

DYNAMIC RANGE IMPROVEMENT THROUGH MULTIPLE EXPOSURES

Mark A. Robertson, Sean Borman, and Robert L. Stevenson

Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

ABSTRACT

This paper presents an approach for improving the effective dynamic range of cameras by using multiple photographs of the same scene taken with different exposure times. Using this method enables the photographer to accurately capture scenes that contain a high dynamic range, i.e., scenes that have both very bright and very dark regions. The approach requires an initial calibration, where the camera response function is determined. Once the response function for a camera is known, high dynamic range images can be computed easily. The high dynamic range output image consists of a weighted average of the multiply-exposed input images, and thus contains information captured by each of the input images. From a computational standpoint, the proposed algorithm is very efficient, and requires little processing time to determine a solution.

1. INTRODUCTION

Intensity values of scenes in the real world can have a very broad *dynamic range*. This is particularly true for scenes that have areas of both low and high illumination, such as when there are transitions between sun-lit areas and areas in shadow. Unfortunately, all image capture devices have a limited dynamic range. For digital cameras, the dynamic range is limited by the charge-coupled devices (CCD) and analog-to-digital conversion (ADC); film characteristics limit the dynamic range of traditional cameras.

When capturing a scene which contains a dynamic range that exceeds that of the camera, there will be a loss of detail in either the lowlight areas, the highlight areas, or both. One may vary the exposure to control which light levels will be captured, and hence which light levels will be lost due to saturation of the camera's dynamic range. Here, we will consider only variation of the exposure *time*, i.e., the duration for which the light sensing element (CCD or film) is exposed to light from the scene. By increasing the exposure time, one may get a better representation of lowlight

areas, at the cost of losing information in areas of high illumination; an example of this is shown in fig. 1(h). Similarly, by using a reduced exposure time, one may sacrifice lowlight detail in exchange for improved detail in areas of high illumination; this is demonstrated in fig. 1(d). However, if the photographer desires an accurate representation of both low- and high-light areas of the scene, and the dynamic range of the scene exceeds that of the camera, then it is futile to adjust the exposure time—detail will definitely be lost, and varying the exposure time merely allows some control over where the loss occurs.

Increasing the dynamic range of images using multiple exposures of a scene has been examined before [1, 2, 3, 4]. However, there are several limitations with these methods. In [2, 3, 4] a linear response is assumed for the capture device, i.e., the observed pixel values are linearly related to the amount of input light, or irradiance. Since cameras in general (and consumer cameras in particular) will not provide a linear response, the methods in [2, 3, 4] are not widely applicable. These same three methods also estimate a high dynamic range pixel value based on only one input pixel, namely, that input pixel which was taken with the highest exposure time yet remained non-saturated. The reason for using only the one input pixel is due to quantization effects: Since observed pixel values arise from quantization of the product of the exposure time and the light present, the quantization error for the light will be lower for higher exposure times. However, rather than using only one input pixel, a better method would take advantage of all data present, and compute an estimate which gives higher weights to pixel values obtained with higher exposure times.

In [1], the response function of the camera is estimated, thus satisfactorily dealing with one limitation of the aforementioned methods. Furthermore, an average of all input pixels is used to determine the high dynamic range light values. However, [1] does not give higher weights to pixel values taken at higher exposures, and thus this method leaves room for improvement.

