Journal of Electronic Imaging 12(2), 219–228 (April 2003).

Estimation-theoretic approach to dynamic range enhancement using multiple exposures

Mark A. Robertson
Sean Borman
Robert L. Stevenson
University of Notre Dame
Department of Electrical Engineering
Notre Dame, Indiana 46556

Abstract. *We present a new 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 high dynamic range by using a device with low dynamic range, which allows the capture of scenes that have both very bright and very dark regions. We approach the problem from a probabilistic standpoint, distinguishing it from the other methods reported in the literature on photographic dynamic range improvement. A new method is proposed for determining the camera's response function, which is an iterative procedure that need be done only once for a particular camera. With the response function known, high dynamic range images can be easily constructed by a weighted average of the input images. The particular form of weighting is controlled by the probabilistic formulation of the problem, and results in higher weight being assigned to pixels taken at longer exposure times. The advantages of this new weighting scheme are explained by comparison with other methods in the literature. Experimental results are presented to demonstrate the utility of the algorithm.* © *2003 SPIE and IS&T.* [DOI: 10.1117/1.1557695]

1 Introduction

Intensity values of real-world scenes can have a very wide dynamic range. This is particularly true for scenes that have areas of both low and high illumination, such as transitions between sunlit areas and areas in shadow, or when a light source is visible in the scene. Unfortunately, all image capture devices have a limited dynamic range. For digital cameras, the dynamic range is limited by properties of the charge-coupled device (CCD) and analog-to-digital conversion (ADC); film characteristics limit the dynamic range of traditional cameras.

When capturing a scene containing a dynamic range that exceeds that of the camera, there will be a loss of detail in either the low-light areas, the high-light 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. This work only considers 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. Variation of the aperture is not considered due to the effects of aperture on depth of field. By increasing the exposure time, one may get a better representation of low-light areas at the cost of losing information in areas of high illumination; an example of this is shown at the top of Fig. 1. Similarly, by using a reduced exposure time, one may sacrifice low-light detail in exchange for improved detail in areas of high illumination, as demonstrated toward the bottom of Fig. 1. 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.

Examination of a scene's histogram offers further insight into the problem. Suppose that a scene has an intensity histogram, as shown in Fig. 2, which has concentrations of intensities around relatively dark and relatively bright levels, with maximum intensity $I_{\max}$. For simplicity, assume that the output of the camera is a linear function of input exposure, and that a uniform quantizer with K levels is used to produce the digital output. A photographer might adjust the exposure settings such that $1/3I_{\max}$ maps to saturation, which emphasizes the dark regions of the scene. Doing this yields quantization intervals of $1/3K^{-1}I_{\max}$. If the photographer wants to capture the bright portions of the scene as well, he or she might reduce the exposure, such that $2/3I_{\max}$ maps to saturation. Doing this captures a larger range of intensity values than the previous exposure setting. However, the quantization intervals are now $2/3K^{-1}I_{\max}$—the dark regions are captured, but information is lost due to coarser quantization. We propose a method for combining data from multiple exposures to form an image with improved dynamic range, which takes advantage of the favorable qualities of each of the individual exposures.

There are a number of situations where one will have multiple exposures of a scene available. As already mentioned, a photographer might wish to accurately capture one particular scene and intentionally vary the exposure, such that the set of images, taken as a whole, captures the

Paper 20066 received Aug. 10, 2000; revised manuscript received Jan. 7, 2002; accepted for publication Sep. 17, 2002.
1017-9909/2003/$15.00 © 2003 SPIE and IS&T.

Fig. 1 "Studio" scene. Eight pictures of a static scene taken at different exposure times, photographed using a Nikon E2N digital camera with aperture f/6.7. The original image resolution is 1280 $\times$1000. The exposure times are, from brightest to darkest, $\frac{1}{8}$, $\frac{1}{15}$, $\frac{1}{30}$, $\frac{1}{60}$, $\frac{1}{125}$, $\frac{1}{250}$, $\frac{1}{500}$, and $\frac{1}{1000}$ s.

entire dynamic range of the scene. Unintentional situations also arise: when using cameras with automatic exposure (or gain) control, a particular exposure (gain) might be automatically selected for one portion of the scene. However, if one pans the camera to a different part of the scene (with the intention of, for example, creating a panoramic picture by stitching multiple pictures together), the automatic exposure or gain can change, especially for scenes with both very bright and very dark areas. The same possibility exists for panning a video camera across a scene. In the general case of camera zooming, panning, tilting, and rotating about its optical axis, it would be necessary to register the images to bring them into proper alignment with each other. This work does not consider this general case, but rather restricts itself to static scenes with no (or very little) camera motion; see Ref. 1 for details of registration assuming projective transformation among the images. The issue of automatic gain will be briefly addressed again in Secs. 2 and 3.

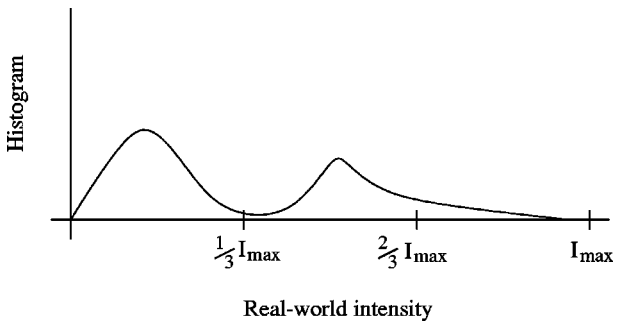

Fig. 2 Example histogram of real-world scene intensities, with arbitrary intensity units.

