TAMPERE UNIVERSITY OF TECHNOLOGY
Degree Programme in Information Technology

Ossi Pirinen

High Dynamic Range Imaging in Luminance-Chrominance Space

Master of Science Thesis

Examiners: Dr. Atanas Gotchev
Dr. Alessandro Foi
Subject approved by departmental council on 9.10.2006

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Tietotekniikan koulutusohjelma
Ossi Pirinen: High Dynamic Range Imaging in Luminance-Chrominance Space
Diplomityö, 61 sivua
Elokuu 2007
Pääaine: Signaalinkäsittely
Tarkastajat: Dr. Atanas Gotchev, Dr. Alessandro Foi
Avainsanat: High Dynamic Range, HDR, Tone Mapping, Luminance-Chrominance

Korkean dynaamisen alueen kuvantaminen on moderni ratkaisu digitaalikameroiden sensoriteknologian rajoituksiin. Aihetta on käsitelty harmaasävy- sekä RGB-kuvien osalta melko syvällisesti. Kumpikaan näistä lähestymistavoista ei kuitenkaan käsittele väritietoa sen ansaitsemalla tarkkuudella. Tässä diplomityössä esitellään uusi, tehokas lähestymistapa värin käsittelyyn korkean dynaamisen alueen kuvantamisessa.

Olemassaolevista tekniikoista poiketen kehitetään menetelmä koko korkean dynaamisen alueen kuvantamisprosessin siirtämiseksi luminanssi-krominanssi -väriavaruuteen. Menetelmä on laskennallisesti tehokas, sekä välttää olemassaolevien, perinteisten menetelmien värikanavien erillisestä käsittelystä johtuvat värivirheongelmat. Tämä saavutetaan tunnustamalla luminanssin ja krominanssin erilaiset luonteet sekä käsittelemällä niitä asianmukaisesti. Käytännössä perinteisistä menetelmistä tuttu kameravaste luodaan ainoastaan luminanssikanavalle. Korkean dynaamisen alueen luminanssi kootaan vasteen avulla. Krominanssit kootaan perustuen värinformaatiovetoiseen painotusmenetelmään. Työssä osoitetaan menetelmän tuottavan luonnollisia, silmää miellyttäviä kuvia jotka on asianmukaisesti muunnettu matalan dynaamisen alueen näyttölaitteita varten. Myös esitellyn menetelmän parempi kohinansieto osoitetaan sekä visuaalisesti että numeerisesti.

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Information Technology
Ossi Pirinen : High Dynamic Range Imaging in Luminance-Chrominance Space
Master of Science Thesis, 61 pages
August 2007
Major: Signal Processing
Examiners: Dr. Atanas Gotchev, Dr. Alessandro Foi
Keywords: High Dynamic Range, HDR, Tone Mapping, Luminance-Chrominance

High dynamic range (HDR) imaging is a modern solution to the problem of sensor capability limitations in digital imaging. The issue has been addressed for gray scale as well as RGB (red, green, blue) images. Color is however not taken into account adequately profoundly in neither. This thesis presents a novel and efficient approach to color in high dynamic range (HDR) imaging.

In contrast to state-of-the-art methods, the complete HDR imaging process is moved from RGB to a luminance-chrominance color space. The result is a computationally efficient technique which avoids possible color distortions originating from the three RGB color channels processed separately. To achieve this, a camera response function is built for the luminance channel only and the HDR luminance is weighted and composed accordingly, while for the chrominance channels a saturation level dependent weighting is applied. It is demonstrated that the described technique yields natural and pleasant to perceive tone-mapped images and is also more robust to noise.

PREFACE

This thesis was done at the Transforms and Spectral Methods Group in the Tampere University of Technology, Institute of Signal Processing. First and foremost I would like to thank my thesis supervisor Doctor Atanas Gotchev. I would also like to extend my gratitude for Dr. Alessandro Foi, who played a central role in brainstorming many of the ideas crucial for this thesis. Thanks also to numerous friends for their tips and constructive criticism. I also express my appreciation to my wife for pushing me forward when needed.

Tampere, August 7, 2007

Ossi Pirinen
Yliopistonkatu 39 D 38
33500 Tampere
Tel. 040-5882435

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

CCD - Charge-Coupled Device
HDR - High Dynamic Range
HVS - Human Visual System
ICC - International Color Consortium
JPEG - Joint Photographic Experts Group
LDR - Low Dynamic Range
NRSTD - Normalized Standard Deviation (of noise)
SNR - Signal to Noise Ratio
sRGB - Standard RGB (Red, Green, Blue)
TIFF - Tagged Image File Format

1. INTRODUCTION

Dynamic range is the term used in many fields to describe the ratio between the highest and the lowest value of a variable quantity. In imaging and especially in display technology dynamic range is also known as the contrast ratio or, simply contrast, which denotes the brightness ratio between black and white pixel visible on the screen at the same time. For natural scenes the dynamic range is the ratio between the density of luminous intensity of the brightest sunbeam and the darkest shadow. In photography the unit of luminance, cd/m^2 (candelas per square metre), is also known as a 'nit'.

It is very difficult to draw the exact line between low and high dynamic range (HDR and LDR, respectively). A coarse division into 'high' and 'low' dynamic range can however be made based on, for example, the usual differentiation ability of a digital camera; while a digital camera is, with the use of shutter speed and aperture combinations, able to adapt to a range of lighting conditions, the gamut (a complete extent or range) of the capture is limited by the relation of saturation vs. lowest differentiable current of the sensor wells. The captured image is exposed according the light characteristics of an area of the scene. This area can be exposed correctly but depending on the scene other areas may be over- or underexposed. Usually the output of a consumer level digital camera is scaled so that zero represents black and 255 white. This is the intensity resolution of a conventional 8 bit digital imaging system. Now if an imaging system significantly exceeds this limitation, it can be said to have high dynamic range. The number of quantization steps, 256 for an 8-bit system, while not so relevant in context of the definition of dynamic range, should be taken into account as well, as one would be able to devise an utterly useless two-step coding with close to infinite dynamic range.

Visual (digital) representation of natural scenes has reached a point where spatial resolution is no longer an issue and greater realism is achieved by either adding the third dimension or utilizing a more and more realistic gamut of light and color. The latter is achieved through a group of techniques commonly known as HDR imaging. This thesis approaches the problem of HDR, also titled by some as 'the new color-TV', in luminance-chrominance space.

Throughout the history of photography, the limits in the ability to capture and represent the gamut of light in a scene have presented a problem. Historically, the

limitations have been met in both capturing and representation and partly overcome in the latter [2]. During the course of the years photographers have become to see these limitations as a feature creating artistic effects, rather than a problem. Everyone who has ever used a camera is, regardless of the personal opinion over the nature of the issue, familiar with the situation of shooting a photo indoors with a window looking out to a sunlit scene in the field-of-view. If the indoor part of the scene is correctly exposed, the window with the outdoor scene is surely over exposed. This is a direct result of the cameras limited capacity to distinguish between different amounts of light. Figures 1.1, 1.2 and 1.3 illustrate this situation and the solution suggested by HDR imaging techniques. The conventional LDR snapshot of Figure 1.1 has a very limited dynamic range, which is seen as clipped bright parts as well as under exposed dark parts. The HDR representation of Figure 1.2 has all the information but only a fraction of it is visible on an LDR media, such as the print-out. The tone mapped HDR image of Figure 1.3 boasts the greatest amount of visible information. Tone mapping is thusly a mere brightness (luminance) range compressor using one of numerous techniques to bring the information of an image into displayable range. A further advantage is the notable noise suppression compared with the LDR frame. These properties are further discussed in this thesis.

Figure 1.1: An LDR snapshot of an HDR scene.

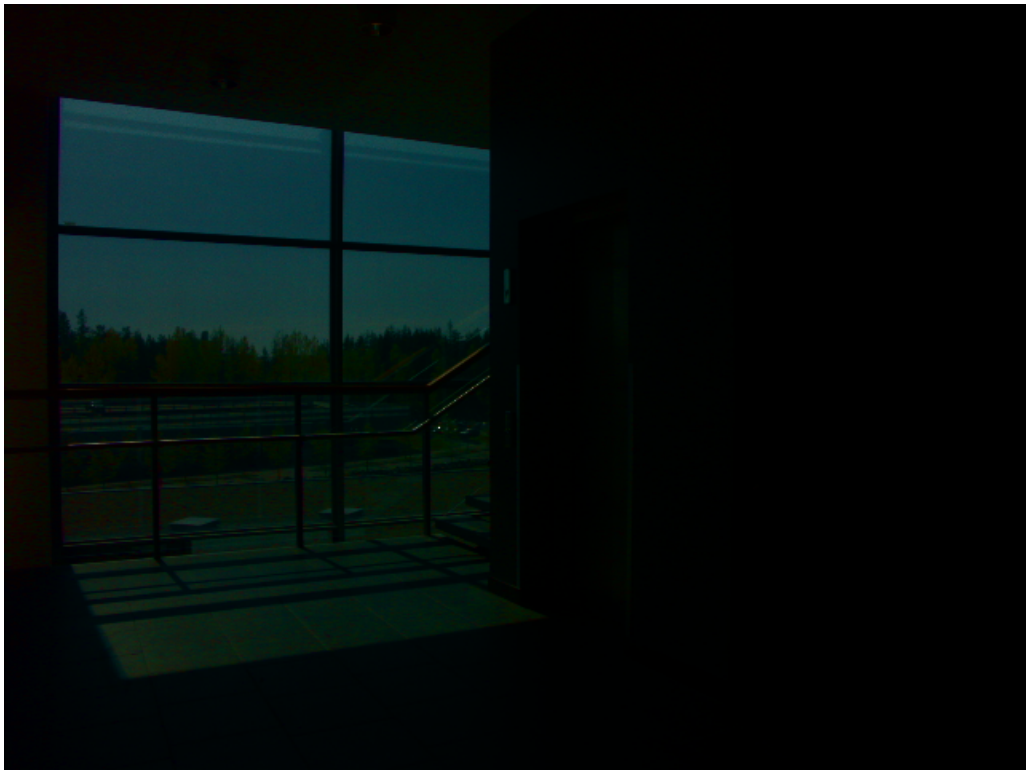

Figure 1.2: Unmapped HDR image. The need for tone mapping is evident.

1.1 Definition of the Problem

The problem of high dynamic range imaging is two-fold. First, how to capture the true luminances, and the full chromatic information, of a scene and second, how to faithfully represent this information on a device that is incapable of reproducing the actual luminances of neither the darkest nor the brightest spots, nor the true colors. The first half of the issue is commonly referred to as high dynamic range *composition* or *recovery* and the latter as *compression*, or more commonly, *tone mapping*. In this thesis the terms used are composition for the acquisition process and tone mapping for the displaying part.

HDR images can be acquired by capturing real scenes or by rendering 3D computer graphics, using techniques like radiosity and ray tracing. HDR images of natural scenes can be acquired either by capturing multiple images of the same scene with different exposures or utilizing only recently introduced special cameras able to directly capture HDR data [24]. The multiple exposure method relies on combining the different exposures into a single image, spanning the whole dynamic range of the scene [18], [5]. In this thesis the focus is on the approach of multiple capture composition. The image composition step requires a preliminary calibration of the camera response [5], [19].

The second part of the problem then is to appropriately tone map the so-obtained

Figure 1.3: Tone mapped HDR image showing a superior amount of information compared with the two previous figures.

HDR image back to a low-dynamic range (LDR) display. Tone mapping techniques range in complexity from a simple gamma-curve to sophisticated histogram equalization methods and complicated lightness perception models [15], [6], [13]. Recently, also display devices equipped with extended dynamic range have started to appear [25].

State-of-the-art HDR imaging methods have been developed for RGB color models. However, techniques working in luminance-chrominance seem more meaningful and preferable for a number of reasons. *First*, decorrelated color spaces offer better compressibility. As a matter of fact, the near totality of image compression techniques store images in some luminance-chrominance space. When one starts with already-compressed multiple-exposure LDR images, it is more efficient to compose the HDR image directly in the same color space. The resulting HDR image is then better suited for compression and, if to be displayed, it can be mapped to sRGB during the tone mapping stage. *Second*, any HDR technique operating in RGB space requires post-composition white balancing since the three color channels undergo parallel transformations. While the white balancing would yield perceptually convincing colors, they might not be the true ones. For the sake of hue preservation and better compression, it is beneficial to opt for a luminance-chrominance space, even if the input data is in RGB, e.g. an uncompressed TIFF image. *Third*, the

luminance channel, being a weighted average of the R, G, and B channels, enjoys a better signal-to-noise ratio (SNR), which is crucial if the HDR imaging process takes place in noisy conditions. In this thesis, the problem of HDR imaging in a generic luminance-chrominance space is addressed, efficient algorithms for HDR image composition and tone mapping are proposed, and the benefits of such an approach emphasized.

1.2 Organization of Thesis

The thesis is organized as follows. In Section 2 a brief overview of the main properties of relevant color models is given, with main focus being on luminance-chrominance color spaces. This section features also the notations used throughout the thesis. Existing methods and technologies for HDR composition, compression and rendition are discussed in Section 3. HDR image composition, including camera response definition and processing the luminance and chrominance channels, is presented in Section 4. An adapted tone mapping method for luminance-chrominance HDR images is presented in Section 5. Examples demonstrating the viability of the approach are given in Section 6 and finally some conclusions are drawn in Section 7.

2. LUMINANCE-CHROMINANCE COLOR SPACES

Generally speaking, the retina has three types of color photo-receptor cells, known as cones. These respond to radiation with a different spectral response, i.e. different cones generate the physical sensation of different colors, which is combined by the human visual system (HVS) to form the color image. From this it is intuitive to describe color with three numerical components, a tri-component vector. All the possible values of this vector then form a vector space called a color space or color model. The three components can be defined in various meaningful ways, which leads to definition of different color spaces [22], [24].

2.1 CIE XYZ Color Space

One of the first mathematically defined color spaces was the CIE XYZ color space, which is based on the 1931 standard observer color curves. To explicitly relate one color model to another, color space transformations are defined. Linear transforms can be defined as matrix multiplication operators while the definitions for nonlinear transforms are much more complex. The vast majority of color space transformations are defined to and from the CIE XYZ, even though to/from RGB transformations are nowadays usually defined as well.

From the XYZ tristimulus values we get the chromaticity coordinates as

$$\begin{aligned} x &= \frac{X}{X+Y+Z} \\ y &= \frac{Y}{X+Y+Z} \\ z &= \frac{Z}{X+Y+Z} \end{aligned}$$

of which z can be defined as $z = 1-x-y$ and thus only the x and y need to be kept. Now from x and y we get the CIE xy chromaticity diagram given in Figure 2.1. It is noted that the chromaticity diagram represents every realizable color with a point within a well defined boundary. This boundary represents the pure primary source colors. Thus the chromaticity diagram can be used as a starting point in defining other color spaces.

