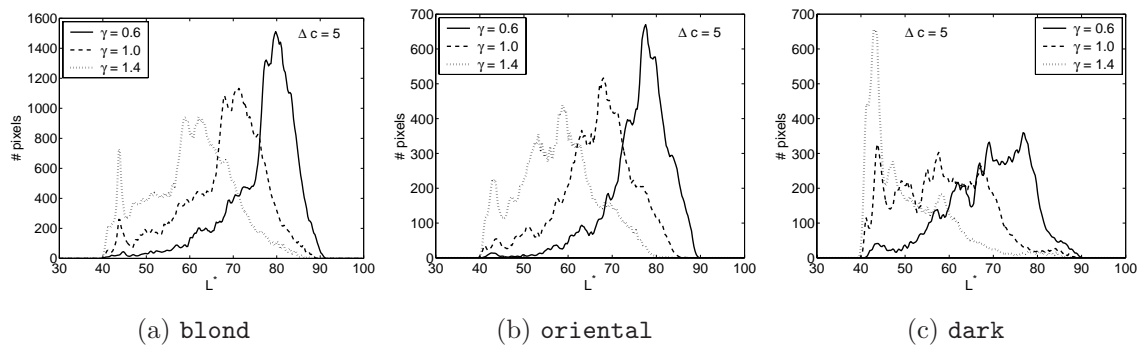
Figure 10.6: The distribution of the lightness in the column around the center of mass in the $h_{center} = 0.63$ hue plane, for $\Delta c = 5$.

values of gamma. These results, although not very extensive, do suggest that q may indeed be used as a quality measure. q is a measure that reflects the degree to which the lightness distribution of the color points within a given hue plane is evenly distributed. We might also study the lightness histograms directly. These histograms are shown in figure 10.6. A good histogram should be the "flattest" histogram (note again the similarity with contrast stretching). Although neither of the histograms is very flat, it is clear that for the blond and oriental woman, the histograms for $\gamma > 0.6$ are relatively flat compared to that for $\gamma = 0.6$, whereas the opposite holds true for the dark woman. So there is at least some indication that q is indeed related to the flatness of the lightness histogram.

10.4 Discussion

In this chapter we proposed the idea that an image, in which the colors fill the available gamut space relatively homogeneously, has a higher quality than image

for which this is not the case. The basis for that idea is that a more homogeneous distribution allows for more discriminable nuances.

In earlier chapters we encountered the problem that subjects based their quality response on certain subparts in an image. For the proposed measure, not only these subparts have to be identified, but also the corresponding colors. For instance, the `musicians` image contains many background colors, but we assume that the distribution of these colors do not influence the quality very much. Both the issue of selecting the subparts in the image and identifying the important colors require further study.

This quality measure (q) seems to apply well to gamma manipulation because, in this manipulation, color distribution (and particularly luminance distributions) are stretched and decompressed. The effects of other manipulations, such as chroma change or hue change, have not been investigated yet.

An interesting aspect of q is that the quality of an image depends on the displaying device. Suppose that the quality of an image is optimized for a certain displaying device, then what happens if the image is displayed on another device, which can display more lighter colors? Clearly this will affect the potential for manipulating gamma, and hence, the quality of the image.

Chapter 11

General conclusions and discussion

This thesis discussed objective quality measures for printed images. More formally, the major goal of this thesis can be defined as:

Finding an objective measure for the quality of a printed image that corresponds to perceptual image quality.

In relation to this major goal a number of secondary goals were defined:

- Improving our insight in the complex subject of perceptual quality.

- Determining possible relationships between perceptual attributes and image quality

- Developing methods for image enhancement in color images.

The research can be broadly divided into three parts. After discussing theoretical and methodological issues, we first discussed subject experiments regarding sharpening and sharpness. We proposed measures for perceptual sharpening and smoothing, and also for perceptual sharpness. In the second part, subject experiments regarding color changes in images are discussed. In these experiments, the achromatic color distribution and the chroma of the images is changed. In the last part, we propose a new, generic method, that allows grey image processing without affecting color rendering, that is, without interfering with the gamut of the apparatus in question.

We have also proposed an idea for a quality measure that depends on the gamut of the apparatus. This quality measure should be applied to images which are changed in the luminance or lightness distribution.

There is a difference between sharpness and gamma manipulation. For sharpening, the *perceptual* attribute is quite well defined [1]. We defined objective measures

---

[1] Although one can question how well sharpness is defined. Bech et.al. investigated the sub attributes of perceptual sharpness [Bech et al., 1996]

for sharpening, smoothing and sharpness. For gamma manipulation, on the other hand, gamma is the objective parameter that is used to produce different stimuli. The perceptual attribute that is changed is something like the amount of contrast, together with the amount of light.

In the following sections, we will discuss the results of this study, following the division into three different parts (described above).

11.1   Sharpening, smoothing and sharpness

The first perceptual attribute studied was sharpness. We started out by studying the effect of sharpening and smoothing filters. These measures depend on the relation between the gradient magnitude of the pixels in the original image and the filtered image, respectively. The resulting sharpening and smoothing measure correlated reasonably well with human perception. Because the subjects had less trouble with different levels of smoothing than with different levels of sharpening, the just noticeable difference of the smoothing measure must be smaller than that of the sharpening measure.

In these experiments, problems arise for images in which parts of the image require different sharpening and/or smoothing values, as is the case for the complex `bicycle` image. This image may be somewhat too artificial and too complex for these experiments. The image is a photograph of a still life, but this scene contains several test charts, which gives the image an artificial look. Because some parts of the image seem more important for the image quality than others, the subjects based their overall decision on the evaluation of different parts of the image. This problem may be solved by an approach in which the measure is based on the results of subparts of the image. A complicating factor for this approach is that it is unclear how to select the different subparts of the image.

The preference of the different images was also tested as a measure for quality. For these images, subjects prefer images in which the smoothing is low and the sharpening is high. However, this may not apply to noisy images, as smoothing is often used (and appreciated) on such images to reduce the noise. We did not perform experiments in which noise is added to the images, because heavy noise does not occur often in printed images.

We also studied the sharpness of images, as a feature of the image itself, rather than in relation to some original image. We proposed a sharpness measure, for which lines and edges are modeled as Gaussian profiles with different widths. Lines and edges in the image are located, and their sharpness is determined by fitting the Gaussian line or edge profile to the Gaussian derivative signature. The sharpness measures studied depended on the widths that were found. A topic for further

research is to study the distribution of the widths and amplitudes to see if other measures that correlate to the sharpness of images can be defined.

All experiments were based on three different filters: Gaussian smoothing filters, unsharp masking filters and anisotropic diffusion filters. For both the sharpening and smoothing and the sharpness measures, we would expect that the results would also apply for certain other filters. However, these filters should alter the image in such a way, that the result is more or less "natural" to an observer. We think that for filters for which this is not the case, such as morphological filters, the proposed sharpening and smoothing measure cannot be used.

The sharpness measures are based on Gaussian models of the lines and edges. The Gaussian smoothing filters fully comply with this model, the Gaussian unsharp masking filters somewhat, and the anisotropic filters only partly. Because the anisotropic filters only change parts in the image where there is not much structure, this filter probably does not influence the sharpness measure. However, since we used the anisotropic filter in combination with the Gaussian unsharp masking filter, we could vary the sharpness of the images.

Concluding, the proposed measures for sharpening, smoothing and sharpness correlate to perceptual sharpening, smoothing and sharpness. This means that these measures can be used to obtain an objective prediction of the attribute in question.

11.2 Color changes: gamma manipulation of luminance and chroma scaling

The second perceptual attribute we studied was color, in particular the effect of two different color distribution manipulations of the image. The first manipulation was a non-linear manipulation of the luminance values using a gamma transformation. The chromaticity of the colors was kept constant. The second manipulation was a scaling of the chroma of the points, which influences the perception of the colorfulness.

The relation between the physical gamma (the stimulus) and the perceptual response ("perceived gamma") was found to be consistent with a stimulus-response function which is more or less linear at the lower gamma range, and has a decreasing slope towards the higher gamma values. Using the well known relation between luminance and lightness (which is for instance used to determine the lightness in CIELAB and CIELUV), this indicates that the subjects evaluate the local contrast rather than the mean luminance, when they are asked to determine which image is perceptually halfway between the image with maximum and minimum gamma. We have not tested this hypothesis.

We also studied the perceptual quality and naturalness of these images. In contrast to the literature [Roufs, 1989], we found for some images distinct differences between the subjects responses. So, depending on the type of image, individual preferences may differ, even when explicitly looking for a uniform perceptual quality.

For all images a gamma was found for which the image quality was highest and a gamma for which the image naturalness was highest. The value of this optimum gamma differed between the images. For all images either decreasing or increasing gamma with respect to the optimum gamma had the same (negative) effect on perceived quality.

We found that the quality and naturalness of all images depended on gamma. The dependency on gamma is influenced by the image that is evaluated. This could be due to the fact that the images are very different in lightness distribution. The content was also very different, some images contained many so-called memory colors [2] and some had only a few memory colors. The gamma manipulation of the luminance also has an effect on the chroma, and therefore on the colorfulness of the image. Some images, for instance images that do not contain many colors, may benefit from more luminance contrast, but for other images the colorfulness may be more important than the luminance contrast.

Using linear regression, we found that we could predict the quality of the images using the discriminability measure of the usefulness-naturalness model [Janssen and Blommaert, 2000b] and the measured perceptual naturalness. The usefulness-naturalness model uses a different naturalness measure, based on the colors of skin, grass or sky. We could not use that measure, because not all images in our set contained these colors.

In the second experiment we evaluated quality perception for images that were scaled in chroma and also transformed by a gamma transformation of the luminance. In this experiment, the subjects did behave as one group, yielding closely corresponding responses to the experimental stimuli.

As expected, both the gamma and chroma scaling have an effect on the perceptual quality. However, the gamma setting has a greater influence on the perceptual quality than the chroma scaling. An explanation may be found in the fact that in natural environments gamma manipulation does not occur naturally, whereas chroma scaling may occur. Another possible explanation may be found in intrinsic differences in the stimulus production. The gamma manipulation is non-linear with fixed minimum and maximum luminance, whereas the chroma manipulation is achieved by a linear scaling factor, in combination with a fixed minimum. An interesting experiment would be to test stimuli where the chroma and luminance are treated in the same way. These stimuli may be produced in two different ways. The

---

[2]Memory colors are colors that are recalled in association with familiar objects in long-term memory [Bartleson, 1968], such as the green of grass or the yellow of a banana.

first is that both the luminance and the chroma are scaled linearly. This happens for instance in natural environments if the light changes from twilight to broad daylight. However, to scale the luminance linearly either the black point or the white point has to be changed. Especially for printed images this may degrade the quality of the images. The second way is to use (non-linear) gamma manipulation for both the luminance and chroma. However, if a gamma manipulation of the luminance is not natural, then a gamma manipulation of the chroma will probably also be unacceptable.

A manipulation that does occur in natural images is local contrast stretching of the luminance. Where the gamma manipulation changes the colors globally, local contrast stretching can be used to change the image locally. In this way the colored parts of the image can remain colored and the more achromatic parts of the image can gain contrast. An interesting question is whether local contrast stretching will affect the naturalness of the image. The technique of contrast stretching is already used in the field of photography, where the colors in the shade are manipulated differently than the colors in full light.

Using the discriminability measure of the usefulness-naturalness model [Janssen and Blommaert, 2000b] and the measured perceptual naturalness, we could not convincingly predict the quality of the gamma and chroma manipulated images (only 44% of the variance was explained).

Concluding, manipulation of the color distribution has a significant effect on perceptual quality. At least in the case of gamma manipulations, subject responses are predictable in the sense that their effect on perceived quality shows a clear maximum; such predictability can be used for creating an overall objective quality measure. On the other hand, matters are complicated by the occurrence of an effect of individual preference.

11.3   Gamut limited image processing and quality

In the last part of this thesis we studied the relationship between the quality of an image and the gamut of the displaying device. First, we proposed a way of using grey value image processing for color images, while taking the gamut, that is, the envelope of colors in color space that determine the limitations of the displaying device, into account. We also discuss a quality measure to find the best image of a series of images that only differ with respect to gamma or any other manipulation that effects the lightness values within the image.

The result of an image processing step can be seen as a point for point correction of the original image. We propose that the maximum and minimum lightness correction should depend on the available gamut space in a chosen direction. The

proposed lightness correction is a function of the original color point, the desired lightness correction and the maximum and minimum lightness corrections.

Because of the monotonic character of the lightness correction, all points that are different before the lightness correction remain different. However, in the case of lightness compression, differences that are physically still present, may nevertheless become invisible when they fall below the visual threshold.

The method was tested for two grey value image processing algorithms: sharpening and contrast improvement using gamma manipulation. For both applications, the recipes for which the chroma was kept constant gave the best results. We expected this for the gamma manipulation, because the lightness corrections are large for all points. Reducing the chroma of all points results in images that are too "greyish", whereas increasing the chroma may result in colors that have too much chroma with respect to their lightness, causing them to appear fluorescent. That the chroma reduction or increase was also visible for the sharpening experiment was surprising, because here the lightness correction is small for most points, and the chroma was varied proportionally to the lightness correction. However, this "small" chroma correction is perceptually quite visible.

In general, we conclude that for these two applications the gamut-limiting process is only advisable when the limiting is applied to the lightness dimension, rather than the chroma dimension.

In the last chapter we speculate on a good approach for developing a quality measure. We concluded that improvements in image quality within the gamut are best done in such a way that manipulations should mainly be confined to the achromatic domain. This leaves lightness (or luminance), rather than hue and chroma as the best choice for quality improvement.

We speculate that an image in which the colors fill the available gamut space relatively homogeneously have a higher quality than an image for which this is not the case. The basis for that idea is that a more homogeneous distribution allows for more discriminable nuances.

The proposed quality measure calculates the ratio of smallest and largest distance between the center of mass of the color points relative to the upper and lower boundary of the gamut, respectively. This measure is one for a center of mass that is located exactly between the upper gamut boundary and the lower gamut boundary. We consider this position optimal because the lightness nuances surrounding that point can be stretched over equal ranges of lighter and darker tones. For all other center of masses, the measure is smaller than one.

We tested such a model for three subparts of the `musicians` image. These preliminary results are promising.