One of the first researchers to investigate improved dynamic range imaging was Wyckoff,[2] who worked with film rather than digital images. Wyckoff's special film contained multiple layers that responded differently with respect to exposure, but the same spectrally. Thus each layer would have a different speed and could, for example, capture a different portion of the intensity histogram in Fig. 2—one layer might accurately capture the bright regions, one layer might accurately capture the dark regions, and a third layer might capture intensities between the two extremes. Although each layer had identical spectral characteristics, the final image could be viewed in a pseudocolor fashion by printing each layer as one of the three primary colors.

The first report of digitally combining multiple pictures of the same scene to improve dynamic range appears to be Mann.[3] Algorithmic detail that is lacking from Ref. 3 is provided in a later publication,[4] where Mann and Picard explicitly examine the situation where multiple pictures, each of different exposures, are taken of a scene. They provide a method of merging these multiple exposures to form a single image with an effective dynamic range greater than that of the camera. By making use of certainty functions, which give a measure of the confidence in an observation, Mann and Picard weight the observations from the various exposures to provide the final image. The certainty function for a particular camera is computed as the derivative of the camera response function, which results in low confidence for pixel values near extremes, and higher confidence for pixel values between these extremes. More detail of certainty functions is provided in later sections.

Madden[5] also examined the dynamic range problem, specifically for the case of CCD capture devices. Using direct CCD output allowed Madden to assume a linear response function for the camera, i.e., the observed output value is linearly related to the input exposure. Madden takes multiple pictures of the same scene while varying the exposure time, and uses these multiply-exposed images to construct the final high dynamic range image. To determine the value of a high dynamic range pixel, information is used from only that input image taken at the highest exposure in which the pixel of interest was not saturated. The author justifies this by pointing out that pixels observed at higher exposure times have less quantization noise than do pixels taken at lower exposure times.

Yamada, Nakana, and Yamamoto[6,7] studied the dynamic range problem in the context of vision systems for vehicles. The authors use multiple exposures of a scene, and assume a linear response for the CCDs. The authors pick the final pixel output based only on the observation with the longest exposure time that is not saturated. While not explicitly

giving justification for using only data from the highest nonsaturated exposure, the implicit justification is the same as that mentioned before—to reduce quantization error.

Moriwaki[8] examined the dynamic range enhancement of color images. The author uses multiple exposures of a static scene, and also assumes a linear CCD response. The method employed is similar to Yamada, Nakana, and Yamamoto, in that the color values for a pixel are taken only from the observation pixel with the highest exposure time that was not saturated.

Chen and Mu[9] suggest using a cut-and-paste method for increasing dynamic range, where blocks of the final image are taken from blocks of the input images in a manual manner. The authors propose this interactive method to avoid more complicated, and perhaps nonlinear, processing. This technique is obviously very limited, and any computational advantage is clearly lost when one considers the computational resources available today.

Debevec and Malik[10] offer a more advanced method of increasing image dynamic range using multiple exposures. Rather than assuming a linear camera response, they assume an arbitrary response that is determined as part of the algorithm. The final output pixel is given as a weighted average of the input pixels taken at different exposures. The algorithm gives higher weight to input data that are nearer to the mean of the input pixel range (128 for 8-bit data), and less weight to the input data that are near to the extremes of the input pixel range (0 and 255 for 8-bit data).

There are several limitations of the algorithms just described. In Ref. 9, the requirement of human intervention is an obvious drawback. The work of Ref. 2 required special film, and is not suitable for digital imagery. In Refs. 5 through 8, linear camera response functions are all required. While one might argue that this is justified due to the linear nature of CCDs[11] there are still potential problems. First, one is strictly limited to using only linear capture devices, which precludes the possibility of using images scanned from film. Second, while consumer digital cameras do typically use CCDs, there is no guarantee of a linear response—for while the actual CCDs may be linear, the camera manufacturer is likely to introduce nonlinearities prior to output to make the image more visually pleasing.

There is also a fundamental limitation when an algorithm determines the light values using data from only one input source, rather than using all input data. Recall that the main motivation for using only the highest nonsaturated input pixel is to try to minimize quantization error. Taking the approaches in Refs. 5–8 would indeed make perfect sense if quantization were the only source of noise in the image capture process. However, there are other sources of noise present in the image capture process, and it makes more sense to compute an estimate that takes advantage of all available data. If there is higher confidence in data values taken at higher exposures, then these data should be weighted more heavily in the estimation process.

An averaging process is indeed what is done in Refs. 4 and 10 However, in Ref. 4, a parametric form is assumed for the response function, restricting the application of the method to a limited number of situations. Furthermore, the weighting procedure does not take into consideration the quantization effects. While Ref. 10 does form an estimate of the nonlinear response function, the weighting process for the output values does not take into consideration the quantization effects discussed previously, and thus leaves room for improvement.

We propose a new method of increasing the dynamic range of images by using multiple exposures; the method is an extension of work first presented by the authors in Ref. 12. The probabilistic formulation of the problem results in a solution that satisfactorily deals with the problems of the algorithms reported before. In particular, the response function of the image capture device is estimated, thus creating versatility in our algorithm that is lacking in algorithms that assume a linear or parametric response. Estimation of the high-dynamic range pixel takes advantage of all available data by performing a weighted average. Proper weights arise from the problem formulation, allowing data from higher exposure times to be weighted more heavily.

Section 2 introduces the observation model for this work. Section 3 gives the maximum likelihood solution of the high dynamic range image for a known camera response, which includes situations such as those in Refs. 5–8. For unknown camera response, Sec. 4 discusses how the response function can be estimated. Experimental results are presented in Sec. 5, followed by concluding remarks in Sec. 6.

2 Observation Model

Assume there are N pictures taken of a static scene, with known exposure times $t_i, i=1,\ldots,N$. Each image consists of M pixels, and 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 is to determine the underlying light values or irradiances, denoted by x_j, that gave rise to the observations y_{ij}. Note that the N images must be properly registered, so that for a particular a, the light value x_a contributes to $y_{ia}, i=1,\ldots,N$. For this work, a normalized cross-correlation function[13] is used as the matching criterion to register images to 1/2-pixel resolution. Since this work considers only still imagery, the images are registered assuming a global translational model. For the more general case of motion imagery, one would need to consider more complicated motion models (e.g., affine or projective transformations).

