Chapter 1

Color High Dynamic Range Imaging: Algorithms for Acquisition and Display

OSSI PIRINEN

OptoFidelity Ltd
Tampere, Finland
Email: ossi.pirinen@optofidelity.com

ALESSANDRO FOI

Department of Signal Processing
Tampere University of Technology
Tampere, Finland
Email: alessandro.foi@tut.fi

ATANAS GOTCHEV

Department of Signal Processing
Tampere University of Technology
Tampere, Finland
Email: atanas.gotchev@tut.fi

1.1 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'.

The problem of high dynamic range (HDR) imaging is two-fold. First, how to capture the true luminances, and the full chromatic information of an HDR scene possibly with capturing device with a dynamic range smaller than the scene's 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 chapter the terms used are composition for the acquisition process and tone mapping for the displaying part. Figures 1.1, 1.2 and 1.3 illustrate the HDR imaging problem and the solution suggested by HDR imaging techniques. The conventional low dynamic range (LDR) snapshot of Figure 1.1 has a very limited dynamic range, which is seen as clipped bright parts as well as underexposed 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.

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 [1]. The multiple exposure method relies on combining the different exposures into a single image, spanning the whole dynamic range of the scene [2], [3]. The image composition step requires a preliminary calibration of the camera response [3], [4]. Though HDR capture and display hardware are both briefly discussed, the focus of this chapter is on the approach of multiple capture composition.

The second part of the problem then is to appropriately tone map the so-obtained HDR image back to an LDR display. Tone mapping techniques range in complexity from a simple gamma-curve to sophisticated histogram equalization methods and complicated lightness perception models [5], [6], [7]. Recently, also display devices equipped with extended dynamic range have started to appear [8].

HDR imaging methods have been originally developed for RGB color models. However, techniques working in luminance-chrominance seem more meaningful and preferable for a number of reasons. *First*,

Figure 1.1: Single exposure from a clearly HDR scene.

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 chapter, the problem of HDR imaging in a generic luminance-chrominance space is addressed, efficient algorithms for HDR image composition and tone mapping are presented, and the functionality of such an approach under various conditions studied.

The chapter is structured as follows. Section 1.2 provides general information concerning color spaces, as well as the notations used in the chapter. Techniques essential to HDR acquisition, namely camera response calibration, are discussed in Section 1.3. A brief overview of HDR sensors is also included in Section 1.3. Section 1.4 focuses on the creation of HDR images. Techniques for both monochromatic and color HDR are presented. Section 1.5 deals with displaying (tone mapping) HDR images. Display devices capable of presenting HDR content are also discussed. Section 1.6 contains examples and discussion. Conclusions are drawn in Section 1.7.

Figure 1.2: HDR image linearly scaled to fit the normal 8-bit display interval. The need for tone mapping is evident.

1.2 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 [9], [1].

1.2.1 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 [10]. All the standardized RGB (Red, Green, Blue) color spaces can be defined by giving the CIE XYZ 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, [11]. 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

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

implementation simplicity, the RGB coordinates are usually limited to $[0\ 1]^3$ (floating point) or $[0\ 255]^3$ (8-bit unsigned integer).

1.2.2 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 the human visual system (HVS). These features can be intuitively formalized through the introduction of luminance-chrominance color models.

Throughout the chapter 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 [9].

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.

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).

The Opponent Color Space

The opponent color space, presented in e.g. [9] 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

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

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, $[\zeta^Y \; 0 \; 0]\,\mathbf{B} = \mathbf{z}_{\text{gray}}(\mathbf{x})$, 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.

1.2.3 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 for 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

$$\mathbf{A}_{yuv} = \begin{bmatrix} 0.30 & -0.17 & 0.50 \\ 0.59 & -0.33 & -0.42 \\ 0.11 & 0.50 & -0.08 \end{bmatrix} \quad \mathbf{B}_{yuv} = \mathbf{A}_{yuv}^{-1} = \begin{bmatrix} 1 & 0 & 1.4020 \\ 1 & -0.3441 & -0.7141 \\ 1 & 1.7720 & 0 \end{bmatrix}.$$

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.

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.

The hue can be defined as $H = \arctan \frac{\zeta^U}{\zeta^V}$ and the saturation as $S = \sqrt{(\zeta^U)^2 + (\zeta^V)^2}$ or, normalized with the value ζ^Y, as $S = \frac{\sqrt{(\zeta^U)^2+(\zeta^V)^2}}{\zeta^Y}$. 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, e.g. in [12] and while the presented ones are among the more frequently noted ones, they are chosen here first and foremost for the sake of simplicity.

1.3 HDR Acquisition

In this section, techniques essential for successful HDR acquisition are presented. As imaging devices directly capable of capturing HDR content are scarcely available on the consumer market, the means for capturing HDR content with a normal camera are given. In addition a quick cross-section of the fast-evolving field of imaging sensor tecnologies directly capable of capturing HDR scenes is presented.

1.3.1 Camera Response Function

The most essential step in 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. Conversely, knowing the camera response is crucial for determining the irradiance of the source scene from the given image data. Essentially two techniques exist for the response calibration. Debevec and Malik approach defined in [3] extends the previous work done by Mann and Picard in [2]. 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 later in this section in detail.

The other calibration method found in the literature was presented by Mitsunaga and Nayar and described in [4]. 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.

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

$$\mathfrak{z}(\mathbf{x}) = f(e(\mathbf{x})) = f(E(\mathbf{x})\,\Delta t), \tag{1.1}$$

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. [3])

$$\ln E(\mathbf{x}) = g(\mathfrak{z}(\mathbf{x})) - \ln \Delta t, \tag{1.2}$$

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 can then be used for the linearization of the input values in all subsequent HDR compositions of the same device.

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. [13]. 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 increasing 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 [3]

$$\sum_{j=1}^{P}\sum_{i=1}^{N} w_{\mathsf{cam}}(\zeta_i^Y(\mathbf{x}_j))\left[g(\zeta_i^Y(\mathbf{x}_j)) - \ln E_i - \ln \Delta t_i\right]^2 + \lambda \int_0^1 w_{\mathsf{cam}}(\zeta) g''(\zeta)^2 d\zeta, \tag{1.3}$$

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 chapter 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 [3]. If instead of processing directly one of the sensor's RGB outputs, a combination ζ_i^Y, (e.g. luminance) is considered, it should be noted that some pixels in this combination can include overexposed components

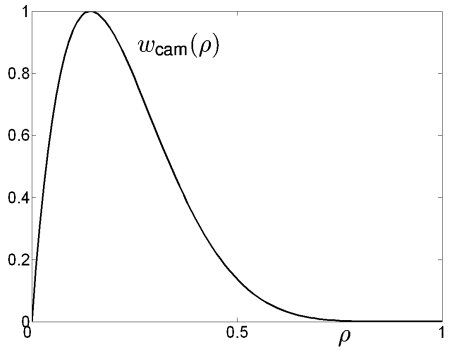


Figure 1.4: Weight function used for the luminance values in the camera response function definition.