A problem encountered in all experiments was that subjects based their quality response on subparts in an image. For the proposed quality measure, we need to identify these subparts, and also the colors of interest in these subparts. For instance, an important subpart in the `musicians` image is the `blond` girl. The skin tones in this subpart affect the quality, the background colors do not. Both the issue of selecting the subparts in the image and identifying the important colors require further study.

According to the proposed quality measure, the quality of an image depends on the gamut of the displaying device. If an image, optimized for one device, is displayed on another device with a different gamut, it does not have to be optimal. This dependency on the gamut may be an interesting topic for a new study.

11.4 Overall conclusions

The major goal of this study was to find an objective measure for the quality of a printed image that corresponds to perceptual image quality. On first sight, one might assume that the result of such a measure is an analytical formula, where measurements lead to a numerical quality estimation. However, this is not as simple as it seems. In chapter 3 we discussed some analytical quality measures. Most of these are *fidelity* measures, which determine the quality or degradation with respect to an original image instead of the quality of an image itself. For other quality measures the type of images on which the measure can be applied is small. And, finally, some quality measures can be written as a formula, but then the values of one or more variables have to be estimated using experiments.

We studied image quality of images reproduced by a printer. A major difference between a printer and a monitor is the shape of the gamut. Nevertheless, standard image enhancement techniques do not take the gamut into account. In the last part of this thesis we evaluate how we can incorporate the gamut into achromatic enhancement of *color* images and image quality. In this way, some of the differences between images displayed on a monitor and printed images can be taken into account.

Appendix A

Proof: constant h_{uv}^* is a plane in XYZ space

In this appendix we prove that a constant hue plane in CIELUV results in a plane in XYZ. We start with giving the transformation equations from XYZ to CIELUV

$$
\begin{aligned}
u' &= \frac{4X}{X + 15Y + 3Z} = \frac{4x}{-2x + 12y + 3} \\
v' &= \frac{9Y}{X + 15Y + 3Z} = \frac{9y}{-2x + 12y + 3}
\end{aligned}
\tag{A.1}
$$

$$
\begin{aligned}
u^* &= 13L * (u' - u_n') \\
v^* &= 13L * (v' - v_n') \\
h_{uv}^* &= \tan^{-1}(v^*/u^*) \\
C_{uv}^* &= \sqrt{u^{*2} + v^{*2}}
\end{aligned}
\tag{A.2}
$$

in which u_n' and v_n' are the u' and v' coordinates of a reference white.

In order for h_{uv}^* to be constant, v^*/u^* should be constant. So,

$$
\begin{aligned}
(v' - v_n')/(u' - u_n') &= K_1 \\
v' - K_1 u' &= v_n' - K_1 u_n' \\
\frac{9y - 4Kx}{-2x + 12y + 3} &= K_2 \\
9y - 4K_1 x &= -2K_2 x + 12K_2 y + 3K_2 \\
(9 - 12K_2)y + (2K_2 - 4K_1)x - 3K_2 &= 0
\end{aligned}
\tag{A.3}
$$

which is of the form $ax + bY = c$. So, keeping h_{uv}^* constant, for any given luminance value Y results in points lying on a line in the x, y plane. In the XYZ color space

these lines construct a plane. We can easily see that this plane goes through the black-and-white axis, by inserting $u' = u'_n$ and $v' = v'_n$ in the third equation. So the plane that we select, that is the plane spanned by the searched point and the line between black and white, indeed has the same hue for every point.

Appendix B

Transfer functions for sharpening and smoothing operations

The transfer functions of the sharpening and smoothing operations used in chapter 5, 6 and 9 are given in figure B.1. Note that no transfer function can be given for the anisotropic diffusion filter, because this filter is not spatially invariant.

Gaussian smoothing

$$g_\sigma$$
$$\tilde{G}_\sigma$$

Extreme sharpening

$$\delta(x) - g_\sigma$$
$$1 - \tilde{G}_\sigma$$

Real unsharp masking

$$\delta(x) + \alpha(\delta(x) - g_\sigma)$$
$$1 + \alpha(1 - \tilde{G}_\sigma)$$

Extreme Gaussian sharpening

$$-\nabla^2 g_\sigma$$
$$f^2 \cdot \tilde{G}_\sigma$$

Laplacian smoothing

$$\delta(x) + \nabla^2 g_\sigma$$
$$1 - f^2 \cdot \tilde{G}_\sigma$$

Gaussian unsharp masking

$$\delta(x) - k\nabla^2 g_\sigma$$
$$1 + kf^2 \cdot \tilde{G}_\sigma$$
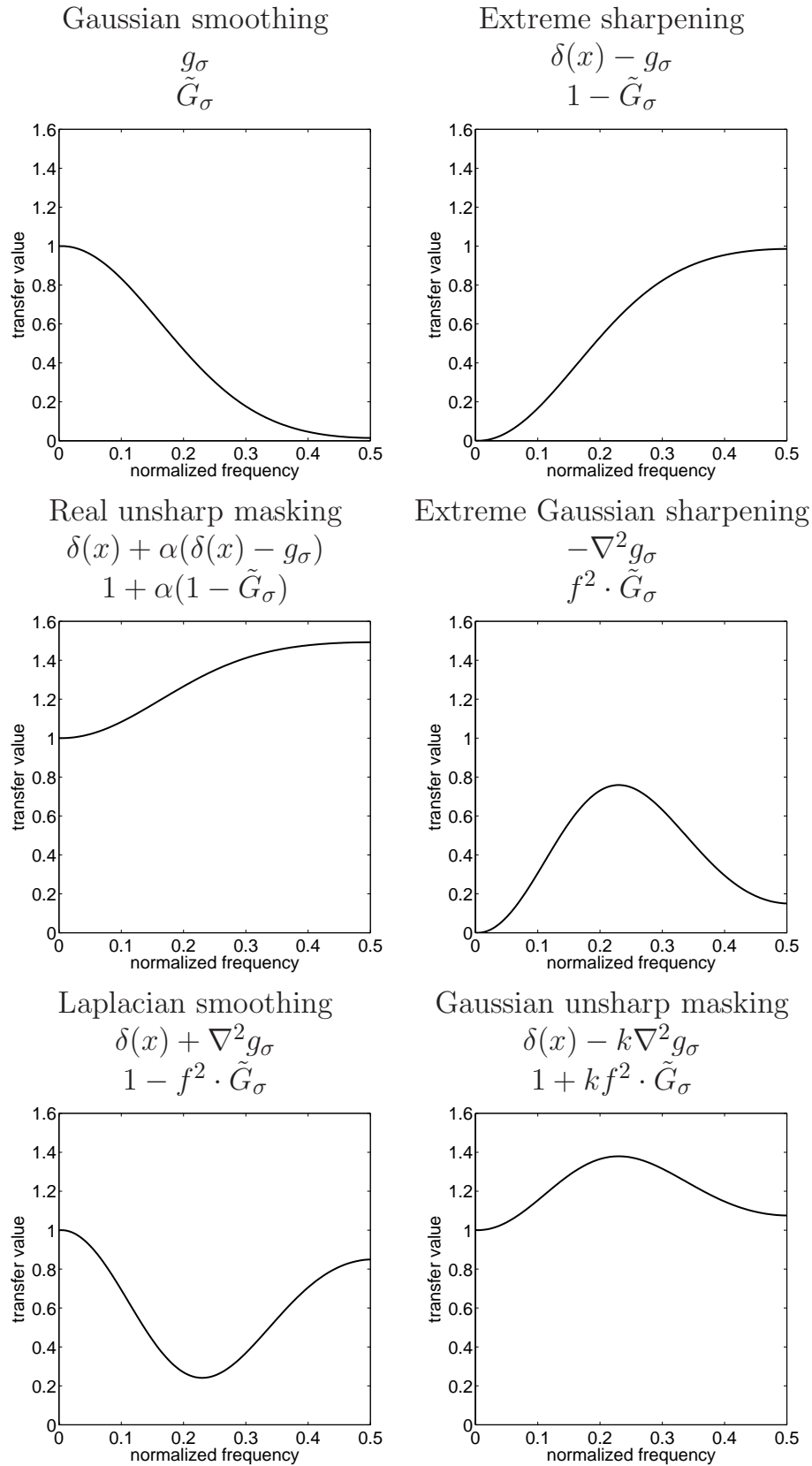
Figure B.1: The transfer functions of the sharpening and smoothing operations used. $\sigma = 1$, $\alpha = 0.3$ and k is 1.

Appendix C

Visualizations of the gamut
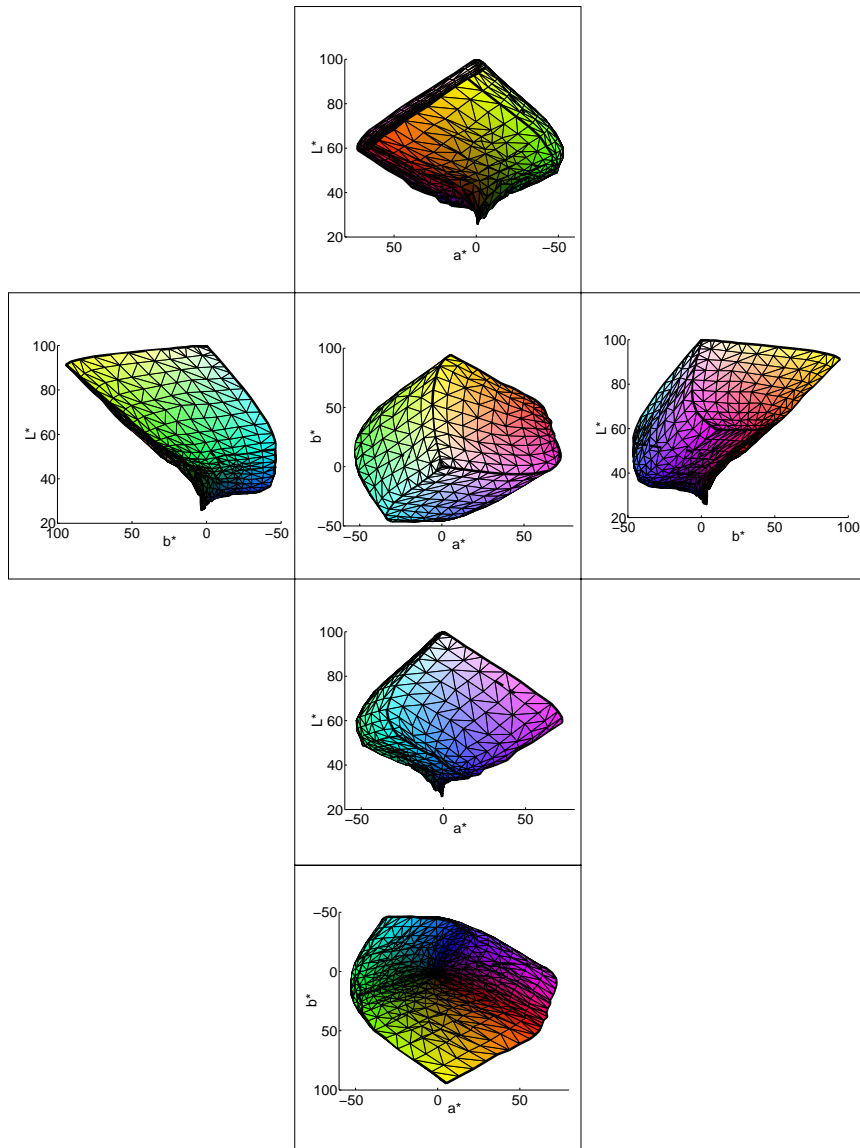
Figure C.1: The gamut of the EPSON printer plotted in the CIELAB space for 6 different viewing points, corresponding to the 6 different faces of a cube. These images are made using the relations in the CMY_{EPSON} space. The solid lines indicate that two of the three inks are 0 or 100%.
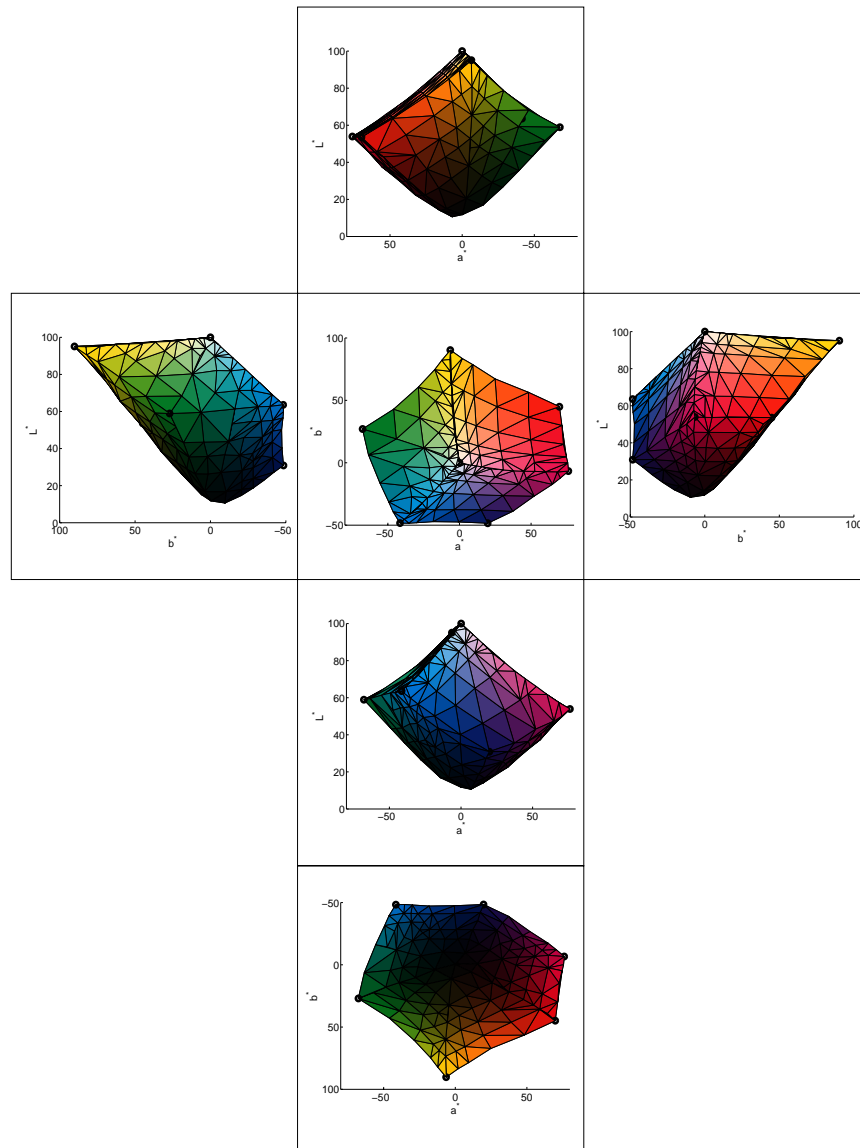
Figure C.2: The gamut of the ISO data plotted in the CIELAB space for 6 different viewing points, corresponding to the 6 different faces of a cube. These images are made using the relations in the $CIELAB_{ISO}$ space (method 1). The circles indicate that C, M, and Y are either 0 or 100%. For all these points K is 0.