There is a response function, denoted here by $f(\cdot)$, that maps exposure values to the observed output data. Since only the exposure time is being varied, the exposure values that are arguments of $f(\cdot)$ are products of time and irradiance, $t_i x_j$. Note that the camera response function is actually the composition of various functions, depending on the method of image capture. For a digital camera, $f(\cdot)$ might consist of the composition of the linear CCD response function, analog-to-digital conversion, and any nonlinear processing added by the camera manufacturer. For an analog camera, $f(\cdot)$ would consist of the composition of the film's response function, the response function of the printing process (if the images are scanned from prints, rather than from the actual film), and the response function of the scanning device, which itself consists of the composition of several more functions. Here, the only concern is the overall composite response function $f(\cdot)$ and not any of its individual elements.

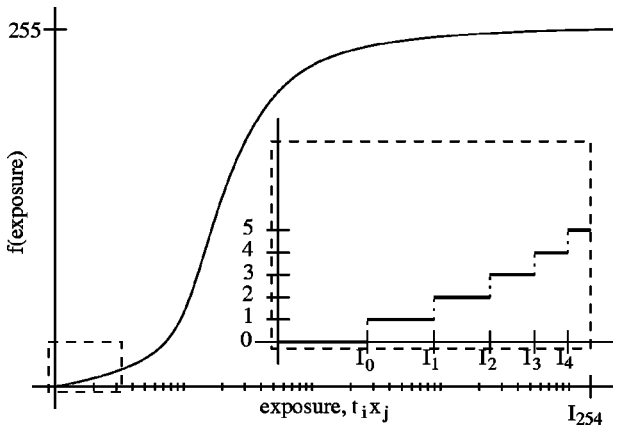

Fig. 3 Example camera response function, $f(\cdot)$, with logarithmic exposure scale. The inset shows a close-up view near the origin and demonstrates the discrete nature of $f(\cdot)$.

Since only the exposure time is being varied, the quantity contributing to the output value y_{ij} will be $t_i x_j$. To account for image capture noise, an additive noise term N_{ij}^c is introduced, which also contributes to the observed pixel values. Depending on the system used to capture the image, N_{ij}^c could come from a variety of sources, such as photon shot noise, dark current noise, and noise in the analog-to-digital conversion process, for example. The quantity $t_i x_j + N_{ij}^c$ is then mapped by the camera's response function $f(\cdot)$ to give the observed output values

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

Since y_{ij} are digital numbers, $f(\cdot)$ maps the nonnegative real numbers representing exposures $\Re^+ = [0,\infty)$ to an interval of integers, $\mathcal{O} = \{0,\ldots,255\}$ for 8-bit data. Without loss of generality, this work assumes the image data are 8 bits. The camera response function is explicitly written as

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

and thus $f(\cdot)$ is defined in terms of the 255 numbers I_m, $m = 0,\ldots,254$. Figure 3 shows an example $f(\cdot)$. For a linear response function, such as in Refs. 5 through 8, the I_m values would be evenly spaced (on a linear scale); in general, however, this will not be true.

In practice, due to camera (or film and scanner) characteristics and noise, the useful range of observations may not encompass the entire range [0,255]. For example, in the absence of light, one would expect the camera output to be 0, but in reality it may be some other small number, such as 3. Similarly, for an exceedingly bright light source, one would expect an output of 255, when in practice one might only observe a value of 253. Such behavior could result from a variety of causes, such as CCD dark current or lossy compression employed by a digital camera. To account for these scenarios, the histogram of an input dataset is examined during estimation of the response function. For a set of exposures that includes both highly underexposed and highly overexposed images (as would be recommended when estimating a response function), there should be significant peaks near 0 and 255 that indicate a zero point and a saturation point. The response function of Eq. (2) is thus modified, such that allowable output pixels belong to this slightly restricted range. Section 5 shows example histograms with limited output ranges.

The shown model assumes that only the exposure time varies between images. However, if one is using a camera (still or video) with automatic exposure and gain control, the model could be modified, such that the t_i values incorporate both exposure time and gain. The t_i would then no longer strictly represent time, but would instead represent composite gains that multiply the irradiances x_j. This work does not explicitly consider such situations, since it is assumed that the photographer is using a camera that allows full control over exposure.

3 High Dynamic Range Image with Known Response Function

In some situations, the response function of the image capture system is known. If one has access to direct CCD output, then one knows that the response is a linear function of exposure,[11] as was the case in Refs. 5 through 8. This section shows how to obtain high dynamic range image data with known response function. For the general situation, where direct CCD output is unavailable or where a film camera is used, Sec. 4 shows how to obtain the response function for arbitrary image capture systems. Once the response function is known, then the methods of this section can be applied directly.

The goal is to estimate the irradiances x_j with a dynamic range higher than that of the original observations. If the function $f(\cdot)$ is known, a mapping from $\mathcal{O}$ to $\Re^+$ can be defined as

$$f^{-1}(y_{ij}) = t_i x_j + N_{ij}^c + N_{ij}^q = I_{y_{ij}}. \tag{3}$$

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 here 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 Eq. (3),

$$I_{y_{ij}} = t_i x_j + N_{ij}. \tag{4}$$

The noise term N_{ij} consists of the noise term introduced in Sec. 2, as well as the dequantization uncertainty term N_{ij}^q. Note that accurately characterizing the noise terms N_{ij} would be extremely difficult, as it would require detailed knowledge of the specific image capture process being employed. One would have to characterize each of the separate noise sources that compose N_{ij}, which would be a complicated task that would have to be performed each time a different image capture system is used. Furthermore, if different noise models are found for different capture devices, then entirely different estimators would result. Rather than attempt this, the N_{ij} will be modeled as zero-

mean independent Gaussian random variables, with variances σ_{ij}^2. The Gaussian approximation is valid due to the potentially large number of noise sources present: all the noise sources inherent to acquiring digital images, e.g., dark current noise, photon shot noise, amplifier noise, and ADC noise; if a traditional camera is used, there is noise inherent to film, e.g., photon shot noise and film grain; and the dequantization noise N_{ij}^q.