This paper proposes a new method of increasing the dy-

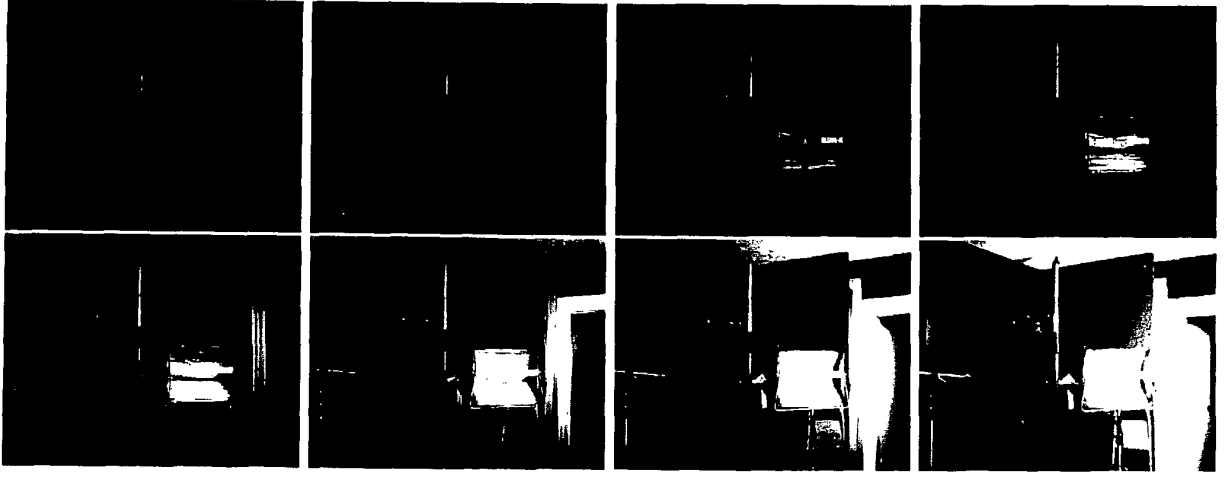


Figure 1:

a	b	c	d
e	f	g	h

 Eight pictures of a static scene taken at different exposure times. The camera used was a Nikon E2N, and the image resolution is 640×500 . The exposure times are (a) 1/1000 sec. (b) 1/500 sec. (c) 1/250 sec. (d) 1/125 sec. (e) 1/60 sec. (f) 1/30 sec. (g) 1/15 sec. (h) 1/8 sec.

dynamic range of images by using multiple exposures. This method finds satisfactory ways of dealing with the problems of the algorithms reported above. In particular, the response function of the image capture device is estimated, thus creating a versatility in our algorithm compared to those algorithms that assume a linear response. Furthermore, the problem of how heavily to weight the various input pixels is addressed.

Section 2 introduces the observation model for this work. The maximum likelihood solution of the high dynamic range image for known camera response is given in Section 3. For unknown camera response, Section 4 discusses how the response function can be estimated. Experimental results are presented in Section 5, followed by concluding remarks in Section 6.

2. OBSERVATION MODEL

Assume there are N pictures taken of a static scene, with known exposure times $t_i, i = 1, \dots, N$. The j^{th} pixel of the i^{th} exposed image will be denoted y_{ij} ; the set $\{y_{ij}\}$ represents the known observations. The goal here is to determine the underlying light values or irradiances, which we denote by x_j , that gave rise to the observations y_{ij} .

Since only the exposure time is being varied, the amount of light contributing to the output value y_{ij} will be $t_i x_j$. To account for image capture noise, we introduce an additive noise term, N_{ij}^c , which also contributes to the observed pixel values. The quantity $t_i x_j + N_{ij}^c$ is then mapped by the

camera's response function $f(\cdot)$ to give the output values

$$y_{ij} = f(t_i x_j + N_{ij}^c). \quad (1)$$

Since y_{ij} are digital numbers, $f(\cdot)$ maps the positive real numbers $\mathbb{R}^+ = [0, \infty)$ to an interval of integers, $\mathcal{O} = \{0, \dots, 255\}$ for 8-bit data. Without loss of generality, this paper assumes the image data is 8 bits. We explicitly write the camera response function as

$$f(z) = \begin{cases} 0 & \text{if } z \in [0, I_0] \\ m & \text{if } z \in (I_{m-1}, I_m], m = 1, \dots, 254 \\ 255 & \text{if } z \in (I_{254}, \infty), \end{cases} \quad (2)$$

and thus $f(\cdot)$ is defined in terms of the 255 numbers I_m , $m = 0, \dots, 254$. For a linear response function, such as in [2, 3, 4], the I_m values would be evenly spaced; in general, however, this will not be true.