Appendix D

Color stimuli
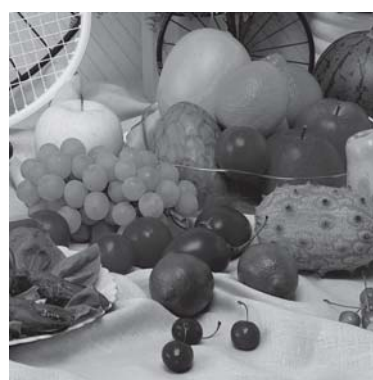
(a) $\gamma = 0.6$ (b) $\gamma = 1.0$ (c) $\gamma = 1.4$

(d) $\gamma = 0.6$ (e) $\gamma = 1.0$ (f) $\gamma = 1.4$

Figure D.1: Examples of the gamma manipulation on both black-and-white images (top) and color images (bottom). The original is the image for which γ is 1.0 (b and e). The input image is the right lower corner of the `bicycle` image.
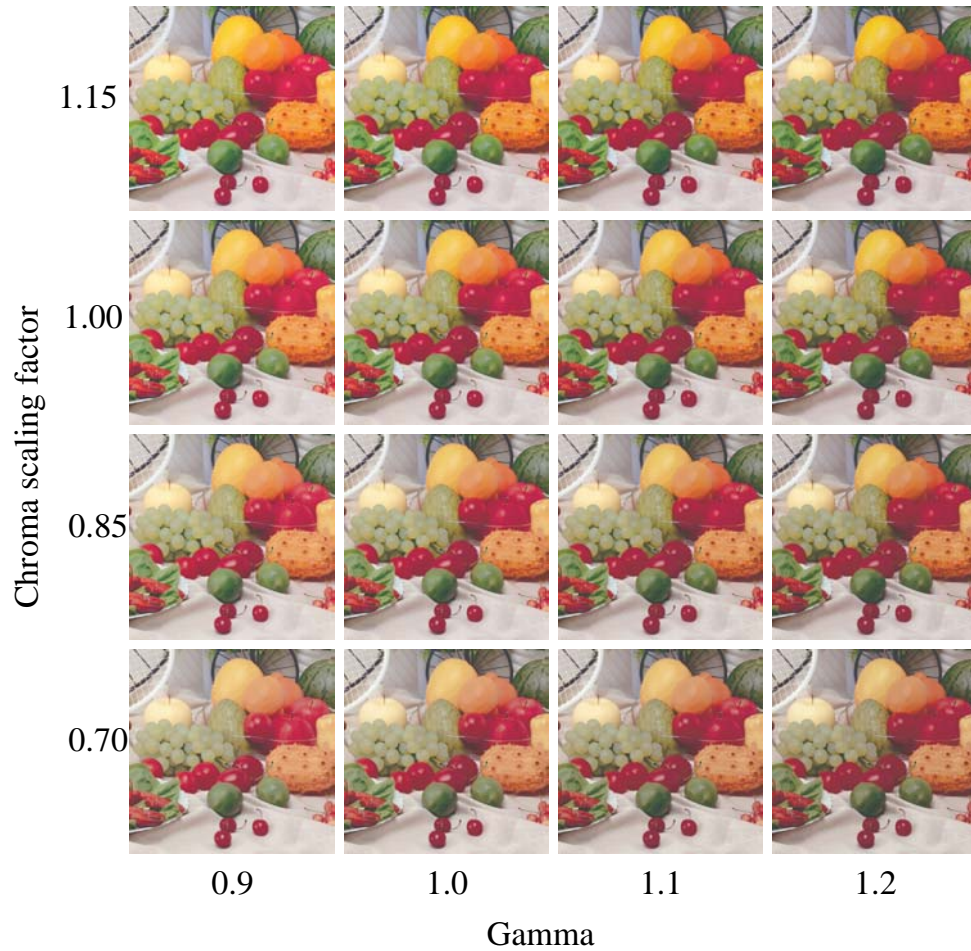
Figure D.2: Examples of the gamma manipulation and the chroma scaling. The original is the image for which γ is 1.0 and s_{chroma} is 1.0. The input image is a part of the `bicycle` image.

Appendix E

Results of gamut-limited manipulations

| $\gamma = 0.6$ | $\gamma = 0.8$ | $\gamma = 1.0$ | $\gamma = 1.2$ | $\gamma = 1.4$ |

Figure E.1: The result for normal gamma manipulation, without taking the gamut into account, using the first gamma manipulation.

| $\gamma = 0.6$ | $\gamma = 0.8$ | $\gamma = 1.0$ | $\gamma = 1.2$ | $\gamma = 1.4$ |

Figure E.2: The results for the $C^* = constant$ recipe, using the first gamma manipulation.

| $\gamma = 0.6$ | $\gamma = 0.8$ | $\gamma = 1.0$ | $\gamma = 1.2$ | $\gamma = 1.4$ |

Figure E.3: The results for the $C^*/L^* = constant$ recipe, using the first gamma manipulation.

$\gamma = 0.6$ $\gamma = 0.8$ $\gamma = 1.0$ $\gamma = 1.2$ $\gamma = 1.4$

Figure E.4: The results for the *mapping towards black* recipe, using the first gamma manipulation.

$\gamma = 0.6$ $\gamma = 1.0$ $\gamma = 1.4$ $\gamma = 2.0$
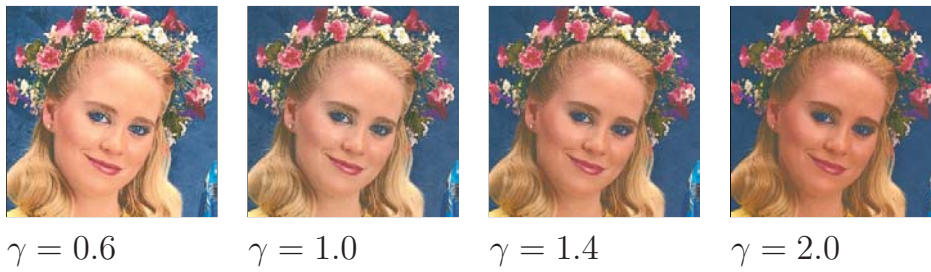
Figure E.5: The results for the *away from black* recipe, using the first gamma manipulation.

$\gamma = 0.6$ $\gamma = 0.8$ $\gamma = 1.0$ $\gamma = 1.2$ $\gamma = 1.4$

Figure E.6: The results for the *mapping over the cusp point* recipe, using the first gamma manipulation.

$\gamma = 0.6$ $\gamma = 0.8$ $\gamma = 1.0$ $\gamma = 1.2$ $\gamma = 1.4$

Figure E.7: The result for normal gamma manipulation, without taking the gamut into account, using the second gamma manipulation. $L^*_{point} = L^*_{average}$.

$\gamma = 0.6$ $\gamma = 0.8$ $\gamma = 1.0$ $\gamma = 1.2$ $\gamma = 1.4$
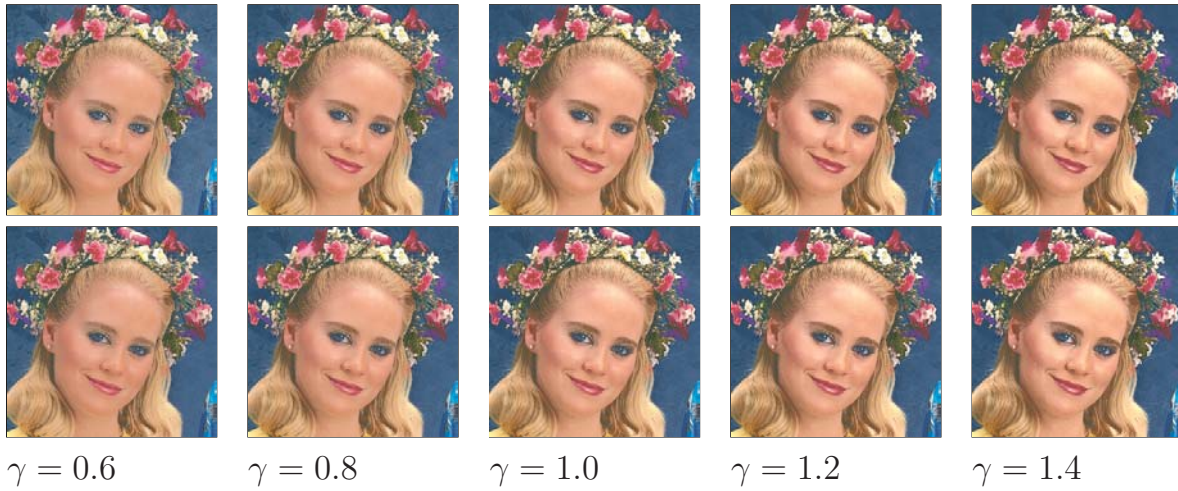
Figure E.8: The results for the $C^* = constant$ recipe, using the second gamma manipulation. In the top row $L^*_{point} = L^*_{average}$, in the bottom row $L^*_{point} = L^*_{cusp}$.

$\gamma = 0.6$ $\gamma = 0.8$ $\gamma = 1.0$ $\gamma = 1.2$ $\gamma = 1.4$
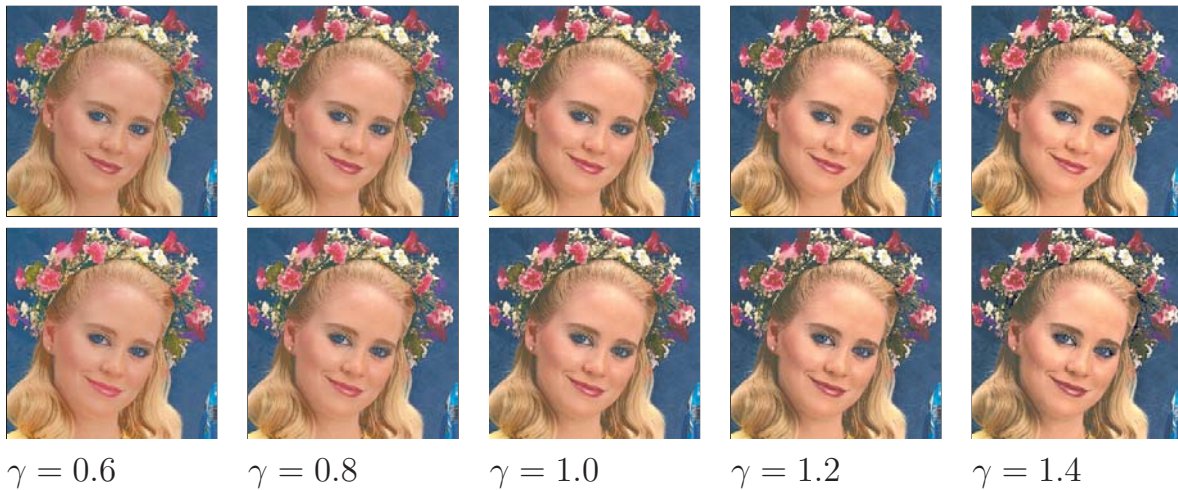
Figure E.9: The results for the $C^*/L^* = constant$ recipe, using the second gamma manipulation. In the top row $L^*_{point} = L^*_{average}$, in the bottom row $L^*_{point} = L^*_{cusp}$.

$\gamma = 0.6$ $\gamma = 0.8$ $\gamma = 1.0$ $\gamma = 1.2$ $\gamma = 1.4$
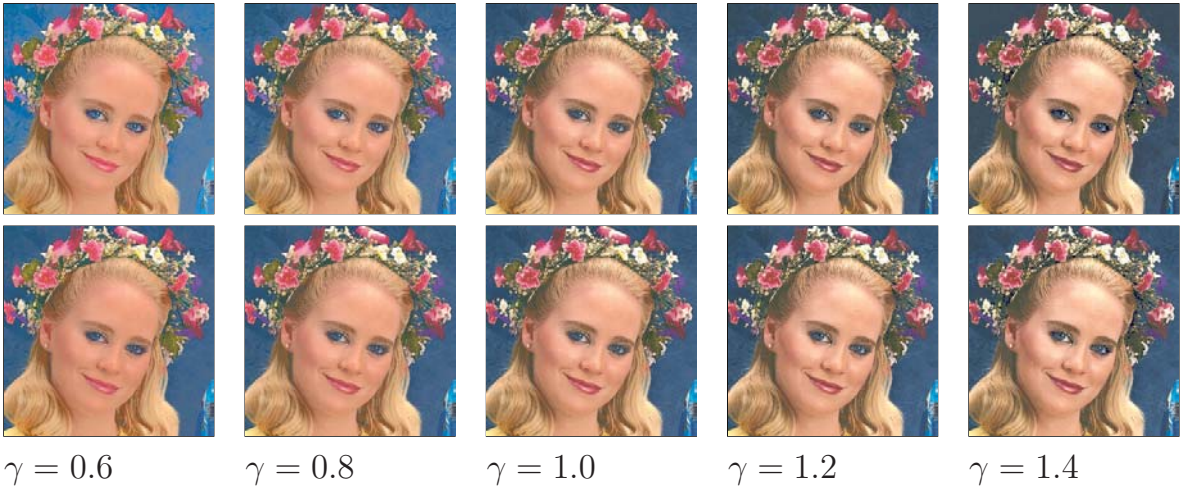
Figure E.10: The results for the *mapping over the cusp point* recipe, using the second gamma manipulation. In the top row $L^*_{point} = L^*_{average}$, in the bottom row $L^*_{point} = L^*_{cusp}$.

