

A Perceptual Framework for Contrast Processing of High Dynamic Range Images

Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel
MPI Informatik

Abstract

In this work we propose a framework for image processing in a visual response space, in which contrast values directly correlate with their visibility in an image. Our framework involves a transformation of an image from luminance space to a pyramid of low-pass contrast images and then to the visual response space. After modifying response values, the transformation can be reversed to produce the resulting image. To predict the visibility of suprathreshold contrast, we derive a transducer function for the full range of contrast levels that can be found in High Dynamic Range images. We show that a complex contrast compression operation, which preserves textures of small contrast, is reduced to a linear scaling in the proposed visual response space.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms; I.4.2 [Image Processing and Computer Vision]: Enhancement—Greyscale manipulation, sharpening and deblurring

Keywords: visual perception, high dynamic range, contrast processing, tone mapping

1 Introduction

Operations on image contrast or image gradient have recently attracted much attention in the fields of lightness determination [Horn 1974], tone mapping [Fattal et al. 2002], image editing [Perez et al. 2003; Agarwala et al. 2004], image matting [Sun et al. 2004], and color-to-gray mapping [Gooch et al. 2005]. However, all these works focus mainly on image processing aspects without considering perceptual issues. In this work we incorporate perceptual issues by deriving a framework for processing images in perceptually linearized visual response space.

The overview of our framework is shown in Figure 1. Pixel luminance values of an image are first transformed to physical contrast values, which are then transduced to response values of the Human Visual System (HVS). The resulting image is then modified by altering the response values, which are closely related to a subjective impression of contrast. The modified response values can later be converted back to luminance values using an inverse transformation. As an application of our framework we demonstrate two tone mapping methods which can effectively compress dynamic range without losing low-contrast information. We show that a complex contrast compression operation, which preserves textures of small contrast, is reduced to a linear scaling in our visual response space.

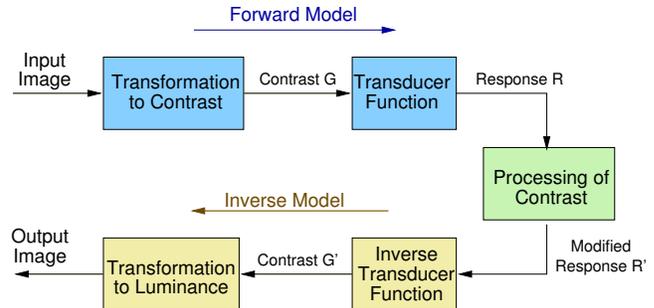


Figure 1: Data flow in the proposed framework of the perceptual contrast processing.

Several models that compute the representation of images as the response of the HVS have been proposed (e.g. [Pattanaik et al. 1998; Fairchild and Johnson 2003]). These models either operate in the luminance domain or decompose an image into band-pass limited contrast channels, similar to those proposed by Peli [1990]. The drawback of the latter approach is that band-pass limited contrast tends to result in severe halo artifacts when each band is modified separately. On the other hand, the methods operating on gradients or contrast were shown to be free from such artifacts. Therefore, we base our framework on a low-pass pyramid of contrast values, which ensures that all spatial frequencies are modified simultaneously. Another major issue that we address in our framework is the perception of High Dynamic Range (HDR) images. To ensure that this aspect is well modelled, we base our framework on data that accounts for high contrast.

In Section 2 we review less well known psychophysical data that was measured for high-contrast stimuli. We also describe a model for suprathreshold contrast discrimination. In Section 3 we introduce the components of our framework, in particular a multi-scale representation of low-pass physical contrast and a transducer function designed for HDR data. As an application of our framework, we propose two tone mapping methods in Sections 4 and 5. Details on how the framework can be implemented efficiently are given in Section 6. We discuss strengths and weaknesses of the proposed framework in Section 7. Finally, we conclude and discuss future work in Section 8.

2 Background

In the following two sections we review some fundamentals of the perception of contrast and summarize the results of a study on the HVS performance in contrast discrimination for HDR images. We use this contrast discrimination characteristic to derive our contrast processing framework.

$\Delta W(W)$ – function of threshold contrast discrimination
W – contrast expressed as Weber fraction (see Table 2)
G – contrast expressed as logarithmic ratio (see Table 2)
$G_{i,j}^k$ – contrast between pixel i and j at the k 'th level of a Gaussian pyramid
L_i^k – luminance of the pixel i at k 'th level of a Gaussian pyramid
x_i^k – \log_{10} of luminance L_i^k
$T(W)$ – transducer function
R – response of the HVS scaled in JND units
\hat{R} – modified response R

Table 1: Used symbols and notation

Simple Contrast $C_s = \frac{L_{max}}{L_{min}}$	
Weber Fraction $W = \frac{\Delta L}{L_{min}}$	
Logarithmic Ratio $G = \log_{10}(\frac{L_{max}}{L_{min}})$	
Michelson Contrast $M = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}$	
Signal to Noise Ratio $SNR = 20 \cdot \log_{10}(\frac{L_{max}}{L_{min}})$	

Table 2: Definitions of contrast and the stimuli they measure.