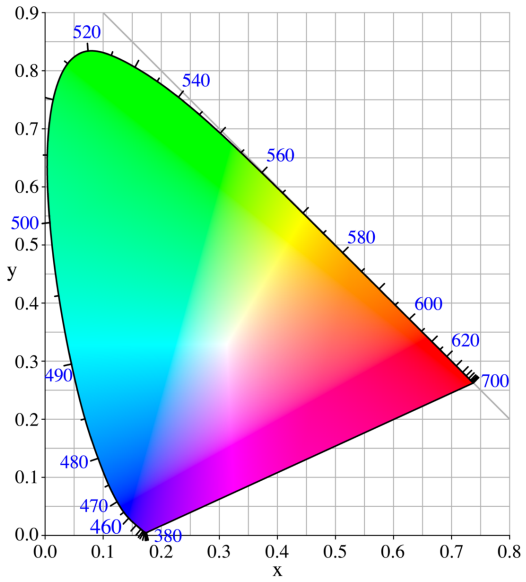

Figure 2.1: The CIE chromaticity diagram defined by the two chromaticity components, x and y.

2.2 RGB Color Spaces

An RGB tristimulus color space can be defined as the Cartesian coordinate system based model with red, green and blue primary spectral components [11]. All the standardized RGB (Red, Green, Blue) color spaces can be defined by giving the chromaticities (x and y) of each primary color and the white reference point. The reference white point serves to define the white "color" inside the gamut of the defined color space. RGB model is the most often used representation in computer graphics and multimedia, and different RGB spaces (primary color and white point combinations) have been standardized for different applications, [26]. What they all have in common is the idea to mix red, green and blue primaries in different relations to produce any given color inside the gamut of that space. All the RGB color models can be categorized under the definition *physiologically inspired color models* because the three primaries have been designed with the idea of matching the three different types of cones in the human retina. For implementation simplicity, the RGB coordinates are usually limited to $[0\ 1]^3$ (floating point) or $[0\ 255]^3$ (8-bit unsigned integer). These limited primaries then define the RGB color solid, which is conveniently shaped like a cube illustrated in Figure 2.2. The diagonal from origin $(0, 0, 0)$ to (max, max, max) defines the gray scale line (if all RGB components are equal, the color is gray, increased intensity moves the value from dark to light).

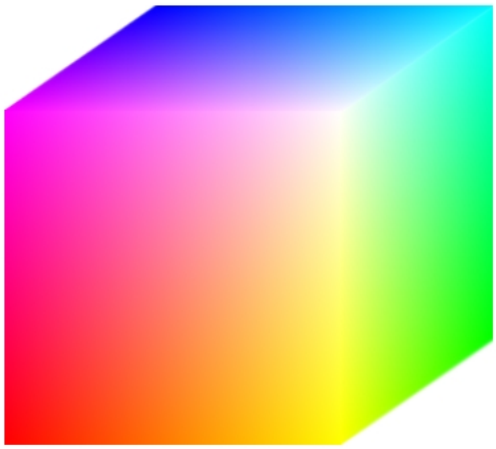

Figure 2.2: The RGB color gamut, shaped like a cube.

2.3 Luminance-Chrominance Color Spaces

Even though technology has steered the selection of the color model used in computer graphics toward the RGB models, that is not the *psychophysical* color interpretation. Perceptual features, namely brightness (luminance), saturation and hue, form the color sensation in HVS (human visual system). These features can be intuitively formalized through the introduction of luminance-chrominance color models.

Consider a generic luminance-chrominance color space linearly related to the RGB space. Loosely speaking, such a space is characterized by an achromatic luminance component, which corresponds to the gray scale part of the image, and two chrominance components, which are orthogonal to gray.

Throughout the thesis the following notations are used. The luminance and the two chrominance channels are denoted by Y, U, and V, respectively. Let $\mathbf{z} = [z^R, z^G, z^B]$ be an image in the RGB space and $\boldsymbol{\zeta} = [\zeta^Y, \zeta^U, \zeta^V]$ be the same image in the luminance-chrominance space (Roman letters are used to denote images in RGB and the corresponding Greek letters to denote images in luminance-chrominance). Transformation of the image from RGB to luminance-chrominance is defined in matrix form as $\boldsymbol{\zeta} = \mathbf{z}\mathbf{A}$, where the matrix $\mathbf{A}$ is normalized in such a way that if $\mathbf{z}(\cdot) \in [0,1]^3$ then $\boldsymbol{\zeta}(\cdot) \in [0,1] \times [-0.5, 0.5]^2$. Because of this constraint, the first column of $\mathbf{A}$ has all elements positive, $a_{j,1} \geq 0$. It is further assumed that $\sum_{j=1}^{3} a_{j,1} = 1$, thus ensuring that the $[0,1]$ range of the luminance component is fully utilized. Examples of such luminance-chrominance spaces are the "*opponent*" , the *YUV/YCbCr*, and the YIQ color spaces [22].

Luminance-chrominance transformations become particularly significant when the image $\mathbf{z}$ is corrupted by some independent noise. In fact, because of the typical correlation among z^R, z^G, z^B, one can observe that the luminance ζ^Y has noticeably

higher signal-to-noise ratio (SNR) than the two chrominances ζ^U and ζ^V and than any of the individual RGB components z^R, z^G, z^B.

It is well known that natural color images exhibit a high correlation between the R, G, and B channels. Thus, it can be observed that the luminance Y contains most of the valuable information (edges, shades, objects, texture patterns, etc.) and the chrominances U and V contain mostly low-frequency information (considering compressed data, these channels very often come from undersampled data).

2.3.1 The Opponent Color Space

The opponent color space is based on the 1964 color opponency theory by a German physiologist Ewald Hering. The theory suggests two pairs of opponent colors, red-green and yellow-blue, that cannot be perceived simultaneously. This theory was supported by color naming experiments where reddish-green and yellowish-blue tones were not identified. This led Hering to presume three opponent channels, red-green, yellow-blue and black-white (achromatic luminance). In fact, Hering was one of the first to separate luminance and two chrominances in a color model. Unlike the RGB tristimulus model, driven by an intent of modelling the retinal color stimulus response, the opponent color space is based on more central mechanisms of the brain. The opponent processes are acquired through a transformation of the cone responses.

The transformation matrices for the opponent color space are

$$\mathbf{A}_{opp} = \begin{bmatrix} \frac{1}{3} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{-1}{2} \\ \frac{1}{3} & \frac{-1}{2} & \frac{1}{4} \end{bmatrix}$$

and

$$\mathbf{B}_{opp} = \mathbf{A}_{opp}^{-1} = \begin{bmatrix} 1 & 1 & \frac{2}{3} \\ 1 & 0 & \frac{-4}{3} \\ 1 & -1 & \frac{2}{3} \end{bmatrix}$$

It can be noted, that the second and third columns of matrix $\mathbf{A}_{opp}$ have zero mean. This is equivalent to the inner product between the chrominance basis vectors and a vector corresponding to a gray pixel (for which $z^R = z^G = z^B$) always being zero. It means that gray is orthogonal to the chrominance components and that the inverse color transformation matrix $\mathbf{B} = \mathbf{A}^{-1}$ has the elements of its first row all equal, $b_{1,1} = b_{1,2} = b_{1,3}$. Since $1 = \sum_{j=1}^{3} a_{j,1} b_{1,j}$ and $\sum_{j=1}^{3} a_{j,1} = 1$, it is obtained that $b_{1,1} = b_{1,2} = b_{1,3} = 1$. This means that the luminance component ζ^Y can be directly treated as a gray scale component of the RGB image $\mathbf{z}$, because the inverse transformation of the luminance component, $[\zeta^Y \ 0 \ 0]\,\mathbf{B} = \mathbf{z}_{\mathsf{gray}}(\mathbf{x})$, is a gray scale

image.

Luminance-chrominance transformations can be considered as special color decorrelating transforms. In particular, up to a diagonal normalization factor, the matrix $\mathbf{A}_{opp}$ is nothing but a 3×3 DCT transform matrix. Further, it is noted that the columns of $\mathbf{A}_{opp}$ are respectively a mean filter, a finite derivative filter, and a second derivative filter.

2.3.2 The YUV Color Space

The YUV color space was first designed for the purposes of television broadcasting with the intent of minimizing the bandwidth requirements. It is used among others in Europe (PAL) and its equivalent in the American broadcasting system is the YIQ color space. They both fill the definition of an opponent color space because they consist of a luminance channel and two color difference channels. As a sidenote it can be mentioned that the usual scheme in compressing a YUV signal is to downsample the chromatic channels with a factor of two or four each so that chromatic data occupies at most half of the total video bandwidth. This can be done without apparent loss of visual quality as the human visual system is far less sensitive to spatial details in the chrominances than in the luminance. The same compression approach is also utilized among others in the well-known JPEG image compression standard.

The transformation matrices for YUV color space are

$$\mathbf{A}_{yuv} = \begin{bmatrix} 0.30 & -0.17 & 0.50 \\ 0.59 & -0.33 & -0.42 \\ 0.11 & 0.50 & -0.08 \end{bmatrix}.$$

and

$$\mathbf{B}_{yuv} = \mathbf{A}_{yuv}^{-1} = \begin{bmatrix} 1 & 0 & 1.4020 \\ 1 & -0.3441 & -0.7141 \\ 1 & 1.7720 & 0 \end{bmatrix}.$$

Because of the similar nature of the color models, the orthogonality properties given in the previous section for opponent color space hold also for the YUV space. One should note that these properties do not depend on the orthogonality of the matrix $\mathbf{A}$ (in fact the three columns of $\mathbf{A}_{yuv}$ are not orthogonal), but rather on the orthogonality between a constant vector and the second and third columns of the matrix.

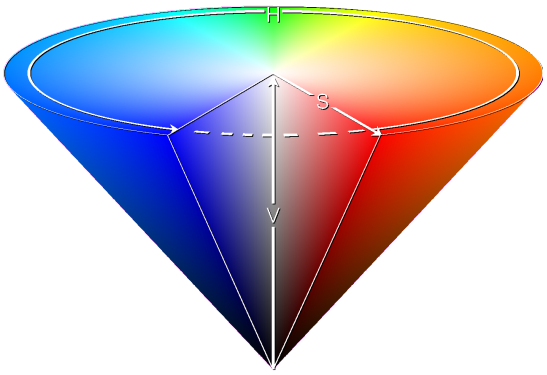

Figure 2.3: Illustration of the HSV color cone. The hue and saturation define the angle and length of a planar vector.

2.3.3 The HSV Color Space

The HSV color space is a representative of the class of perceptual color spaces. The name stands for the three components that form the space, hue, saturation and value (intensity, brightness). This type of color models are categorized under the umbrella of perceptual color spaces for a reason; while an untrained observer can hardly form an image of a color based on individual RGB tristimulus components, everybody can form a color based on its hue (tint, tone) and saturation. Additionally the value (intensity) describes the lightness of the color, ranging from black (zero) to white (one).

Intuitively the achromatic value component is equal to the luminance, while hue and saturation can be defined as $H = \arctan \frac{\zeta^U}{\zeta^V}$ and $S = \sqrt{(\zeta^U)^2 + (\zeta^V)^2}$, respectively. These can be interpreted as the angular component and the length of a planar vector. Thus the triplet Y, H, and S corresponds to a luminance-chrominance representation with respect to cylindrical coordinates. It is noted that multiple definitions of both hue and saturation can be found in the literature and while the presented ones are among the more frequently noted ones, they are chosen here first and foremost for the sake of simplicity. An illustration of the HSV color space utilizing this interpretation can be found in Figure 2.3.

3. BASICS OF MULTIPLE EXPOSURE HDR

The history of high dynamic range imaging is relatively short. The first articles dealing with the creation of HDR images were published in the beginning of the 1990's ([17], [18]) and based partially on these (among others), in the year 1997, what now is the core of HDR imaging in RGB space was defined in [5]. The vast majority of research done in the field has focused on optimizing the parameters (number of exposures, exposure times, etc.), e.g. [20] or developing different tone mapping methods, e.g. [6]. Though the general consensus is that parallel treatment of the RGB channels causes distortions in the colors, only very limited attention has been focused on addressing these problems. In one of the few studies, ICC (International Color Consortium) color profiles are suggested to be used in attempt to reach more realistic color reproduction in context with RGB HDR imaging [10]. While HDR composition techniques have practically been in a standstill, tone mapping has taken huge leaps after the 1997 release of the Debevec and Malik paper.

3.1 Problem Statement

State-of-the-art techniques for HDR imaging have been developed generically for the RGB color space. The problem of HDR imaging and the state of the art techniques are discussed in [24]. The basic structure of the multiple exposure HDR approach is the same independent of the technique. The starting point is an image sequence that fills the qualifications given later in this thesis. With this prerequisite filled, the general line-up of the process, illustrated in the block diagram of Figure 3.1, is as follows. *First*, a response curve for each color channel of the camera needs to be solved. This needs to be done only once for a given imaging device. *Second*, based on this response curve and the exposure time, the different exposures can be linearized and the radiance map acquired utilizing a weighted average of the linearized data. *Third*, the acquired HDR data needs to be stored in an efficient manner. This step is not necessary if the data is composed only for direct display. *Fourth*, the HDR data needs to be compressed so that it can be displayed on a conventional LDR display or for example a print-out. The compression of the visual range, generally known as tone mapping, can be skipped if the image is to be displayed on an HDR display device.

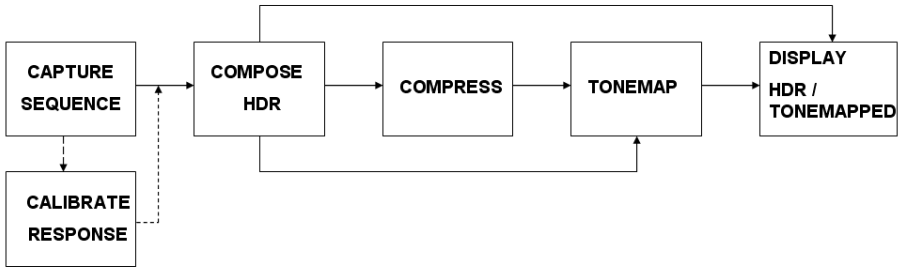

Figure 3.1: The block diagram of the general multiple exposure HDR imaging pipeline. Different paths indicate alternatives in the process, such as displaying a tone mapped HDR image on a normal display or alternatively displaying an HDR image on an HDR capable display.

3.2 RGB HDR Composition

The most essential step in RGB HDR image acquisition is the definition of the camera response function. This function defines the relation between the scene irradiances and the camera pixel output and can therefore be used to linearize the output data. Essentially two techniques exist for the response calibration. Debevec and Malik approach defined in [5] extends the previous work done by Mann and Picard in [18]. The basic principle of their method is that by capturing frames of a still scene in different exposures, one is actually sampling the camera response function at each pixel. The Debevec and Malik technique is relatively simple and robust, and is therefore presented here in detail.