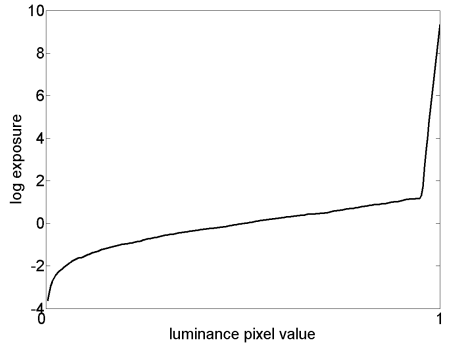


Figure 1.5: The estimated inverse response function g for the luminance channel of Nikon COOLPIX E4500.

without reaching the upper limit of the range of the combination. 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 [3]) 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 1.4. An example of a camera response function for the luminance solved with this method is illustrated in Figure 1.5. 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. As discussed in [14] 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.

1.3.2 HDR Sensors

The importance and the potential of HDR imaging are further highlighted by the introduction of HDR hardware. Novel camera sensor technologies have recently been unveiled by a few companies. For the direct capture of HDR data several sensor designs exist [1], [15], [16]. Most HDR sensor designs focus on enhancing the dynamic range by adding to the maximum amount of light the sensor can capture without saturating,

rather than enhancing sensitivity for low light levels. The work of [15] studies four different architectures: time-to-saturation, multiple capture, asynchronous self-reset with multiple capture, and synchronous self-reset with residue readout. Time-to-saturation and multiple capture offer extended dynamic range by various methods of varying exposure time. Varying exposure time does exactly what one would assume; it adapts the pixel-wise integration time to pixel photocurrent enabling the use of long exposure times for small photocurrents and shorter exposure times for larger ones. In asynchronous self-reset the pixel voltage is sampled at regular intervals and pixels reset autonomously on saturation. Sychronous self-reset requires on-pixel counter to keep track of reset count. At the end of integration, the number of resets is combined with the residual voltage to estimate the total photocurrent.

None of the abovementioned technologies have yet been made available to the general public. A couple of digital single-lens reflex (DSLR) cameras however are. Fujifilm's S5 Pro and its new CCD-array is the first attempt of bringing direct extended dynamic range sensors 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 is saturated. The dynamic range improvement takes place at the bright end of the scene and the total dynamic range is according to Fuji comparable to traditional film. Another example of HDR imaging in a consumer camera is the Pentax K-7. Though not featuring a true HDR sensor, the 2009 released DSLR is to the authors' best knowledge the first commercial camera to provide in-camera HDR capture and tone mapping. An automated bracketing and composition program is made available, and the user is presented with a tone mapped HDR image.

1.4 Algorithms for HDR Composition

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], [2]) 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 [3]. The vast majority of research done in the field has focused on optimizing the parameters (number of exposures, exposure times, etc.), e.g. [18] or developing different tone mapping methods, e.g. [6]. Though parallel treatment of the RGB channels causes distortions in the colors (see e.g. [3] and examples in Section 1.6), 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 [19]. 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].

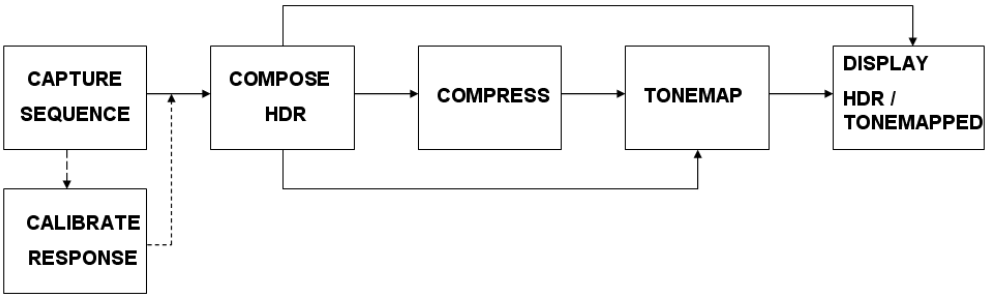


Figure 1.6: The block diagram of the general multiple exposure HDR imaging pipeline. Different paths indicate alternatives in the process, such as displaying a tone mapped HDR image on a normal display or alternatively displaying an HDR image on an HDR capable display.

1.4.1 Monochromatic HDR Composition

The problem of HDR imaging and the state-of-the-art techniques are discussed in [1]. 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 in Section 1.3.1. With this prerequisite filled, the general line-up of the process, illustrated in the block diagram of Figure 1.6, 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.

With the camera response function solved as described in Section 1.3.1 and the exposure times for each captured frame known, the logarithmic HDR radiance map for a monochromatic channel C can then be composed as a weighted sum of the camera output pixels values as follows.

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

In case of color images, the RGB channels are treated in parallel. This assumes that the interactions between channels are negligible, which is, as even the author of [3] 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 provided that the source sequence is very well aligned and noise is negligible.

1.4.2 Luminance-Chrominance HDR Composition

An alternative approach to the parallel processing of the three RGB channels 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 method described here treats the luminance channel similarly to the conventional approach for a single channel of the RGB triplet and presents a more sound method for the composition of color.

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.

Luminance Component Composition

The HDR luminance component is obtained by a pixelwise weighted average of the pixel log irradiances defined according to Eq. (1.2) 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 1.7. The logarithmic HDR luminance is obtained as

$$\ln \tilde{\zeta}^Y(\mathbf{x}) = \frac{\sum_{i=1}^N w^Y(\zeta_i^Y(\mathbf{x}))(g(\zeta_i^Y(\mathbf{x})) - \ln \Delta t_i)}{\sum_{i=1}^N w^Y(\zeta_i^Y(\mathbf{x}))}. \tag{1.4}$$

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.

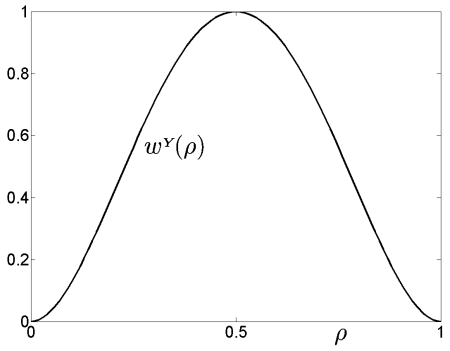


Figure 1.7: Weight function w^Y used for the luminance channel composition.

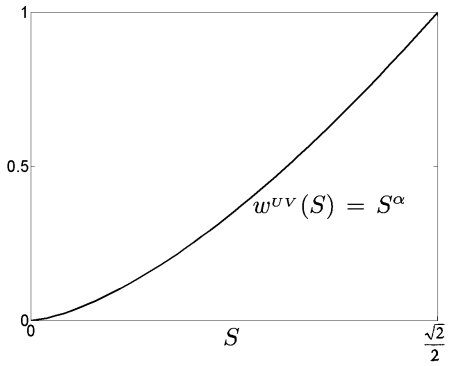


Figure 1.8: Weight function w^{UV} used for the composition of the chrominance channels.