3. HIGH DYNAMIC RANGE IMAGE WITH KNOWN RESPONSE FUNCTION

We wish to estimate the irradiances x_j with a dynamic range higher than that of the original observations. The observations are in the range space \mathcal{O} of the response function $f(\cdot)$, and the desired result belongs to the domain of $f(\cdot)$, \mathbb{R}^+ . If the function $f(\cdot)$ is known, as is assumed in [2, 3, 4], we can define a mapping from \mathcal{O} to \mathbb{R}^+ as

$$\begin{aligned} f^{-1}(y_{ij}) &= t_i x_j + N_{ij}^c + N_{ij}^q \\ &= I_{y_{ij}}. \end{aligned} \quad (3)$$

When determining $f^{-1}(m)$, one knows only that it belongs to the interval $(I_{m-1}, I_m]$. The N_{ij}^q noise term above accounts for the uncertainty in assigning $f^{-1}(m) = I_m$, and is a dequantization error. One should keep in mind that $f^{-1}(\cdot)$ is not a true inverse, since $f(\cdot)$ is a many-to-one mapping.

Rewriting (3),

$$I_{y_{ij}} = t_i x_j + N_{ij}. \quad (4)$$

The noise term N_{ij} consists of the noise term introduced in Section 2, as well as the dequantization uncertainty term N_{ij}^q . The N_{ij} terms will be modeled as zero-mean independent Gaussian random variables, with variances σ_{ij}^2 .

Note that accurately characterizing the variances σ_{ij}^2 would be extremely difficult, as it would require detailed knowledge of the specific image capture device being used. One would have to characterize each of the separate noise sources that compose N_{ij} , which would be a complicated task. This process would need to be performed each time a different device is used to capture the image data. Therefore, rather than attempting this, the variances will be chosen heuristically.

It will be convenient in the following to replace the variances with weights, $w_{ij} = 1/\sigma_{ij}^2$. The concept of weights is intuitive, and serves to ease the notational burden. The weights are chosen based on our confidence that the observed data is accurate. Here, an approach similar to that of [1] is taken. The response function of a camera will typically be steepest, or most sensitive, towards the middle of its output range, or 128 for 8-bit data. As the output levels approach the extremes, 0 and 255, the sensitivity of the camera typically decreases. For this reason, a weighting function will be chosen such that values near 128 are weighted more heavily than those near 0 and 255. The function chosen here is a Gaussian-like function,

$$w_{ij} = w_{ij}(y_{ij}) = \exp\left(-4 \cdot \frac{(y_{ij} - 127.5)^2}{(127.5)^2}\right), \quad (5)$$

but scaled and shifted so that $w_{ij}(0) = w_{ij}(255) = 0$, and $w_{ij}(127.5) = 1.0$. This choice of weighting function implies that we have very low confidence in the accuracy of pixel values near 0 and 255, while we have high confidence in the accuracy of pixel values near 128. The weighting function is shown in fig. 2.

From (4), $I_{y_{ij}}$ are independent Gaussian random variables, and the joint probability density function can be written as

$$P(\mathbf{I}_y) \propto \exp\left\{-\sum_{i,j} w_{ij} (I_{y_{ij}} - t_i x_j)^2\right\}. \quad (6)$$

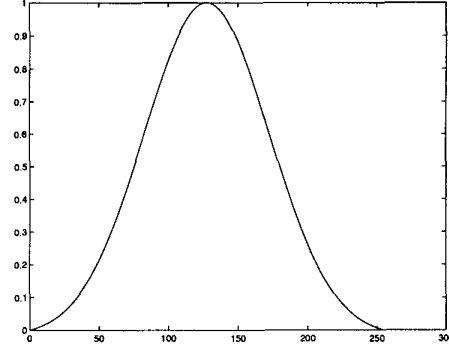


Figure 2: Weighting function, w_{ij} .