Note that even with the Gaussian simplifying approximation, the noise variances σ_{ij}^2 would be difficult to characterize accurately. Again, detailed knowledge of the image capture process would be required, and the noise characterization would have to be performed each time a different image capture device is used. Alternatively, one could attempt to characterize the noise experimentally; however, this would be a burdensome task to perform with every image capture system. Therefore, rather than attempting either of these approaches, the variances will be chosen based on certainty functions.

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 the confidence that observed data are accurate. An approach is taken here that is similar to that of Mann and Picard.[4] Recall from Sec. 1 that the certainty function for a particular camera is determined by taking the derivative of the response function. In areas where the response function is approximately flat, its certainty will be very small. Similarly, high certainty functions occur when the response function is steep. For a response function, such as the example from Fig. 3, the certainty function is approximately zero near the extrema, and large near the middle of its response range. This work uses the certainty function as the weighting function.

To determine the certainty function, the derivative of the response function is taken with a logarithmic exposure axis. Note that the response function $f(\cdot)$ is not a continuous function, so a numerical approximation must be calculated instead. Here, the response function is approximated by a cubic spline with a moderate number of knots, from which the derivative is easy to calculate. To guarantee that zero weight is given to the extreme pixel values, the cubic spline is constructed such that the first derivative is zero at the two end points. Again, note that the cubic spline approximation is performed with a logarithmic exposure axis. The certainty function is then normalized, such that its maximum value is unity. Once the certainty function $c(\cdot)$ is determined, the weighting function becomes a function of the pixel observations and the certainty function: $w_{ij}=w(y_{ij})=c(I_{y_{ij}})$. Section 5 provides experimental examples of determining the certainty function.

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

$$P(\mathbf{I_y}) \propto \exp\left\{ -\frac{1}{2}\sum_{i,j} w_{ij}(I_{y_{ij}}-t_i x_j)^2 \right\}. \tag{5}$$

A maximum-likelihood (ML) approach is taken to find the high dynamic range image values. The maximum-likelihood solution finds the values x_j that maximize the probability in Eq. (5). Maximizing Eq. (5) is equivalent to minimizing the negative of its natural logarithm, which leads to the following objective function to be minimized:

$$O(\mathbf{x}) = \sum_{i,j} w_{ij}(I_{y_{ij}}-t_i x_j)^2. \tag{6}$$

Equation (6) is easily minimized by setting the gradient $\nabla O(\mathbf{x})$ equal to zero. (Note that if t_i were unknown, one could jointly estimate x_j and t_i by arbitrarily fixing one of the t_i, and then performing an iterative optimization of Eq. (6) with respect to both x_j and t_i.) This yields

$$\hat{x}_j = \frac{\Sigma_i w_{ij} t_i I_{y_{ij}}}{\Sigma_i w_{ij} t_i^2}, \tag{7}$$

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 Eq. (7). Thus this method takes advantage of the quantization effects utilized in Refs. 5 through 8. However, here a noise-reducing averaging is being performed, which utilizes data from all input pixels.

Equation (7) 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, so that in the future the results just presented can be applied directly.

4 Unknown Response Function

Except in very specialized situations, the camera response function will not be known, and must be estimated. To determine the response function uniquely, the 255 values I_m, $m=0,\ldots,254$ must be found.

At first glance, one may consider directly using the objective function in Eq. (6) to determine the I_m values needed to define the response function. Note that to estimate the I_m values from Eq. (6), the x_j values are also unknown and need to be estimated simultaneously. Thus, the objective function for the case of an unknown response function is

$$\tilde{O}(\mathbf{I},\mathbf{x}) = \sum_{i,j} w_{ij}(I_{y_{ij}}-t_i x_j)^2. \tag{8}$$

An additional constraint on the response function is required when estimating **I** and **x** together using Eq. (8). This restriction on $f(\cdot)$ is in regard to scale. The goal here is not to determine absolute irradiance values, so issues relating to physical units are avoided. It is sufficient to determine the high dynamic range image to within a scale factor, for then the range of values found can be mapped to any desired interval. Since the scale of $\hat{x}_j$ is dependent on the scale of I_m, the estimates for the I_m values are constrained, such that $\hat{I}_{128}=1.0$. This is enforced by dividing each $\hat{I}_m$ by $\hat{I}_{128}$.

Since the response function is not yet known, the weighting function (determined by differentiation of the response) is not known either. Rather than jointly attempting

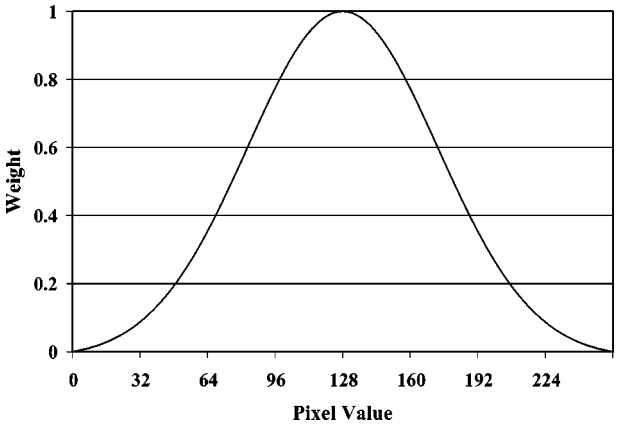

Fig. 4 Weighting function for determination of camera response.

to estimate w_{ij}, I_m, and x_j, the weighting function will be fixed *a priori*. (With so many unknowns, the estimation quickly becomes intractable, especially considering the dependence of w_{ij} on I_m.) Figure 4 shows the weighting function used for determination of the camera response. This weighting function is the same one used in Ref. 12, and similar to the one used in Ref. 10. Note, however, that the weighting functions used in Refs. 10 and 12 were used for determining both response function and image estimates, whereas here the weighting function of Fig. 4 is being used to estimate the response function, while the certainty function is being used to estimate actual high dynamic range images.

