Journal of Electronic Imaging 12(3), 459–469 (July 2003).

High dynamic range imaging for digital still camera: an overview

S. Battiato
A. Castorina
M. Mancuso
STMicroelectronics
Advanced System Technology
Catania Laboratory
Imaging and Multimedia Mobile Group
Catania, Italy
E-mail: sebastiano.battiato@st.com

Abstract. *We present a collection of methods and algorithms able to deal with high dynamic ranges of real pictures acquired by digital engines e.g., charge-coupled device (CCD/CMOS) cameras. An accurate image acquisition can be challenging under difficult light conditions. A few techniques that overcome dynamic range limitations problems are reported. The presented methods allow the recovery of the original radiance values of the final 8-bit-depth image starting from differently exposed pictures. This allows the capture of both low- and high-light details by merging the various pictures into a single map, thus providing a more faithful description of what the real world scene was. However, in order to be viewed on a common computer monitor, the map needs to be compressed and requantized while preserving the visibility of details. The main problem comes from the fact that the contrast of the radiance values is usually far greater than that of the display device. Various related techniques are reviewed and discussed.* © 2003 SPIE and IS&T.
[DOI: 10.1117/1.1580829]

1 Introduction

Scenes from the real world can exhibit a broad range of light variations. This is particularly true for scenes containing both areas of low and high illumination, like a dark room with a sunlit window. Also, the presence of objects with reflective or opaque surfaces contributes to the variety of local light levels. It is clear that in such scenes, the *dynamic range*, where the term refers to the ratio of the highest and lowest recorded level of light, can be very high. Thus it is impossible to obtain an adequate representation of a scene having a wide dynamic range by using a film or a digital still camera (DSC) as a recording system. There will be an information loss due to the limited dynamic range of the used device. In the case of a DSC, the dynamic range is usually described in terms of the ratio between the maximum charge that the sensor can collect (*full well capacity*) and the minimum charge that just overcomes sensor noise (*noise floor*). The light values captured by a CCD or CMOS sensor are usually finally quantized (linearly or not) in a range between [0:255] to produce an 8 bit per pixel storage, even if more bits were initially used by the sensor analog to digital converter (ADC) to cover the input signal. This yields a coarse resolution interpretation of the physically, continuously varying values present in the real world. Information loss is observed in highly illuminated areas (where all light variations are mapped to the same value, and thus become saturated) and in dimly illuminated areas (where information is overridden by sensor-noise-producing pixel values floating nearly between 0 and 1 in a random fashion). Today most DSCs incorporate some kind of automatic gain control (AGC) acting accordingly with a *metering* technique[1,2] to allow the users to take pictures without worrying about the exposure settings of the device. In the ''manual mode,'' it is possible to selectively adjust camera settings (the so-called *exposure bracketing*), until a subjective satisfactory representation is obtained. It is clear that these techniques allow only a slight control on where information loss will occur. Furthermore, once portions of the dynamic range have been definitively lost, there is no way to restore them, even using some sort of postprocessing technique.[3] However, if multiple pictures of the same scene are acquired using different exposure settings, each of them will reveal different details, covering a wider dynamic range than the one that would have been captured with a single shot: pictures taken with short exposure times provide highlight information and vice versa. Multiple shots are acquired with the aim of merging them into a single image of increased dynamic range (see Ref. 4 for a merging technique used to increase image resolution). Multiple pictures are used to obtain an estimation of the original quantities of light falling on the sensor cells, providing an approximation of what the original scene was, called a *radiance map*. These maps somehow go further than a usual 8 bit image, since information is expressed in arbitrary ranging, floating point values. To be displayed on a common CRT monitor of limited dynamic range with finite

Paper JEI 02020 received Feb. 20, 2002; revised manuscript received Aug. 20, 2002, and Jan. 27, 2003; accepted for publication Jan. 29, 2003. This paper is a revision of a paper presented at the SPIE conference on Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications III, January 2002, San Jose, California. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 4669.
1017-9909/2003/$15.00 © 2003 SPIE and IS&T.

available intensity values, the maps need to be compressed and requantized. In the following sections, some techniques to build and represent *radiance maps* are described and compared. Some experimental results are also presented.