Test image

(a) Sharpening within the gamut

(c) Mapping towards black and white

(e) mapping away from black and white

Normal sharpening

(b) Gamut limited sharpening

(d) Mapping halfway towards black and white
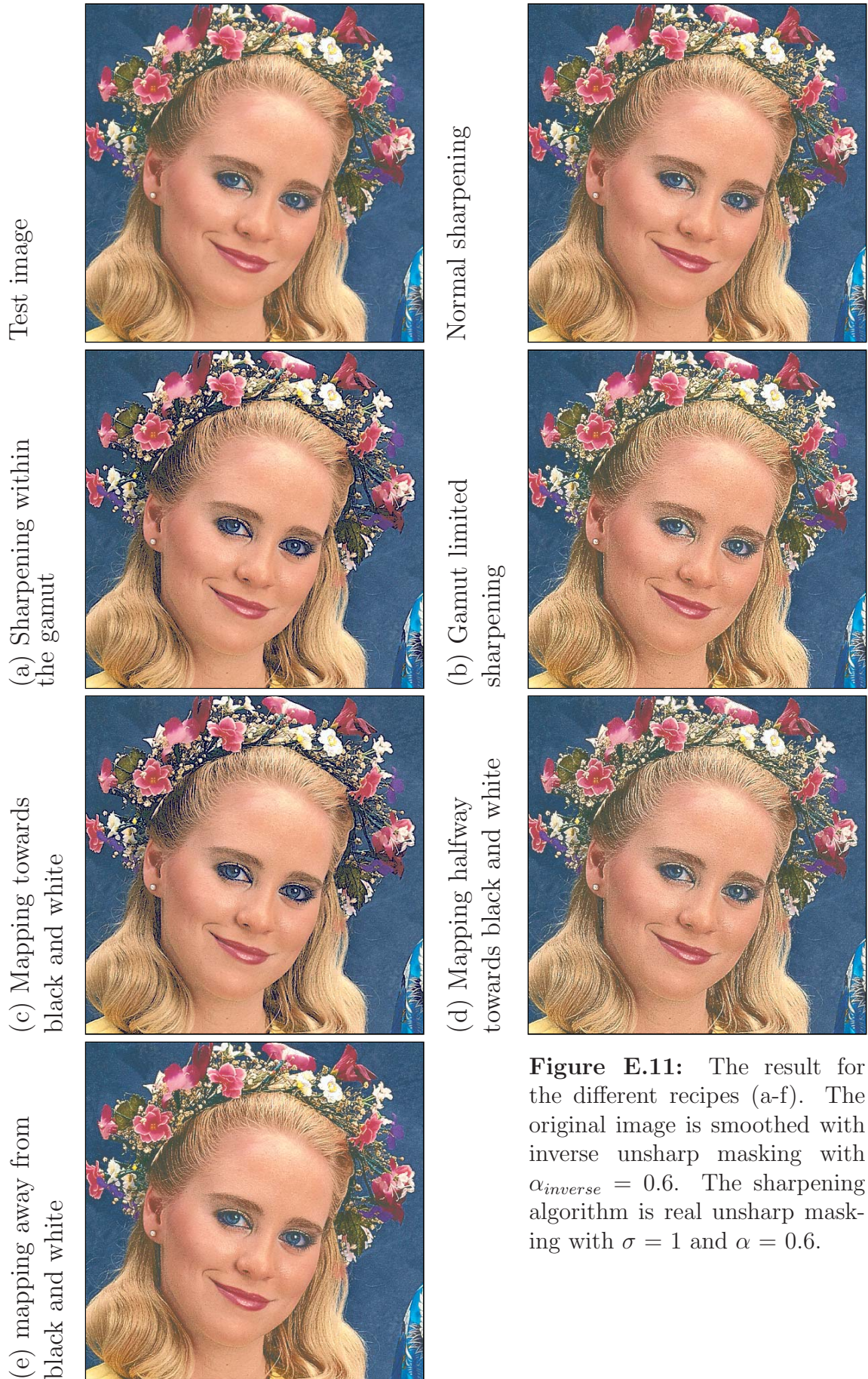
Figure E.11: The result for the different recipes (a-f). The original image is smoothed with inverse unsharp masking with $\alpha_{inverse} = 0.6$. The sharpening algorithm is real unsharp masking with $\sigma = 1$ and $\alpha = 0.6$.

Test image $\alpha = 0.6$

$\alpha = 0.9$ $\alpha = 1.2$
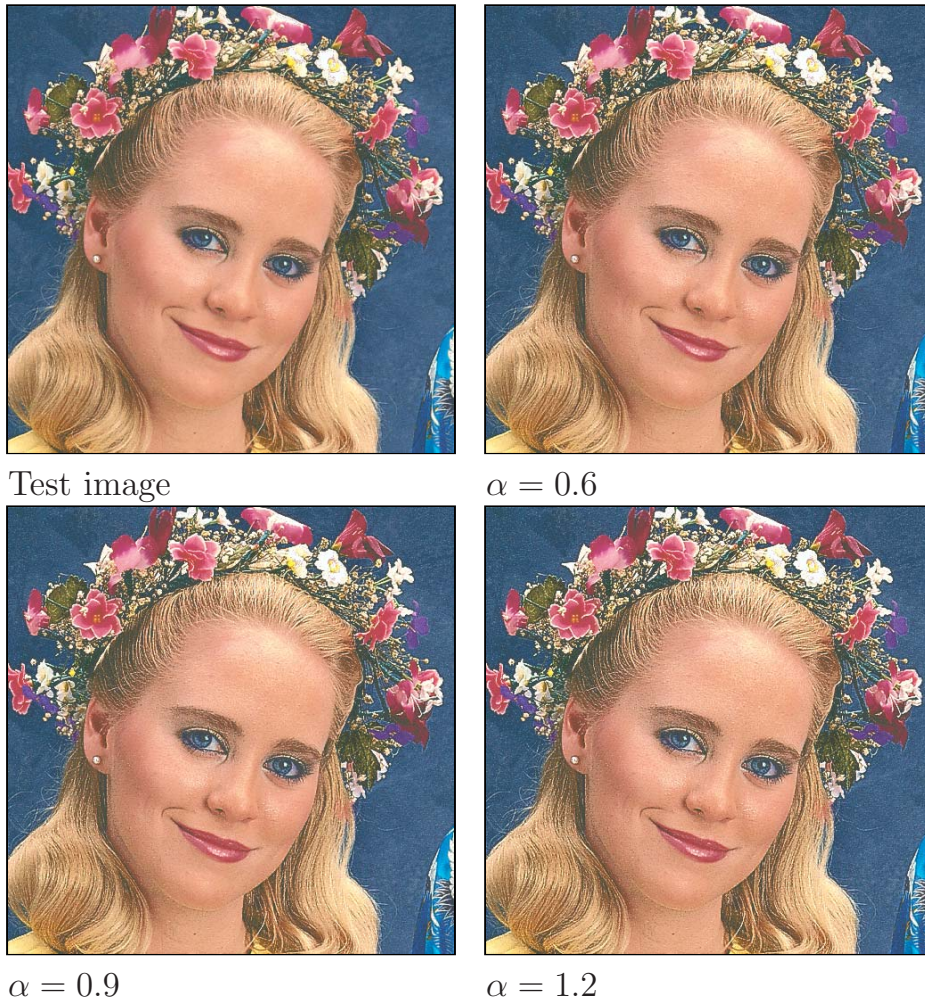
Figure E.12: The result for normal (not gamut-limited) sharpening for different values of α, with $\sigma = 1$. The original image is smoothed with inverse unsharp masking with $\alpha_{inverse} = 0.6$. The sharpening algorithm is real unsharp masking.

Test image

$\alpha = 0.6$

$\alpha = 0.9$

$\alpha = 1.2$
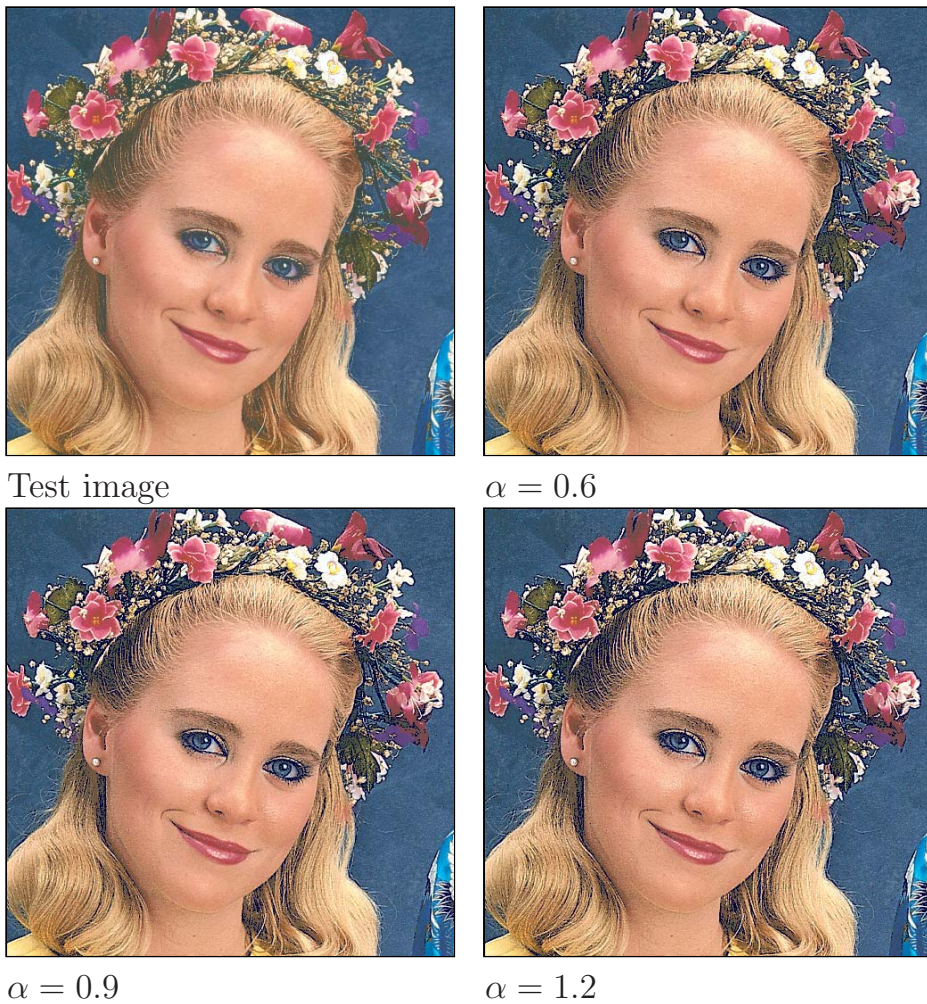
Figure E.13: The result for *sharpening within the gamut* (a) for different values of α, with $\sigma = 1$. The original image is smoothed with inverse unsharp masking with $\alpha_{inverse} = 0.6$. The sharpening algorithm is real unsharp masking.

Test image $\alpha = 0.6$

$\alpha = 0.9$ $\alpha = 1.2$
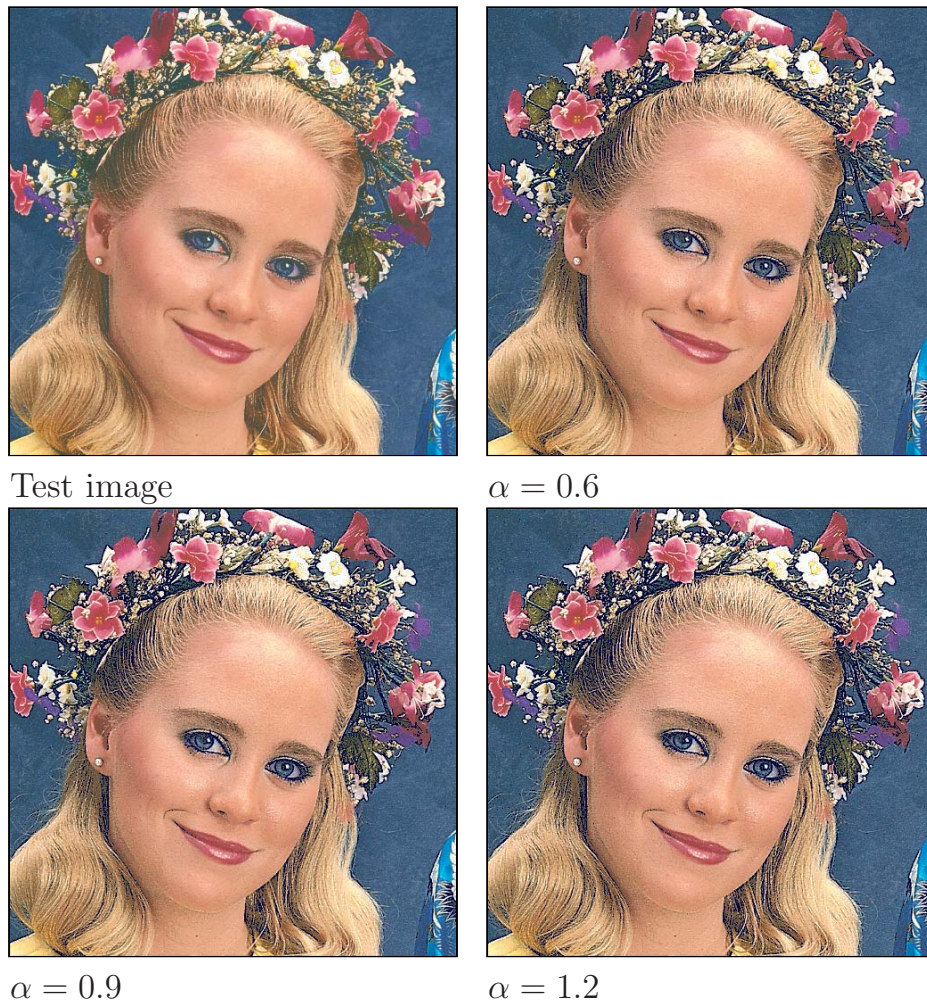
Figure E.14: The result for *mapping towards black and white* (c) for different values of α, with $\sigma = 1$. The original image is smoothed with inverse unsharp masking with $\alpha_{inverse} = 0.6$. The sharpening algorithm is real unsharp masking.