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 is illustrated in Figure 1.8. 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

$$\tilde{\zeta}^C(\mathbf{x}) = \frac{\sum_{i=1}^N w^{UV}(S_i(\mathbf{x}))\zeta_i^C(\mathbf{x})}{\sum_{i=1}^N w^{UV}(S_i(\mathbf{x}))}, \tag{1.5}$$

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

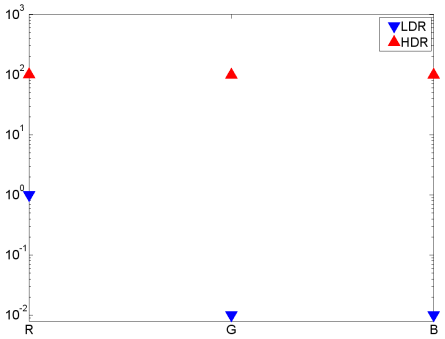


Figure 1.9: The illustration of the desaturating effect of the luminance-chrominance HDR composition.

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.

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 $[-0.5\ 0.5]$ 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 1.2.2 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 1.9.

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

$$\mu(\mathbf{x}) = \frac{\tilde{\zeta}^Y \sum_{i=1}^N w^{UV}(S_i(\mathbf{x}))}{\sum_{i=1}^N w^{UV}(S_i(\mathbf{x}))\,\zeta_i^Y(\mathbf{x})}. \tag{1.6}$$

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. (1.6) and Eq. (1.5) 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 1.10. 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 1.11 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 luminance-chrominance method.

1.5 Algorithms for HDR Display

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.

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 for instance in Figure 1.12. The same scene is shown in Figure 1.13 tone mapped using the method described in Section 1.5.2. Again the age-old phrase

Figure 1.10: The HDR image is composed in luminance-chrominance space and transformed directly to RGB for tone mapping.

Figure 1.11: 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.

holds, a picture is worth a thousand words: an HDR image is useless on an LDR display device without tone mapping.

Figure 1.12: An HDR image of a stained glass window from Tampere Cathedral, Finland. The image is linearly scaled for display.

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 by Devlin et al. [6].

Figure 1.13: 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 1.5.2.

1.5.1 Tone-Mapping of Monochromatic HDR

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 [1] 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 tone mapping operators (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 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 neighbor-

hoods. 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 [20]. 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. 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 Section 1.5.2.

1.5.2 Tone-Mapping of Luminance-Chrominance HDR

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 chapter 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 presented.

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.

Anchoring-Based Compression

The first method is based on an anchoring theory of lightness perception and was presented for RGB in [7]. The core idea of the method is to divide the luminance image into frameworks based on the theory of lightness perception [21]. 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 1.14.

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 Section 1.6.

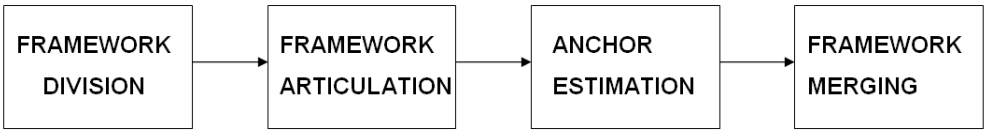


Figure 1.14: The block diagram of the anchoring based tone mapping process. The blocks represent the four central stages of the procedure.

Histogram-Based Compression

The other method presented here is based on a relatively simple histogram adjustment technique. It coarsely approximates the method described in [5]. It is noted however that most of the sophisticated ideas introduced in [5] 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 ($10^{-4}cd/m^2$). 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. 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 Section 1.6.

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

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

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

It can be noted that $\mathring{\mathbf{z}}_{\text{gray}}(\mathbf{x})$ 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

$$\mathring{\mathbf{z}}(\mathbf{x}) = \mathring{\mathbf{z}}_{\text{gray}}(\mathbf{x}) + \delta(\mathbf{x})\,\mathbf{z}_{\text{chrom}}(\mathbf{x}) \in [0,1]^3. \tag{1.7}$$

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

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

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 (1.7) to hold. Figure 1.15 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

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

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.

1.5.3 HDR Display Hardware

Obviously even the most elaborate tone mapping algorithms cannot match the sensation of directly viewing a high dynamic range scene. The limitations in dynamic range of modern displays are based on partly

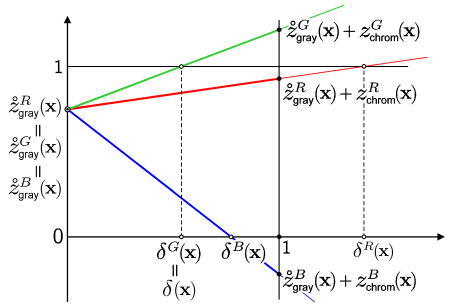


Figure 1.15: Illustration of the definition of the chromatic tone mapping parameter δ.

the limited maximum presentable intensity and more importantly, the minimum presentable intensity. Conventional LCDs and even plasma screens both suffer from the phenomenon of some light being present in pixels that should in fact be completely black. For LCDs this is due to the LCD screen not being able to completely block the backlight from reaching the observer. For plasma screens grayish blacks are caused by the precharge induced in the plasma needed to achieve an acceptable response time.

In recent years some advances have beed made in display technology. The paper [8] presents the fundamentals of a conceptual adaptive backlight LCD display. In such a display the backlight is implemented with an LED matrix. Such backlight can then be locally dimmed and areas of the backlight can even be completely turned off to achieve true black in some image areas while simultaneously displaying maximum intensity in others, yielding a theoretical infinite contrast ratio. The adaptive backlight was later formulated into two technology packages by Dolby. Dolby Contrast and Dolby Vision offer different levels of backlight control and accuracy and are at the time of writing waiting to be licensed by the first display manufacturer. Meanwhile numerous display manufacturers have introduced LED-based LCD screens of their own, significantly extending the dynamic range (contrast ratio) of the conventional LCD TV. Many of the caveats of current display technologies will be antiquated by the maturation of OLED and similar technologies where all pixels control their own intensity and can be turned on or off independently of neighboring image elements.

1.6 Examples

In this section we present examples of HDR scenes imaged and visualized with methods described in this chapter. We consider both real and synthetic image data. Special attention is focused on the effects of noise and misalignment, both realistic components of distortion when HDR scenes are imaged with off-the-shelf components. Extensive numerical and visual comparison against state-of-the-art RGB methods can be found in [14].

1.6.1 HDR from Real LDR Data

To display the basics of luminance-chrominance HDR composition, a sequence of 28 frames was captured with a remotely operated Canon EOS 40D set on a tripod. The images were captured with ISO 100 and have thus very low noise. The image sequence is composed into HDR in opponent color space with the methods described in this chapter. First a response function is solved for the camera as described in Section 1.3.1. Using the solved response function an HDR image in luminance-chrominance (in this case, opponent) space is acquired as described in Section 1.4.2. The exposure times for the captured sequence range from 0.01s, to 5.00s. Extracts of the sequence can be found in Figure 1.17. The histogram for the R,G and B channels of the HDR image is shown on logarithmic axis is Figure 1.16. The tone mapping results of the two methods described in this chapter in Figures 1.18 and 1.19 and an adaptive logarithmic tone mapping result of the RGB transformed HDR image done with the method described in [22] in Figure 1.20. When tone mapping with the method of [22] implemented in [23], the following parameters are used for the tone mapping of all images: exposure adjustment 0.64, bias 0.77, shadow luminance 1.00, contrast 0.00.