The successful camera response calibration requires a perfectly aligned image sequence with negligible noise. The exposures should be as dense as possible to guarantee that the camera response function is sampled at maximally short intervals. If these prerequisites are filled, the logarithmic camera response function g_{R} is solved by minimizing the mean-squared error over the following objective function.

$$\mathcal{O} = \sum_{j=1}^{P}\sum_{i=1}^{N} w_{\mathsf{R}}(z_i(\mathbf{x}_j)) \left[g_{\mathsf{R}}(z_i(\mathbf{x}_j)) - \ln E_i^R - \ln \Delta t_i\right]^2 + \lambda \int_0^1 w_{\mathsf{R}}(z) g''(z)^2 dz,$$

where P is the number of pixels used from each image, N is the number of exposures (images) used, g_{R} is the unknown target logarithmic camera response function for the red channel (the green and the blue channel are treated similarly), E_j^R are the unknown constant pixels irradiance values at each pixel j, and Δt_i is the exposure time of frame i. The weight function w_{R} is a simple hat function defined as

$$w_{\mathsf{R}}(z) = \begin{cases} z - Z_{min} & for \quad z \leq \frac{1}{2}(Z_{min} + Z_{max}) \\ Z_{max} - z & for \quad z > \frac{1}{2}(Z_{min} + Z_{max}) \end{cases}$$

where Z_{min} and Z_{max} are the minimum and maximum obtainable pixels values respectively. The weights are used to reduce the effect of clipped and over exposed pixels. The second derivative on the right hand side is used to guarantee the smoothness of the solution and the amount of smoothing is controlled by the parameter λ. The equation is solved using singular-value decomposition, leading to an enumerated table of discrete values. The RGB channels are treated in parallel. This assumes that the interactions between channels are negligible, which is, as even the authors admit, problematic to defend. As a result to the parallel treatment of the three channels, color distortions are in many cases introduced to the composed image. These have to be corrected by post-composition white-balancing which in turn may lead to colors that are not faithful to the original ones. Nevertheless, the approach works well enough to produce satisfactory results.

With the solved camera response function and known exposure times, the logarithmic HDR radiance map can then be composed as a weighted sum of the camera output pixels values as follows.

$$\ln E_i^R = \frac{\sum_{i=1}^{N} w_{\mathsf{R}}(z_i(\mathbf{x}_j))(g_{\mathsf{R}}(z_i(\mathbf{x}_j)) - \ln \Delta t_i)}{\sum_{i=1}^{N} w_{\mathsf{R}}(z_i(\mathbf{x}_j))}$$

The other calibration method found in the literature was presented by Mitsunaga and Nayar and described in [19]. The main difference to the Debevec and Malik technique is that instead of filling an enumerated table, polynomial coefficients for the response function are approximated. This enables the solution of not only a camera response function but also exact exposure ratios which are essential in combining HDR images from source sequences whose aperture and shutter speed are not known exactly.

3.3 HDR Image Compression Methods

If intended for archiving etc, the resulting HDR image then has to be compressed and stored somehow. The increased bit-depth of HDR images (up to 96 bits per pixel or more) can obviously not be stored by conventional formats designed for the storage of 24 bits per pixel image data. Moreover, the majority of 24-bit image encodings are output-referred, meaning that the colors of the image are associated with a specific output device. HDR images are preferably stored in a scene-referred format, as their pixels are directly related to an irradiance of the scene, be it real or artificial. As HDR images by definition always include colors that are unpresentable in a normal display device, it would be senseless to save them in an output-referred format.

This has motivated the development of numerous HDR image storage formats,

of which perhaps the most used are Industrial Light and Magic's openEXR [12], Gregory Ward's Radiance [28] and Logluv TIFF [14]. The openEXR file format has been developed since 1999 and was made publicly available in 2002. It is capable of storing data in either 16-, 24- or 32-bits per channel modes. The coding of the pixel values is exponential; for example in 16-bits per channel mode one bit is used to store the sign, five for the exponent and ten for the mantissa. The senior of the HDR file encodings, Radiance was first introduced as a part of the Radiance lighting and simulation system as early as 1989. Since then the file format has found wide acceptance as a format suitable for HDR photography and image-based lighting. The image data is stored as RGBE (red, green, blue, exponent) or XYZE (CIE X, Y and Z with exponent). The exponent is common for the three channels. Differing from the two, Logluv TIFF is an extension of the standard TIFF library. As the name suggests, image data is stored as a logarithmic luminance and CIE chromaticity coordinates (u,v), described in Section 2.1. The encoding comes in two variants, 24-bit and 32-bit, the difference between the two being in the uniform stepsize.

Also emerging are a few JPEG based HDR storage methods (see e.g. [16]). Their usefulness is in the fact that they do not only provide a medium for storing the full luminance and chrominance ranges of an HDR image, but also an approximation of these ranges for the conventional LDR display device. In other words, while the JPEG based HDR formats include an extended dynamic range, if opened on a LDR device they provide the application a predefined tone mapped image instead of the full HDR data that is provided for more sophisticated HDR enabled applications. Among the existing compression methods, the JPEG based ones are the only truly lossy compressors. The sub-band compression pipeline utilized in the scheme is shown in Figure 3.2. As seen in the figure, the method uses a TMO (tone mapping operator) to generate a tone mapped representation of the scene-referred HDR image. The tone mapped image is stored as standard JPEG/JFIF data and a ratio image (RI) is generated to preserve the scene-referred HDR data. The ratio image is stored as a 64-Kbyte marker. This marker is neglected by naïve applications and used to create a scene-referred combination by more sophisticated ones. The ratio-image is gray-scale and could therefore be derived directly from the luminance channel. Additionally, as JPEG stores image data in a luminance-chrominance space, these JPEG based storage formats are an attractive alternative for the storage of luminance-chrominance HDR data.

3.4 RGB HDR Tone Mapping

The problem of tone mapping is essentially one of compression with preserved visibility. A linear scene-to-output mapping of an HDR image produces results shown

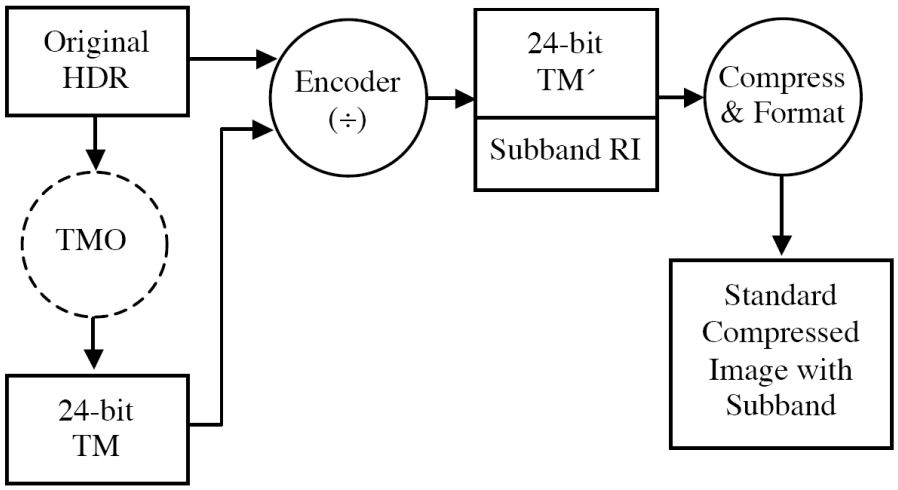

Figure 3.2: The JPEG based HDR image sub-band encoding pipeline. (Reprinted from [16])

for instance in Figure 3.3. The same scene is shown in Figure 3.4 tone mapped using the method described in Section 5.2.1. Again the age-old phrase holds, a picture is worth a thousand words: an HDR image is useless on an LDR display device without tone mapping.

The history of tone mapping dates back to much longer than the introduction of HDR imaging. As the dynamic range of the film has always exceeded that of the photographic paper of the era, manual control has been necessary in the development process. The darkroom was introduced in the late 19th century and since that time, a technique called dodging and burning has been used to manipulate the exposures of the photographic print.

The same idea has later been transported into controlling the visibility of digital HDR images. In this context the operation is known as tone mapping. The very basic aim of tone mapping or *tone reproduction* (the two terms are interchangeable) is to provoke the same response through the tone mapped image that would be provoked by the real scene. In other words, matching visibility. In the darkroom the procedure was of course done manually, but for digital purposes the ideal solution would fit every problem and not need any human interaction. So far the optimal tone mapping method has not been invented and one has to choose the method dependent on the problem. Many of the methods require parameter adjustment, while some are automatic. The results of the tone mapping depend highly on the chosen method as well as the parameters used. New tone mapping approaches are introduced almost monthly and the aim of the scientific community is, as usual, to develop more and more universal approaches well suited for the majority of problems. Some of the major tone mapping algorithms were reviewed in the 2002 state-of-the-art review

Figure 3.3: An HDR image of a stained glass window from Tampere Cathedral, Finland. The image is linearly scaled for display.

by Devlin et al. [6].

Tone mapping methods can be divided in various ways. One of the most frequently used divisions is the *local* versus *global* division. Additionally, methods operating in some transform domain can be classified in their own classes. The division in [24] is exactly this: Global, local, Frequency domain and gradient domain operators are distinguished as their own classes.

The global methods are the most simple class of TMOs. They all share a couple of inherent properties: the same mapping is applied for all the pixels and the tone mapping curve is always monotonic in nature. The mapping has to be monotonic in order to avoid disturbing artifacts. This imposes a great limitation on the compression/visibility preservation combination. As the usual target is to map an HDR scene into the range of standard 8-bit representation, only 256 distinct brightness values are available. Global methods excel generally in computational complexity, or the lack-thereof. As the same mapping is applied on all the pixels, the operations can be done very efficiently. On the other hand, for scenes with very high dynamic range, the compression ability of the global class may not be sufficient.

Local methods are able, to an extent, to escape the limitations met with global TMOs. In general local methods do not relay on image-wide statistics. Instead, every pixel is compressed depending on its luminance value and the values of a local neighborhood. Often local methods try to mimic properties of the HVS; the eye is known to focus locally on an area of a scene forming an independent adaptation

Figure 3.4: A tone mapped HDR image of a stained glass window from Tampere Cathedral, Finland. The tone mapping was done with the method described in Section 5.2.1.

based on the local neighborhood contents. As a result, the cost of more flexible compression is in the computational complexity. The number of necessary computations goes up with the number of local adaptation neighborhoods. Also the higher, local compression of scene brightness may at times lead to halo-like artifacts around objects.

The transform domain operators are distinguished from global as well as local methods by the fact that they operate on the data in some domain other than the conventional spatial one. Frequency domain operators compress data, as the name suggests, utilizing a frequency-dependent scheme. The first digital tone mapping operator was published by Oppenheim et. al. already in 1968 and it was a frequency domain one [21]. Many of the properties of modern frequency domain operators are inherited from Oppenheim's original approach. Gradient domain operators rely on the notion that a discrimination between illuminance and reflectance is for many scenes relatively well approximated by the gradient approach. This is supported by the notion that an image area with a high dynamic range usually manifest a large gradient between neighboring pixels. The follow-up is a tone mapping operator functioning on the differentiation domain, using gradient manipulation for dynamic range reduction.

The majority of tone mapping methods work on the spatial domain and are therefore categorized under either local or global umbrella, depending on the nature of the compression. A few examples of distinct tone mapping approaches are given

Figure 3.5: In reading order, the results of a global, local, frequency domain and gradient domain tone mappings.

to display the vast differences between different classes. Figure 3.5 illustrates in reading order the tone mapping results of a global, a local, a frequency domain and a gradient domain method, respectively presented in [27], [4], [21] and [8]. Finally it is noted that as there are numerous methods for tone mapping of HDR scenes and their basic functionality is, apart from the core idea of range compression, very different from one method to another, it is not meaningful to give a detailed description of an example implementation. All the methods are well described in the literature and the methods implemented for the luminance-chrominance approach of this thesis are described in detail in Chapter 5.

3.5 HDR Hardware

The importance and the potential of HDR imaging are further highlighted by the introduction of HDR hardware. Both camera and display technologies have recently been unveiled by a few companies. For the direct capture of HDR data several experimental capture devices exist, [24]. Fujifilm's new CCD-array is however the first attempt of bringing direct extended dynamic range capture to consumer, albeit professional, cameras. It improves the dynamic range of the captured photograph by including two sensor wells for each pixel on the sensor. The two wells have different sensitivities, where the less sensitive well starts reacting only when the normal well

Figure 3.6: Brightside Technologies HDR LCD display.

is saturated. In the display front the Brightside (recently acquired by Dolby Laboratories) HDR display, shown in Figure 3.6 uses adaptive backlighting to produce quote: "peak luminance in excess of $3,000$ $candela/m^2$" as well as "a contrast ratio in excess of 200,000:1". While Brighside's display is the first commercialized HDR display, there are other, competitive technologies in development, [25]. The introduction of these as well as other technological advances brings HDR imaging and display closer and closer to being as everyday as for example the color-TV, and thus further underlines the need for cost-effective, easy way to produce HDR content.

4. LUMINANCE-CHROMINANCE HDR IMAGE COMPOSITION

As stated in Section 3, the parallel processing of the three RGB channels leads in many cases to colors that are unfaithful to the original ones. An alternative approach is rooted in the psychophysically inspired color models where the image is formed based on one component describing luminance and two components containing the chromatic information. Alone the luminance channel forms a gray scale representation of the scene, and as such, contains most of the information associated with an image. The two chrominances are nearly meaningless alone, and as such contain complementary information that is not per se crucial. This fact is also exploited in color image compression where the chromatic components are often downsampled compared to the luminance. The completely different nature of the luminance and the chrominances motivates a new approach for HDR composition. The new method treats the luminance channel similarly to the state-of-the-art approach for a single channel of the RGB triplet and suggests a novel method for the composition of color.

4.1 Problem Statement

Let us consider a set of images $\boldsymbol{\zeta}_i = [\zeta_i^Y, \zeta_i^U, \zeta_i^V], i = 1, \ldots, N$ in a luminance-chrominance space, captured with different exposure times Δt_i and with LDR, assuming $\boldsymbol{\zeta}(\mathbf{x}) \in [0,1] \times [-0.5, 0.5]^2$, where $\mathbf{x} = [x_1, x_2]$ is a pixel coordinate. The goal is to obtain a single HDR image $\tilde{\boldsymbol{\zeta}} = \left[\tilde{\zeta}^Y, \tilde{\zeta}^U, \tilde{\zeta}^V\right]$ in the same color space. In the setting, the luminance and chrominance channels are treated separately. A pre-calibrated camera response function is used for the luminance channel, whereas a saturation-driven weighting is applied for the chrominance channels.