Test image $\alpha = 0.6$

$\alpha = 0.9$ $\alpha = 1.2$
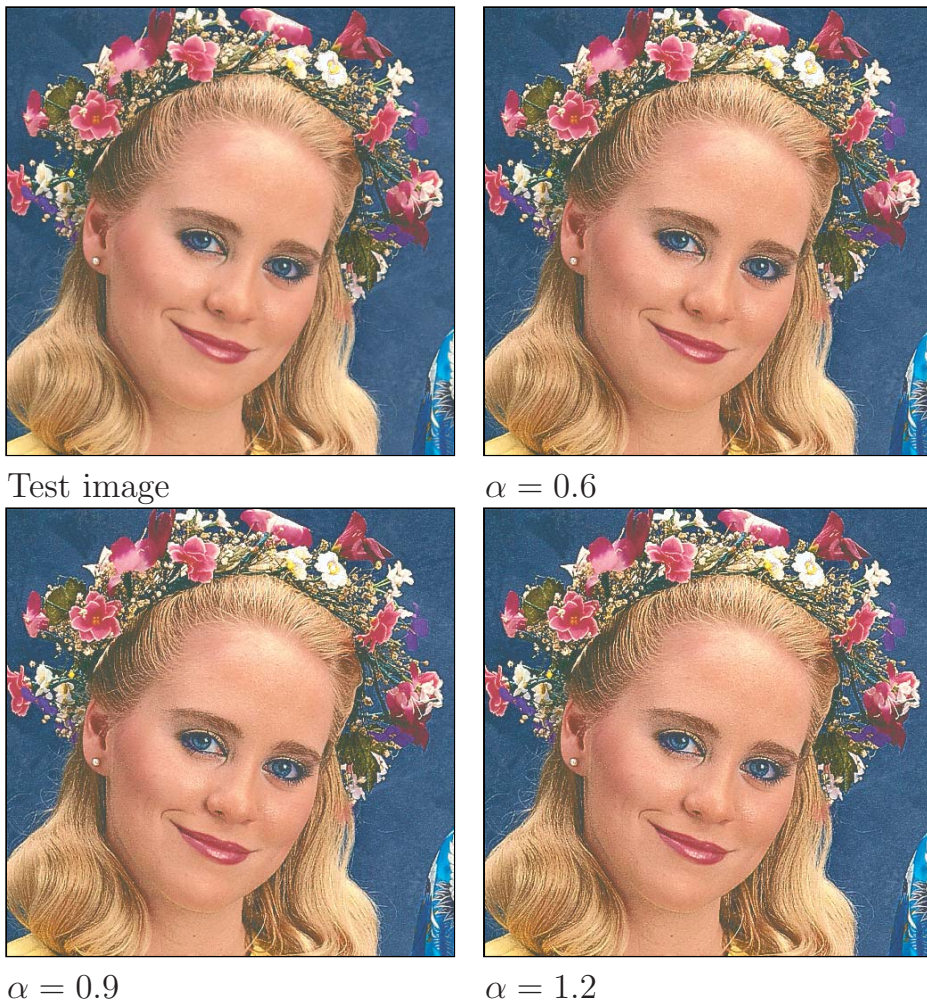
Figure E.15: The result for *mapping away from black and white* (e) for different values of α, with $\sigma = 1$. The original image is smoothed with inverse unsharp masking with $\alpha_{inverse} = 0.6$. The sharpening algorithm is real unsharp masking.

Test image

$\sigma = 0.5$

$\sigma = 1.0$

$\sigma = 2.0$
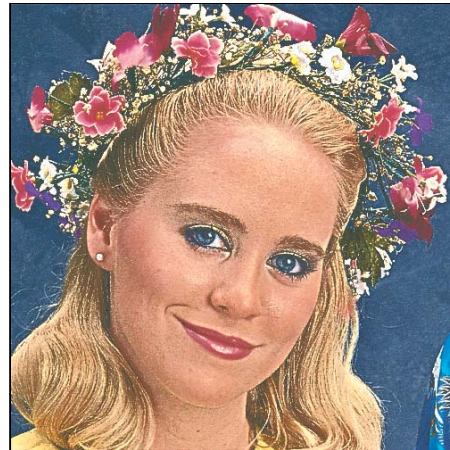
$\sigma = 4.0$

Figure E.16: The result for *gamut limited sharpening* (b) for different values of σ. Note that the value of α is set by the algorithm. The original image is smoothed with inverse unsharp masking with $\alpha_{inverse} = 0.6$. The sharpening algorithm is real unsharp masking.

Appendix F

Gamma experiment: Friedman two-way analysis of variance and Kendall Coefficient of Concordance

Using the Friedman two-way analysis of variance by ranks and the Kendall Coefficient of Concordance two hypotheses are tested. These null hypotheses are that the k samples have been drawn from the same population, and that there are no correlations between the individuals [Siegel and Castellan, 1988, StatSoft, 2000]. The methods are discussed in subsection 4.4.2.

**Table F.1:** The Friedman rank order coefficient and the Kendall Coefficient of Concordance for the quality experiment. The values for which the null hypothesis cannot be rejected are given in italic.

		Friedman	p <	Kendall
portrait	color	*17.30*	*0.10*	*0.26*
cafe	color	**47.59**	**0.00**	**0.72**
basket	color	**30.21**	**0.00**	**0.50**
bicycle	color	*16.51*	*0.12*	*0.25*
musicians	color	**40.41**	**0.00**	**0.61**
portrait	bw	**41.59**	**0.00**	**0.63**
cafe	bw	**44.67**	**0.00**	**0.68**
basket	bw	*14.97*	*0.18*	*0.22*
bicycle	bw	**36.67**	**0.00**	**0.56**
musicians	bw	**43.07**	**0.00**	**0.65**

The results are shown in table F.1. It was found that for the `basket`, `cafe` and `musicians` images both null hypotheses could be rejected for the color images, meaning that the rank ordering by the subjects was not random and the responses by the subjects were not independent. For the other two images, this was only found

Table F.2: The Friedman rank order coefficient and the Kendall Coefficient of Concordance for subsets of the subjects. For the ranges in this table the null hypothesis could not be rejected for the total range. The values for which the null hypothesis cannot be rejected are given in italic.

			subjects	Friedman	p <	Kendall
portrait	color	quality	1 2 3	**20.33**	**0.04**	**0.62**
portrait	color	quality	4 7 8	**29.41**	**0.00**	**0.89**
bicycle	color	quality	1 2 3	*10.38*	*0.50*	*0.31*
bicycle	color	quality	4 7 8	*14.69*	*0.20*	*0.45*
basket	bw	quality	1 2 3	*9.77*	*0.55*	*0.30*
basket	bw	quality	4 7 8	*7.56*	*0.75*	*7.56*
basket	bw	naturalness	1 2 3	**23.72**	**0.01**	**0.72**
basket	bw	naturalness	4 7 8	**21.00**	**0.03**	**0.64**
portrait	color	quality	1 4 7	*14.08*	*0.23*	*0.43*
portrait	color	quality	2 3 8	*7.26*	*0.78*	*0.21*
bicycle	color	quality	1 4 7	**23.82**	**0.01**	**0.72**
bicycle	color	quality	2 3 8	**28.33**	**0.00**	**0.86**
basket	bw	quality	1 4 7	*13.21*	*0.28*	*0.40*
basket	bw	quality	2 3 8	*11.00*	*0.44*	*0.33*
basket	bw	naturalness	1 4 7	**21.00**	**0.03**	**0.63**
basket	bw	naturalness	2 3 8	**29.82**	**0.00**	**0.90**
portrait	color	quality	1 3 7	**20.44**	**0.04**	**0.62**
portrait	color	quality	2 4 8	*5.21*	*0.92*	*0.16*
bicycle	color	quality	1 3 7	*6.64*	*0.83*	*0.20*
bicycle	color	quality	2 4 8	**20.44**	**0.04**	**0.62**
basket	bw	quality	1 3 7	**23.41**	**0.02**	**0.71**
basket	bw	quality	2 4 8	**26.23**	**0.01**	**0.79**
basket	bw	naturalness	1 3 7	*15.26*	*0.17*	*0.46*
basket	bw	naturalness	2 4 8	**28.18**	**0.00**	**0.85**

if the subjects were split in two groups. In table F.2, the Friedman rank order coefficient and the Kendall Coefficient of Concordance are given for several subgroups of subjects. It can be seen that for all ranges the set of subjects can be divided into two subsets of images that are significant. However, these subsets differ for the different ranges.

For the naturalness responses the results are shown in table F.3 For all images, except the basket black-and-white image, both null hypotheses could be rejected.

Table F.3: The Friedman rank order coefficient and the Kendall Coefficient of Concordance for the naturalness experiment. The values for which the null hypothesis cannot be rejected are given in italic.

		Friedman	$p <$	Kendall
portrait	color	**21.49**	**0.03**	**0.33**
cafe	color	**44.49**	**0.00**	**0.67**
basket	color	**43.03**	**0.00**	**0.65**
bicycle	color	**31.05**	**0.00**	**0.47**
musicians	color	**31.64**	**0.00**	**0.47**
portrait	bw	**40.59**	**0.00**	**0.62**
cafe	bw	**43.41**	**0.00**	**0.66**
basket	bw	*19.26*	*0.06*	*0.29*
bicycle	bw	**21.44**	**0.03**	**0.32**
musicians	bw	**39.03**	**0.00**	**0.59**

Appendix G

Gamma & chroma experiment: Friedman two-way analysis of variance and Kendall Coefficient of Concordance

Using the Friedman two-way analysis of variance by ranks and the Kendall Coefficient of Concordance two hypotheses are tested. The first null hypothesis is that the 16 images in this experiment have been drawn from the same population. The second null hypothesis is that there are no correlations between the subjects [Siegel and Castellan, 1988, StatSoft, 2000]. The methods are discussed in subsection 4.4.2.

The results are given in table G.1. It was found that for all experiments both null hypothesis could be rejected for all subjects, meaning that the rank ordering returned by the subjects was not random and that the responses by the subjects were not random.

Table G.1: The Friedman rank order coefficient and the Kendall Coefficient of Concordance for the quality (upper part) and the naturalness (lower part) experiment. Note that values for which the null hypothesis cannot be rejected are given in italic.

Quality			
	Friedman	p <	Kendall
portrait	124.29	0.00	0.24
cafe	86.69	0.00	0.52
basket	54.22	0.00	0.36
bicycle	70.95	0.00	0.30
Naturalness			
	Friedman	p <	Kendall
portrait	90.63	0.00	0.40
cafe	77.48	0.00	0.34
basket	62.92	0.00	0.28
bicycle	58.05	0.00	0.21

Bibliography

[Ahumada, 1993] Ahumada, Jr., A. (1993). Computational image quality metrics: a review. *SID Digest*, 24:305–308.

[Allen, 1980] Allen, E. (1980). Colorant formulation and shading. In Grum, F. and Bartleson, C. J., editors, *Optical radiation measurements*, volume 2, chapter 7. Academic Press, NY.

[ANSI, 1995] ANSI (1995). Graphic technology - color characterization data for type 1 printing. Technical Report ANSI CGATS TR 001-1995, NPES.

[van Asselt, 1997] van Asselt, R. J. (1997). Line and edge characterization using mallat wavelets. Master's thesis, Delft University of Technology.

[Balasubramanian, 1995] Balasubramanian (1995). Colorimetric modeling of binary color printers. In *IEEE International Conference on Image Processing 2*, pages 2327–2330.

[Barten, 1989] Barten, P. G. J. (1989). The square root integral (SQRI): a new metric to describe the effect of various display parameters on perceived image quality. In Rogowitz, B. E. and Allebach, J. P., editors, *Human Vision, Visual Processing and Digital Display VII*, volume 1077, pages 73–82. SPIE.

[Barten, 1999] Barten, P. G. J. (1999). *Contrast sensitivity of the human eye and its effects on image quality*. PhD thesis, Technische Universiteit Eindhoven.

[Bartleson, 1968] Bartleson, C. J. (1968). Memory colors of familiar objects. *Journal of Optical society of America*, 50:73–79.

[Bech et al., 1996] Bech, S., Hamberg, R., Nijenhuis, M., Teunissen, C., Looren de Jong, H., Houben, P., and Pramanik, S. K. (1996). The RaPID perceptual image description method (RaPID). In Rogowitz, B. E. and Allebach, J. P., editors, *Human Vision and Electronic Imaging*, volume 2657, pages 317–328. SPIE.

[Berns, 1992] Berns, R. S. (1992). Color WYSIWYG: a combination of device colorimetric characterization and appearance modeling. *SID Digest*, pages 549–552.

[Borg and Groenen, 1997] Borg, I. and Groenen, P. (1997). *Modern Multidimensional Scaling, theory and applications*. Springer, New York, NY.

[Bourgin, 1998] Bourgin, D. (1998). Color spaces FAQ. *http://www.cc.iastate.edu/olc_answers/ packages/graphics/color-space.faq.html*.

[Catté et al., 1992] Catté, F., Lions, P.-L., Morel, J.-M., and Coll, T. (1992). Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 29(1):182–193.

[CIE, 1995] CIE (1995). Colorimetry, second edition. Technical Report no. 15.2, CIE.

[Daly, 1993a] Daly, S. (1993a). Quantitative performance assessment of an algorithm for the determination of image fidelity. *Society for information display international symposium digest of technical papers*, 24:317–320.

[Daly, 1993b] Daly, S. (1993b). The visible differences predictor: An algorithm for the assessment of image fidelity. In Watson, A. B., editor, *Digital images and human vision*, chapter 14, pages 179–206. M.I.T. Press, London, England.

[Daniels et al., 1997] Daniels, C. M., Giorgianni, E. J., and Fairchild, M. D. (1997). The effect of surround on perceived lightness contrast of pictorial images. In *The fifth color imaging conference: Color science, systems and applications*, pages 12–16. IS&T/SID.

[Davies and Rose, 1993] Davies, I. R. L. and Rose, D. (1993). Automated image quality assessment. In Allebach, J. P. and Rogowitz, B. E., editors, *Human vision, visual processing and digital displays IV*, volume 1913, pages 15–26. SPIE.