2.1 Contrast

The human eye shows outstanding performance when comparing two light patches, yet it almost fails when assessing the absolute level of light. Such an effect can be achieved in a ganzfeld, an experimental setup where the entire visual field is uniform. In fact, it is possible to show that the visual system cannot discern mean level variations unless they fluctuate in time or with spatial signals via eye movements, thus having a higher temporal frequency component. Low sensitivity to absolute luminance can be easily explained by the adaptation of the HVS to the real world conditions. Because the HVS is mostly sensitive to relative luminance ratios (contrast) rather than absolute luminance, the effect of huge light changes over the day is reduced and therefore we perceive the world in a similar way regardless of the light conditions. This and other sources of evidence strongly suggest that the perception of contrast (difference between two light stimuli) is the fundamental ability of the HVS.

Many years of research on contrast have resulted in several definitions of contrast, some of them listed in Table 2. The variety of contrast definitions comes from the different stimuli they measure. For example, the Michelson contrast [Michelson 1927] is commonly used to describe a sinusoidal stimulus, while the Weber fraction is often used to measure a step increment or decrement stimulus. In the next section we show that certain contrast definitions are more suitable for describing the performance of the HVS than others.

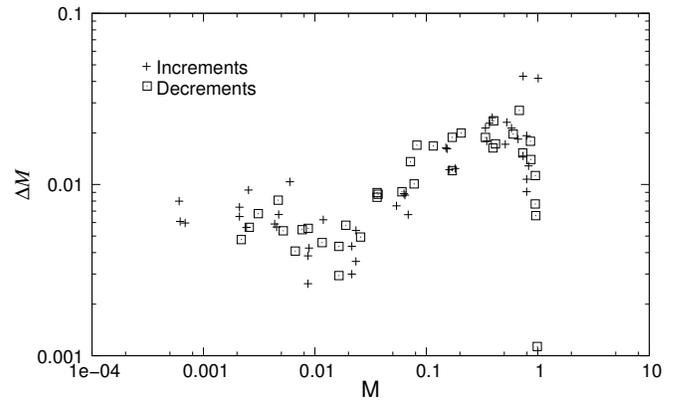


Figure 2: Contrast discrimination thresholds plotted using Michelson contrast M . Michelson contrast does not give a good prediction of the discrimination performance, especially for high contrast.

2.2 Contrast Discrimination

Contrast detection and *contrast discrimination* are two of the most thoroughly studied perceptual characteristics of the eye [Barten 1999]. The contrast discrimination threshold is the smallest visible difference between two nearly identical signals, for example two sinusoidal patterns that differ only in their amplitudes. In the case of contrast detection, only the presence of a signal has to be detected, i.e. the smallest contrast for which a sinusoidal signal becomes visible on a uniform field has to be found. Detection can be considered as a special case of discrimination when the masking signal has zero amplitude. Contrast discrimination is associated with the suprathreshold characteristics of the HVS and in particular with *visual masking*. Contrast detection, on the other hand, describes the performance of the HVS for subthreshold and threshold stimulus, which can be modelled by the *Contrast Sensitivity Function* (CSF), the *threshold versus intensity* function (t.v.i), or Weber's law for luminance thresholds.

Since suprathreshold contrast plays a dominant role in the perception of HDR images, we will consider contrast discrimination data (suprathreshold) in detail and simplify the character of contrast detection (threshold). Although discrimination thresholds of the HVS have been thoroughly studied in psychophysics for years, most of the measurements consider only small contrast levels up to 50% of the Michelson contrast. Such limited contrast makes the usefulness of the data especially questionable in the case of HDR images, for which the contrast can easily exceed 50%. The problem of insufficient scale of contrast in psychophysical experiments was addressed by Whittle in [1986]. By measuring detection thresholds for the full range of visible contrast, Whittle showed that the discrimination data plotted with Michelson contrast does not follow increasing slope, as reported in other studies (refer to Figure 2). He also argued that Michelson contrast does not describe the data well. Figure 2 shows that the data is very scattered and the character of the threshold contrast is not clear, especially for large contrast values. However, when the same data is plotted using the contrast measure $W = \Delta L/L_{min}$, the discrimination thresholds for all but the smallest contrast values follow the same line on log-log plot, which resembles Weber's law, but for suprathreshold contrast: $\Delta W/W = c$ (see Figure 3). To model suprathreshold contrast thresholds, we fit Whittle's original data [1986, Figure 2] to a continuous function:

$$\Delta W(W) = 0.0928 \cdot W^{1.08} + 0.0046 \cdot W^{-0.183} \quad (1)$$

The shape of the fitted function is shown as a solid line in Fig-

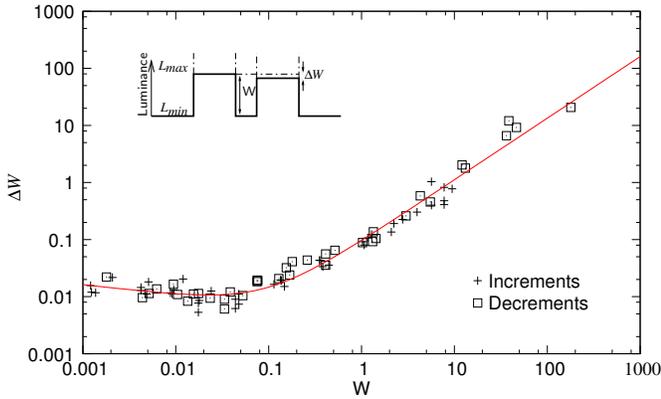


Figure 3: Contrast discrimination thresholds plotted using contrast W . Inset: the stimulus used to measure increments.