4.2 Camera Response

Knowing the camera response is essential for determining the irradiance of the source scene from the given image data. With the pixel exposure $e(\mathbf{x})$ defined as the product between exposure time Δt and irradiance $E(\mathbf{x})$, $e(\mathbf{x}) = E(\mathbf{x})\Delta t$, the generic pixel output $\mathfrak{z}(\mathbf{x})$ is

$$\mathfrak{z}(\mathbf{x}) = f(e(\mathbf{x})) = f(E(\mathbf{x})\Delta t),$$

where f is the function describing the response of the output to the given exposure. Therefore, the irradiance can be obtained from the pixel output by the formal expression (see e.g. [5])

$$\ln E(\mathbf{x}) = g(\mathfrak{z}(\mathbf{x})) - \ln \Delta t, \tag{4.1}$$

where $g = \ln f^{-1}$ is the (inverse) camera response function. This function can be estimated from a set of images of a fixed scene captured with different exposure times. Images for the calibration sequence are assumed to be perfectly aligned and shot under constant illumination. Noise should be negligible. For the camera response function calibration, a much larger set of images (i.e. a denser set of exposures) than typically available for the HDR composition, is used. The camera response function is estimated (calibrated) only once for each camera. The estimated function is then used for the linearization of the input values in all subsequent HDR compositions of the same device.

Conventionally, camera response functions are estimated separately for the R, G, and B channels, i.e. using these channels as the output $\mathfrak{z}$. There exist a number of techniques developed for this purpose, e.g., [5], [19]. In this thesis, a single response function is estimated for the luminance channel only by replacing the output $\mathfrak{z}$ with ζ^Y. In practice, this can be done adapting any of the techniques developed for RGB data. Some specific modifications are however required in order to treat the luminance.

Because of underexposure (which produces dramatically low SNR) and overexposure (which results in clipping the values which would otherwise exceed the dynamic range) not all pixels from the given set of images should be used for the calibration of the camera response function, nor for the subsequent composition, with equal weights. Near the minimum of the value interval of an imaging device, the information is distorted by numerous noise components, most influential of which is the poissonian photon shot noise, see e.g. [3]. As the number of photons hitting a well on the sensor within a given time interval is a poissonian distributed random process, at low enough light levels the variations in the number of photons become substantial in comparison to the mean level, effectively rendering the current created in the well completely unreliable. In addition to this, components like thermal and read-out noise play some role in corrupting the low-level sensor output. In the vicinity of the maximum of the value interval, clipping starts to play a part. Naturally a saturated pixel can only convey very limited amount of information about the scene. The described phenomena motivate the penalization of under- and overexposed pixels in both the camera response calibration and the HDR image composition phase. Additionally, only pixels having monotonically strictly increasing values between under- and overexposure, throughout the sequence, can be considered to be valid in the camera response calibration. This is a natural idea, since the pixels used

for response calibration are assumed to come from a sequence shot with decreasing exposure time. As the scene is assumed to be static and the pixel irradiance constant, the only aberration from the strict monotonical increase in pixel value can come from noise. Noisy pixels should not be used for camera response calibration, as erraneous response function would induce a systematic error in the subsequent HDR compositions. Using these criteria for pixel validity, the function g can be fitted by minimizing the quadratic objective function [5]

$$\sum_{j=1}^{P}\sum_{i=1}^{N} w_{\mathsf{cam}}(\zeta_i^Y(\mathbf{x}_j))\left[g(\zeta_i^Y(\mathbf{x}_j)) - \ln E_i - \ln \Delta t_i\right]^2 + \lambda \int_0^1 w_{\mathsf{cam}}(\zeta) g''(\zeta)^2 d\zeta, \quad (4.2)$$

where P is the number of pixels used from each image and w_{cam} is a weight function limiting the effect of the under- and overexposed pixels. The regularization term on the right-hand side uses a penalty on the second derivative of g to ensure the smoothness of the solution. In this thesis the preference is to use a relatively high pixel count (e.g. 1000 pixels) in order to minimize the need for regularization in the data fitting. The system of equations is solved employing singular-value decomposition, in a fashion similar to that of [5]. As this method deals with the luminances ζ_i^Y, the sensor output is not processed directly, but rather a combination of its components is considered. This combination can include overexposed components without reaching the upper limit of its dynamic range. Such pixels have to be penalized in the processing. To ensure a more powerful attenuation of the possibly overexposed pixels, instead of a simple triangular hat function (e.g. as in [5]) an asymmetric function of the form $w_{\mathsf{cam}}(\rho) = \rho^\alpha(1-\rho)^\beta$ where $0 \leq \rho \leq 1$ and $1 \leq \alpha < \beta$, is used. An example of such a weight function is given in Figure 4.1. An example of a camera response function for the luminance solved with this method is illustrated in Figure 4.2. When looking at the camera response one can observe an abrupt peak towards the right end of the plot. This is a result of the instability of the system for overexposed pixels. It becomes evident that such pixels cannot be directly included in the subsequent composition step. Instead they have to be penalized with weights.

It may seems like the described overexposure problem is a complication of the luminance-chrominance approach. It is however emphasized that the issue of overexposure being present also in pixels whose value is lower than the maximum is not absent even for the conventional approaches dealing in the RGB space. *First*, the R, G, and B channels in the images which are used in the calibration never coincide with the original RGB components of the sensor (this is the actual element limiting the dynamic range in most camera systems). In fact, several intermediate processing stages (which include color matrixing, Bayer pattern color interpolation, white balance, etc.) are typically present. These operations combine the original RGB

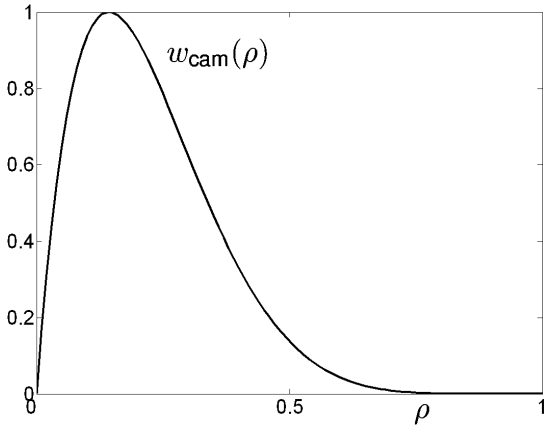

Figure 4.1: Weight function used for the luminance values in the camera response function definition.

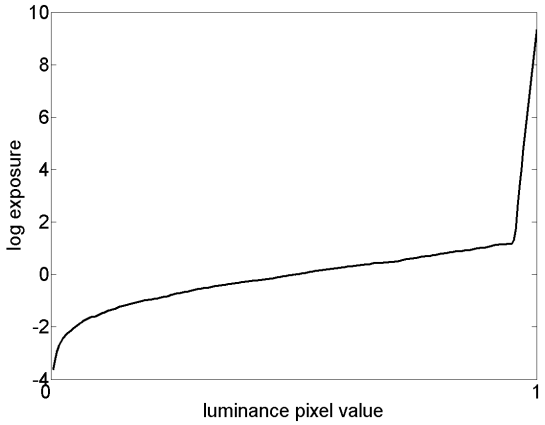

Figure 4.2: The estimated inverse response function g for the luminance channel of Nikon COOLPIX E4500.

components of the raw-data and can thus introduce clipped (overexposed) components into pixels whose final output value is still lower than the maximum. *Second*, if the data has been precedently compressed (e.g. JPEG), the found values might not exactly coincide with the original (uncompressed) ones. *Finally*, let us note that even in the RAW data some correlation between adjacent pixels in the same rows of the CMOS/CCD sensor is introduced during the read-out process. Especially in the case of overexposure, also blooming can be present. Therefore, processing luminance data does not introduce any fundamental complication to the value or meaning of the camera-response function. By using a properly selected weight function one can effectively deal with the wider overexposure region at the end of the dynamic range. Additionally, since the luminance channel always benefits from an increased SNR when compared with the R,G,B channels, the procedure for the luminance channel is less sensitive to noise than the corresponding ones in RGB space.

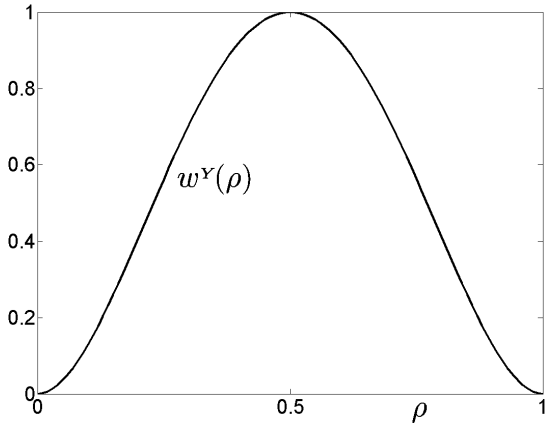

Figure 4.3: Weight function w^Y used for the luminance channel composition.

4.3 Luminance Component Composition

The HDR luminance component is obtained by a pixelwise weighted average of the pixel log irradiances defined according to Eq. (4.1) as $\ln E_i(\mathbf{x}) = g(\zeta_i^Y(\mathbf{x})) - \ln \Delta t_i$. As observed in the previous section, pixels whose value is close to 0 or 1, carry little valuable information, respectively because of low SNR (underexposure) and clipping (overexposure). Therefore, these pixels are penalized employing weights during the composition. As a weighting function a polynomial function $w^Y(\rho) = \rho^\alpha(1-\rho)^\beta$, where $0 \leq \rho \leq 1$ and $\alpha = 2$, $\beta = 2$ is used, thus ensuring a smaller impact of the under- or overexposed pixels. An example of such a weight function for the luminance is given in Figure 4.3. The logarithmic HDR luminance is obtained as

$$\ln \tilde{\zeta}^Y(\mathbf{x}) = \frac{\sum_{i=1}^N w^Y(\zeta_i^Y(\mathbf{x}))(g(\zeta_i^Y(\mathbf{x})) - \ln \Delta t_i)}{\sum_{i=1}^N w^Y(\zeta_i^Y(\mathbf{x}))}. \tag{4.3}$$

Because of the nature of the camera response function g, the HDR luminance is obtained in logarithmic scale. After employing the natural exponential, the resulting values are positive, normally spanning $[10^{-4}\ 10^4]$ thus being truly high dynamic range.

4.4 Chrominance Components Composition

For the chrominance components no camera response is defined. Instead, the chrominances are weighted in relation to the level of color saturation. The higher the color saturation, the more the pixel contains valuable chromatic information, and thus the higher the weight. This is motivated by the fact that when a pixel is over- or underexposed it is always less saturated than it would be at the correct exposure. More specifically, $w^{UV}(S) = S^\alpha$, where $\alpha > 1$. In exhaustive experiments, $\alpha = 1.5$ has been found to be a good choice. A saturation based chrominance weight function

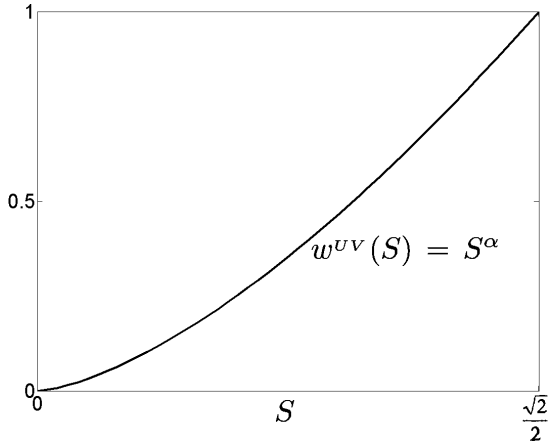

Figure 4.4: Weight function w^{UV} used for the composition of the chrominance channels.

is illustrated in Figure 4.4. To guarantee the color preservation, the same weights are used for both chromatic components and any chromatic component $C \in \{U, V\}$ is composed as

$$\tilde{\zeta}^C(\mathbf{x}) = \frac{\sum_{i=1}^{N} w^{UV}(S_i(\mathbf{x}))\zeta_i^C(\mathbf{x})}{\sum_{i=1}^{N} w^{UV}(S_i(\mathbf{x}))}, \tag{4.4}$$

where S_i denotes the saturation of $\boldsymbol{\zeta}_i$. It is pointed out that being a convex combination of the input chrominances, the range of $\tilde{\zeta}^C(\mathbf{x})$ is again in $[-0.5, 0.5]$. However, because of averaging, the possible number of distinct chrominance values is remarkably higher than in the original source sequence.

Another intuitively interesting approach would be to use the luminance-dependent weights w^Y for not only the luminance but also for the chrominance channels. Loosely this approach would be similar to the proposed saturation-driven weighting because of the fact that color saturation typically decreases with the luminance moving closer to the extremes. However, in experiments it was found that saturation-based weighting provides a better reproduction of color especially for saturated highly exposed or low-lit details.

4.4.1 HDR Saturation Control

After the composition the luminance component, or more precisely, the radiance map has a range much higher than the original [0 1]. The newly acquired pixel irradiances span an extended range, while the chrominance component composition, being a convex combination of the components, results in a denser sampling of the original range. This presents a problem, because luminance-chrominance color spaces represent color as difference to the gray luminance component. If the HDR image would directly be mapped to RGB for display, the result would be an effectively gray scale image as the differences represented by chrominances spanning

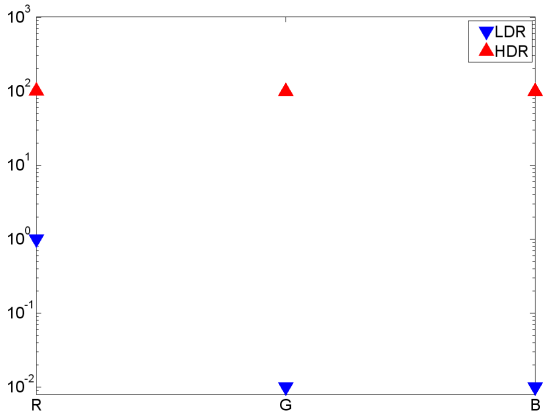

Figure 4.5: The illustration of the desaturating effect of the luminance-chrominance HDR composition.

$[-0.5\ 0.5]$ cannot form saturated colors in relation to a luminance component with maximum values up to 10^5. This can be practically demonstrated with the following example. Let us assume a red pixel in the RGB color space. The tri-component vector describing this pixel is then of the form $z = [1\ 0\ 0]$. If this pixel is transformed into for example the opponent color space using the matrix A_{opp} defined in Section 2 the resulting luminance-chrominance pixel is $\zeta = [\frac{1}{3}\ \frac{1}{2}\ \frac{1}{4}]$. Let us further assume that the HDR composition process results in a pixel irradiance of 100 and the chrominances of ζ. Then the HDR pixel is defined in opponent color space as $\tilde{\zeta} = [100\ \frac{1}{2}\ \frac{1}{4}]$. The application of the inverse transformation leads into an RGB pixel $\tilde{z} = [100\frac{2}{3}\ 99\frac{2}{3}\ 99\frac{2}{3}]$. The resulting values form a gray pixel, as the relative differences between components are not transmitted through the luminance increase. This is of course unwanted, as the correct result would be $\tilde{z} = [100\ 0\ 0]$. The phenomenon is illustrated in Figure 4.5.