[Dijk et al., 2000] Dijk, J., Verbeek, P. W., Walraven, J., and Young, I. T. (2000). The effect of gamma on the perception of printed color images. In *First international conference on color in graphics and image processing*, pages 7–13.

[Dijk et al., 1999] Dijk, J., Young, I. T., and Walraven, J. (1999). The effect of gamma on printed images. In Rogowitz, B. E. and Pappas, T. N., editors, *Human vision and electronic imaging IV*, volume 3644, pages 290–299. SPIE.

[van Dijk, 1997] van Dijk, A. M. (1997). *Image representation and Compression using Steered Hermite Transforms*. PhD thesis, Eindhoven University of Technology.

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

[Engeldrum, 1999a] Engeldrum, P. G. (1999a). The image quality circle. *http://www.imcotek.com/iqc01.htm*.

[Engeldrum, 1999b] Engeldrum, P. G. (1999b). Image quality modeling: where are we? In *IS&T's 1999 PICS Conference*, pages 251–255.

[Evans, 1974] Evans, R. M. (1974). *The perception of color*. John Wiley and Sons, NY.

[Farrell, 1999] Farrell, J. E. (1999). Image quality evaluation. In MacDonald, L. W. and Luo, M. R., editors, *Colour imaging: vision and technology*, chapter 15, pages 315–337. Wiley and Sons.

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

[Ford, 1999] Ford, A. M. (1999). Determination of compressed image quality. In MacDonald, L. W. and Luo, M. R., editors, *Colour imaging: vision and technology*, chapter 16, pages 315–337. Wiley and Sons.

[Freeman and Adelson, 1991] Freeman, W. T. and Adelson, E. H. (1991). The design and use of steerable filters. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906.

[Gilchrist et al., 1983] Gilchrist, A. L., Delman, S., and Jacobsen, A. (1983). The classification and integration of edges as critical to the perception of reflectance and illumination. *Perception and psychophysics*, 33(5):425–436.

[van Ginkel, 2002] Ginkel, M. van (2002). *Image Analysis using Orientation Space based on Steerable Filters*. PhD thesis, Delft University of Technology, Delft, The Netherlands.

[Granger and Cupery, 1972] Granger, E. M. and Cupery, K. N. (1972). An optical merit function (sqf), wich correlates with subjective image quality judgments. *Photographic science and engineering*, 16:221–230.

[Hardeberg and Schmitt, 1998] Hardeberg, J. Y. and Schmitt, F. (1998). Color printer characterization using a computational geometry approach. In Buckley, R., editor, *Recent progress in color management and Communications*, pages 88–91. IS&T.

[Hardeberg et al., 1999] Hardeberg, J. Y., Schmitt, F., Brettel, H., Crettez, J.-P., and Maitre, H. (1999). Multispectral image acquisition and simulation of illuminant changes. In MacDonald, L. W. and Luo, M. R., editors, *Colour imaging: vision and technology*, chapter 8, pages 145–164. Wiley and Sons.

[Hering, 1878] Hering, E. (1964 (original work published in 1878)). *Outlines of a theory of the light sense*. Harvard University Press, Boston, MA.

[Horn and Bachman, 1978] Horn, B. K. and Bachman, B. L. (1978). Using synthetic images to register real images with surface models. *Communications of the ACM*, 21:914–924.

[Hunt, 1977] Hunt, R. G. W. (1977). The specification of color appearance-1: concept and terms. *Color research and application*, 2(2):55–68.

[Hunt, 1987] Hunt, R. W. G. (1987). *The reproduction of color in photography, printing and television*. Fountain, England.

[Hunt, 1991] Hunt, R. W. G. (1991). *Measuring colour, second edition*. Ellis Horwood Limited, Chichester, England.

[Hurvich and Jameson, 1955] Hurvich, L. M. and Jameson, D. (1955). Some quantitative aspects of an opponent-colors theory-2: Brightness, saturation and hue in normal and dichromatic vision. *Journal of the optical society of america*, 45:602–616.

[ICC, 1998] ICC (1998). Specification ICC.1:1998-09 file format for color profiles. Technical report, International Color Consortium.

[Jacobsen and Gilchrist, 1988] Jacobsen, A. and Gilchrist, A. L. (1988). The ratio principle holds over a million-to-one range of illumination. *Perception and psychophysics*, 43:1–6.

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

[Janssen and Blommaert, 2000a] Janssen, T. J. W. M. and Blommaert, F. J. J. (2000a). A computational approach to image quality. *Displays*, 21(4):129–142.

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

[Johnson and Fairchild, 2000] Johnson, G. M. and Fairchild, M. D. (2000). Sharpness rules. In *8th Color Imaging Conference*, pages 24–30. IS&T/SID.

[Judd and Wyszecki, 1975] Judd, D. B. and Wyszecki, G. (1975). *Color in business, science and industry*. John Wiley and Sons.

[Kang, 1997] Kang, H. R. (1997). *Color technology for electronic imaging devices*. SPIE optical engineering press.

[Katoh et al., 1999] Katoh, N., Ito, M., and Ohno, S. (1999). Three-dimensional gamut mapping using various color difference formulae and color spaces. *Journal of electronic imaging*, 8(4):365–379.

[Katsulai and Arimizu, 1981] Katsulai, H. and Arimizu, N. (1981). Evaluation of image fidelity by means of the fidelogram and level mean-square error. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(3):337–347.

[Katz, 1911] Katz, D. (1935 (original work published in 1911)). *The world of colour (P. Kegan Trans.).* Trench, Trubner & Co., London.

[Kayargadde, 1995] Kayargadde, V. (1995). *Feature extraction for image quality prediction.* PhD thesis, Technische universiteit Eindhoven.

[Knutsson and Granlund, 1983] Knutsson, H. and Granlund, G. (1983). Texture analysis using two-dimensional quadrature filters. In *IEEE Computer society workshop on computer architecture for pattern analysis and image database management*, Pasadena.

[Kubelka, 1948] Kubelka, P. (1948). New contributions to the optics of intensely light-scattering materials: Part 1. *Journal of Optical Society of America*, 38:448–457.

[Kuwahara et al., 1976] Kuwahara, M., Hachimura, K., Eiho, S., and Kinoshita, M. (1976). *Digital processing of biomedical images*, pages 187–203. Plenum Press, New York, NY.

[Lane, 2003] Lane, D. M. (2003). Hyperstat online: an introductory statistics book and online tutorial for help in statistics. *http://davidmlane.com/hyperstat/.*

[Lindeberg, 1993] Lindeberg, T. (1993). On scale selection for differential operators. In *Proc. 8th Scandinavian Conference on Image Analysis*, pages 857–866.

[Lubin, 1993] Lubin, J. (1993). The use of phychophysical data and models in the analysis of display system performance. In Watson, A., editor, *Digital images and human vision*, chapter 13, pages 163–178. M.I.T Press, London, England.

[Lubin, 1995] Lubin, J. (1995). A visual discrimination model for imaging system design and evaluation. In Peli, E., editor, *Vision models for target detection and recognition*, pages 245–283. World Scientific Publishing Co., Singapore.

[Lucassen and Walraven, 1990] Lucassen, M. P. and Walraven, J. (1990). Evaluation of a simple model for color monitor recalibration. *Color research and Application*, 15:321–326.

[Luo, 1998] Luo, M. R. (1998). Color science. In Sangwine, S. J. and Horne, R. E. N., editors, *The colour imaging processing handbook*, chapter 3, pages 26–66. Chapman and Hall.

[Luo and Hunt, 1998] Luo, M. R. and Hunt, R. W. G. (1998). The structure of the CIE 1997 colour appearance model (CIECAM97s). *Color research and application*, 23(3):138–146.

[Mallat and Hwang, 1992] Mallat, S. and Hwang, W. L. (1992). Singularity detection and processing with wavelets. *IEEE transactions on information theory*, 38(2):617–643.

[Marmolin, 1986] Marmolin, H. (1986). Subjective MSE measures. *IEEE Transactions on Systems, Man and Cybernetics*, 16(3):486–489.

[Martens and Meesters, 1998] Martens, J. B. and Meesters, L. (1998). Image dissimilarity. *Signal Processing*, 70:155–176.

[Morovic, 1998] Morovic, J. (1998). The colour image processing handbook. In Sangwine, S. and Horne, R., editors, *The colour imaging processing handbook*, chapter Colour management for the graphic arts, pages 332–357. Chapman and Hall, London, UK.

[Morovic and Luo, 2000] Morovic, J. and Luo, M. R. (2000). Calculating medium and image gamut boundaries for gamut mapping. *Color research and application*, 25(6):394–401.

[Nachmias and Sansbury, 1974] Nachmias, J. and Sansbury, R. V. (1974). Grating contrast: discrimination may be better than detection. *Vision Research*, 14:1039–1042.

[Neugebauer, 1937] Neugebauer (1937). Die theoretischen grundlagen des mehrfarbendruck. *Zeitschrift fur wissenschaftliche Photographie, Photophysik und Photochemie*, 36:73–89.

[Niessen et al., 1994] Niessen, W. J., ter Haar Romeny, B. M., and Viergever, M. A. (1994). Numerical analysis of geometry-driven diffusion equations. In ter Haar Romeny, B., editor, *Geometry-driven diffusion in computer vision*, pages 393–410, Dordrecht. Kluwer Academic Publishers.

[Nishikawa et al., 1965] Nishikawa, S., Massa, R. J., and Mott-Smith, J. C. (1965). Area properties of television pictures. *IEEE transactions on information theory*, 11(3):348–352.

[Okabe et al., 1992] Okabe, A., Boots, B., and Sugihara, K. (1992). *Spatial tessellations: Concepts and applications of voronoi diagrams*. John Wiley and Sons.

[Peli, 1990] Peli, E. (1990). Contrast in complex images. *Journal of Optical Society of America A*, 7(10):2032–2040.

[Perona et al., 1994] Perona, P., Shiota, T., and Malik, J. (1994). Anisotropic diffusion. In ter Haar Romeny, B., editor, *Geometry-driven diffusion in computer vision*, pages 73–92, Dordrecht. Kluwer Academic Publishers.

[Press et al., 1988] Press, W. H., Flannery, B. P., Taukolsky, S. A., and Vetterling, W. T. (1988). *Numerical recipes in C*. Cambridge University Press, Cambridge.

[de Ridder, 2001] de Ridder, D. (2001). *Adaptive measures of image processing.* PhD thesis, Delft University of Technology.

[de Ridder et al., 1998] de Ridder, D., Duin, R. P. W., Verbeek, P. W., and van Vliet, L. J. (1998). On the application of neural networks to non-linear image processing tasks. In *Proceedings International Conference on Neural Information Processing 1998 vol. I*, pages 161–165.

[de Ridder et al., 1995] de Ridder, H., Blommaert, F. J. J., and Fedorovskaya, E. A. (1995). Naturalness and image quality: chroma and hue variation in color images of natural scenes. In Rogowitz, B. E. and Allebach, J. P., editors, *Human Vision, Visual processing and digital display VI*, volume 2411, pages 51–61. SPIE.

[Robertson, 1977] Robertson, A. R. (1977). The CIE 1976 color difference formulae. *color research and application*, 2:7–11.

[Roufs, 1989] Roufs, J. A. J. (1989). Brightness contrast and sharpness, interactive factors in perceptual image quality. In Rogowitz, B. E., editor, *Human vision, visual processing and digital display*, volume 1077, pages 66–72. SPIE.

[Roufs, 1992] Roufs, J. A. J. (1992). Perceptual image quality: concepts and measurements. *Philips J.Res.*, 47:35–62.

[Roufs et al., 1994] Roufs, J. A. J., Koselka, V. I. F., and van Tongeren, A. A. A. M. (1994). Global brightness contrast and the effect on perceptual image quality. In Rogowitz, B. E. and Allebach, J. P., editors, *Human Vision, Visual Processing and Digital Display V*, volume 2179, pages 80–89. SPIE.

[Sagawa, 1999] Sagawa, K. (1999). Visual comfort to colored images evaluated by saturation distribution. *Color research and application*, 24(5):313–321.

[Sangwine and Horne, 1998] Sangwine, S. J. and Horne, R. E. N. (1998). *The colour imaging processing handbook.* Chapman and Hall, London, UK.

[van der Schaaf, 1998] van der Schaaf, A. (1998). *Natural image statistics and visual processing.* PhD thesis, University of Groningen.

[Sharma and Trussell, 1997] Sharma, G. and Trussell, H. J. (1997). Digital color imaging. *IEEE transactions on image processing*, 6(7):901–932.

[Sibson, 1981] Sibson, R. (1981). A brief description of natural neighbour interpolation. In Barnett, V., editor, *Interpreting Multivariate data*, pages 21–36. Wiley and Sons, NY, USA.

[Siegel and Castellan, 1988] Siegel, S. and Castellan, N. J. (1988). *Non parametric statistics for the behavioral sciences.* McGraw-Hill, New York, NY.

[Sjöberg, 1987] Sjöberg, L. (1987). Psychometric considerations in the dimensional analysis of subjective picture quality. *Displays*, 8(4):210–212.

[StatSoft, 2000] StatSoft (2000). *STATISTICA for Windows [Computer program manual]*. StatSoft, Inc, 2300 East 14th Street, Tulsa.

[Stone et al., 1988] Stone, M. C., Cowan, W. B., and Beatty, B. C. (1988). Color gamut mapping and the printing of digital color images. *ACM transactions on Graphics*, 7(4):249–292.

[Torgerson, 1958] Torgerson, W. S. (1958). *Theory and methods of scaling*. Wiley and sons, NY.

[Viggiano, 1990] Viggiano, J. A. S. (1990). Modeling the color of multi colored halftones. In *TAGA proceedings*, pages 44–62. TAGA.

[van Vliet, 1993] van Vliet, L. J. (1993). *Grey scale measurements in multi-dimensional images*. PhD thesis, Delft University of Technology.

[Walraven, 1992] Walraven, J. (1992). Color basics for the display designes. In Widdel, H. and Post, D. L., editors, *Color in electronic displays*, chapter 1.1, pages 3–38. Plenum Press, New York.