Figure 3. In Section 3.2 we use the above function rather than Whittle’s original model $\Delta W/W = c$ to properly predict discrimination thresholds for lower contrast values. But before we utilize the above discrimination function, we have to consider whether it can be generalized for different stimuli and spatial frequencies. In a later study Kingdom and Whittle [1996] showed that the character of the suprathreshold discrimination is similar for both a square-wave and sine-wave patterns of different spatial frequencies. This is consistent with other studies that show little variations of suprathreshold contrast across spatial frequencies [Georgeson and Sullivan 1975; Barten 1999].

3 Framework for Perceptual Contrast Processing

In the next two sections we introduce a framework for image processing in a visual response space. Section 3.1 proposes a method for transforming complex images from luminance to physical contrast domain (blocks *Transform to Contrast* and *Transform to Luminance* in Figure 1). Section 3.2 explains how physical contrast can be converted into a response of the HVS, which is a perceptually linearized measure of contrast (blocks *Transducer Function* and *Inverse Transducer Function* in Figure 1).

3.1 Contrast in Complex Images

Before we introduce contrast in complex images, let us consider the performance of the eye during discrimination of spatially distant patches. We can easily observe that contrast can be assessed only locally for a particular spatial frequency. We can, for example, easily see the difference between fine details if they are close to each other, but we have difficulty distinguishing the brighter detail from the darker if they are distant in our field of view. On the other hand, we can easily compare distant light patches if they are large enough. This observation can be explained by the structure of the retina, in which the foveal region responsible for the vision of fine details spans only about 1.7 visual degrees, while the parafoveal vision can span over 160 visual degrees, but has almost no ability to process high frequency information [Wandell 1995]. When seeing fine details in an image, we fixate on a particular part of that image and employ the foveal vision. But at the same time the areas further apart from the fixation point can only be seen by the parafoveal

vision, which can not discern high frequency patterns. The contrast discrimination for spatial patterns with increasing separation follows Weber’s law when the eye is fixed to one of the patterns and this is the result of the increasing eccentricity of the other pattern [Wilson 1991]. Therefore, due to the structure of the retina, the distance at which we can correctly assess contrast is small for high frequency signals, but grows for low frequency signals.

While several contrast definitions have been proposed in the literature (refer to Table 2), they are usually applicable only to a simple stimulus and do not specify how to measure contrast in complex scenes. This issue was addressed by Peli in [1990], in which he noticed that the processing of images is neither periodic nor local and therefore the representation of contrast in images should be quasi-local as well. Drawing analogy from the center-surround structures in the retina, he proposed to measure contrast in complex images as a difference between selected levels of a Gaussian pyramid. However, the resulting difference of Gaussians leads to a band-pass limited measure of contrast, which tends to introduce halo artifacts at sharp edges when it is modified. To avoid this problem, we introduce a low-pass measure of contrast. We use a logarithmic ratio G as the measure of contrast between two pixels, which is convenient in computations since it can be replaced with the difference of logarithms. Therefore, our low-pass contrast is defined as a difference between a pixel and one of its neighbors at a particular level of a Gaussian pyramid, which can be written as:

$$G_{i,j}^k = \log_{10}(L_i^k/L_j^k) = x_i^k - x_j^k \quad (2)$$

where L_i^k and L_j^k are luminance values for neighboring pixels i and j at a particular level k of the Gaussian pyramid (for brevity we denote $x_i^k = \log_{10}L_i^k$). For a single pixel i there are two or more contrast measures $G_{i,j}^k$, depending on how many neighbouring pixels j are considered. Note that both L and x cover a larger and larger area of an image when moving to the coarser levels of the pyramid. This way our contrast definition takes into account the quasi-local perception of contrast, in which fine details are seen only locally, while variations in low frequencies can be assessed for the entire image.

Equation 2 can be used to transform luminance to contrast. Now we would like to perform the inverse operation that restores an image from the modified contrast values \hat{G} . The problem is that there is probably no image that would match such contrast values. Therefore, we look instead for an image whose contrast values are close but not necessarily exactly equal to \hat{G} . This can be achieved by the minimization of the distance between a set of contrast values \hat{G} that specifies the desired contrast, and G , which is the contrast of the actual image. This can be formally written as the minimization of the objective function:

$$f(x_1, x_2, \dots, x_N) = \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^M (G_{i,j}^k - \hat{G}_{i,j}^k)^2 \quad (3)$$

with regard to the pixel values x_i^1 on the finest level of the pyramid. An efficient solution of the above equation is given in Section 6.

Note that restoring images from contrast does not produce halo artifacts as long as the solution of the minimization problem comes close enough to the objective goal \hat{G} and the modified contrast \hat{G} has the same sign as the original. The problem may appear only for pixel pairs of very low contrast, for which the error of the minimization procedure may be large enough to reverse the ratio between them. This can be alleviated if the square difference terms in Equation 3 are weighted by a factor that increases for lower magnitude of the contrast $\hat{G}_{i,j}^k$.