Obviously, if one wants to display or process the HDR data on a device with RGB HDR input, a *saturation control* method has to be introduced. For this purpose the following approach can be used. Let us define the scalar proportionality factor $\mu(\mathbf{x})$ between the HDR luminance $\tilde{\zeta}^Y$ and the weighted average of the LDR luminances $\zeta_i^Y(\mathbf{x})$ with weights $w^{UV}(S_i(\mathbf{x}))$ as

$$\mu(\mathbf{x}) = \frac{\tilde{\zeta}^Y \sum_{i=1}^{N} w^{UV}(S_i(\mathbf{x}))}{\sum_{i=1}^{N} w^{UV}(S_i(\mathbf{x}))\, \zeta_i^Y(\mathbf{x})}. \tag{4.5}$$

In other words, $\mu(\mathbf{x})$ is a pixel-wise scaling parameter defining the ratio between the weighted average of the original pixel values and the pixel irradiances obtained through the HDR composition process. Now, the HDR image in RGB space is

obtained by the normalized inverse color transformation

$$\tilde{\mathbf{z}}(\mathbf{x}) = \tilde{\boldsymbol{\zeta}}(\mathbf{x}) \begin{bmatrix} 1 & 0 & 0 \\ 0 & \mu(\mathbf{x}) & 0 \\ 0 & 0 & \mu(\mathbf{x}) \end{bmatrix} \mathbf{B}.$$

Here the normalization is realized by multiplication against the diagonal matrix $diag(1, \mu(\mathbf{x}), \mu(\mathbf{x}))$ which scales the two chrominances $\tilde{\zeta}^U$ and $\tilde{\zeta}^V$ yielding a value of saturation which matches the full dynamic range achieved by $\tilde{\zeta}^Y$. Indeed, the weights $w^{UV}(S_i(\mathbf{x}))$ in Eq. (4.5) and Eq. (4.4) are exactly the same. From the diagonal scaling matrix it can be observed that the value of $\mu(\mathbf{x})$ does not have any influence on the luminance of $\tilde{\mathbf{z}}(\mathbf{x})$. Likewise, from the definition of hue, $H = \arctan \frac{\zeta^U}{\zeta^V}$ it is clear that the hue is left intact and the only thing altered is the saturation.

Though the need for saturation control cannot be visualized on conventional displays let alone printout, in full HDR, the issue can be conveyed through the chain of RGB tone mapping. This is visualized in Figure 4.6. The shown image is composed in luminance-chrominance space and transformed to RGB for tone mapping. Due to reasons explained above, virtually all color is lost in the transformation. In Figure 4.7 the same luminance-chrominance HDR image is transformed into RGB utilizing the saturation control described here. The colors are reproduced faithfully underlining the hue preservation properties of the proposed method.

Figure 4.6: The atlantic coastline of Rota, Spain, shot during the author's honeymoon. The HDR image is composed in luminance-chrominance space and transformed directly to RGB for tone mapping.

Figure 4.7: The atlantic coastline of Rota, Spain, shot during the author's honeymoon. The HDR image is composed in luminance-chrominance space and transformed to RGB utilizing the presented saturation control. Subsequent tone mapping is done in RGB.

5. TONE MAPPING LUMINANCE-CHROMINANCE HDR IMAGES

Tone mapping is an HDR imaging technique used to approximate the visibility of the tones of an HDR image on an LDR media, such as LCD and CRT displays or print-outs [24]. Essentially, tone mapping compresses the contrast of a scene to fit into the displayable range of a media while preserving details and color.

5.1 Problem Statement

The dynamic range of an HDR image often spans more than 5 orders of magnitude, of which a conventional display is able to visualize a maximum of 2 orders of magnitude. This presents a problem, as while HDR images are becoming more and more available, HDR displays lag behind. The problem is then to fit the greater dynamic range of an HDR image into the limited gamut of a display device. The most simple solution is to linearly scale the data and while a simple linear scaling very rarely produces acceptable results, applying gamma curves or some more sophisticated mapping procedures will likely do better.

A number of tone mapping methods working in RGB space exist. These techniques can easily be adapted for the luminance range reduction. However, the term 'tone mapping' would be questionable in this context, as tone is usually used in connection with color. In this thesis such a luminance range reduction operator is denoted as $\mathcal{T}$. Its output is a luminance image with range $[0, 1]$, i.e. $\mathcal{T}\left(\tilde{\zeta}^{Y}\right)(\cdot) \in [0, 1]$. As for the chromatic channels, a simple, yet effective approach is suggested.

5.2 Luminance Component Compression

For the compression of the luminance channel, two global luminance range reduction operators are presented. The selection is limited to global operations simply because thus far, local operators have not been able to produce results faithful to the original scene. It should be noted that majority of tone mapping methods developed for RGB can be applied more or less directly for the compression of the luminance channel as if it were a gray scale HDR image. As such, the continuous development of tone mapping methods for RGB HDR images benefits also the compression of luminance-chrominance HDR data.

5.2.1 Anchoring Based Compression

The first method is based on an anchoring theory of lightness perception and was presented for RGB in [13]. The core idea of the method is to divide the luminance image into frameworks based on the theory of lightness perception [9]. This theory states that in order to relate the luminance values to perceived lightness, one or more mappings between the luminance and perceived values on the gray scale, *an anchor*, have to be defined. A block diagram of the method is given in Figure 5.1.

The division into frameworks is done utilizing the standard K-means algorithm and is based on the histogram of $\log_{10}$ luminance. Preliminary centroids are based within one $\log_{10}$ unit of each other. In subsequent steps the clusters based on these centroids are fused *first* if the distance between two centroids becomes less than one and *second* if the cluster contains no pixels at all or no pixels with probability more than 0.6 of belonging into that cluster. The probability is calculated as a function of distance from the cluster centroid. When this algorithm is left to converge, the result is a division into usually no more than three frameworks with similar intensities. At the last stage of the framework division the acquired frameworks are articulated based on the dynamic range of an individual framework. A framework with a high dynamic range (above one $\log_{10}$-unit) has the maximum articulation, and the articulation goes down to zero as the dynamic range goes down to zero. The articulation is imposed on the frameworks to make sure a framework consisting of background does not play a significant role in the computation of the net lightness. At the penultimate stage the frameworks are anchored so that in an individual framework, the 95th percentile is mapped as white. Ultimately the reduced range luminance is calculated by subtracting the anchored frameworks from the original luminance.

In terms of detail visibility, the anchoring based method is superior to the other one presented in this thesis. Even scenes with extremely high contrast are squeezed into the range of a consumer level LCD display. For some scenes this may lead to results appearing slightly unnatural, as global contrast is reduced very severely. The fact that the described method allows no manual tuning serves as both a drawback and a benefit; for most images majority of the scene is brought visible in a believable manner but for images containing extremely high dynamic range, the lack of global contrast may at times lead to results appearing slightly too flattened. Results acquired employing this method are presented in Chapter 6.

5.2.2 Histogram Based Compression

The other method presented here is based on a relatively simple histogram adjustment technique. It coarsely approximates the method described in [15]. It is noted

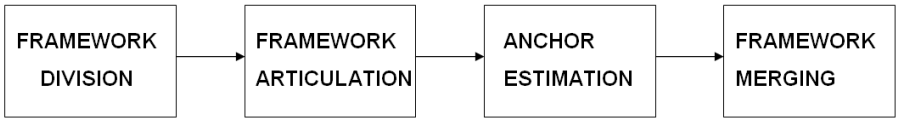

Figure 5.1: The block diagram of the anchoring based tone mapping process. The blocks represent the four central stages of the procedure.

however that most of the sophisticated ideas introduced in [15] are sacrificed in the implementation for speed and simplicity. In the histogram adjustment method the first step is to clip the HDR luminance image values to the lower limit of the HVS ($10^{-}4cd/m^2$). Then the luminance image can be downsampled with a factor of eight (the downsampling has to be preceded by a corresponding low-pass filtering to avoid aliasing). The downsampling is a simplistic approximation of the foveal fixation phenomenon encountered in adaptation of the eye. Then a luminance histogram is calculated from the downsampled image. A cumulative distribution is then defined as $P(b) = \frac{\sum_{b_i<b} f(b_i)}{\sum_{b_i} f(b_i)}$, where $f(b_i)$ is the frequency count for the bin number i. Now if the goal was to equalize the probability of each brightness value, this could be achieved by histogram equalization

$$B_{disp} = \log(L_{dmin}) + [\log(L_{dmax}) - \log(L_{dmin})]\, P(b)$$

where L_{dmax} and L_{dmin} are the maximum and minimum display luminances respectively and B_{disp} is the output display brightness. However, while histogram equalization compresses dynamic range in sparsely populated regions of the histogram, it also expands contrast in highly populated zones resulting in exaggerated contrast. This can be avoided by imposing a linear ceiling on the contrast produced by the method. The linear ceiling can be defined as

$$f(b) \leq frac\sum_{b_i} fb_i \Delta b \log(L_{dmax}) - \log(L_{dmin})$$

where Δb is the histogram bin stepsize. This ceiling has to be imposed on the histogram equalization in an iterative manner, because when values exceeding the ceiling are truncated, the histogram samplecount $\sum_{b_i} fb_i$ is altered thus altering also the ceiling. Pseudocode for the ceiling can be given in a way similar to that of [15] as follows.

```
tolerance = 2.5% of histogram total
repeat
{
  trimmings = 0
  compute the new histogram total T
  if T < tolerance then
    return FALSE
  else for each histogram bin i do
    compute the ceiling
    if f(bi) > ceiling then
    {
      trimmings += f(bi) - ceiling
      f(bi) = ceiling
    }
} until trimmings <= tolerance
```

Then the display mapping is obtained by using the newly acquired luminance histogram (or more specifically, its cumulative distribution) in the histogram equalization formula defined above. Results of the method described here can be found in Chapter 6.

5.3 Chrominance Component Compression

The sRGB gamut does not allow the rendition of very dark or very bright vivid and saturated colors which exist in real scenes and which are captured in HDR images. Therefore there exists a need for chromatic tone mapping. In this approach, in order to get faithful colors that fit into the sRGB gamut the hue is kept intact by sacrificing saturation. Introducing a scaling factor δ for the two chrominances will then not change the hue, but will scale down the saturation. The scheme that is used to guarantee legal sRGB values is embedded in the color space transformation itself and described as follows.

Let $\mathbf{B} = \mathbf{A}^{-1}$ be the luminance-chrominance to RGB transformation matrix and define the gray (achromatic) image and its chromatic complement image in RGB space by

$$\mathring{\mathbf{z}}_{\mathsf{gray}}(\mathbf{x}) = \left[\mathring{z}^R_{\mathsf{gray}}(\mathbf{x}) \;\; \mathring{z}^G_{\mathsf{gray}}(\mathbf{x}) \;\; \mathring{z}^B_{\mathsf{gray}}(\mathbf{x})\right] = \begin{bmatrix} \mathcal{T}\left(\tilde{\zeta}^Y\right)(\mathbf{x}) \\ 0 \\ 0 \end{bmatrix}^T \mathbf{B},$$

$$\mathbf{z}_{\mathsf{chrom}}(\mathbf{x}) = \left[z^R_{\mathsf{chrom}}(\mathbf{x}) \;\; z^G_{\mathsf{chrom}}(\mathbf{x}) \;\; z^B_{\mathsf{chrom}}(\mathbf{x}) \right] = \begin{bmatrix} 0 \\ \tilde{\zeta}^U(\mathbf{x}) \\ \tilde{\zeta}^V(\mathbf{x}) \end{bmatrix}^T \mathbf{B}.$$

It can be noted that $\mathring{\mathbf{z}}_{\mathsf{gray}}(\mathbf{x})$ is truly a gray image because in RGB to luminance-chrominance transforms $b_{1,1} = b_{1,2} = b_{1,3}$. Then a map $\delta \geq 0$ is needed, such that

$$\mathring{\mathbf{z}}(\mathbf{x}) = \mathring{\mathbf{z}}_{\mathsf{gray}}(\mathbf{x}) + \delta(\mathbf{x})\, \mathbf{z}_{\mathsf{chrom}}(\mathbf{x}) \in [0,1]^3 . \tag{5.1}$$

It can be defined by $\delta(\mathbf{x}) = \min\{1, \delta^R(\mathbf{x}), \delta^G(\mathbf{x}), \delta^B(\mathbf{x})\}$ where

$$\delta^R(\mathbf{x}) = \begin{cases} \frac{\mathring{z}^R_{\mathsf{gray}}(\mathbf{x})}{-z^R_{\mathsf{chrom}}(\mathbf{x})} & \text{if} \;\; z^R_{\mathsf{chrom}}(\mathbf{x}) < 0 \\ \frac{1-\mathring{z}^R_{\mathsf{gray}}(\mathbf{x})}{z^R_{\mathsf{chrom}}(\mathbf{x})} & \text{if} \;\; z^R_{\mathsf{chrom}}(\mathbf{x}) > 0 \\ 1 & \text{if} \;\; z^R_{\mathsf{chrom}}(\mathbf{x}) = 0 \end{cases}$$

and δ^G and δ^B are defined analogously. Thus, $\delta(\mathbf{x})$ is the largest scalar smaller or equal to one, which allows the condition (5.1) to hold. Figure 5.2 illustrates the definition of $\delta(\mathbf{x})$. From the figure it is easy to realize that the hue, i.e. the angle of the vector, of $\mathring{\mathbf{z}}(\mathbf{x})$ is not influenced by δ, whereas the saturation is scaled proportionally to it. Roughly speaking, the low dynamic range image $\mathring{\mathbf{z}}(\mathbf{x})$ has colors which have the same hue as those in the HDR image $\tilde{\boldsymbol{\zeta}}$ and which are desaturated as little as is needed to fit within the sRGB gamut.

It is now obvious that the tone mapped LDR image can be defined in luminance-chrominance space as

$$\mathring{\boldsymbol{\zeta}}(\mathbf{x}) = \left[\mathcal{T}\left(\tilde{\zeta}^Y\right)(\mathbf{x}) \quad \delta(\mathbf{x})\, \tilde{\zeta}^U(\mathbf{x}) \quad \delta(\mathbf{x})\, \tilde{\zeta}^V(\mathbf{x}) \right].$$

The tone mapped luminance-chrominance image $\mathring{\boldsymbol{\zeta}}$ can be compressed and stored directly with an arbitrary method (for example, DCT-based compression, as in JPEG), and for display transformed into RGB using the matrix $\mathbf{B}$. It is demonstrated that this approach yields lively, realistic colors.