2 Radiance Map Recovery

When the observed scene contains a wide dynamic range exceeding the one of the recording medium, there is no way to prevent information loss in high-light or shadowed details. Once details are lost (due to saturation or because information has been completely overwritten by noise), even the application of postprocessing operations[3] is completely useless. Besides this, it is possible to increase the dynamic range by using the final 8 bit depth information coming from multiple, differently exposed frames of the same scene. Every image of the set provides a different measurement of the light quantity falling on the sensor array (or on the film), with each measure being more or less reliable depending on the exposure: high-light measures are accurate in images with low exposure and vice versa. The exposure X is defined as the product,

$$X = I \cdot t, \tag{1}$$

where I is the radiance and t the time of exposure to light. When a pixel value Z is observed on the captured image, the only information available is that it comes from some scene radiance I (at a certain exposure setting) mapped to Z by some unknown function f. This mapping is seldom linear; pixel brightness values are not true measurements of the relative radiance values. If a pixel value has twice the value of another one, it rarely comes from the double of the radiance value. Nonlinearity is more evident for film-based systems, but even if digital image detectors are designed to produce voltages that are linearly related to irradiance values, the mapping process from these voltages to the storage medium often introduces nonlinearity, which is caused, for example, by some form of gamma mapping,[5] A/D conversion, sensor noise, etc. Furthermore, often DSC cameras have a built-in nonlinear mapping to mimic a film-like response, since nonlinearity produces better images when viewed on low-contrast displays. The more evident nonlinearity can be found at the saturation point of the response curve (the *shoulder* region), where all radiance values are mapped to white, and in the lower portion of the curve (the *toe* region), where all radiance values are mapped to a near-black value. By using differently exposed images, the measured brightness values change with the exposure, while radiance values remain constant. Because of this, once the imaging system transfer f function is known, it is possible to use its inverse to estimate the underlying (properly scaled) radiance values and build an accurate estimation of the original scene values: the so-called *radiance map*. In the following sections, some of the main techniques developed for this purpose are examined.

2.1 *High Dynamic Range (HDR) Image Recovering Using Exposure Times*

The algorithm described by Debevec and Malik[6] considers a sequence of N digitized pictures, representing the same scene acquired at known exposure durations Δt_j, $j = 1,2,...,N$. The input images are used to estimate the camera response function, and then to reconstruct the high dynamic range *radiance map*. Denoting with Z_{ij} the i'th pixel value of the j'th image, it is possible to relate pixel values to exposures with the following equation:

$$Z_{ij} = f(I_i \cdot \Delta t_j), \tag{2}$$

where I_i are the unknown radiance values. Assuming that f is monotonic (this should be obviously true, because an increase in radiance values always produces an increased or equal recorded value), its inverse can be obtained as follows:

$$f^{-1}(Z_{ij}) = I_i \cdot \Delta t_j \tag{3}$$

taking the logarithm of both sides and substituting $\log f^{-1}(Z_{ij})$ with $g(Z_{ij})$, we finally obtain:

$$g(Z_{ij}) = \log[f^{-1}(Z_{ij})] = \log(I_i) + \log(\Delta t_j). \tag{4}$$

The problem consists in estimating the unknown values I_i and the different finite values (Z values range in [0:255]) that g can assume. This leads to minimizing the following objective function:

$$O = \sum_{i=1}^{P} \sum_{j=1}^{N} \{w(Z_{ij})[g(Z_{ij}) - \log(I_i) - \log(\Delta t_j)]\}^2 + \lambda \sum_{z=Z_{\min}+1}^{z=Z_{\max}-1} [w(z)g''(z)]^2, \tag{5}$$

where N and P are the number of images and the number of pixels of the images, respectively, λ is a smoothing parameter, and $w(Z)$ is a weighting function to anticipate the smoothness of the curve in the middle region (for example, a triangular curve centered at 128). More details for both the parameter λ and the use of the second derivatives of $g(Z)$ can be found in Ref. 6. A further constraint added to the linear system, to fix a suitable scale factor, imposes $g(128) = 0$. Equation (5) can be solved using singular value decomposition (SVD), but of course complexity makes it impossible to use all of the available pixel values. In practice only a good subset of the pixel values is used. The pixels considered should meet the following constraints:

1. to have a good spatial distribution all over the images
2. to be extracted from *low variance* regions where noise and high radiance variations are less severe
3. to have a good tonal repartition: they should well sample the input range.