A form of Gauss-Seidel relaxation[14] is used to determine the solution. 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, Eq. (8) is first to be minimized with respect to each I_m. Then the scale restriction mentioned before is enforced. Finally, Eq. (8) is minimized with respect to each x_j, which constitutes one iteration of the algorithm.

Estimates for the variables of interest at the l'th iteration are denoted as $\hat{\mathbf{I}}^{(l)}$ and $\hat{\mathbf{x}}^{(l)}$. The initial $\hat{\mathbf{I}}^{(0)}$ is chosen as a linear function, with $\hat{I}^{(0)}_{128}=1.0$. The initial $\hat{\mathbf{x}}^{(0)}$ is chosen according to Eq. (7), using the initial linear $\hat{\mathbf{I}}^{(0)}$.

First, to minimize with respect to I_m at the l'th iteration, the partial derivative of Eq. (8) with respect to I_m is taken and set equal to zero. This yields

$$\hat{I}^{(l)}_m=\frac{\Sigma_{(i,j)\in E_m} w_{ij}t_i\hat{x}^{(l-1)}_j}{\Sigma_{(i,j)\in E_m} w_{ij}}, \tag{9}$$

where the index set E_m is defined as

$$E_m=\{(i,j):y_{ij}=m\}, \tag{10}$$

the set of indices such that m was observed for the input images. However, since w_{ij} is constant for $(i,j)\in E_m$, Eq. (9) simplifies to

$$\hat{I}^{(l)}_m=\frac{1}{\mathrm{Card}(E_m)}\sum_{(i,j)\in E_m} t_i\hat{x}^{(l-1)}_j. \tag{11}$$

$\mathrm{Card}(E_m)$ is the cardinality of E_m, i.e., the number of times m was observed. Equation (11) is applied for each $\hat{I}^{(l)}_m, m=0,\ldots,254$.

After scaling the response function, such that $\hat{I}^{(l)}_{128}=1.0$, minimization is performed with respect to each $x_j, j=1,\ldots,M$ by applying Eq. (7),

$$\hat{x}^{(l)}_j=\frac{\Sigma_i w_{ij}t_i\hat{I}^{(l)}_{y_{ij}}}{\Sigma_i w_{ij}t_i^2}. \tag{12}$$

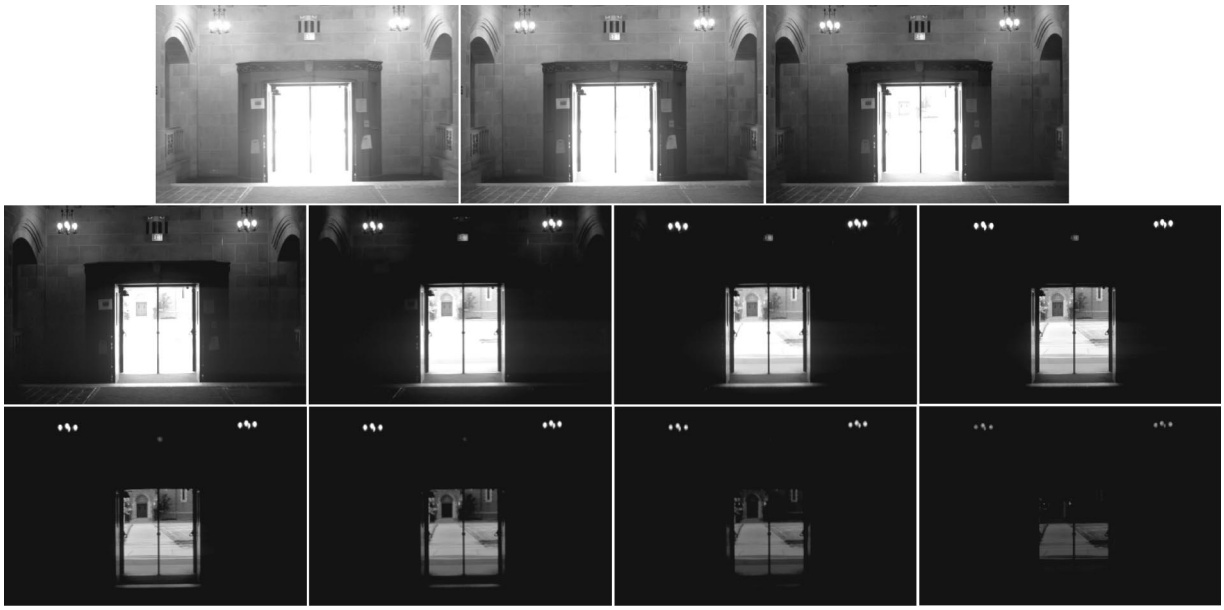

Fig. 5 "Cushing" scene. Eleven pictures from inside Cushing Hall at the University of Notre Dame, taken with a Nikon FM camera with aperture f/5.6. The original image resolution is 450×300. The exposure times are, from brightest to darkest, 1, $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{15}$, $\frac{1}{30}$, $\frac{1}{60}$, $\frac{1}{125}$, $\frac{1}{250}$, $\frac{1}{500}$, and $\frac{1}{1000}$ s.