[Walraven et al., 1990] Walraven, J., Enroth-Cugell, C. H., Hood, D. C., MacLeod, D. I. A., and Schnapf, J. L. (1990). The control of visual sensitivity: receptoral and post-receptoral processes. In Spillman, L. and Werner, J. S., editors, *Visual perception: the neurophysiological foundations*. Academic press, San Diego, CA.

[Webster et al., 1993] Webster, A. A., Jones, C. T., Pinson, M. H., Voran, S. D., and Wolf, S. (1993). An objective video quality assessment system based on human perception. In Rogowitz, B. E. and Allebach, J. P., editors, *Human Vision, Visual Processing and Digital Display IV*, volume 1913, pages 15–26. SPIE.

[Westheimer, 1986] Westheimer, G. (1986). The eye as an optical instrument. In Boff, K. R., Kaufman, L., and Thomas, J. P., editors, *Handbook of perception and human performance*. Wiley and Sons, New York, NY.

[Winer, 1970] Winer, B. J. (1970). *Statistical priciples in experimental design*. Mc-Graw_Hill, London, UK.

[Wyble and Berns, 2000] Wyble, D. R. and Berns, R. S. (2000). A critical review of spectral models applied to binary color printing. *Color research and application*, 25(1):4–19.

[Wyszecki and Stiles, 1982] Wyszecki, G. and Stiles, W. S. (1982). *Color Science (2nd edition)*. Wiley, New York.

[Xia et al., 1999] Xia, M., Saber, E., Sharma, G., and Tekalp, A. M. (1999). End-to-end color printer calibration by total least squares regression. *IEEE transactions on image processing*, 8(5):700–716.

[Yendrikhovskij, 1998] Yendrikhovskij, S. N. (1998). *Color reproduction and the naturalness constraint.* PhD thesis, Technische Universiteit EIndhoven.

[Yendrikhovskij et al., 1999] Yendrikhovskij, S. N., Blommaert, F. J. J., and de Ridder, H. (1999). Towards perceptually optimal colour reproduction of natural scenes. In MacDonald, L. W. and Luo, M. R., editors, *Colour imaging: vision and technology*, chapter 18, pages 363–382. Wiley and Sons.

[Young et al., 1998] Young, I. T., Gerbrands, J. J., and Van Vliet, L. J. (1998). Image processing fundamentals. In Madisetti, V. K. and Williams, D. B., editors, *The Digital Signal Processing handbook*, chapter 51, pages 1 – 81. IEEE Press and CRC Press.

[Yule, 1944] Yule, J. A. C. (1944). Unsharp marks. *Photographic Journal*, 84:321–327.

[Zhang et al., 1997] Zhang, X., Silverstein, D. A., Farrell, J. E., and Wandell, B. A. (1997). Color image quality metric S-CIELAB and its application on halftone texture visibility. In *Compcon97 digest of papers*, pages 44–48.

[Zhang and Wandell, 1996] Zhang, X. and Wandell, B. A. (1996). A spatial extension of CIELAB for digital color image reproduction. *Proceedings of the SID symposiums*, pages 731–734.

[Zhang and Wandell, 1998] Zhang, X. and Wandell, B. A. (1998). Color image fidelity metrics evaluated using image distortion maps. *Signal processing*, 70(3):201–214.

Summary

This thesis is concerned with the topic of perceptual image quality. Our visual communication is mainly based on images, either natural or synthetic. Humans are able to judge whether an image is a good image. However, for many applications it would be faster to estimate the quality of images using a computer. The major goal of this thesis is to find an objective measure for the quality of a printed image that corresponds to perceptual image quality.

In addition to this major goal some closely related, secondary goals are defined as

Improving insight in the complex topic of perceptual quality. When estimating the quality, does a subject evaluate the image as a whole or only sub-parts of the image? How do current ideas about the relation between naturalness and quality hold up for various sets of printed images?

Determining the relation between perceptual attributes and image quality. How does the quality depend on certain perceptual attributes of the image, such as sharpness and color distribution?

Developing methods for color image processing. How can we optimize tools used for grey-value images, when these are applied to color images?

In chapters 2, 3 and 4 the theoretical background and experimental approach for this study are presented. In chapter 2 we give a short overview of how color can be quantified and what color variables are used. Chapter 3 reviews some of the current image quality models. A distinction is made between image fidelity models, where the quality of an image is related to a reference or an ideal model, and image quality models, where the quality is modeled directly, independent of any physical reference. In chapter 4 the methods and materials used in our experiments are presented. The topics discussed in that chapter include: how to characterize and calibrate a printer, what images can be used for this kind of experiments, and how to design experiments in which the measuring instrument is the human observer.

We have performed four groups of experiments, which are described in chapter 5, 6, 7 and 8. In chapter 5 we report new measures for sharpening and smoothing. These measures depend on the relation between the gradient magnitude of the pixels in the original image and the filtered image, respectively. The resulting sharpening

and smoothing measure correlated reasonably well with human perception. This means that these measures can be used to obtain an objective prediction of sharpening and smoothing. Because the subjects had less trouble with different levels of smoothing than with different levels of sharpening, the just noticeable difference of the smoothing measure must be smaller than that of the sharpening measure. The preference of the different images was also tested as a measure for quality. For these images, subjects prefer images in which the smoothing is low and the sharpening is high.

In these experiments, problems arise for images in which parts of the image require different sharpening and/or smoothing values. Because some parts of an image seemed more important for the image quality than others, the subjects based their overall decision on the evaluation of different parts of the image. This problem may be solved by an approach in which the measure is based on the results of subparts of the image. A complicating factor for this approach is, that it is unclear how to select the different subparts of the image.

In chapter 6 we report a new measure for describing the sharpness of an image, that is, as a feature of the image itself, rather than in relation to some original image. We proposed a sharpness measure for which lines and derivatives of edges are modeled as Gaussian profiles with different width. Lines and edges in the image are located, and their sharpness is determined by fitting the Gaussian line or edge profile to the Gaussian derivative signature. The sharpness measures depend on the width that is found. The resulting sharpness measures correlated reasonably well with human perception, so these measures can be used to obtain an objective prediction of the perceptual sharpness.

All sharpness and sharpening experiments were based on three different filters: Gaussian smoothing filters, unsharp masking filters and anisotropic diffusion filters. For all measures, we would expect that the results would also apply for certain other filters. Of course, these filters should alter the image in such a way, that the result is more or less "natural" to an observer.

The second perceptual attribute we studied was color, in particular the effect of two different color distribution manipulations of the image. In chapter 7 we report experiments on the effect of changing the luminance distribution of the image by a gamma manipulation. The chromaticity of the colors was kept constant. We studied the perceptual quality and naturalness of these images. In contrast to the literature [Roufs, 1989], we found for some images distinct differences between the subject responses. So, depending on the type of image, individual preferences may differ, even when explicitly looking for a uniform perceptual quality.

We found, as expected, that the quality and naturalness of all images depended on gamma. The dependency on gamma is influenced by the image that is evaluated. This could be due to the fact that the images are very different in lightness distri-

bution. The content was also very different, some images contained many so-called memory colors [1] and some had only a few memory colors. The gamma manipulation of the luminance also has an effect on the chroma, and therefore on the colorfulness of the image. Some images, for instance images that do not contain many colors, may benefit from more luminance contrast, but for other images the colorfulness may be more important than the luminance contrast. Using linear regression, we found that we could predict the quality of the images using the discriminability measure of the usefulness-naturalness model [Janssen and Blommaert, 2000b] in combination with the measured perceptual naturalness. The usefulness-naturalness model uses a naturalness measure, based on the memory colors of skin, grass or sky. We could not use that measure, because not all images in our set contained these colors. In conclusion, manipulation of the color distribution has a significant effect on perceptual quality. Subject responses are predictable in the sense that their effect on perceived quality shows a clear maximum; such predictability can be used for creating an overall objective quality measure. On the other hand, matters are complicated, because of individual preference.

In chapter 8 we present results of experiments in which both the chroma was scaled and the luminance distribution was varied by a gamma manipulation. In this experiment, the subjects did behave as one group, yielding closely corresponding responses to the experimental stimuli. As expected, both the gamma and chroma scaling affected perceptual quality. The gamma setting has a greater influence on the perceptual quality than the chroma scaling in this experiment. This can be caused by perceptual more relevant differences in gamma setting than in chroma scaling. Another explanation may be found in the fact that in natural environments gamma manipulation does not occur naturally, whereas chroma scaling may occur. Another possible explanation may be found in intrinsic differences in the stimulus production. The gamma manipulation is non-linear with fixed minimum and maximum luminance, whereas the chroma manipulation is achieved by a linear scaling factor, in combination with a fixed minimum. Using the discriminability measure of the usefulness-naturalness model [Janssen and Blommaert, 2000b] and our measure of subjective naturalness, we could not convincingly predict the quality of the gamma and chroma manipulated images (only 44% of the variance was explained). So, although this manipulation of the color distribution also has a significant effect on perceptual quality, the subject responses are not so predictable. On the other hand, effects of individual preferences were not found.

We studied image quality of images reproduced by a printer. However, images may seem very different when displayed on a monitor. One of the major differences between a printer and a monitor is the gamut, that is, the envelope of colors in color space that determine the limitations of the displaying device.[2] Nevertheless,

---

[1] Memory colors are colors that are recalled in association with familiar objects in long-term memory [Bartleson, 1968], such as the green of grass or the yellow of a banana.

[2] Another major difference is that a monitor generates its own colors, whereas the colors on a print are made by reflecting the light. This difference is not taken into account.

standard image enhancement techniques do not take the gamut into account. In the last part of this thesis we evaluate how we can incorporate the gamut into achromatic enhancement of *color* images (chapter 9) and how this affects image quality (chapter 10). In this way, some of the differences between images displayed on a monitor and printed images can be taken into account.

In chapter 9 a generic method is described to use grey value image processing algorithms for color image processing. The grey value algorithm is applied to the lightness component of the color image in such a way that the colors of the image can still be rendered. The result of an image processing step can be seen as a point for point correction of the original image. We propose that the maximum and minimum lightness correction should depend on the available gamut space in a chosen direction. The proposed lightness correction is a function of the original color point, the desired lightness correction and the maximum and minimum lightness corrections. Because of the monotonic character of the lightness correction, all points that are physically different before the lightness correction remain different. However, in the case of lightness compression, differences that are physically still present, may nevertheless become invisible when they fall below the visual threshold.

The method was tested for two grey value image processing algorithms: sharpening and contrast improvement using gamma manipulation. For both applications, the recipes for which the chroma was kept constant gave the best results.

In chapter 10 an outline for a quality measure is suggested. This quality measure is useful to select the best image of a series of images that only differ in the distribution of the lightness. We found in chapter 9, that improvements in image quality within the gamut are best done in such a way that manipulations should mainly be confined to the achromatic domain. This leaves lightness (or luminance), rather than hue and chroma as the best choice for quality improvement.

We speculated that an image in which the colors fill the available gamut space relatively homogeneously have a higher quality than an image for which this is not the case. The basis for that idea is that a more homogeneous distribution allows for more discriminable nuances. The homogeneity can be expressed in a number varying between 0 and 1. The measure is one for a center of mass that is located exactly between the upper gamut boundary and the lower gamut boundary. We consider this position optimal because the lightness nuances surrounding that point can be stretched over equal ranges of lighter and darker tones. For all other center of masses, the measure is smaller than one. The results of preliminary tests with such a model on three small images are promising.

Samenvatting

Het onderwerp van dit proefschrift is perceptieve beeldkwaliteit. Onze visuele communicatie is vooral gebaseerd op beelden, zowel op natuurlijke als op synthetische beelden. Mensen zijn uitstekend in staat om te beoordelen of de kwaliteit van een beeld ook echt goed is. Voor veel applicaties zou het echter sneller en goedkoper zijn om de kwaliteit van een beeld met een computer te bepalen. Het hoofddoel van dit proefschrift is om een objectieve maat voor de kwaliteit van een geprint beeld te vinden, die goed correleert met perceptieve beeldkwaliteit.

Naast dit hoofddoel zijn een aantal gerelateerde doelstellingen bepaald:

Het verbeteren van het inzicht in het complexe onderwerp van perceptieve beeldkwaliteit. Is de kwaliteitsbepaling van personen gebaseerd op het gehele beeld of op bepaalde gedeelten van het beeld? Hoe goed zijn bestaande ideeën over the relatie tussen natuurlijkheid en kwaliteit toepasbaar voor verschillende sets van geprinte beelden?

Het bepalen van de relatie tussen perceptieve attributen en beeldkwaliteit. Hoe hangt de kwaliteit van perceptieve variabelen af, zoals van scherpte en kleurdistributie?

Het ontwikkelen van methodes voor kleurenbeeldbewerking. Hoe kunnen we bestaande beeldbewerkingmethoden voor grijswaardenbeelden zo goed mogelijk toepassen op kleurenbeelden?

In de hoofdstukken 2, 3 en 4 worden de theoretische achtergronden en de experimentele aanpak van deze studie besproken. In hoofdstuk 2 wordt een korte introductie gegeven over hoe kleur gekwantificeerd kan worden en worden de gebruikte kleurenvariabelen geïntroduceerd. In hoofdstuk 3 worden een aantal bestaande beeldkwaliteitsmodellen besproken. Er wordt hier onderscheid gemaakt tussen betrouwbaarheidsmodellen, waar de kwaliteit van een beeld afhangt van een referentiebeeld of een origineel beeld, en beeldkwaliteitsmodellen, waarmee de kwaliteit van een beeld onafhankelijk van een fysieke referentie "gemeten" kan worden. In hoofdstuk 4 worden de methodes en materialen gepresenteerd die gebruikt zijn in de experimenten, waarin de mens het meetinstrument is. In dit hoofdstuk wordt ook besproken hoe een printer gekarakteriseerd en gekalibreerd kan worden en wat voor

beelden gebruikt worden in de experimenten.

Er werden vier experimenten uitgevoerd, die worden besproken in de hoofstukken 5, 6, 7 en 8. In hoofdstuk 5 bespreken we nieuwe maten voor opscherping en effening. Deze maten zijn afhankelijk van de relatie tussen de magnitude van de gradiënt van de pixels in het originele beeld enerzijds en de magnitude van de gradiënt van de pixels in het gefilterde beeld anderzijds. De gevonden opscherpings- en effenings- maten bleken goed genoeg te correleren met de menselijke perceptie, om gebruikt te kunnen worden als voorspellers van de subjectief waargenomen verschillen tussen de testbeelden.