3.2 Transducer Function

A transducer function predicts the hypothetical response of the HVS for a given physical contrast. As can be seen in Figure 1, our framework assumes that the processing is done on the response rather than on the physical contrast. This is because the response closely corresponds to the subjective impression of contrast and therefore any processing operations can assume the same visual importance of the response regardless of its actual value. In this section we would like to derive a transducer function that would predict the response of the HVS for the full range of contrast, which is essential for HDR images.

Following [Wilson 1980] we derive the transducer function $T(W)$ based on the assumption that the value of the response R should change by one unit for each Just Noticeable Difference (JND) both for threshold and suprathreshold stimuli. However, to simplify the case of threshold stimuli, we assume that $T(0) = 0$ and $T(W_{threshold}) = 1$, or $T^{-1}(0) = 0$ and $T^{-1}(1) = W_{threshold}$ for the inverse transducer function $T^{-1}(R)$, where the detection threshold is approximated with the value commonly used for digital images $W_{threshold} = 0.01$ [Wyszecki and Stiles 2000, Section 7.10.1]. For a suprathreshold stimulus we approximate the response function T by its first derivative:

$$\Delta T \approx \frac{dT(W)}{dW} \Delta W(W) = 1 \quad (4)$$

where $\Delta W(W)$ is the discrimination threshold given by Equation 1. The above equation states that a unit increase of response R (right hand side of the equation) should correspond to the increase of W equal to the discrimination threshold ΔW for the contrast W (left side of the equation). The construction of function $R = T(W)$ is illustrated in the inset of Figure 4. Although the above equation can be solved by integrating its differential part, it is more convenient to solve numerically the equivalent differential equation:

$$\frac{dT^{-1}(R)}{dR} = \Delta W(T^{-1}(R)) \quad (5)$$

for the inverse response function $T^{-1}(R) = W$ and for the boundary condition $T^{-1}(1) = W_{threshold}$. W is a non-negative Weber fraction (refer to Table 2) and R is the response of the HVS. Since the function T^{-1} is strictly monotonic, finding the function T is straightforward. We numerically solve Equation 5 to find the transducer function $T(W) = R$ shown in Figure 4.

Before we can use the transducer function in our framework (*Transducer Function* and *Inverse Transducer Function* in Figure 1) we have to unify contrast measures. Note that in the previous section we used logarithmic ratio G for computational efficiency. Contrast can be easily converted between G and W units using the formulas:

$$W = \begin{cases} 10^G - 1 & \text{if } G \geq 0 \\ -10^{-G} - 1 & \text{otherwise} \end{cases} \quad (6)$$

and

$$G = \begin{cases} \log_{10}(W + 1) & \text{if } W \geq 0 \\ -\log_{10}(-W + 1) & \text{otherwise} \end{cases} \quad (7)$$

The above formulas make sure that the sign of contrast is properly handled. Since the sign contains important information, it should also be preserved when converting contrast W to response R and then later response R to contrast W .

The transducer function derived in this section has a similar derivation and purpose as the Standard Grayscale Function from the DICOM standard [DICOM 2001] or the capacity function

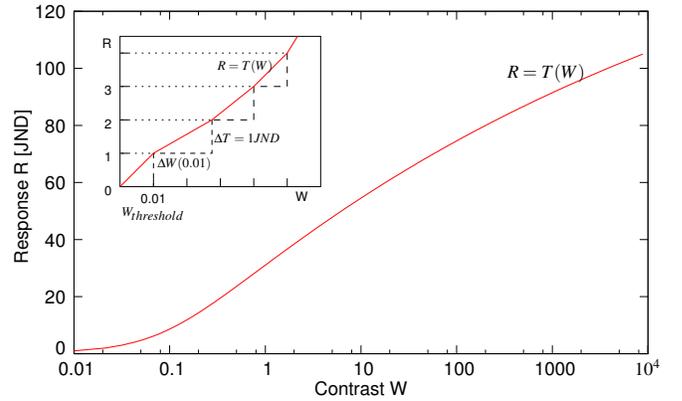


Figure 4: Transducer function derived from the contrast discrimination data [Whittle 1986]. The transducer function can predict the response of the HVS for the full range of contrast. The inset depicts how the transducer function is constructed from the contrast discrimination thresholds $\Delta W(W)$.

in [Ashikhmin 2002]. The major difference is that the transducer function operates in the contrast domain rather than in the luminance domain. It is also different from other transducer functions proposed in the literature (e.g. [Wilson 1980; Watson and Solomon 1997]) since it is based on the discrimination data for high contrast and operates on contrast measure W . This makes the proposed formulation of the transducer function especially suitable to HDR data. The derived function also simplifies the case of the threshold stimuli and assumes a single detection threshold $W_{threshold}$. Such a simplification is acceptable, since our framework focuses on suprathreshold rather than threshold stimuli.

4 Application: Contrast Mapping

In previous sections we introduce our framework for converting images to perceptually linearized contrast response and then restoring images from the modified response. In this section we show that one potential application of this framework is to compress the dynamic range of HDR images to fit into the contrast reproduction capabilities of display devices. We call this method contrast mapping instead of tone mapping because it operates on contrast response rather than luminance.