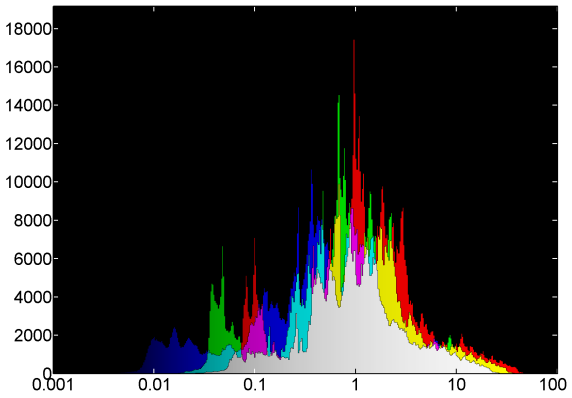


Figure 1.16: The histograms of the HDR image composed in luminance-chrominance space and transformed into RGB. Histogram is displayed in logarithmic scale.

1.6.2 HDR from Synthetic LDR Data

The HDR image acquired as described in Section 1.6.1, can also be transformed into RGB space as described in Section 1.4.2. The HDR image transformed into RGB space shall then function as 'ground truth' for subsequent synthetic composition examples. The HDR image is assumed to represent the true scene irrandiance E, with an HDR pixel value $E(\mathbf{x}) = 1$ equal to a scene irradiance of $1\ W/m^2$. Where comparison with RGB techniques is provided, the tone mapping described in [22] is used to tone map images composed both in RGB and in luminance-chrominance to allow for relevant visual comparison of the results.

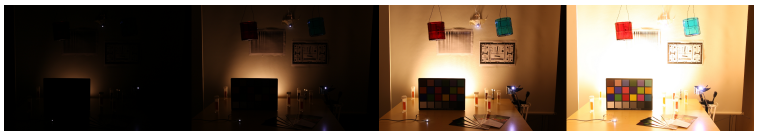


Figure 1.17: Four exposures from the real LDR sequence used to compose an HDR image. The exposure times of the frames are 0.01, 0.0667, 0.5, 5 seconds.

Figure 1.18: The HDR image composed in luminance-chrominance space from real data, tone mapped with the histogram adjustment technique presented in Section 1.5.2.

Generating Synthetic LDR Data

In practice, the LDR frames are captured sequentially, one after the other, each time with a different exposure time. Particularly with a hand-held camera, this results in some misalignement between the frames. Unless the scene is perfectly static, misalignment can occur even when the camera is held on a tripod, simply because minor changes in the scene content (movement of the subjects, etc.). In our simulations, to create realistic misalignement of the LDR frames, the HDR image has been sampled using linear interpolation after a set of geometrical deformations. These include rotation, translation, as well as parabolic stretch, faithfully mimicking the apparent nonrigid deformation produced by the optical elements of the camera. Although we use random deformation parameters to generate each frame, we quantify the actual misalignment by tracking the position of the three light-emitting diodes (LEDs, used as markers in the HDR scene) over the sequence of frames. In particular, for each sequence of LDR frames, we compute the standard deviation of the position of the center of each of the three LEDs. The numbers reported in Table 1.1 are the average variances for position of the three LEDs.

A synthetic LDR image sequence is obtained by simulating the acquistion of the above HDR image into LDR frames. The LDR frames are produced through the equations (1.1)-(1.2) applied separately on the R,G

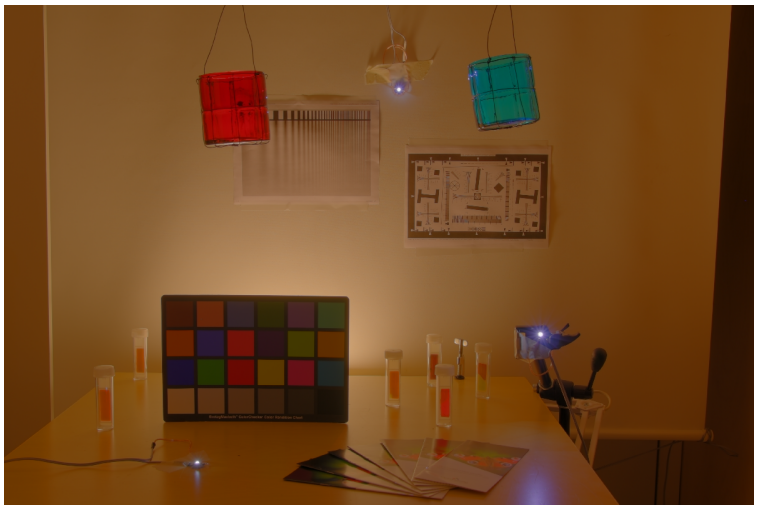


Figure 1.19: The HDR image composed in luminance-chrominance space from real data, tone mapped with the anchoring technique presented in Section 1.5.2.

and B channels. The camera response function f used in the equations is defined as

$$f\left(e\left(\mathbf{x}\right)\right)=\max\left\{0,\min\left\{1,\left\lfloor\left(2^{8}-1\right)\kappa e\left(\mathbf{x}\right)\right\rceil\left(2^{8}-1\right)^{-1}\right\}\right\}. \tag{1.8}$$

$\kappa=1$ is a fixed factor representing the acquisition range of the device (full well) and the $\lfloor\cdot\rceil$ brackets denote the rounding to the nearest integer, thus expressing 8-bit quantization. This is a simplified model which, modulo the quantization, corresponds to a linear response of the LDR acquisition device.

Noise is also introduced into the so-obtained LDR images. Aiming at realism, we follow a signal-dependent noise model of the sensor [24],[25]. More precisely, noise corrupts the term $\kappa e\left(\mathbf{x}\right)$ in the acquisition equation (1.8), which thus becomes

$$f\left(e\left(\mathbf{x}\right)\right)=\max\left\{0,\min\left\{1,\left\lfloor\left(2^{8}-1\right)\varepsilon\left(\mathbf{x}\right)\right\rceil\left(2^{8}-1\right)^{-1}\right\}\right\}, \tag{1.9}$$

where $\varepsilon\left(\mathbf{x}\right)=\kappa e\left(\mathbf{x}\right)+\sqrt{a\sigma\left(\kappa e\left(\mathbf{x}\right)\right)+b}\eta\left(\mathbf{x}\right)$ and η is standard Gaussian noise, $\eta\left(\mathbf{x}\right)\sim\mathcal{N}(0,1)$. We use the parameters a=0.004, b=0.02^2, which correspond to the noise model of a Fujifilm FinePix S9600 digital camera at ISO1600 [25]. The selected noise level is relatively high. Such a high noise level is intentionally selected to clearly visualize the impact of noise on the HDR compositions.