Here, a maximum-likelihood (ML) approach will be taken to find the high dynamic range image values. The maximum-likelihood solution finds the x_j values which maximize the probability in (6). Maximizing (6) is equivalent to minimizing the negative of its natural logarithm, which leads to the following objective function to be minimized:

$$O(\mathbf{x}) = \sum_{i,j} w_{ij} (I_{y_{ij}} - t_i x_j)^2. \quad (7)$$

Equation (7) is easily minimized by setting the gradient $\nabla O(\mathbf{x})$ equal to zero. This yields

$$\hat{x}_j = \frac{\sum_i w_{ij} t_i I_{y_{ij}}}{\sum_i w_{ij} t_i^2}, \quad (8)$$

the desired high dynamic range image estimate. Note that data from images taken with longer exposure times are weighted more heavily, as indicated by the t_i term in the numerator of (8). Thus this method takes advantage of the quantization effects utilized in [2, 3, 4]; however, here a noise-reducing averaging is being performed, which utilizes data from *all* input pixels.

Equation (8) requires that the I_m values (i.e., the response function) be known. In general, however, the response function is not known. The following section describes how to determine the response function in an initial calibration step. When using the same camera in the future, the calibration step is unnecessary, and (8) may be directly applied.

4. FOR UNKNOWN RESPONSE FUNCTION

Except in very specialized situations, the camera response function will *not* be known, and must be estimated before attempting to use (8). To uniquely determine the response function, the 255 values $I_m, m = 0, \dots, 254$ must be found.

At first glance, one may consider directly using the objective function in (7) to determine the I_m values needed to define the response function. Note that in order to estimate the I_m values from (7), the x_j values are also unknown and need to be estimated simultaneously (if the x_j 's were already known, there would be little need to estimate the I_m 's!) Thus, the objective function for the case of unknown response function is

$$\tilde{O}(\mathbf{I}, \mathbf{x}) = \sum_{i,j} w_{ij} (I_{y_{ij}} - t_i x_j)^2. \quad (9)$$

An additional constraint on the response function is required when estimating \mathbf{I} and \mathbf{x} together using (9). This restriction on $f(\cdot)$ is in regard to scale. Note that the scale of the high dynamic range image estimates \hat{x}_j is arbitrary—in order to be viewed, these \hat{x}_j need to be mapped to a usable range, $\{0, \dots, 255\}$ for 8-bit display. Since the scale of \hat{x}_j is directly dependent on the scale of I_m , we will constrain the estimates for the I_m values such that $\hat{I}_{128} = 1.0$. This is enforced by dividing each of the \hat{I}_m 's by \hat{I}_{128} .

A form of Gauss-Seidel relaxation will be used to determine the response function. Seidel relaxation minimizes an objective function with respect to a single variable, and then uses these new values when minimizing with respect to subsequent variables. Here, (9) will first be minimized with respect to each I_m . Then the restriction mentioned above is enforced. Finally, (9) will be minimized with respect to each x_j . This will constitute one iteration of the algorithm.

The initial $\hat{\mathbf{I}}$ for the first iteration is chosen as a linear function, with $\hat{I}_{128} = 1.0$. The initial $\hat{\mathbf{x}}$ for the first iteration is chosen according to (8), using the initial linear $\hat{\mathbf{I}}$.

First, to minimize with respect to I_m , the partial derivative of (9) with respect to I_m is taken and set equal to zero. This yields

$$\hat{I}_m = \frac{1}{\text{Card}(E_m)} \sum_{(i,j) \in E_m} t_i x_j, \quad (10)$$

where the index set E_m is defined as

$$E_m = \{(i, j) : y_{i,j} = m\}, \quad (11)$$

the set of indices such that m was observed for the input images. $\text{Card}(E_m)$ is the cardinality of E_m , i.e., the number of times m was observed.

After scaling the response function such that $\hat{I}_{128} = 1.0$, minimization is performed with respect to each x_j . This merely involves using (8). This completes one iteration of the algorithm, and the process is repeated until some convergence criterion is met. The convergence criterion used here is for the rate of decrease in the objective function to fall below some minimum threshold.

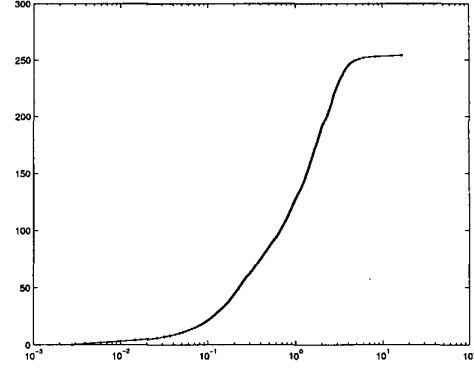


Figure 3: Camera response function, $f(\cdot)$.