Tone mapping algorithms try to overcome either the problem of the insufficient dynamic range of a display device (e.g. [Tumblin and Turk 1999; Reinhard et al. 2002; Durand and Dorsey 2002; Fattal et al. 2002]) or the proper reproduction of real-world luminance on a display (e.g. [Pattanaik et al. 1998; Ashikhmin 2002]). Our method does not address the second issue of trying to make images look realistic and natural. Instead we try to fit to the dynamic range of the display so that no information is lost due to saturation and at the same time, small contrast details, such as textures, are preserved. Within our framework such non-trivial contrast compression operation is reduced to a linear scaling in the visual response space. Since the response $R_{i,j}^k$ is perceptually linearized, contrast reduction can be achieved by multiplying the response values by a constant l :

$$\hat{R}_{i,j}^k = R_{i,j}^k \cdot l \quad (8)$$

where l is between 0 and 1. This corresponds to lowering the maximum contrast that can be achieved by the destination display. Since



Figure 5: The results of the contrast mapping algorithm. The images from left to right were processed with the compression factor $l = 0.1, 0.4, 0.7, 1.0$. After the processing images were rescaled in the \log_{10} domain to use the entire available dynamic range.

the contrast response R is perceptually linearized, scaling effectively enhances low physical contrast W , for which we are the most sensitive, and compresses large contrast magnitudes, for which the sensitivity is much lower. The result of such contrast compression for the Memorial Church¹ image is shown in Figure 5.

In many aspects the contrast compression scheme resembles the gradient domain method proposed by Fattal et al. [2002]. However, unlike the gradient method, which proposes somewhat ad-hoc choice of the compression function, our method is entirely based on the perceptual characteristic of the eye. Additionally, our method can avoid low frequency artifacts as discussed in Section 7.

We tested our contrast mapping method on an extensive set of HDR images. The only visible problem was the magnification of the camera noise on several HDR photographs. Those pictures were most likely taken in low light conditions and therefore their noise level was higher than in the case of most HDR photographs. Our tone mapping method is likely to magnify camera noise if its amplitude exceeds the threshold contrast $W_{threshold}$ of the HVS. Therefore, to obtain good results, the noise should be removed from images prior to the contrast mapping.

In Figure 8 we compare the results of our method with other tone mapping algorithms. Our contrast mapping method produces very sharp images without introducing halo artifacts. Sharpening is especially pronounced when the generated images are compared to the result of linear scaling in the logarithmic domain (see Figure 9).

5 Application: Contrast Equalization

Histogram equalization is another common method to cope with extended dynamic range. Even if high contrast occupies only a small portion of an image, it is usually responsible for large dynamic range. The motivation for equalizing the histogram of contrast is to allocate dynamic range for each contrast level relative to the space it occupies in an image. To equalize a histogram of contrast responses, we first find the Cumulative Probability Distribution Function (CPDF) for all contrast response values in the image

$R_{i,j}^k$. Then, we calculate the modified response values:

$$\hat{R}_{i,j}^k = \text{sign}(R_{i,j}^k) \cdot \text{CPDF}(|R_i^k|) \quad (9)$$

where $\text{sign}()$ equals -1 or 1 depending on the sign of the argument and $|R_i^k|$ is a root-mean-square of the contrast response between a pixel and all its neighbors:

$$|R_i^k| = \sqrt{\sum_{j=1}^M R_{i,j}^k{}^2} \quad (10)$$

The histogram equalization scheme produces very sharp and visually appealing images, which may however be less natural in appearance than the results of our previous method (see some examples in Figures 8 and 9). Such a tone mapping method can be especially useful in those applications, where the visibility of small details is paramount. For example, it could be used to reveal barely visible details in forensic photographs or to improve the visibility of small objects in satellite images.

6 Implementation Details

In this section we give an efficient solution to the optimization problem stated in Section 3.1 and explain how we handle color and the dynamic range of images before display.

The major computational burden of our method lies in minimizing the objective function given in Equation 3. Since the objective function reaches its minimum when all its derivatives $\frac{\partial f}{\partial x_i}$ equal 0 for $i = 1, \dots, N$, the problem can be reformulated as solving the set of linear equations $A \cdot X = B$, where X is the resulting image. To further limit computational complexity, we consider only the closest neighbors of each pixel: the contrast given by Equation 2 is computed between a pixel and its four neighbors within the same level of a Gaussian pyramid. To solve the set of linear equations effectively, we use the *biconjugate gradient method* [Press et al. 2002]. The method requires computing the product $\Psi = A \cdot X$ for the iteratively refined image X . The product can be efficiently calculated by the following recursive formula:

$$\Psi^k(X^k) = X^k \times \mathcal{L} + \text{upsample}[\Psi^{k+1}(\text{downsample}[X^k])] \quad (11)$$

¹Memorial Church image courtesy of Paul Debevec.