Composition from synthetic LDR frames

By generating the LDR frames synthetically, we have exact knowledge of the camera response. Thus, the logarithmic inverse g of the camera response function f used in the composition is the logarithm $g(x)=\ln x$. In the case of synthetic LDR frames, when processing in RGB, this response is used for all channels, processed in parallel, to compose the original scene. When processing synthetic LDR frames

Figure 1.20: The HDR image composed in luminance-chrominance space from real data, tone mapped in RGB with the adaptive logarithmic technique presented in [22].

in a luminance-chrominance space, the same inverse response function is used for the luminance channel, while chrominances are processed as described in Section 1.4.2. As RGB composition is essentially an inverse of the acquisition process followed to obtain the synthetic LDR frames, the RGB composition result is, apart from some minute quantization-induced differences, perfect. For luminance-chrominance this cannot be expected, since the luminance channel is always obtained as a combination of the RGB channels which due to clipping reduces the accuracy of the response. As we shall see however, the reduced accuracy does not compromise the actual quality of the composed HDR image. Instead, the luminance-chrominance approach leads to much more accurate composition when the frames are degraded by noise or misaligned.

As we deal with imperfect image data corrupted by noise, we use the normalized saturation definition given in Section 1.2.3 modified with a regularization term τ. The regularized saturation is then obtained as $S = \frac{\sqrt{(\zeta^U)^2+(\zeta^V)^2}}{\sqrt{(\zeta^Y)^2+\tau^2}}$. For all experiments, we used $\tau = 0.1$. Without regularization, such saturation would become unstable at low luminance values, eventually resulting in miscalculation of the weights for the composition of the chrominances.

Experiments and discussion

Examples are provided as follows: the trivial case of HDR composition from perfect synthetic frames is provided for reference. Further, two different levels of misalignment are presented with and without added noise. Also a perfectly aligned case is presented with added noise. For every case, images composed in both LCR and RGB space are presented, tone mapped with the method of [22] in RGB space. This demonstrates the presented luminance-chrominance composition's ability to function under imperfect conditions. The parameters for the tone mapping algorithm are the same for all cases: exposure adjustment 0.64, bias 0.77,

shadow luminance 1.00, contrast 0.00.

As can be witnessed in Figures 1.22a and 1.22b, the HDR compositions done from uncorrupted frames are, except a slight overall tonal difference, identical. Both LCR and RGB perform very well with synthetic frames not subject to any degradation. Figures 1.22c and 1.22d show tone mapped HDR images composed of frames subjected to simulated camera shake. The average variance of the LED positions used to measure sequence stability is 0.2289 pixels, as shown in Table 1.1. The left-hand luminance-chrominance composed frame is again visually very pleasing, whereas some clear color artifacts can be seen on high contrast edges in its right-hand RGB counterpart. Especially on the upper edge of the color chart some reddish and blueish artifacts can be witnessed. This is further illustrated in figure 1.23, which displays magnified fragments of Figure 1.22. Similar artifacts can also be seen on the upper-right edge of the white-screen. These artifacts are caused by the parallel processing of the RGB channels. In Figures 1.22e and 1.22f, the composition results are shown for images degraded by misalignment (average LED position variance 0.2833 pixels) and noise. Though masked somewhat by the noise, similar color artifacts as in the middle row can be witnessed for example in the upper right corner of the color chart and in the vertical shelf edge found in the left side of the scene. It is also evident that in both the luminance-chrominance and the RGB composition, noise is visible. It is however the dark parts of the RGB composition that suffer from this noise the most.

Effects very similar to Figures 1.22a-f can be seen in Figures 1.24a-f. On the first row, composition results for a synthetic sequence degraded by noise alone are shown. Again the comparison favors the luminance-chrominance composed image shown on the left. On the second row, the source sequence is degraded by camera shake with the average variance of 2.6858 pixels. The luminance-chrominance representation, apart from slight blurriness, seems visually acceptable. In the RGB image the color artifacts are very much visible on almost all of the edges. Among the worst are the greenish artifacts of the upper-right edge of the white-screen, further illustrated in figure 1.25, which displays magnified fragments of Figure 1.24 and the red distortions on the wires of the LED at the upper part of the scene. Figures 1.24e and 1.24f show the composition results for a sequence degraded by noise and camera shake with average variance of 1.5339 pixels. Again, both noise and color artifacts are very much present in the left-hand RGB composition result, whereas the right-hand luminance-chrominance composed image handles the imperfect conditions visibly significantly better.

It is interesting to comment about the blueish colored noise visible on the darker parts of the HDR images produced by the RGB composition of noisy LDR frames. First, as we can see in Figure 1.21, these are areas which remain rather dark even in the frame with the longest exposure ($\Delta t = 4.096$ s). Second, we can observe in Figure 1.20 that the scene is dominated by a cream yellow cast, essentially because of the tone of the lamp used for lighting. Thus, in each LDR frame, in these areas the blue component is the one with both the lowest intensity and the poorest signal-to-noise ratio. Because of the way weights are defined,

Sequence	Average LED position variance (pixels)
Reference	0
1st degraded, no noise	0.2289
2nd degraded, added noise	0.2833
Only added noise	0
3rd degraded, no noise	2.4859
4th degraded, added noise	1.5339

Table 1.1: Average LED position variances for the source image sequences of Tables 1.22 and 1.24.

Figure 1.21: The four synthetic LDR frames used to create the HDR images shown tone mapped in the first row of 1.22. The exposure times used in creating the frames are 0.064, 0.256, 1.024, 4.096 seconds.

for the composition of these dark areas, only the longest exposed frame contributes with significant weights. This situation is quite different from that of the other parts of the image, which are instead produced as an average of two or more frames, one of which is properly exposed. In particular, for the darkest components, it is the right tail of the noise distribution which is awarded larger weights (see Figure 1.7). Moreover, because of the clipping at zero in Equation (1.9), the noise distribution itself is asymmetric. This results in a positive bias in the composition of the darker parts, which causes the blueish appearence of the noise over those parts.