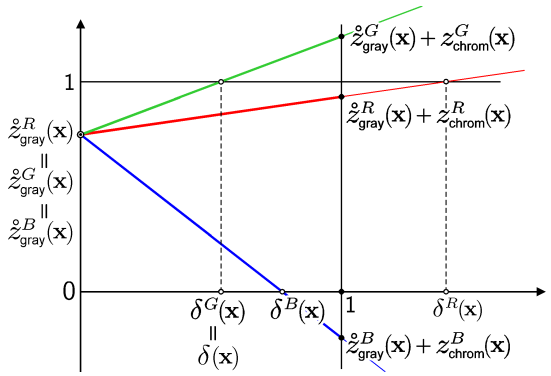

Figure 5.2: Illustration of the definition of the chromatic tone mapping parameter δ.

6. RESULTS

A detailed phase-by-phase example of the introduced HDR imaging process is given. All the phases of the process are considered and corresponding results provided. Subsequently the proposed HDR imaging approach in luminance-chrominance space is compared against established techniques working in the RGB space. In addition to the case of noise-free, appropriately registered LDR images, the cases of LDR images corrupted by certain amount of noise as well as noise-free misaligned sequences are considered. In capturing the LDR source images, two different cameras were used. It is emphasized that the focus in visual evaluation of the methods should be in the perception of hue, rather than brightness which is strongly affected by the selection of a tone mapping operator.

6.1 HDR Imaging Process in Luminance-Chrominance Space

The process of HDR imaging in luminance-chrominance space is structurally divided into three parts. The camera response calibration, HDR image composition and HDR image tone mapping. It is noted that while the camera response needs only be calibrated once, it is the most influential part of the process and as such demands thorough coverage in this section as well. The process is done for images shot with a Fujifilm FinePix S5600 digital camera. The sequence used for camera response calibration contains a total of 17 frames shot with maximally dense exposures $\Delta t = [\frac{1}{10}\ \frac{1}{13}\ \frac{1}{15}\ \frac{1}{20}\ \frac{1}{25}\ \frac{1}{30}\ \frac{1}{40}\ \frac{1}{50}\ \frac{1}{60}\ \frac{1}{80}\ \frac{1}{100}\ \frac{1}{125}\ \frac{1}{160}\ \frac{1}{200}\ \frac{1}{250}\ \frac{1}{320}\ \frac{1}{400}]$. The exposures are scene dependent and were selected so that a maximum number of pixels in all images would fit between under- and overexposure. The pixel locations selected by the procedure described in Section 4.2 are shown on the tone mapped image of the calibration sequence in Figure 6.1 and the corresponding response curve is given in Figure 6.1. The response curve was calibrated with a λ equal to 10, because the calibration sequence is dense enough to discourage heavy smoothing.

After defining the camera response, HDR images can be composed according to the procedure described in Sections 4.3 and 4.4. The image sequence shown in Figure 6.3 contains 4 images with exposure times $\Delta t = [\frac{1}{8}\ \frac{1}{60}\ \frac{1}{250}\ \frac{1}{1000}]$. The exposures are approximately quadrupled through the sequence, which in most cases results in dense enough frames to contain all the detail of the scene. The contribution of each LDR image is shown in Figure 6.4 where red, green, blue and violet correspond to the four

Figure 6.1: Pixel locations used for the camera response calibration imposed on the tone mapped HDR image of the calibration sequence.

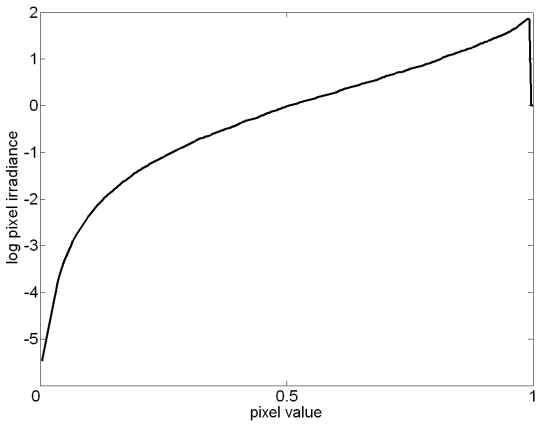

Figure 6.2: The logarithmic camera response function for a Fujifilm FinePix S5600

Figure 6.3: The four image sequence used for the HDR composition.

exposures from the longest to the shortest. The image is clearly divided into four zones which illustrate the limited range of usable pixels found in each LDR frame. However from the four frame sequence a good coverage of the scene can be formed, as shown by the successful HDR composition shown in Figure 6.5. The image is scaled for display so that the brightest area is well visible. The HDR pixel irradiances are shown in the false color image of Figure 6.6. The value range is [0.4205 1803.8] which corresponds to a dynamic range of approximately 3.6 zones. Both the HDR representations underline the necessity of luminance range compression, as the image is not visualizable in any particularly meaningful way in its HDR form.

As one of the most central aims of HDR imaging is to match the visibility of a scene and its digital representation on an imaging device, and as the HDR image illustrated on an LDR display or printout clearly does not fulfill this aim, additional processing is required. The *tone mapping* techniques introduced in Section 5 are deviced to serve this purpose. The HDR image shown in Figure 6.5 is tone mapped using the two approaches described in Section 5. The image compressed following the histogram equalization based method is shown in Figure 6.7 and the image acquired employing the method based on the anchoring theory of lightness perception is given in Figure 6.8. Both of the operators used are global and as such their results are more or less comparable. It can be noted that the perception based method yields better compression of the luminance channel bringing almost all the regions in the image visible. The downside of high compression is a grayish, sort of pale look which is a direct effect of the loss of global contrast. The histogram based method compresses the luminance noticeably less, preserving the sensation of contrast between the dark interior and the bright exterior and in doing so, looses in detail visibility. All the global tone mapping methods have to compromise between compression and contrast and in most cases, as very often in imaging applications, the order of superiority is highly subjective and objective criterion are hard to come up with.

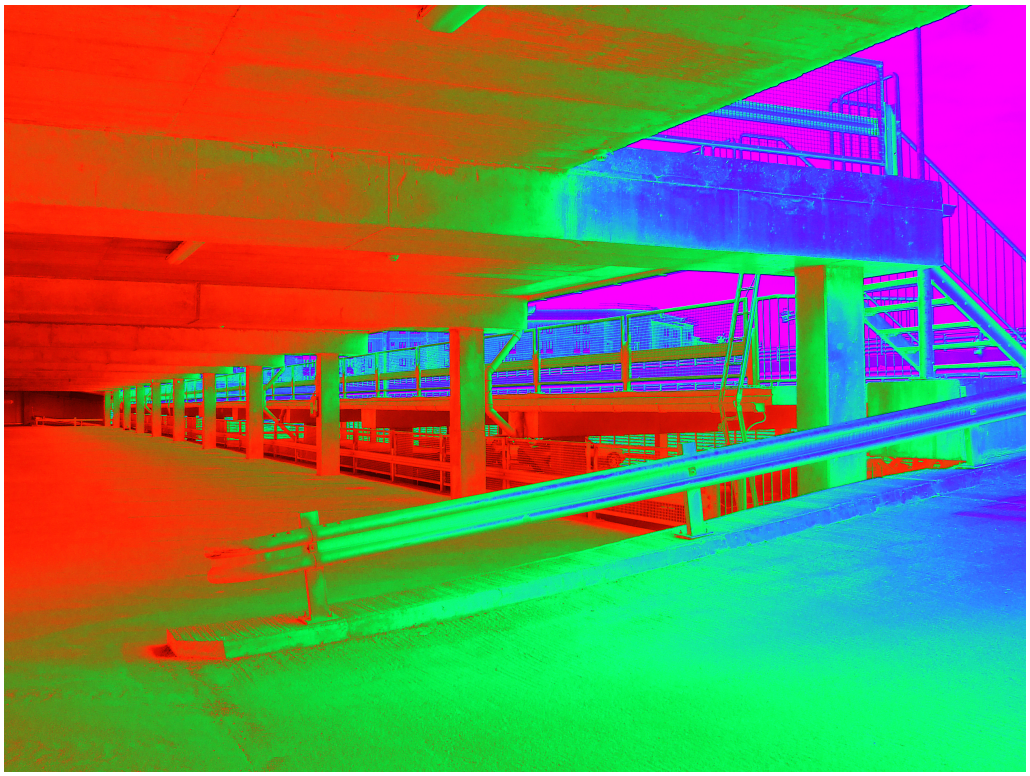

Figure 6.4: The impact of each LDR frame on the composition result shown in different colors. Red, green, blue and violet correspond to the four exposures from the longest to the shortest.

Figure 6.5: The linearly scaled HDR image.

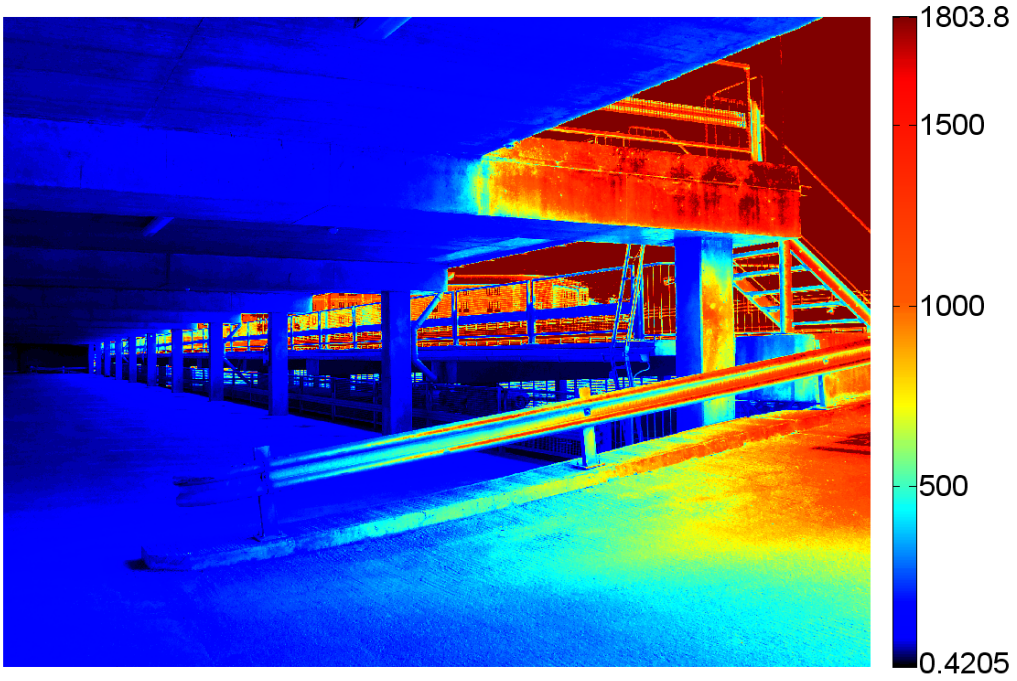

Figure 6.6: The false color representation of the HDR pixel irradiances.

6.2 Noise-free Case

A sequence shot with the lowest possible ISO for a given camera is in this context considered to be noise-free. To demonstrate the hue preservation properties of the described approach, experiments with two sets are given. The first set, $\{\mathbf{z}_i\}_{i=1,2,3}$ was captured with Fujifilm FinePix S5600 with exposure times (seconds) $\{1/5, 1/20, 1/80\}$ (Figure 6.9). The second set, $\{\mathbf{z}_i\}_{i=1,2,3,4,5}$ was captured with Nikon COOLPIX E4500 with exposure times $\{1/4, 1/15, 1/60, 1/250, 1/1000\}$ (Figure 6.12). Using these two sets, HDR images have been composed both in RGB and luminance-chrominance space. The HDR images have been then tone-mapped back to LDR ones for display. Debevec's HDR Shop v.1.0.3 software [1] has been used for the HDR composition in RGB space.

There are a number of tone mapping techniques, most of them requiring careful tuning of a set of input parameters. For the RGB HDR tone mapping the *photographic tone reproduction operator* [24] was chosen, as it is a global tone mapping technique with fully automatic parameter estimation. The tone mapped result of processing the sequences in RGB space are shown in Figures 6.10 and 6.13. In this approach, the sequences are first transformed to the opponent color space. Then, the HDR image composition is performed as described in Section 4. The so-obtained HDR image is tone mapped according to the method described in 5. The operator $\mathcal{T}$ used for reducing the range of $\tilde{\zeta}^Y$ is an adaptation of the histogram adjustment tech-

Figure 6.7: The tone mapping result of the histogram equalization based method.

Figure 6.8: The tone mapping result of the anchoring theory of lightness perception based method.

nique based on the population of local adaptation luminances in a scene, developed in [15].

The results for the sequences are shown in Figures 6.11 and 6.14. Beside some obvious differences in brightness, chromatic distortions which can arise from the RGB processing can be seen when the tone-mapped images are compared with the source sequences in Figures 6.9 and 6.12. The RGB approach seems to systematically fail in reproducing colors at the dark end of the dynamic range of the scene. This is best pronounced at the colorchart cover at the foreground of Figures 6.12, 6.13, and 6.14, and at the wall under the table on the left hand side of the same figures. In the same areas, the luminance-chrominance approach produces lively, faithful to the original, colors. The results are further illustrated in the enlarged details of the above-mentioned parts, shown in Figures 6.15 and 6.16. The detail figures compare the visually best exposure from the source sequence shown in 6.12, the tone mapped RGB HDR image, and the tone mapped luminance-chrominance image. Overall, the colors obtained by processing in luminance-chrominance space are much more faithful, as the hue has been preserved.

Figure 6.9: Three-image source sequence used to compose the HDR images.

6.3 Noisy Case

A further notable advantage of the luminance-chrominance approach appears when the source sequence is degraded by noise. Three different types of results are presented for noisy data. First, a visual assessment of HDR images corrupted with artificial noise is given. Then, same for images shot with a very high ISO setting, thus visibly corrupted by noise from the camera digital imaging sensor, is provided for reference. Third, a numerical comparison of the noise attenuation properties of processing in RGB against processing in two separate luminance-chrominance spaces is done.

Figure 6.10: HDR image composed and tone mapped by processing in RGB color space.

Figure 6.11: HDR image composed and tone mapped by the introduced approach in luminance-chrominance space.

Figure 6.12: Five-image source sequence used to compose the HDR images.

6.3.1 Artificial noise

In simulations, zero-mean Gaussian noise with standard-deviations $\sigma = \frac{5}{255}$, $\frac{15}{255}$, and $\frac{25}{255}$ was added to the second set of source LDR images (Figure 6.12), and an HDR image was composed with the approach described in this thesis as well as with the conventional RGB approach. Scanline of the noisy HDR image ($\sigma = \frac{15}{255}$) imposed on the scanline of the noise-free reference can be seen in Figure 6.19 (RGB) and in Figure 6.20 (luminance-chrominance). The difference in noise levels can also be seen in the tone mapped images, given in Figure 6.17 (RGB) and Figure 6.18 (luminance-chrominance). The luminance-chrominance approach preserves details fairly well in many parts where the RGB approach completely fails. For example details of the colorchart cover at the foreground are completely lost with the RGB approach whereas the new method preserves the visibility much better.