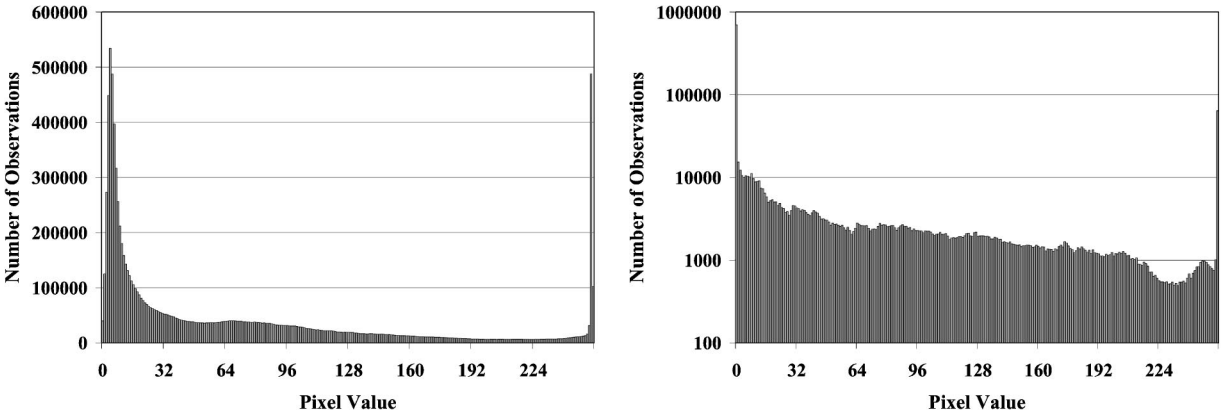

Fig. 6 Histograms of the input datasets. Left: Histogram of the Studio dataset. Right: Histogram of the Cushing dataset (note the logarithmic scale for the Cushing histogram).

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 threshold.

5 Experimental Results

Figures 1 and 5 show series of photographed scenes that have wide dynamic ranges. The photographs in Fig. 1 were taken with a digital camera. Since raw image output was

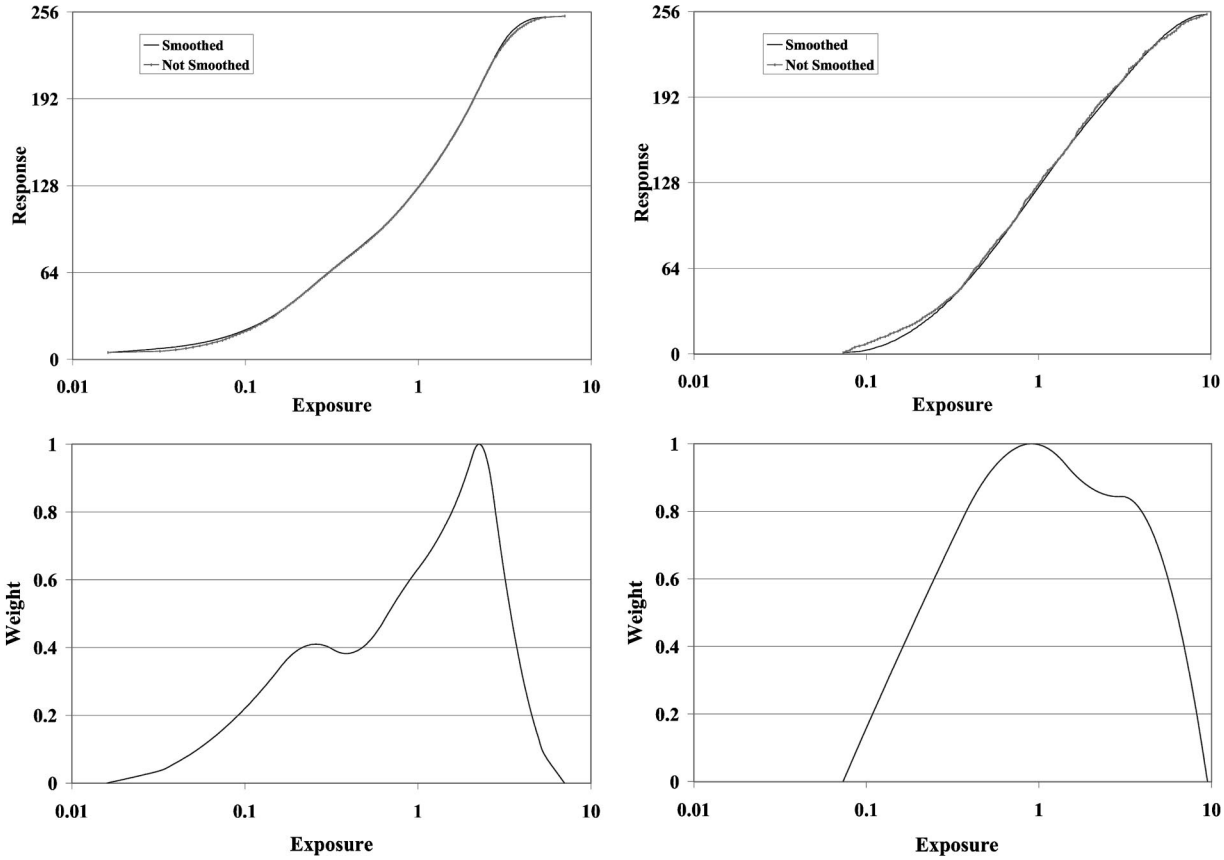

Fig. 7 Results for the Studio (left) and Cushing (right) datasets. Both smoothed and unsmoothed response functions are shown on top, while the new weighting functions (which are the derivatives of the smooth response functions) are on the bottom. In all four plots, the horizontal axis represents exposure on a logarithmic scale. Note the difference between the weighting functions above and that of Fig. 4. The units for the horizontal axis here are exposure, while those of Fig. 4 are pixel observations. The two are related by the response function.