1.7 Conclusions

In this chapter we have presented methods for capture, composition and display of color HDR images. In particular, we have considered composition techniques which effectively allow to produce HDR images using LDR acquisition devices. Composition in luminance-chrominance space is shown to be especially suitable for the realistic imaging case where the source LDR frames are corrupted by noise and/or misalignment. In addition to being more robust to degradations, the luminance-chrominance approach to HDR imaging focuses special attention on the faithful treatment of color. As opposed to traditional methods working with the RGB channels, this method does not suffer from systematic color artifacts in cases of

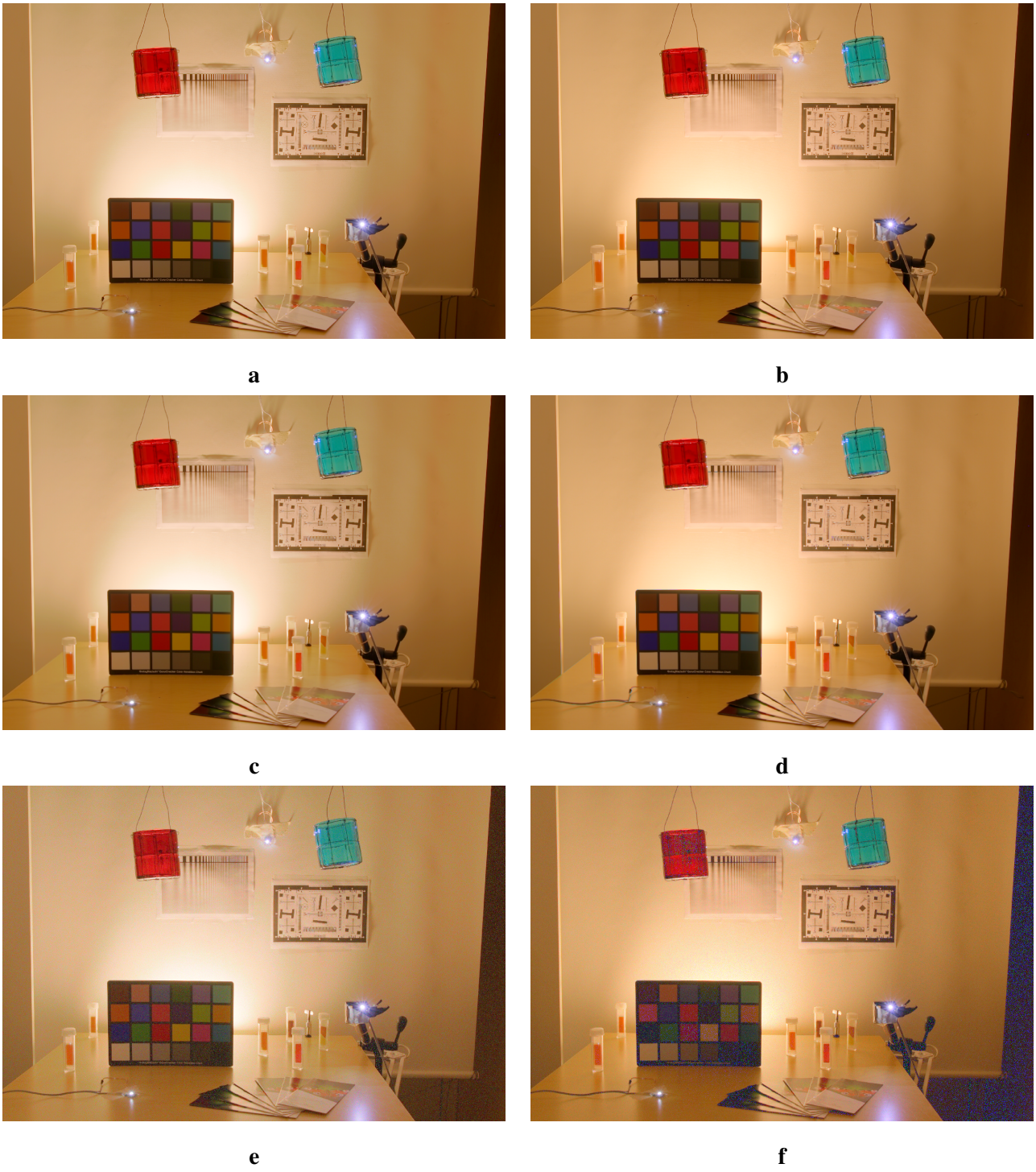


Figure 1.22: Experiments with synthetic data, **a** and **b** are the luminance-chrominance and RGB composed image of synthetic LDR data subject to no degradations; **c** and **d** are composed of a sequence subject to misalignment with measured average LED position variance of 0.2289; **e** and **f** are composed of a sequence subject to misalignment with measured average LED position variance of 0.2833 as well as noise.