6.3.2 Real noise

A sequence with the Nikon COOLPIX E4500 with ISO 800 was also shot to acquire a source sequence that was inherently noisy. Because for the given camera, the native (lowest) ISO is 100, by increasing the ISO up to 800 one ends up boosting not only the signal, but also the noise, which becomes very noticeable in most parts of the image. For this sequence two details are shown; one from the darker part (Figure 6.21) and one from the lighter (Figure 6.22). From the detail sequences the presence of the blue noise can be distinguished especially in the darker frames. The notable noise amplification of the RGB approach is a consequence of the separate weighting of the three channels. As the noise is most visible in the blue channel in the dark parts of the image, it is boosted out of proportion by the weighting. Instead, the luminance channel is much less vulnerable to noise and the exaggerated blue noise of the tone mapped Figure 6.23 done by processing in RGB cannot be seen in the luminance-chrominance processing counterpart of Figure 6.24.

Differing from the other tone mapped results shown in the thesis, the tone mapping of both noisy images shown in Figures 6.23 and 6.24 for RGB and luminance

Figure 6.13: HDR image composed and tone mapped by processing in RGB color space.

Figure 6.14: HDR image composed and tone mapped by our approach in luminance-chrominance space.

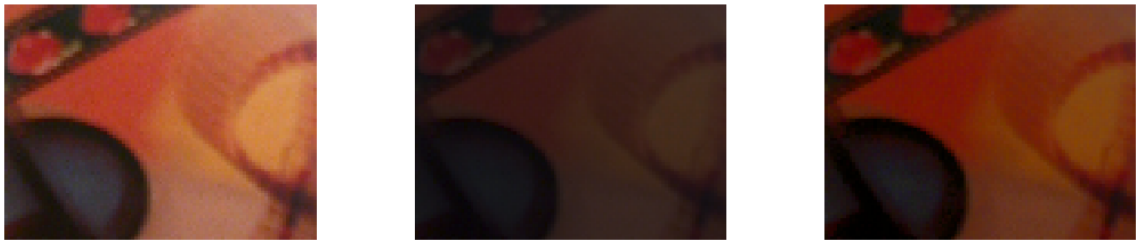

Figure 6.15: An enlarged detail illustrating the color preservation of our method. The images are from left to right as follows: Exposure with best color and detail from the original sequence. Tone mapped RGB HDR image. Tone mapped luminance-chrominance HDR image.

Figure 6.16: Another enlarged detail illustrating the color preservation properties of our method. The images are from left to right as follows: Exposure with best color and detail from the original sequence. Tone mapped RGB HDR image. Tone mapped luminance-chrominance HDR image.

chrominance composition respectively, was done under HDR Shop with the included plugin, implementing the photographic tone reproduction operator introduced in [23]. Default settings were used, except for sharpening and gamma correction, which were not done in the tone mapping process. For the purpose of correctly transforming a luminance-chrominance HDR image into RGB space, the saturation control described in Section 4.4.1 was performed on the image before the standard color space transformation realized by multiplication against the matrix $\mathbf{B} = \mathbf{A}_{opp}^{-1}$.

6.3.3 Objective noise evaluation

Because of the nonlinear nature of HDR processing, the addition of a zero-mean noise to the source sequence $\{\mathbf{z}_i\}_{i=1}^{N}$ does not generally result in the introduction of simple zero-mean errors in the composed HDR image $\{\tilde{\mathbf{z}}_i\}_{i=1}^{N}$ or in the tone mapped image $\{\mathring{\mathbf{z}}_i\}_{i=1}^{N}$. Strictly speaking, the mathematical expectation of the image composed from the noise sequence does not coincide with the image composed from the noise-free sequence. A number of alterations might be present in the underlying signal, such as different scaling due to the wider histogram of the noisy data and shifts in the intensities of darker and brighter areas caused by the asymmetric distribution of the clipped noise. These make it difficult to quantitatively measure the degradation due to noise using reference-based error criteria such as the MSE, PSNR, etc. It should be observed that most of these alterations are not necessarily negative (e.g.,

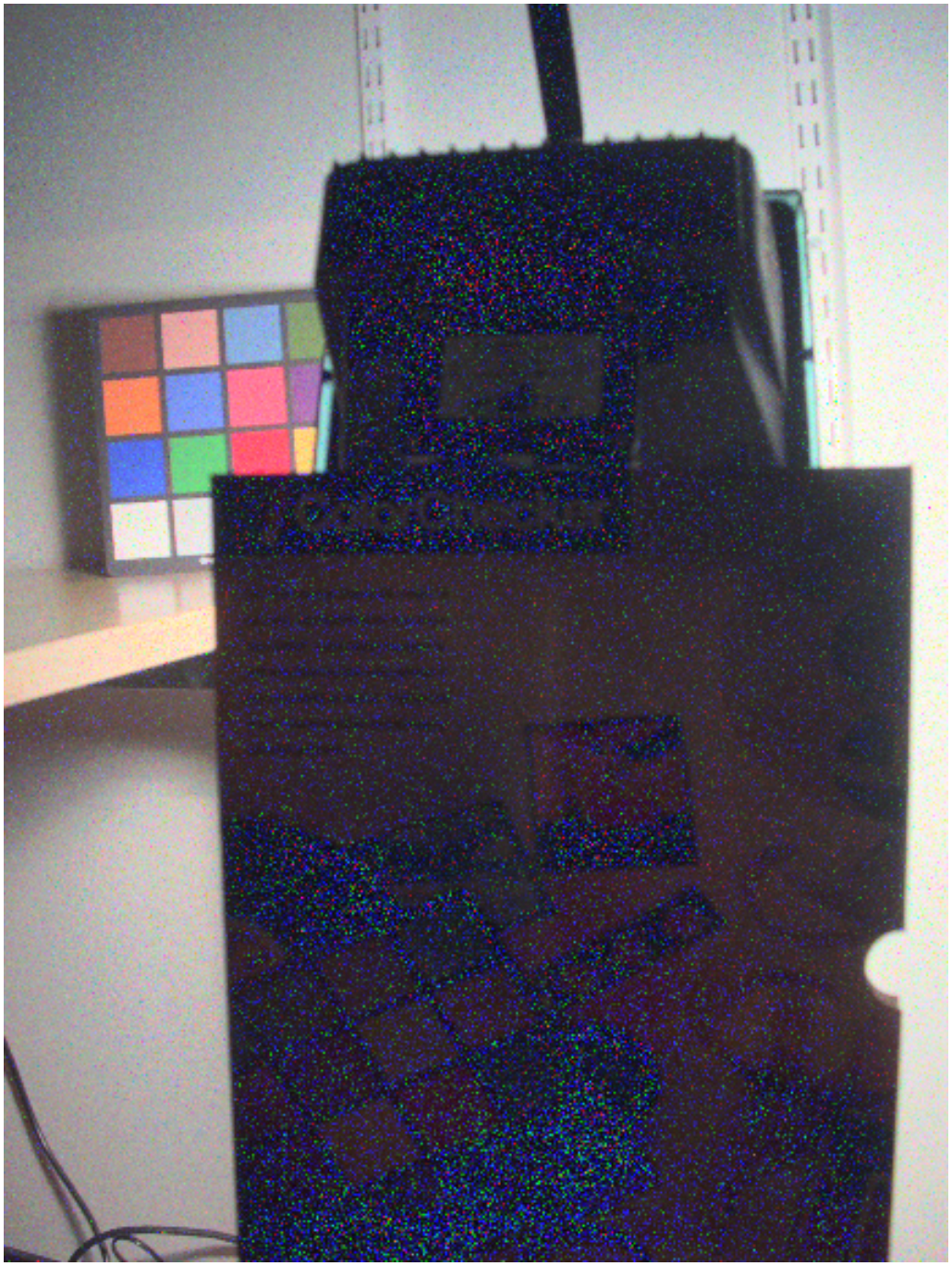

Figure 6.17: HDR image composed and tone mapped from the noisy sequence in RGB color space.

Figure 6.18: HDR image composed and tone mapped from the noisy sequence by our approach in luminance-chrominance space.

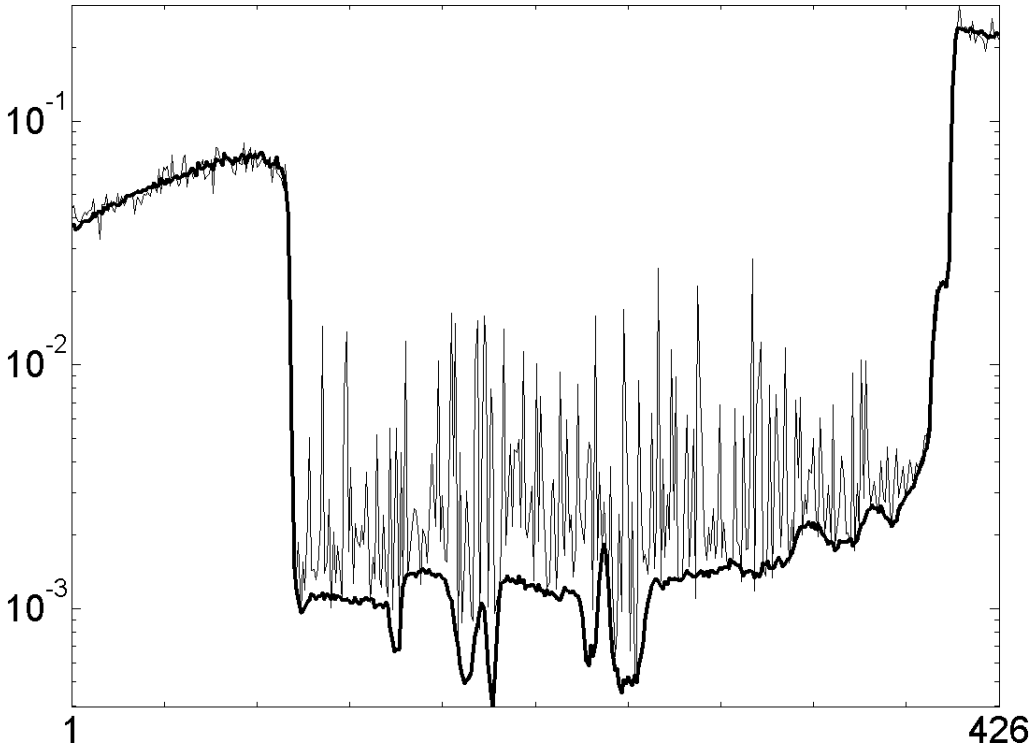

Figure 6.19: Horizontal scanline (450) comparison between the noise-free and the noisy HDR image processed in RGB space.

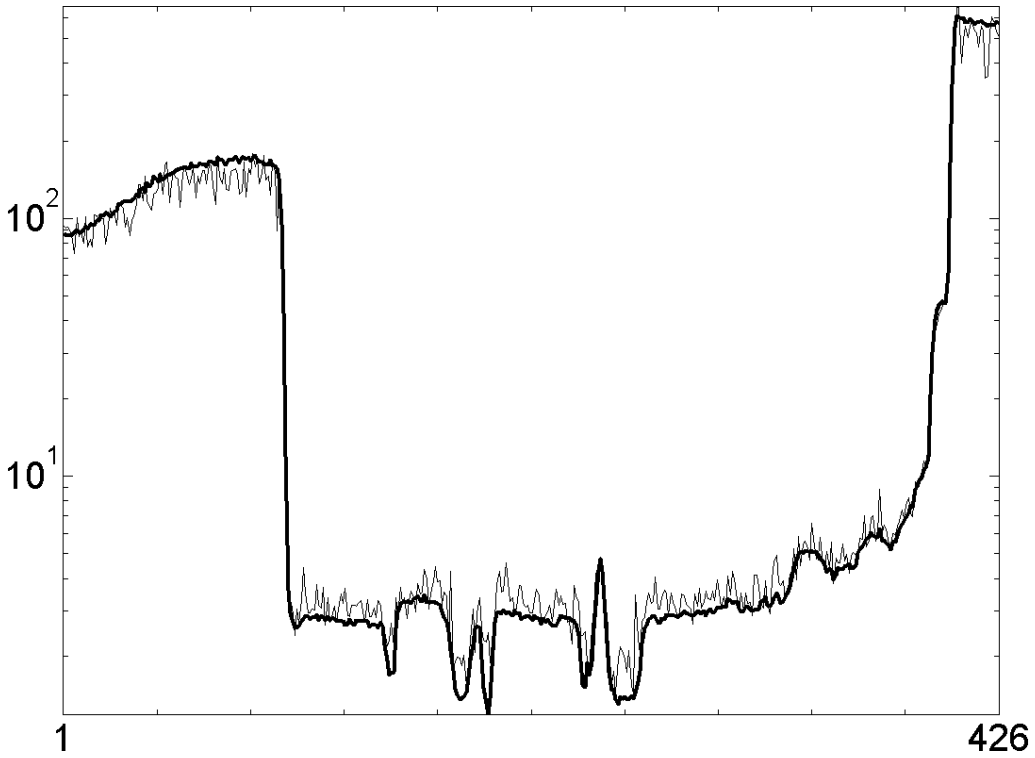

Figure 6.20: Horizontal scanline (450) comparison between the noise-free and the noisy HDR image processed by our approach in luminance-chrominance space.

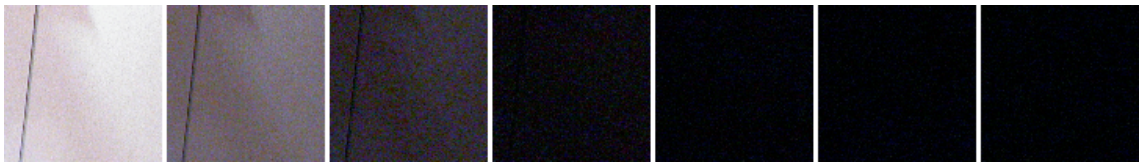

Figure 6.21: A clipped detail of the darker part of the seven-frame noisy sequence shot with ISO 800. The blue noise enhanced by the RGB approach can be seen in the darker frames.

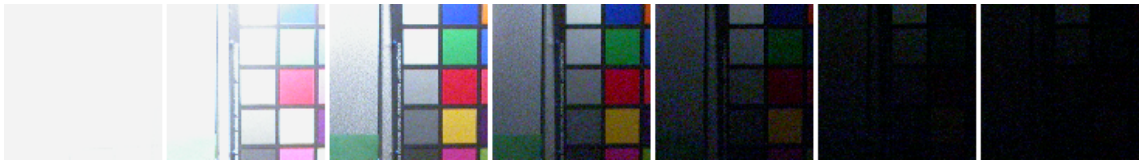

Figure 6.22: A clipped detail of the lighter part of the seven-frame noisy sequence shot with ISO 800. The blue noise enhanced by the RGB approach can be seen in the darker frames.