5. EXPERIMENTAL RESULTS

Figure 1(f) shows a photograph of a scene with a broad dynamic range taken at what might be considered a “normal” exposure setting. From the picture, one notices that there is little detail visible in either the very bright or the very dark regions. Simple contrast stretching in the dark regions results in noisy-looking image areas; contrast stretching in the very bright regions does little good, due to the saturated pixel values. Thus, a scene such as this is an excellent candidate for the algorithm discussed in this paper.

Figure 3 shows the response function determined by the algorithm discussed in this paper when applied to the eight images in fig. 1. Note that the response function is *not* a simple linear function, and thus use of algorithms such as in [2, 3, 4] would have been inappropriate for this camera.

The \hat{x}_j values which were found while determining the response function are the ultimate variables of interest. For viewing purposes, the full range of \hat{x}_j values was linearly mapped into $[0, \hat{I}_{254}]$; the function $f(\cdot)$ was then applied to these values. This has the effect of simulating the output of the camera if the exposure time was set such that the maximum light value of the scene maps to 254. The result is shown in fig. 4(a).

From fig. 4(a), it is apparent that the solution contains accurate high-light information; however, it is not so apparent that the low-light detail is present. To demonstrate that accurate low-light detail is indeed contained in our solution, we can use the \hat{x}_j values to simulate the camera output for a high exposure time. This is done by simply creating an 8-bit image with the j^{th} pixel value equal to $f(t_d \hat{x}_j)$, where t_d is the desired exposure time. This result is shown in fig. 4(b) for $t_d = 1/3$ sec. Note that this is simulating camera output for an exposure time not available on our camera. From the



Figure 4: a b High dynamic range output images. (a) Full dynamic range mapped to eight bits; (b) using high dynamic range image data to simulate camera output for an exposure time of $1/3$ sec.

figure, once sees that low-light detail is also contained in the high dynamic range image data.

The high dynamic range image determined using the method outlined here has several advantages over a single image such as one of those in fig. 1. The image obtained with our method has decreased noise, due to the averaging of pixel values from each of the input images. Furthermore, it contains information in both low- and high-light areas, since the high dynamic range image consists of data from each of the input images. Traditional image processing algorithms (e.g., contrast stretching, histogram equalization, edge detection, etc.) can be applied to the high dynamic range image with more accurate results, due to the increased amount of information present.

For the example results just given, the high dynamic range radiance map was found while estimating the response function. Note, however, that once the response function for a capture device has been determined, this process need not be repeated when using that device in the future. Instead, (8) can be used to directly determine the desired high dynamic range image values.

The convergence of the calibration step has been observed to require approximately five iterations, which amounts to roughly five seconds for eight input pictures of size 640×500 , running on a 180 MHz PC.

6. CONCLUSION

This paper has introduced a method of increasing the effective dynamic range of digital cameras by using multi-

ple pictures of the same scene taken with varying exposure times. If necessary, the method first estimates the response function of the camera. Once the response function is known, high dynamic range images can be directly computed. These high dynamic range images contain accurate representations of both low- and high-light areas in the image, with decreased noise due to averaging of the input images.

7. REFERENCES

- [1] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH 97 Conference Proceedings*, pages 369–378, Aug. 3–8 1997.
- [2] Brian C. Madden. Extended intensity range imaging. Technical report, GRASP Laboratory, University of Pennsylvania, 1993.
- [3] Kousuke Moriwaki. Adaptive exposure image input system for obtaining high-quality color information. *Systems and Computers in Japan*, 25(8):51–60, July 1994.
- [4] Keiichi Yamada, Tomoaki Nakana, and Shin Yamamoto. Wide dynamic range vision sensor for vehicles. In *IEEE Int. Conf. on Vehicle Navigation and Information Systems*, pages 405–408, Aug. 31–Sep. 2 1994.