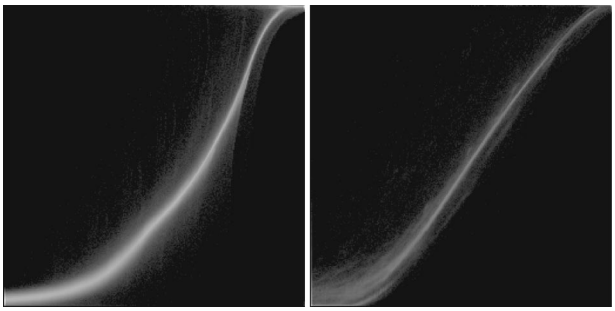

Fig. 8 Two-dimensional response histograms of dimension 256 $\times$256. Left: Studio dataset. Right: Cushing dataset. The indices to the histogram are $[\log(t_i\hat{x}_j), y_{ij}]$, and the histogram counts are shown on a logarithmic scale. Such 2-D histograms provide a visual indication of how well the data match the results. (Note that the right figure is not as ''bright'' as the one on the left due to the higher fraction of saturated observations, as shown in Fig. 6.)

not available for this camera, the images were compressed at the highest JPEG quality setting, and the green color planes were later extracted for use in these tests. The photographs in Fig. 5 were taken with a traditional camera using slide film and scanned using a Leafscan-35 slide-film scanner. The Leafscan-35 allows one to maintain constant exposure times between scans, as well as retaining the black and white points. Exposures of 1/30 s from Fig. 1 and 1/8 s from Fig. 5 represent photographs taken at what might be considered ''normal'' exposure settings. From the figures, 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, scenes such as these are excellent candidates for the algorithm discussed here.

Figure 6 shows histograms for the two input datasets. As described in Sec. 2, the zero and saturation points are taken as the peaks of these two histograms. Limits for the Studio scene were 4 and 254, and limits for the Cushing scene were 0 and 255. For the Studio scene, the large number of observations that are less than 4 are probably due to the nonlinear nature of the JPEG compression.

Figure 7 shows the determined response functions and their derivatives for the Studio and Cushing datasets. The top row of the figure shows both the response functions and their cubic-spline fits. Note that the response functions are not linear functions, and thus the use of algorithms such as described in Refs. 5–8 would be inappropriate for either of these cameras. The bottom row of the figure shows the derivatives of the smoothed response functions, which are the new weighting functions to be utilized when using the algorithm to determine high dynamic range pixels.

A visual aid for judging the quality of the estimates is given in Fig. 8, where 2-D histograms of the result are shown. The plots are constructed based on high dynamic range pixel estimates $\hat{x}_j$, exposure times t_i, and pixel observations y_{ij}. The column index for the histogram is taken as $\log(t_i\hat{x}_j)$, which represents an estimate of the i'th exposure value of the j'th pixel, while the row index is simply taken as y_{ij}. Histogram counts are shown on a logarithmic scale. Ideally, the 2-D histogram would trace out a fine curve that exactly follows the camera response function. However, due to noise, the fine curve is smeared. Also affecting these histograms could be unmodeled components of the image acquisition process, such as the JPEG compression for the Studio scene. The response functions are quite visible in these plots, and bear obvious resemblances to the response functions shown in Fig. 7.

Fig. 9 High dynamic range output images for the Studio scene. Top: Results after gamma modification, $\gamma=0.01$, and $\epsilon=0.001$. Bottom: Results after histogram equalization.

The $\hat{x}_j$ values are the ultimate variables of interest. After the response functions were determined, the final image estimates were formed by the technique in Sec. 3. Displaying these high dynamic range images on devices of limited dynamic range is a nontrivial undertaking. Methods from computer graphics literature can be found in Refs. 15–18. However, the focus of this research is not the display of high dynamic range images, but rather the acquisition of high dynamic range images. The methods used here for visualization of high dynamic range images are not chosen to give the most ''visually pleasing'' image, but rather to demonstrate the results of the proposed algorithm.

Two methods are chosen here for display of the high dynamic range results. The first method is similar to gamma correction, and obeys the input-output relation

Fig. 10 High dynamic range output images for the Cushing scene. Top: Results after gamma modification, $\gamma=0.01$, and $\epsilon=0.0002$. Bottom: Results after histogram equalization.

$$h(z)=\begin{cases} Az, & z\leqslant\epsilon \\ Bz^{\gamma}-C, & z>\epsilon \end{cases}. \tag{13}$$

In this equation, input values z are assumed to be normalized to $[0,1]$. Imposing continuity of $h(z)$ and its first derivative at $z=\epsilon$, as well as forcing $h(1.0)=1.0$, allows Eq. (13) to be uniquely defined by the parameters ϵ and γ. The second method used here for display of high dynamic range images is simple histogram equalization.

Figures 9 and 10 show results for the Studio and Cushing scenes. It is readily apparent that the images shown in these two figures contain significantly more information than any of the single input images from their respective datasets—both bright and dark areas of the scenes have been captured to a single image. Note that the two methods of visualization presented here are relatively simple, and that the photographer would be free to adjust the images to his or her liking by, for example, fine tuning an input-output transformation, or performing an adaptive histogram equalization.

For the example results just given, the response function was first estimated using a large number of training images. 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. Although not shown here, Eq. (7) can be used with the determined response function to estimate the desired high dynamic range image values directly using fewer images.

The high dynamic range images determined using the method outlined here have several advantages over single images such as those in Figs. 1 or 5. The images obtained with the proposed method have decreased noise due to the averaging of pixel values from each of the input images. Furthermore, they contain information in both low- and high-light areas, since the high dynamic range images consist of data from each of the input images. They also have higher weights for pixels at longer exposures, advantages of which were described previously. Traditional image processing algorithms (e.g., contrast stretching, histogram equalization, edge detection, etc.) can be applied to the high dynamic range images with better results due to the increased amount of information present.