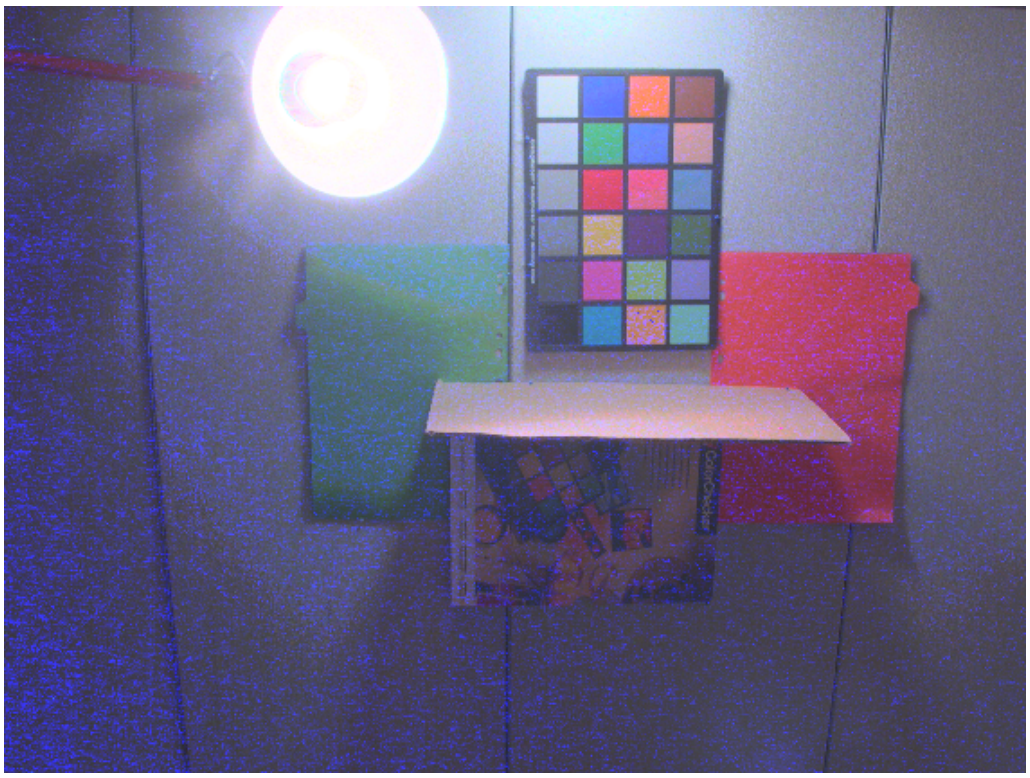

Figure 6.23: HDR image composed and tone mapped from the noisy (ISO 800) sequence in RGB color space.

Figure 6.24: HDR image composed and tone mapped from the noisy (ISO 800) sequence in luminance-chrominance (opponent) color space.

the different scaling of the data is absolutely negligible).

In order to quantify the effect of noise presence, or the "noisiness", of the HDR images composed from noisy source sequences, one has to resort to a *non-reference-based* criterion. In particular, this thesis uses a normalized robust estimator of the standard-deviation ($NRSTD$) of the noise defined as

$$NRSTD\,(\xi) = \frac{median\,\{|\xi \circledast \mathcal{H}|\}}{0.6745} \frac{1}{\max\{\xi \circledast \mathcal{L}\} - \min\{\xi \circledast \mathcal{L}\}}, \tag{6.1}$$

where ξ is the degraded "noisy" image, $\circledast$ denotes the convolution, and $\mathcal{H}$ and $\mathcal{L}$ are a high-pass and a low-pass filtering kernel, respectively, satisfying

$$\int \mathcal{H} = 0, \quad \int \mathcal{H}^2 = 1, \text{ and } \int \mathcal{L} = 1.$$

The following selections are made:
$\mathcal{H} = \psi^T\psi$, with $\psi = [0.035 \;\; 0.085 \;\; -0.135 \;\; -0.460 \;\; 0.807 \;\; -0.333]$ (a Daubechies wavelet), and $\mathcal{L}$ is chose as 7×7 uniform kernel equal to $\frac{1}{49}$. The factor on the left in Eq. (6.1) is essentially the classic robust estimator of the noise standard deviation [7] given as the median of the wavelet detail coefficients' moduli at the finest scale. It assumes that the image is concentrated at the coarser scales and that the finest scale contains

mostly noise. The denominator in the right factor of Eq. (6.1) serves to normalize the estimated standard-deviation with respect to the dynamic range of a smoothed version of ξ, in which the effects of the noise have been substantially suppressed. In order to obtain meaningful and comparable results, the NRSTD is calculated on the luminance component of the HDR images only. This is based on the fact that the image content is differently distributed among the three channels of different color-space representations and the luminance is the most informative one. For the RGB HDR images this component is simply the average of the R, G, and B channels. The results are summarized in Table 6.1. They confirm the visual impression

σ	$NRSTD_{RGB}$	$NRSTD_{lum-chrom}$
0/255	$8.76 \cdot 10^{-5}$	$10.7 \cdot 10^{-5}$
5/255	$3.15 \cdot 10^{-4}$	$2.04 \cdot 10^{-4}$
15/255	$3.83 \cdot 10^{-3}$	$0.93 \cdot 10^{-3}$
25/255	$1.20 \cdot 10^{-2}$	$0.30 \cdot 10^{-2}$

Table 6.1: NRSTD values calculated on the luminance channel of the HDR images acquired with the RGB and luminance-chrominance approaches respectively

given by Figures 6.17 and 6.18 and further reinforced by Figures 6.19 and 6.20: the HDR luminance composed according to Eq. (4.3) is much less noisy than the one composed following the conventional RGB procedures. One might wonder why it is so. After all, the luminance calculated as the average of the three HDR R,G,B channels is also much less noisy than each of these channels. The dramatic difference between the two sets of results can be explained by considering the weights used in the composition. These weights are the crucial element on which all the HDR composition depends. In the RGB approach, channels are processed separately. It means that the weights used for composing, say, the HDR red channel $\tilde{z}^R$ are based on the noisy LDR z_i^R, $i = 1, \ldots, N$, channels only. In the luminance-chrominance approach, the weights are instead defined as functions of the luminance components ζ_i^Y, $i = 1, \ldots, N$, which have about only half of the standard-deviation of the individual components z_i^C, $c = R, G, B$, $i = 1, \ldots, N$. Thus, these weights are less affected by the noise than their counterpart in the RGB processing.

6.4 Misalignment Case

In an ideal case the source LDR sequence is assumed to be perfectly aligned so that in every image the same pixel coordinate corresponds to the same point from the scene. It is very hard to capture a perfectly aligned exposure sequence with a consumer level camera. Computer controlled capturing settings have been proposed, e.g. [20], but the practicality of this is severely hindered by the fact that the setting needs a control computer, normally a laptop, to externally control the camera. The portability of

such a system is highly limited and thus these configurations are practical for mostly creating image archives or similar tasks. For consumer applications the maximum requirement is a tripod and in the optimal case not even that would be necessary. In shooting from a tripod (sub)pixel shifts between the frames are almost always created. An attempt can be made to correct the misalignment by registration. While numerous techniques have been introduced for the registration of misaligned image sequences, only a few are applicable for the case of differently exposed frames. [24] suggests two different methods that could be used for aligning differently exposed frames.

The state of the art HDR imaging techniques working in RGB space seem to be very vulnerable to misaligned sequences. Artifacts are created especially in the thin boundaries of very bright and very dark regions, for example the black lead stripes in stained glass windows, common subjects to HDR imaging. The artifacts created by the RGB approach are usually chromatic in nature and in particular often clearly created of components not present in the neighborhood of the erraneous pixel, while in the case of the luminance-chrominance approach the HDR colors are always a convex combination of the LDR counterparts and as such errors are in many cases not visible at all. To illustrate the problem, four enlarged details of HDR images provided with [24] are presented in 6.25 along with a comparative example between the RGB approach of Debevec's HDR Shop v.1.0.3 software [1] and the proposed approach in a luminance-chrominance space. In both Figure 6.25 and Figure 6.26 the chromatic aberrations in the boundaries between bright and dark areas should be especially noted, whereas in Figure 6.27 the absence of such artifacts in comparison to Figure 6.26 should be noted.

Figure 6.25: Four detail enlargements of RGB misalignment artifacts picked from images provided with [20].

Figure 6.26: An enlargement of an HDR scene composed of misaligned LDR data in RGB space.

Figure 6.27: An enlargement of an HDR scene composed of misaligned LDR data in a luminance-chrominance space.

7. CONCLUSIONS AND FUTURE WORK

In this contribution an efficient method for composing HDR images in luminance-chrominance color spaces was presented along with a method for tone mapping images acquired in this fashion. The proposed luminance-chrominance approach appears especially practical if the data is subject to some further postprocessing such as sharpening. These operations are implemented in consumer cameras, partly to overcome limitations of the optical system of the camera, partly to provide the user with a more attractive photographic result (though maybe very different than the reality). It is well known that sharpening, by boosting the high-frequencies, amplifies the noise. The higher SNR of the luminance ensures that the image can be sharpened without introducing disturbing noise or artifacts. Additionally, the human visual system lacks good sensitivity to high-frequencies in the chrominances. Thus sharpening is mostly needed for the luminance.

The results of the proposed HDR imaging approach are shown solely for the opponent color space. The differences caused by the selection of color space are almost always perceptual and subjective rather than objective. For all the luminance-chrominance color spaces falling under the umbrella defined in Section 2 the same properties hold for noise attenuation and hue preservation. Thus, the results are not presented here to avoid duplication. Generally the presented approach yields more realistic colors with better noise suppression qualities compared to the traditional RGB methods, and does so in a computationally efficient manner. Objective comparison of computational efficiency is difficult due to, among others, differences in implementation and execution environments and optimization of the different methods. However, in terms of complexity, the presented approach is not less efficient than any RGB-based one. In particular, the complexity of the proposed luminance-chrominance image composition is comparable to that of [5], which is arguably the simplest composition method in the literature. Likewise, the proposed luminance-chrominance tone mapping technique is comparable to performing the selected tone mapping operator $\mathcal{T}$ on a single channel (monochromatic image). Moreover, if starting with previously compressed luminance-chrominance images (e.g. JPEG), the conversions from and to RGB are not necessary, further simplifying the application of our methods.

The new approach yields more realistic colors with better noise suppression qual-

ities compared to the traditional RGB methods, and does so in a computationally efficient manner. It is a justified belief that the advantages of the proposed approach shall motivate the migration of HDR processing from RGB to luminance-chrominance space.

REFERENCES

[1] HDR shop: High dynamic range image processing and manipulation. available at http://www.hdrshop.com.

[2] C. J. Bartleson and E. J. Breneman. Brightness reproduction in the photographic process. *Photographic Science and Engineering*, 11(4):254–262, 1967.

[3] A.J. Blanksby, M.J. Loinaz, D.A Inglis, and B.D. Ackland. Noise performance of a color cmos photogate image sensor. In *Electron Devices Meeting, 1997. Technical Digest., International*, pages 205–208, 1997.

[4] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman. Spatially nonuniform scaling functions for high contrast images. In *Proceedings of Graphics Interface '93*, pages 245–253, San Francisco, CA, 1993. Morgan Kaufmann.

[5] P. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. 24th Conf. Comp. Graphics and Interactive Techniques*, pages 369–378, 1997.

[6] K. Devlin, A. Chalmers, A. Wilkie, and W. Purgathofer. Tone reproduction and physically based spectral rendering. In *Proc. EUROGRAPHICS'02*, volume 22, 2002.

[7] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[8] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. In *Proceedings of ACM SIGGRAPH 2002*, 2002.

[9] A. Gilchrist, C. Kossyfidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, and E. Economou. An anchoring theory of lightness perception. *Psychological Review*, 106(4):795–834, Oct 1999.

[10] M. Goesele, W. Heidrich, and H-P. Seidel. Color calibrated high dynamic range imaging with icc profiles. In *Proc. 9th Color Imaging Conference, Scottsdale, USA*, pages 286–290, 2001.

[11] R. C. Gonzales and R. E. Woods. *Digital Image Processing, 2/E.* Prentice-Hall, Upper Saddle River, NJ, 2001.

[12] F. Kainz, R. Bogart, and D. Hess. The openexr image file format. In *SIGGRAPH Technical Sketches*, 2003.

[13] G. Krawczyk, K. Myszkowski, and H. Seidel. Lightness perception in tone reproduction for high dynamic range images. In *Proc. EUROGRAPHICS'05*, volume 24, 2005.

[14] G.W. Larson. Overcoming gamut and dynamic range limitations in digital images. In *IS&T 6th Color Imaging Conference*, 1998.

[15] G.W. Larson, H. Rushmeier, and C. Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Trans. Visual and Computer Graphics*, 3(4):291–306, 1997.

[16] G.W. Larson and M. Simmons. Subband encoding of high dynamic range imagegy. In *First ACM Symposium on Applied Perception in Graphics and Visualization (APVG)*, pages 83–90, 2004.

[17] B. C. Madden. Extended intensity range imaging. Technical report, GRASP Laboratory, University of Pennsylvania, 1993.

[18] S. Mann and R. Picard. Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proc. IS&T 46th Conf.*, pages 422–428, 1995.

[19] T. Mitsunaga and S.K. Nayar. Radiometric self calibration. In *Proc. IEEE Conf. Comp. Vision and Pattern Recognition*, pages xx–yy, 1999.

[20] S. O'Malley. A simple, effective system for automated capture of high dynamic range images. In *Proc. IEEE Int. Conf. Computer Vision Systems, ICVS'06*, 2006.

[21] A. V. Oppenheim, R. Schafer, and T. Stockham. Nonlinear filtering of multiplied and convolved signals. In *Proceedings of the IEEE*, volume 56, pages 1264–1291, 1968.

[22] K. Plataniotis and A. Venetsanopoulos. *Color image processing and applications*. Springer-Verlag New York, Inc., New York, NY, 2000.

[23] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. In *Proceedings of ACM SIGGRAPH 2002*, 2002.

[24] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High Dynamic Range Imaging*. Morgan Kaufmann Publishers, San Francisco, CA, 2006.

[25] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. *ACM Trans. Graph.*, 23(3):760–768, 2004.

[26] S. Süsstrunk, R. Buckley, and S. Swen. Standard rgb color spaces, 1999.

[27] J. Tumblin, J. K. Hodgins, and B. K. Guenter. Two methods for display of high contrast images. *ACM Transactions on Graphics*, 18(1):56–94, jan 1999.

[28] G.J. Ward. The radiance lighting simulation and rendering system. In *Proceedings of '94 SIGGRAPH conference*, pages 459–472, 1994.