Omdat de proefpersonen minder moeite hadden met het onderscheiden van de verschillende niveaus in effening dan in opscherping, concluderen we dat de scha- ling van de twee maten niet perceptief gelijk is. Het nog juist-onderscheidbare- verschil in de effening, uitgedrukt in de effeningsmaat, is kleiner dan het nog-juist- onderscheidbare verschil in opscherping, uitgedrukt in de opscherpingsmaat. Ook hebben we de proefpersonen gevraagd naar het beste plaatje uit een serie van ver- schillende opscherping en effening. De proefpersonen prefereerden beelden waarin de effening klein was en de opscherping groot.

Er ontstonden problemen voor beelden waarin delen van het beeld verschillende opscherpings- of effeningswaarde hadden. De proefpersonen bepaalden hun oordeel op sommige van deze delen. Helaas vonden de proefpersonen niet altijd dezelfde delen van het beeld belangrijk. Door de maat te baseren op een beperkt aantal delen van het beeld kan dit probleem enigszins opgelost worden. Een complicerende factor is dat de belangrijke delen in het beeld dan wel geselecteerd moeten worden.

In hoofdstuk 6 wordt een nieuwe maat voor de scherpte van een beeld beschreven. Met deze maat wordt de scherpte bepaald als kenmerk van het beeld zelf, in plaats van in verhouding tot een origineel beeld. Lijnen en afgeleiden van randen wor- den hier gemodelleerd als Gaussische profielen met verschillende breedte. Nadat de lijnen en randen in het beeld gedetecteerd zijn, wordt de scherpte ervan bepaald door de afgeleide van het profiel te "fitten" aan de afgeleiden van lijnen en ran- den. De voorgestelde scherptemaat wordt bepaald door de gevonden breedte. De resulterende scherptemaat correleerde redelijk goed met de menselijke perceptie, en kan dus gebruikt worden om een objectieve voorspelling van perceptieve scherpte te doen.

Alle scherpte- en opscherpingsexperimenten zijn gebaseerd op drie verschillende filters: Gaussische effeningsfilters, "unsharp masking" filters en anisotrope diffusie filters. We verwachten dat alle voorgestelde maten ook zouden moeten werken voor andere filters, althans wanneer deze filters het beeld zo veranderen, dat mensen het resultaat min of meer natuurlijk vinden.

In hoofdstuk 7 en 8 wordt een tweede perceptieve kenmerk bestudeerd, namelijk

kleur, dat wil zeggen het effect van twee verschillende operaties die de kleurdistributie van het beeld veranderen. In hoofdstuk 7 worden experimenten beschreven waarin het effect van luminantieverandering door gamma-manipulatie centraal staat. Hierbij bleef de chromaticiteit (x, y, "kleur") constant. Met behulp van proefpersonen is de perceptieve kwaliteit van verschillende beelden bepaald. In tegenstelling tot resultaten uit de literatuur [Roufs, 1989], hebben we voor een aantal beelden duidelijke verschillen tussen de voorkeuren van proefpersonen gevonden. Individuele voorkeuren van mensen kunnen dus verschillen, zelfs wanneer het experiment erop is gericht een relatief eenvoudige variabele, gamma, te isoleren.

Zoals verwacht, bleek de kwaliteit en natuurlijkheid van een beeld af te hangen van de gamma waarmee het beeld gemaakt is. Deze afhankelijkheid van gamma verschilt per testbeeld. Dit kan verklaard worden door grote verschillen in de verdeling van de lichtheid in de testbeelden. De inhoud was ook erg verschillend, sommige beelden bevatten veel zogenoemde geheugenkleuren[3], waar andere beelden er maar een paar hadden.

De gamma-manipulatie beïnvloedt de chroma (C^*), en daardoor de kleurigheid van het beeld. Sommige beelden, zoals beelden die weinig kleuren bevatten, kunnen iets winnen met meer luminantiecontrast, maar voor andere beelden kan de kleurigheid juist belangrijker zijn dan het luminantiecontrast.

Met lineaire regressie kon worden aangetoond dat de kwaliteit van de beelden redelijk goed voorspeld kan worden met de onderscheidbaarheidsmaat van het bruikbaarheid-natuurlijkheid-model [Janssen and Blommaert, 2000b], in combinatie met de gemeten perceptieve natuurlijkheid. In het bruikbaarheid-natuurlijkheid-model zit een natuurlijkheidsmaat, die gebaseerd is op de geheugenkleuren van huid, gras en lucht. Omdat niet alle gebruikte beelden tenminste een van deze kleuren gebruikten, konden wij deze maat niet gebruiken. Gebruik makend van de perceptief gemeten natuurlijkheid bleek het effect van gamma op de waargenomen kwaliteit een duidelijk maximum te vertonen, waardoor de oordelen van de proefpersonen enigszins voorspelbaar worden. Deze voorspelbaarheid kan gebruikt worden voor een algemene objectieve kwaliteitsmaat. Aan de andere kant wordt de zaak gecompliceerd door individuele voorkeuren.

In hoofdstuk 8 worden experimenten beschreven waarin de luminantie distributie is veranderd met een gamma-manipulatie en tegelijkertijd de chroma geschaald is. In dit experiment gedroegen de proefpersonen zich wel als een homogene groep, dat wil zeggen, er zijn geen significante verschillen tussen de proefpersonen gevonden. De kwaliteit werd zowel door de gamma- als door de chroma-schaling beïnvloed. De waarde van gamma had echter een veel grotere invloed op de kwaliteit dan de waarde van chroma. Dit zou verklaard kunnen worden door het feit, dat in

---

[3]Geheugenkleuren zijn kleuren die geassocieerd worden met bekende objecten en opgeslagen in zijn het lange-termijngeheugen [Bartleson, 1968], zoals het groen van gras of het geel van een banaan.

natuurlijke omgevingen gamma-manipulatie niet optreedt en chroma-schaling wel. Gamma manipulatie zal dus al snel als onnatuurlijk worden ervaren. Een andere mogelijke verklaring zou gevonden kunnen worden in intrinsieke verschillen bij het maken van de beelden. De gamma-manipulatie is niet lineair, terwijl de minimale en maximale luminantie vast staan. De chroma-manipulatie is een lineaire schaling, waarbij alleen het minimum vaststaat.

In tegenstelling tot de resultaten van het gamma-experiment, lieten de kwaliteits-beoordelingen in het gamma- en chroma-experiment zich niet goed voorspellen door de onderscheidbaarheidsmaat van het bruikbaarheid-natuurlijkheid-model en de gemeten subjectieve natuurlijkheid; niet meer dan 44 % van de variantie kon ver-klaard worden. Kennelijk heeft deze manipulatie van de kleurendistributie ook een significant effect op de perceptieve kwaliteit. De voorkeuren van de proefpersonen waren niet zo voorspelbaar, ondanks het uitblijven van individuele voorkeuren.

Dit onderzoek houdt zich bezig met de beeldkwaliteit van beelden die gerepro-duceerd worden met een printer. Dezelfde beelden kunnen er echter op een beeld-scherm heel anders uitzien, een probleem dat centraal staat bij het ontwikkelen van *color management* systemen. Een van de belangrijkste oorzaken van het verschil in monitor- en printer-beeld is het verschil in kleurenbereik, doorgaans aangeduid als de gamut.[4] De standaard beeldbewerkingtechnieken nemen (de beperkingen van) deze gamut niet mee. In het laatste gedeelte van dit proefschrift hebben we onderzocht hoe de gamut wel meegenomen kan worden in de achromatische verbe-tering van kleurenbeelden (hoofdstuk 9) en hoe dat uitwerkt voor de beeldkwaliteit (hoofdstuk 10). Hierbij gebruiken we lichtheid in plaats van luminantie, omdat lichtheid eenvoudiger samen met kleurtint te gebruiken is.

In hoofdstuk 9 introduceren we een generieke methode waarmee beeldbewerkings-algoritmes voor grijswaardenbeelden gebruikt kunnen worden voor kleurenbeeldbe-werking. Het grijswaarden-algoritme wordt toegepast op de lichtheidcomponent in het beeld op zo'n manier dat het beeld nog steeds met de juiste kleuren weergegeven kan worden. Het resultaat van een beeldbewerkingstap kan gezien worden als een puntsgewijze correctie van het originele beeld. In ons voorstel hangen de maxi-male en minimale lichtheidcorrectie voor een pixel af van de beschikbare gamut in een bepaalde richting. De voorgestelde lichtheidcorrectie wordt bepaald door het originele punt, de gewenste lichtheidcorrectie en de maximale en minimale licht-heidscorrecties. Omdat de lichtheidcorrectie monotoon is, zijn alle punten die voor de beeldbewerkingsstap verschillend waren na deze stap nog steeds verschillend. Verschillen die fysiek nog steeds aanwezig zijn, zouden echter toch niet zichtbaar kunnen zijn als ze onder de visuele drempel vallen.

---

[4]Een ander groot verschil tussen een monitor en een printer is dat een monitor zijn eigen kleuren maakt, terwijl de kleuren op een print ontstaan doordat het omgevingslicht dat op de print valt gereflecteerd wordt. Dit verschil tussen monitors en printers wordt niet meegenomen.

De voorgestelde methode is getest met twee (grijswaarden-)beeldbewerkingsalgorit-mes: opgescherping en contrast-verbetering met behulp van een gamma-manipulatie. Voor beide methodes werd het beste resultaat bereikt als de chroma constant gehouden werd.

In hoofdstuk 10 wordt een aanzet gegeven voor het ontwikkelen van een kwaliteits-maat. Deze kwaliteitsmaat kan gebruikt worden om uit een serie beelden die uit-sluitend verschillen in lichtheidsdistributie, de beste te selecteren. In hoofdstuk 9 laten we zien dat men bij lichtheidsmanipulaties het beste de chroma constant kan houden. Dit betekent dat lichtheid het beste onafhankelijk gemanipuleerd kan worden van kleurtint of chroma.

Het idee achter de kwaliteitsmaat is dat een beeld, waarin de punten de beschikbare gamut ruimte relatief homogeen vullen, een hogere kwaliteit heeft dan een beeld waarvoor dit niet het geval is. Deze veronderstelling gaat ervan uit dat een homo-genere distributie meer onderscheidbare nuances bevat. De mate van homogeniteit kan worden uitgedrukt in een getal dat varieert tussen 0 en 1. De maat heeft de waarde 1 voor een zwaartepunt dat precies tussen de twee randen van de gamut ligt. We beschouwen deze positie als optimaal omdat de lichtheids nuances die om dit punt liggen uitgestrekt kunnen worden over gelijke ranges van lichte en donkere kleuren. Voor alle andere posities van het zwaartepunt is de voorgestelde maat kleiner. De eerste resultaten van dit kwaliteitsmodel op drie kleine beelden zijn veelbelovend.

Curriculum vitae

Judith Dijk was born in Düsseldorf, Germany, on the February 21, 1974. In 1992 she graduated from the Stedelijk Gymnasium Arnhem. In the same year she started a study in Applied Physics at Delft Universiteit of Technology. She obtained her Masters Degree in the Pattern Recognition Group in 1998.

In January 1998 she started a PhD in an IOP beeldbewerking project called "Image quality measures for the perception of printing". This project was a collaboration between the Applied Physics department in Delft and TNO Human Factors in Soesterberg. Her supervisors were prof. dr. Ian T. Young and dr. Piet W. Verbeek in Delft, and dr. Jan Walraven in Soesterberg.

In October 2003 she started working at the TNO Physics and Electronics Laboratory as a research scientist.

Dankwoord

Dit proefschrift werd mede mogelijk gemaakt door...

Een proefschrift en het bijbehorende onderzoek voltooi je niet alleen, daar heb je gelukkig hulp bij. Bij deze wil ik alle mensen bedanken, die me hierbij geholpen hebben. Een aantal mensen wil ik bij naam noemen.

De meeste dank ben ik verschuldigd aan mijn begeleiders. Piet Verbeek, mijn begeleider in Delft, genereerde een bijna oneindige hoeveelheid goede ideeën. Jan Walraven, mijn begeleider in Soesterberg, introduceerde me in de kleurenwereld. Ook na zijn pensionering was Jan altijd bereikbaar voor begeleiding en commentaar op geschreven hoofdstukken. Ted Young, die begon als mijn begeleider maar wegens tijdgebrek de dagelijkse begeleiding aan Piet overgedragen heeft, wil ik bedanken voor de opzet van het project en de kritische opmerkingen bij de voltooiing van mijn proefschrift.

Het beeldbewerkingsgedeelte van mijn onderzoek heb ik uitgevoerd binnen de vakgroep Patroonherkennen aan de TU Delft. Hier moet ik met name de koffiepauzes en de PH-dagjes uit noemen. Behalve nuttig was het in Delft vaak ook erg gezellig. Ik wil mijn kamergenoten Mike, Peter, Bernd, Tuan, Geert, Kees en bijna-kamergenoten Cris en Frank, samen met de overige leden van ph-aio bedanken voor deze leuke tijd. Barbecues op het strand of Mike's dakterras, vrijdags naar het TPKV en het spelen van Kolonisten (na de juiste hoeveelheid sla ?).

Het psycho-fysische gedeelte van mijn onderzoek heb ik gedaan op de afdeling Waarneming van TNO Technische Menskunde. Johan Alferdinck en Marcel Lucassen hebben me geholpen met de opzet van mijn experimenten. Ook de rest van de visuele groep stond altijd open voor vragen en hulp. Mij wisten ze regelmatig te vinden als er een proefpersoon nodig was.

Het onderzoek dat ten grondslag ligt aan dit proefschrift werd uitgevoerd in samenwerking met Océ in Venlo. Ik wil Frans Gaykema, Toon van Dijk, Paul Hofman, Etienne van Dorsselaer, Ruud Janssen en Bas Middelbos bedanken voor hun samenwerking en nuttige adviezen.

Mijn ouders wil ik bedanken voor hun steun. Tenslotte ben ik ook veel dank verschuldigd aan Sander. Zonder zijn steun en motivatie was dit proefschrift nooit afgekomen. Samen met Onno toont hij me dagelijks dat er veel belangrijker dingen in het leven zijn dan promoveren.

Hij maakt het verschil.