where X^k is a solution at the k -th level of the pyramid, the operator \times denotes convolution, \mathcal{L} is the kernel

$$\mathcal{L} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (12)$$

and $upsample[]$ and $downsample[]$ are image upsampling and downsampling operators. The recursion stops when one of the image dimensions is less than 3 pixels after several successive downsamplings. The right-hand term B can be computed using another recursive formula:

$$B^k(\hat{G}^k) = \hat{G}_{:,x}^k \times Dx + \hat{G}_{:,y}^k \times Dy + \\ + upsample[B^{k+1}(downsample[\hat{G}^k])] \quad (13)$$

where \hat{G}^k is the modified contrast at the k -th level of the pyramid, $\hat{G}_{:,x}^k$ and $\hat{G}_{:,y}^k$ are the subsets of contrast values \hat{G}^k for horizontal and vertical neighbors, and Dx , Dy are the convolution kernels:

$$Dx = \begin{bmatrix} 1 & -1 \end{bmatrix} \quad Dy = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (14)$$

If only the first level of the pyramid is considered, the problem is reduced to the solution of Poisson's equation as in [Fattal et al. 2002]. Since the recursive operator from Equation 11 affects an image at all spatial frequencies, there is no need to employ multi-grid techniques to achieve fast convergence to the solution. To account for the boundary conditions, we can pad each edge of an image with a line or column that is a replica of the image edge.

After the contrast processing in our framework, the pixel luminance values can fall within an arbitrary range of luminance. Therefore, the final step of our framework involves scaling the color channels to the range from 0 to 1. To achieve proper saturation of colors, we scale each color channels using the formula:

$$C_{out} = \frac{1}{l_{max} - l_{min}} \cdot (X - l_{min} + s(C_{in} - L_{in})) \quad (15)$$

where C_{in} and C_{out} are the input and output pixel values for the red, green or blue color channel, L_{in} is the input luminance, and X is the result of the optimization (all values are in the logarithmic domain). The parameter s is responsible for the saturation of colors and is usually set between 0.4 and 0.6. If P_k is k -th percentile of X and $d = \max(P_{50} - P_{0.1}, P_{99.9} - P_{50})$, then $l_{min} = P_{50} - d$ and $l_{max} = P_{50} + d$. This way, the average gray level is mapped to the gray level of the display ($r = g = b = 0.5$) and overall contrast is not lost due to a few very dark or bright pixels. Equation 15 is similar to formulas proposed by Tumblin and Turk [1999] but it is given in the logarithmic domain and includes a linear scaling. The resulting color values, C_{out} , can be linearly mapped directly to the pixel values of a gamma corrected (perceptually linearized) display.

7 Discussion

Although this paper only shows a tone mapping as an application of our framework, we expect it can be also used for other image processing algorithms as well. For example, Gooch et al. [2005] proposed a method for converting color images to gray-scale in a such way that important color differences are not lost. For this purpose, they reconstruct a gray-scale image from a set of contrast values, which correspond to color differences between all pixels in the image. This is equivalent to solving Equation 3 for the first level of the pyramid, where the contrast values between all pixels in the

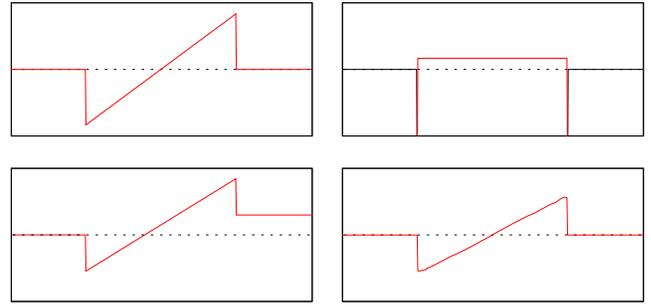


Figure 6: When an original signal (upper left) is restored from attenuated gradients (upper right) by solving Poisson's equation (or integration in 1-D), the flat parts of the restored signal are shifted relative to each other (lower left). However, if the minimization constraints are set for multiple levels of the pyramid as in our proposed method, the flat parts can be accurately restored although the sharp peaks are slightly blurred (lower right).

image are considered instead of a limited number neighbors of each pixel.

The proposed framework is most suitable for those problems where the best solution is a compromise between conflicting goals. For example, in the case of contrast mapping (Section 4), we try to compress an overall contrast by suppressing low frequencies (low frequency contrast has large values and thus is heavily compressed), while preserving details. However, when enhancing details we also lessen compression of overall contrast since details can span a broad range of spatial frequencies (the lower levels of low-pass Gaussian pyramid) including low-frequencies, which are primarily responsible for an overall contrast. The strength of our method comes from the fact that the objective function given in Equation 3 leads to a compromise between the conflicting goals of compressing low-frequency large contrast and preserving small contrast of the details.

The minimization problem introduced in Equation 3 seems similar to solving Poisson's equation in order to reconstruct an image from the attenuated gradients, as proposed by Fattal et al. [2002]. The difference is that our objective function puts additional optimization constraints on the contrast at coarser levels of the pyramid (summation over l), which improves a restoration of low frequency information. When an objective function is limited only to the finest level of the Gaussian pyramid (as it is done in Poisson's equation), the low frequency content may be heavily distorted in the resulting image². This is illustrated on the examples of a 1-D signal in Figure 6 and a tone-mapped image in Figure 7³. In general, the algorithm by Fattal et al. may lead to the reduction (or even reversal) of global low-frequency contrast measured between distant image fragments. High quality and visually pleasing images can be obtained if the compression of contrast is reduced, though with a loss in local contrast enhancement. Our framework can be considered as a generalization of the Fattal et al. approach, whose distinctive feature is that control over the global and local contrasts relies on the minimization problem rather than on user-tuned parameters.