6 Conclusion

This work has introduced a method of increasing the effective dynamic range of digital images by using multiple pictures of the same scene taken with varying exposure times. The technique is based on a probabilistic formulation. 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 data from all the input images.

References

1. S. Mann and R. W. Picard, ''Video orbits of the projective group: A simple approach to featureless estimation of parameters,'' *IEEE Trans. Image Process.* **6**(9), 1281–1295 (Sep. 1997).
2. C. W. Wyckoff, ''An experimental extended exposure response film,'' in *SPIE Newsletter*, pp. 16–20 (June/July 1962).
3. S. Mann, ''Compositing multiple pictures of the same scene,'' *Proc. 46th Annual IS&T Conf.*, Boston, MA, pp. 50–52, May 9–14, 1993.
4. S. Mann and R. W. Picard, ''On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures,'' *IS&T's 48th Annual Conf.* Washington, D.C., pp. 422–428, May 7–11, 1995.
5. B. C. Madden, ''Extended intensity range imaging,'' technical report, GRASP Laboratory, Univ. of Pennsylvania (1993).
6. K. Yamada, T. Nakano, and S. Yamamoto, ''Effectiveness of video camera dynamic range expansion for lane mark detection,'' *Proc. IEEE Conf. Intell. Transportation Syst.*, pp. 584–588, Nov. 9–12, 1997.
7. K. Yamada, T. Nakano, S. Yamamoto, E. Akutsu, and K. Ashi, ''Wide dynamic range vision sensor for vehicles,'' *IEEE Intl. Conf. Vehicle Navigation Info. Syst.*, pp. 405–408, Aug. 31–Sep. 2, 1994.
8. K. Moriwaki, ''Adaptive exposure image input system for obtaining high-quality color information,'' *Systems and Computers in Japan* **25**(8), 51–60 (July 1994).
9. Z. Chen and G. Mu, ''High-dynamic-range image acquisition and display by multi-intensity imagery,'' *J. Imaging Sci. Technol.* **39**(6), 559–564 (Nov/Dec 1995).
10. P. E. Debevec and J. Malik, ''Recovering high dynamic range radiance maps from photographs,'' *SIGGRAPH 97 Conf. Proc., Computer Graphics Annual Conf. Series*, pp. 369–378, Aug. 3–8, 1997.
11. G. C. Holst, *CCD Arrays, Cameras, and Displays*, JCD Publishing and SPIE Optical Engineering Press, Bellingham, WA (1996).
12. M. A. Robertson, S. Borman, and R. L. Stevenson, ''Dynamic range improvement through multiple exposures,'' *Int. Conf. Image Process.* **3**, 159–163 (1999).
13. H.-M. Hang and Y.-M. Chou, ''Motion estimation for image sequence compression,'' in *Handbook of Visual Communications*, H. M. Hang and John W. Woods, Eds., pp. 147–188, Academic Press, New York (1995).
14. J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York (1970).
15. S. N. Pattanaik, J. A. Ferwerda, M. D. Fairchild, and D. P. Greenberg, ''A multiscale model of adaptation and spatial vision for realistic image display,'' *SIGGRAPH 98 Conf. Proc., Computer Graphics Annual Conf. Series*, pp. 287–298 (July 1998).
16. G. W. Larson, H. Reshmeier, and C. Piatko, ''Visibility matching tone

reproduction operator for high dynamic range scenes,'' *IEEE Trans. Vis. Comput. Graph.* **3**(4), 291–306 (1997).
17. J. Tumblin and G. Turk, ''LCIS: A boundary hierarchy for detail-preserving contrast reduction,'' *SIGGRAPH 99 Conf. Proc., Computer Graphics Annual Conf. Series* (1999).
18. C. Schlick, ''Quantization techniques for visualization of high dynamic range pictures,'' in *Photorealistic Rendering Techniques, Proc. 5th Euro-graphics Rendering Workshop*, G. Sakas, P. Shirley, and S. Mueller, editors, pp. 7–20 (1995).

Mark A. Robertson received his BS in electrical engineering from Tri-State University, Angola, Indiana, in 1996, and his MS and PhD degrees in electrical engineering from the University of Notre Dame, Indiana, in 1998 and 2001. While at Notre Dame, he was supported by the Arthur J. Schmitt Fellowship, the Intel Corporation, and the Indiana Space Grant Consortium. He is currently working in the Multi-Sensor Exploitation Branch of the Air Force Research Laboratory in Rome, New York. His current research interests include image and video enhancement, high-quality electronic imaging, and image and video communication.

Sean Borman received his BSc degree in electrical engineering with first class honors from the University of Cape Town, South Africa, in 1992. He completed his MSEE degree at the University of Notre Dame in 1996. After teaching and consulting in South Africa, he returned to Notre Dame to pursue his PhD degree, also in electrical engineering. His professional interests include inverse problems, super-resolution video restoration and motion estimation. He is a recipient of the South African Foundation for Research Development (FRD) award for post-graduate overseas study and a Fulbright Scholarship.

Robert L. Stevenson received his BEE degree (summa cum laude) from the University of Delaware in 1986, and his PhD in electrical engineering from Purdue University in 1990. While at Purdue he was supported by graduate fellowships from the National Science Foundation, DuPont Corporation, Phi Kappa Phi, and Purdue University. He joined the faculty of the Department of Electrical Engineering at the University of Notre Dame in 1990, where he is currently a professor. His research interests include image/video processing, image/video compression, robust image/video communication systems, multimedia systems, ill-posed problems in computational vision, and computational issues in image processing. He is an Associate Editor of the *IEEE Trans. on Image Processing* and the *IEEE Trans. on Circuits and Systems for Video Technology*, and is a former Associate Editor of the *Journal of Electronic Imaging*.