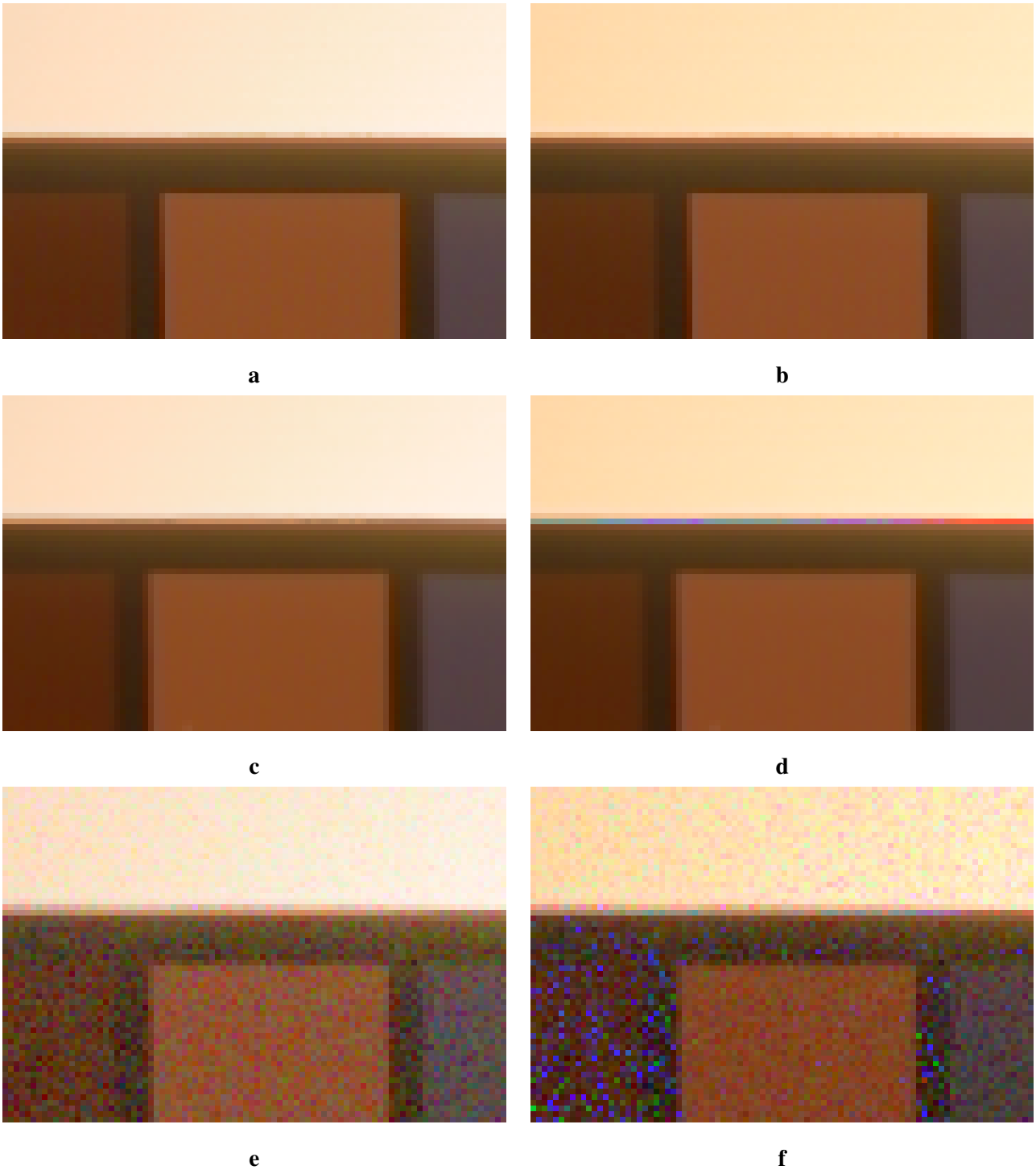


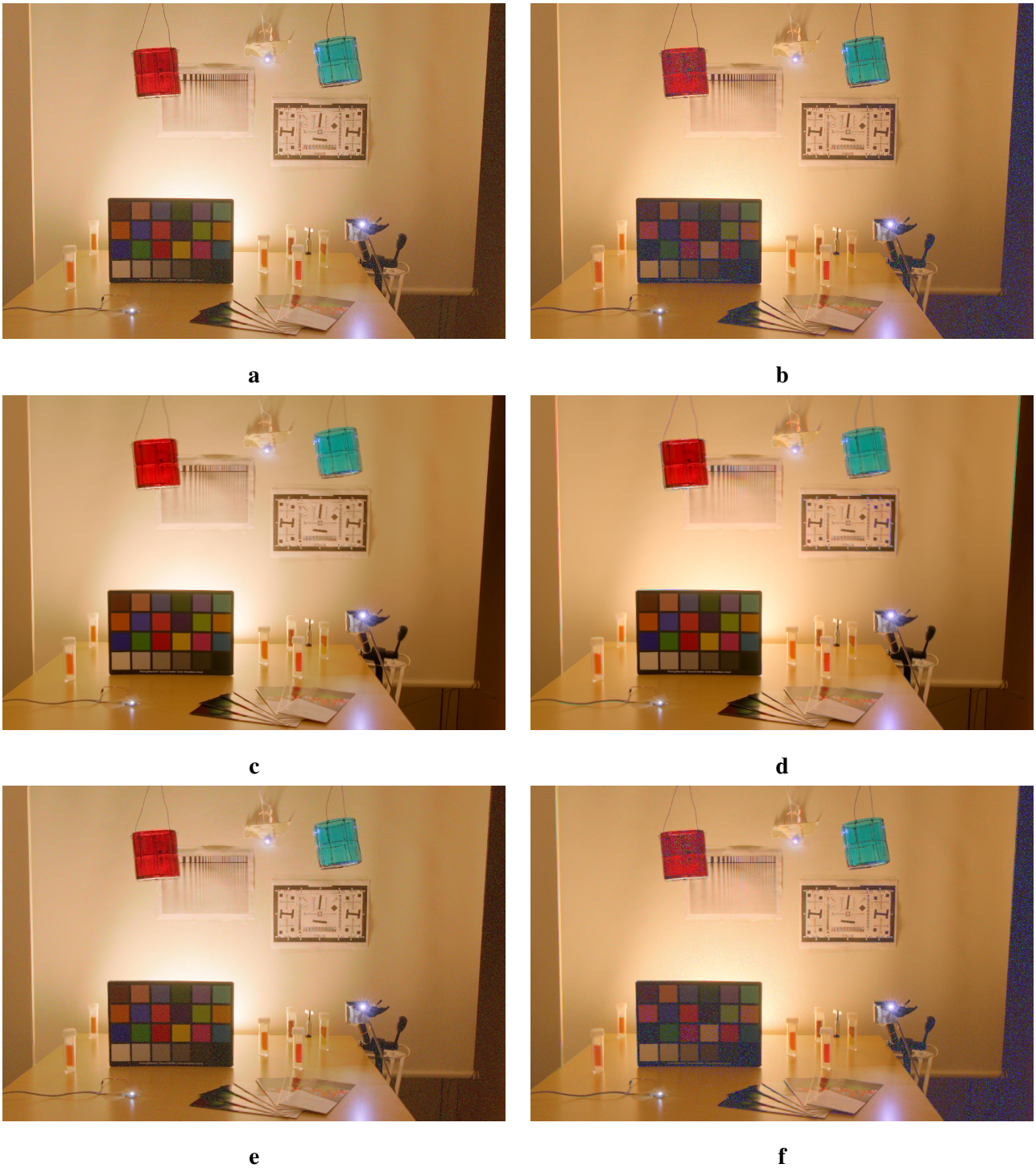
Figure 1.23: Magnified details extracted from the images of 1.22. **a** and **b** are the luminance-chrominance and RGB composed image of synthetic LDR data subject to no degradations; **c** and **d** are composed of a sequence subject to misalignment with measured average LED position variance of 0.2289; **e** and **f** are composed of a sequence subject to misalignment with measured average LED position variance of 0.2833 as well as noise.

misaligned data nor color balancing errors usually induced among other things by imperfectly calibrated camera response.

With the ongoing introduction of HDR acquisition and display hardware, HDR imaging techniques are set to gain an ever important role in all parts of computational photography and image processing. With new applications, both industrial and commercial, introduced nearly daily, it is not far fetched to say that HDR will in more than one meaning be ”the new color TV.”

Acknowledgements

This work was in part supported by OptoFidelity Ltd (www.optofidelity.com) and in part by the Academy of Finland (project no. 213462, Finnish Programme for Centres of Excellence in Research 2006-2011, project no. 118312, Finland Distinguished Professor Programme 2007-2010, and project no. 129118, Postdoctoral Researcher’s Project 2009-2011).



a **b**

c **d**

e **f**

Figure 1.24: Experiments with synthetic data, **a** and **b** are the luminance-chrominance and RGB composed image of synthetic LDR data subject to noise; **c** and **d** are composed of a sequence subject to misalignment with measured average LED position variance of 2.4859; **e** and **f** are composed of a sequence subject to misalignment with measured average LED position variance of 1.5339 as well as noise.