The major limitation of the proposed framework is the computational complexity of the contrast-to-luminance transformation. The

²Loss of low-frequency contrast is also visible in Figure 3 in the paper by Fattal et al. [2002], where low intensity levels of the left and middle peaks in the original image (a) are strongly magnified in the output image (f), so that they eventually become higher than the originally brightest image part on the right side.

³Aeroporto image courtesy of Greg Ward.



Figure 7: The algorithm by Fattal et al. [2002] (top) renders windows panes of different brightness due to the local nature of the optimization procedure. The contrast compression on the multi-scale contrast pyramid used in our method can maintain proper global contrast proportions (bottom).

solution of the minimization problem for 1–5 Mpixel images can take from several seconds up to half a minute to compute on a modern PC. This limits the application of the algorithm to off-line processing. However, our solution is not much less efficient than multi-grid methods (for example [Fattal et al. 2002]), since it solves the linear problem simultaneously for low and high spatial frequencies, and it is computationally much less expensive than some other gradient methods (for example [Gooch et al. 2005]).

8 Conclusions and Future Work

In this paper we have presented a framework for image processing operations that work in the visual response space. Our framework is in many aspects similar to the gradient methods based on solving Poisson’s equation, which prove to be very useful for image and video processing. Our solution can be regarded as a generalization of these methods which considers contrast on multiple spatial frequencies. We express a gradient-like representation of images using physical and perceptual terms, such as contrast and visual response. This gives perceptual basis for the gradient methods and offers several extensions from which these methods can benefit. For instance, unlike the solution of Poisson’s equation, our

pyramidal contrast representation ensures proper reconstruction of low frequencies and does not reverse global brightness levels. We also introduce a transducer function that can give the response of the HVS for the full range of contrast amplitudes, which is especially desired in case of HDR images. Some applications can also make use of the contrast discrimination thresholds, which describe suprathreshold performance of the eye from low to high contrast. As a proof of concept, we implemented tone mapping inside our framework as a simple linear scaling. The tone mapping was shown to produce sharper images than the other contrast reduction methods. We believe that our framework can also find many applications in image and video processing.

In the future, we would like to improve the performance of reconstructing the image from the contrast representation, which would make the framework suitable for real-time applications. We would also like to include color information using a representation similar to luminance contrast. The framework could be extended to handle animation and temporal contrast. Furthermore, the accuracy of our model can be improved for the threshold contrast if the Contrast Sensitivity Function were taken into account in the transducer function. A simple extension is required to adapt our framework to the task of predicting visible differences in HDR images: since the response in our framework is in fact scaled in JND units, the difference between response values of two images gives the map of visible differences. One possible application of such HDR visible difference predictor could be the control of global illumination computation by estimating visual masking [Ramasubramanian et al. 1999; Dumont et al. 2003]. Finally, we would like to experiment with performing common image processing operations in the visual response space.

9 Acknowledgments

We would like to thank Dani Lichinski for helpful discussion on his gradient domain method. We are also grateful for his patience in generating for us images using his tone mapping method.

References

- AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM Trans. Graph.* 23, 3, 294–302.
- ASHIKHMIN, M. 2002. A tone mapping algorithm for high contrast images. In *Rendering Techniques 2002: 13th Eurographics Workshop on Rendering*, 145–156.
- BARTEN, P. G. 1999. *Contrast sensitivity of the human eye and its effects on image quality*. SPIE – The International Society for Optical Engineering, P.O. Box 10 Bellingham Washington 98227-0010. ISBN 0-8194-3496-5.
- DICOM. 2001. Part 14: Grayscale standard display function. In *Digital Imaging and Communications in Medicine (DICOM)*.
- DUMONT, R., PELLACINI, F., AND FERWERDA, J. A. 2003. Perceptually-driven decision theory for interactive realistic rendering. *ACM Transactions on Graphics* 22, 2 (Apr.), 152–181.
- DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. on Graph.* 21, 3, 257–266.
- FAIRCHILD, M., AND JOHNSON, G. 2003. Image appearance modeling. In *IS&T’s Electronic Imaging Conference, SPIE Vol. 5007*, 149–160.

- FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM Trans. on Graph.* 21, 3, 249–256.
- GEORGESON, M., AND SULLIVAN, G. 1975. Contrast constancy: Deblurring in human vision by spatial frequency channels. *J. Physiol.* 252, 627–656.
- GOOCH, A. A., OLSEN, S. C., TUMBLIN, J., AND GOOCH, B. 2005. Color2gray: Saliency-preserving color removal. *ACM Transactions on Graphics (Proc. of SIGGRAPH 2005)* 24, 3.
- HORN, B. 1974. Determining lightness from an image. *Computer Graphics and Image Processing* 3, 1, 277–299.
- KINGDOM, F. A. A., AND WHITTLE, P. 1996. Contrast discrimination at high contrasts reveals the influence of local light adaptation on contrast processing. *Vision Research* 36, 6, 817–829.
- MICHELSON, A. 1927. *Studies in Optics*. U. Chicago Press.
- PATTANAİK, S. N., FERWERDA, J. A., FAIRCHILD, M. D., AND GREENBERG, D. P. 1998. A multiscale model of adaptation and spatial vision for realistic image display. In *Siggraph 1998, Computer Graphics Proceedings*, 287–298.
- PELI, E. 1990. Contrast in complex images. *Journal of the Optical Society of America A* 7, 10, 2032–2040.
- PEREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM Trans. Graph.* 22, 3, 313–318.
- PRESS, W., TEUKOLSKY, S., VETTERLING, W., AND FLANNERY, B. 2002. *Numerical Recipes in C++*, second ed. Cambridge Univ. Press, ch. 2.7, 87–92.
- RAMASUBRAMANIAN, M., PATTANAİK, S. N., AND GREENBERG, D. P. 1999. A perceptually based physical error metric for realistic image synthesis. In *Siggraph 1999, Computer Graphics Proceedings*, 73–82.
- REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Trans. on Graph.* 21, 3, 267–276.
- SUN, J., JIA, J., TANG, C.-K., AND SHUM, H.-Y. 2004. Poisson matting. *ACM Trans. Graph.* 23, 3, 315–321.
- TUMBLIN, J., AND TURK, G. 1999. LCIS: A boundary hierarchy for detail-preserving contrast reduction. In *Siggraph 1999, Computer Graphics Proceedings*, 83–90.
- WANDELL, B. 1995. *Foundations of Vision*. Sinauer Associates, Inc.
- WATSON, A. B., AND SOLOMON, J. A. 1997. A model of visual contrast gain control and pattern masking. *Journal of the Optical Society A* 14, 2378–2390.
- WHITTLE, P. 1986. Increments and decrements: Luminance discrimination. *Vision Research* 26, 10, 1677–1691.
- WILSON, H. 1980. A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics* 38, 171–178.
- WILSON, H. 1991. Psychophysical models of spatial vision and hyperacuity. In *Vision and Visual Dysfunction: Spatial Vision*, D. Regan, Ed. Pan Macmillan, 64–86.
- WYSZECKI, G., AND STILES, W. 2000. *Color Science*. John Wiley & Sons.



Figure 8: Comparison of the result produced by our contrast mapping (top left) and contrast equalization (top right) to those of Durand and Dorsey [2002] (bottom left) and Fattal et al. [2002] (bottom right). *Tahoma* image courtesy of Greg Ward.

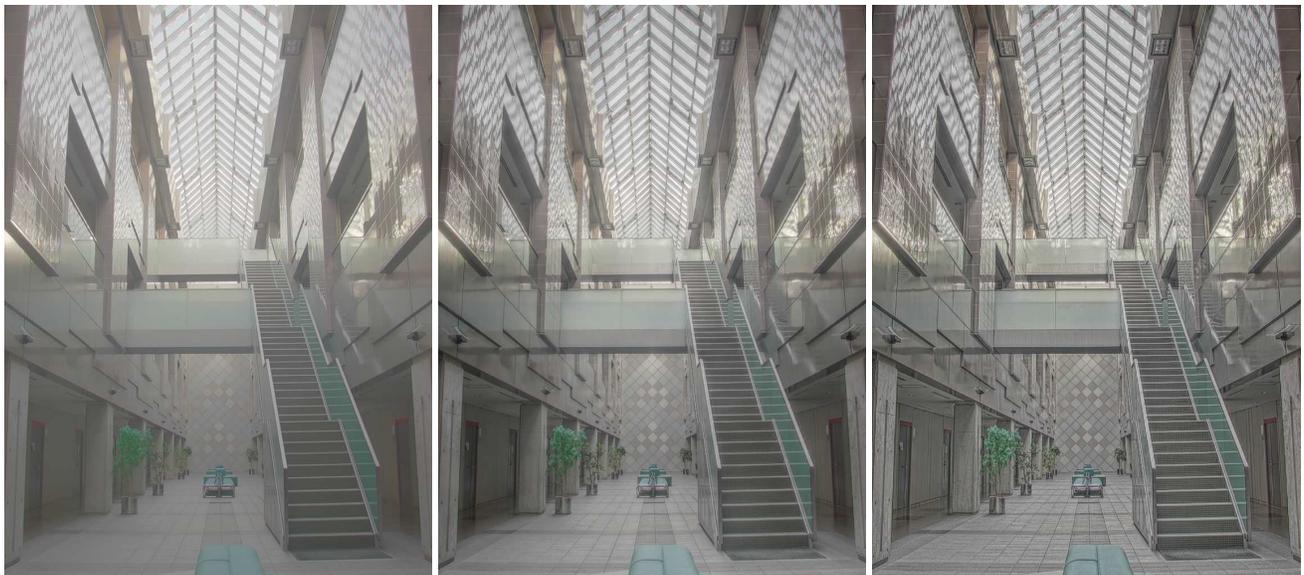


Figure 9: The linear rescaling of luminance in the logarithmic domain (left) compared with two proposed contrast compression methods: contrast mapping (middle) and contrast equalization (right).