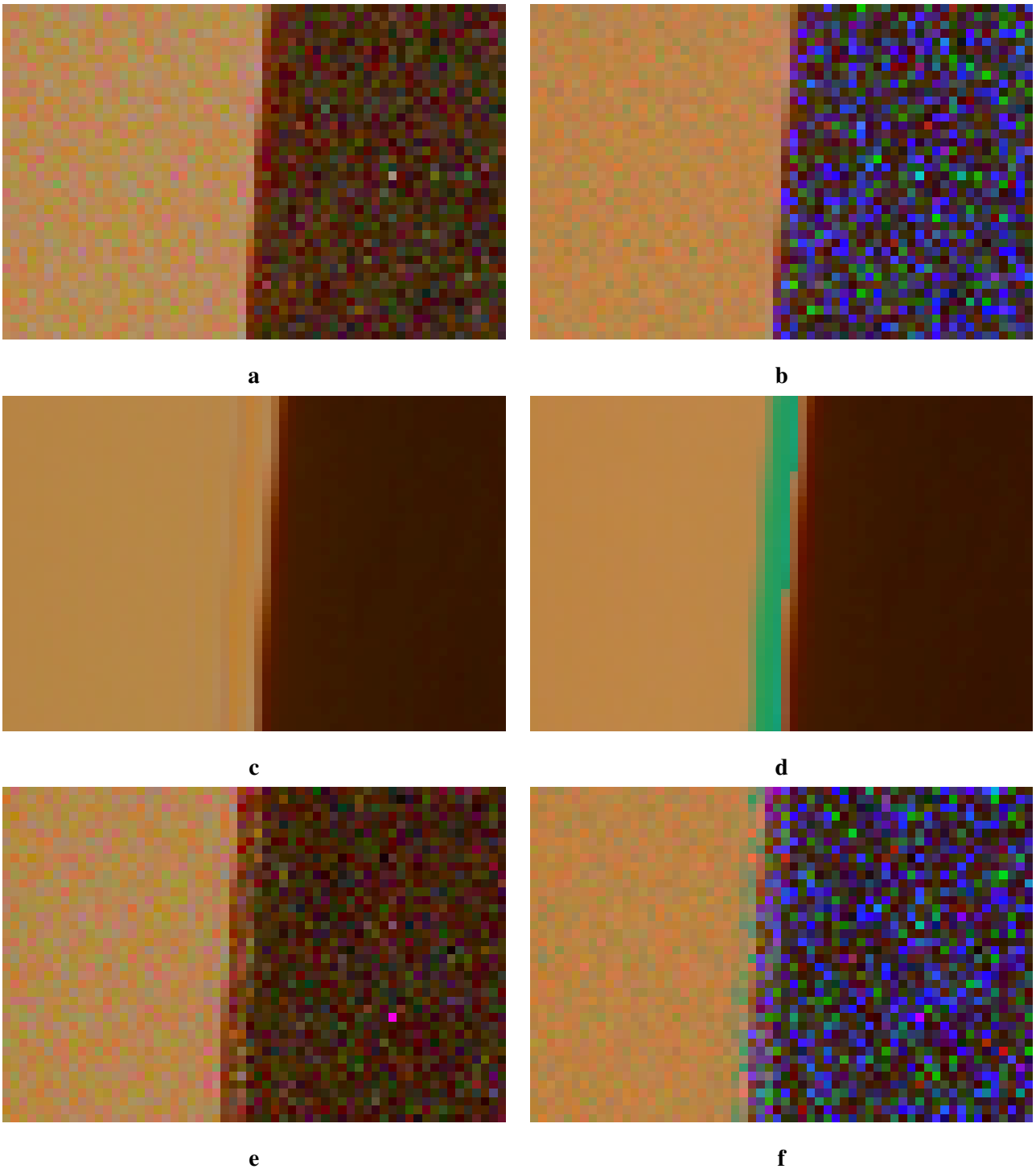


Figure 1.25: Magnified details extracted from the images of 1.24. **a** and **b** are the luminance-chrominance and RGB composed image of synthetic LDR data subject to noise; **c** and **d** are composed of a sequence subject to misalignment with measured average LED position variance of 2.4859; **e** and **f** are composed of a sequence subject to misalignment with measured average LED position variance of 1.5339 as well as noise.

Bibliography

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

[2] S. Mann and R. Picard, "Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," in *Proc. IS&T 46th Conf.*, (Boston, MA, USA), pp. 422–428, May 1995.

[3] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. 24th Conf. Comp. Graphics and Interactive Techniques*, pp. 369–378, August 1997.

[4] T. Mitsunaga and S. Nayar, "Radiometric self calibration," in *Proc. IEEE Conf. Comp. Vision and Pattern Recognition*, pp. –380, June 1999.

[5] G. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Visual and Computer Graphics*, vol. 3, pp. 291–306, October 1997.

[6] K. Devlin, A. Chalmers, A. Wilkie, and W. Purgathofer, "Tone reproduction and physically based spectral rendering," in *Proc. EUROGRAPHICS'02*, vol. 22, (Saarbrucken, Germany), September 2002.

[7] G. Krawczyk, K. Myszkowski, and H. Seidel, "Lightness perception in tone reproduction for high dynamic range images," in *Proc. EUROGRAPHICS'05*, vol. 24, (Dublin, Ireland), September 2005.

[8] 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.*, vol. 23, pp. 760–768, August 2004.

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

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

[11] S. Susstrunk, R. Buckley, and S. Swen, "Standard rgb color spaces," 1999.

[12] A.Ford and A. Roberts, "Colour space conversions." http://herakles.fav.zcu.cz/research/night_road/westminster.pdf, dated Aug 11 1998.

[13] A. Blanksby, M. Loinaz, D. Inglis, and B. Ackland, "Noise performance of a color cmos photogate image sensor," in *Electron Devices Meeting, 1997. Technical Digest., International*, (Washington, DC, USA), pp. 205–208, December 1997.

[14] O. Pirinen, A. Foi, and A. Gotchev, "Color high dynamic range imaging: The luminance-chrominance approach," *International Journal of Imaging Systems and Technology, Special Issue on Applied Color Image Processing*, vol. 17, pp. 152–162, October 2007.

[15] S. Kavusi and A. E. Gamal, "Quantitative study of high-dynamic-range image sensor architectures," in *Proc. Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, vol. V, (San Jose, CA, USA), pp. 264–275, January 2004.

[16] J. Burghartz, H.-G. Graf, C. Harendt, W. Klingler, H. Richter, and M. Strobel, "Hdr cmos imagers and their applications," in *Proc. of 8th International Conference on Solid-State and Integrated Circuit Technology. (ICSICT '06)*, vol. VII, (Shanghai, China), pp. 528–531, October 2006.

[17] B. C. Madden, "Extended intensity range imaging," tech. rep., GRASP Laboratory, University of Pennsylvania, 1993.

[18] 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*, (New York, NY, USA), January 2006.

[19] 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*, pp. 286–290, November 2001.

[20] A. V. Oppenheim, R. Schafer, and T. Stockham, "Nonlinear filtering of multiplied and convolved signals," in *Proceedings of the IEEE*, vol. 56, pp. 1264–1291, August 1968.

[21] 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*, vol. 106, pp. 795–834, October 1999.

[22] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Computer Graphics Forum*, vol. 22, pp. 419–426, July 2003.

[23] "Picturenaut 3.0." available at http://www.hdrlabs.com/picturenaut/index.html.

[24] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. Image Process.*, vol. 17, pp. 1737–1754, October 2008.

[25] A. Foi, "Clipped noisy images: Heteroskedastic modeling and practical denoising," *Signal Processing*, vol. 89, pp. 2609–2629, December 2009. http://dx.doi.org/10.1016/j.sigpro.2009.04.035 .

Index