The construction of the map is simply done by averaging the different radiance estimates, provided by every image of the set. In fact, from Eq. (4):

$$\log(I_i) = g(Z_{ij}) - \log(\Delta t_j). \tag{6}$$

An accurate estimate for I_i, by weighting the different radiance values, is then obtained as follows:

Fig. 1 Four images extracted from a full sequence of a real scene.

$$\log(I_i)=\frac{\Sigma_{j=1}^{N} w(Z_{ij})[g(Z_{ij})-\log(\Delta t_j)]}{\Sigma_{j=1}^{N} w(Z_{ij})}. \quad (7)$$

Note how radiance values having mid-gray values are weighted more heavily, since they can be trusted more in the sense that they come from the middle portion of the curve where the system is more sensitive. Each image, in every point, provides an estimate of the underlying original radiance value, with some of them less or more reliable, depending of the portion of the curve it comes from. The map reconstruction obtained by a suitable weighted average is also used in a similar manner by the following techniques. Figure 1 shows a subset of a sequence (the full set was composed by eight images) used to recover both curve and radiance maps of the real scene. The images have been captured by a conventional film camera and then digitized using a scanner. Figure 2 (left) shows the recovered response curve, and Fig. 3 (left) illustrates the weighting function used.

2.2 *HDR Recovery Using Automatically Estimated Exposure Ratios*

The basic concepts of the algorithm described by Mitsunaga and Nayar[7] are nearly the same and they are briefly described here. Different than Ref. 6, instead of using images with known exposure times, here exposure ratios between them are known. The response curve f is modeled as a high-order polynomial function, which is supposed to be monotonic or semimonotonic, and thus invertible. The only available information is that the recorded final value of the i'th pixel for the j'th image Z_{ij} is related to the scaled scene radiance $I_i e_j$, where e_j is a scaling factor by some mapping in the form

$$Z_{ij}=f(I_i\cdot e_j). \quad (8)$$

Please note that the notation of Eq. (8) is similar to Eq. (2), and that the quantity $I_i e_j$ is something equivalent to the previously defined exposure. Thus, for example, having two images where one has been exposed double the times of the other yields a ratio of 0.5, and we could set $e_1 \approx 0.66$ and $e_2 \approx 1.33$ (since the absolute values of e_j are not important they are fixed such that their mean is 1). Once again we are interested in finding the inverse of Eq. (8):

$$g(Z_{ij})=f^{-1}(Z_{ij})=I_i e_j. \quad (9)$$

A polynomial function of order K is supposed for g:

$$Ie=g(Z)=\sum_{k=0}^{K} c_k Z^k. \quad (10)$$

The problem can thus be stated as recovering the order K and the coefficients c_k. If the ratios $R_{j,j+1}$ between the image pairs $(j,j+1)$ are known, then for a pixel value in position i:

$$\frac{I_i e_j}{I_i e_{j+1}}=\frac{g(Z_{i,j})}{g(Z_{i,j+1})}=R_{j,j+1}. \quad (11)$$

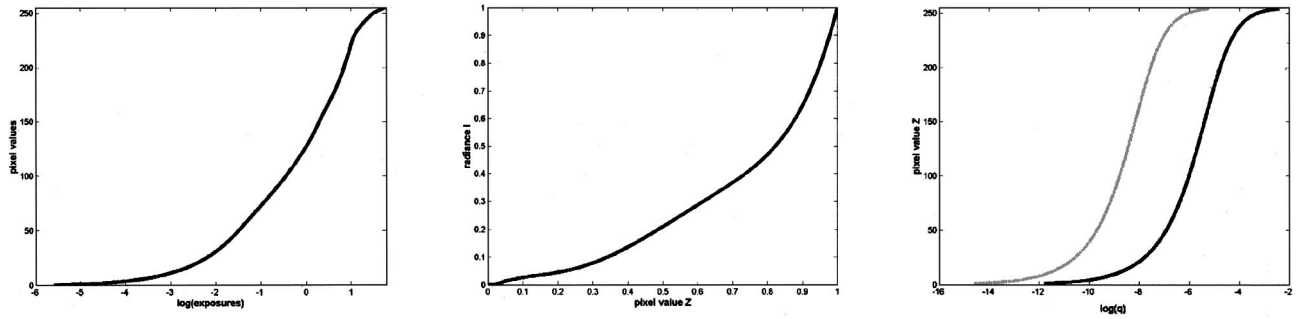

Fig. 2 From left to right, the response curves obtained from the techniques described in Refs. 6, 7, and 8. The first and the last curve are plotted using a logarithmic scale. This representation is consistent with that used in the original research. The last curve shows the effective response functions just for two images (where the second received 16 times more light than the first). All curves refer to the sequence of Fig. 1.

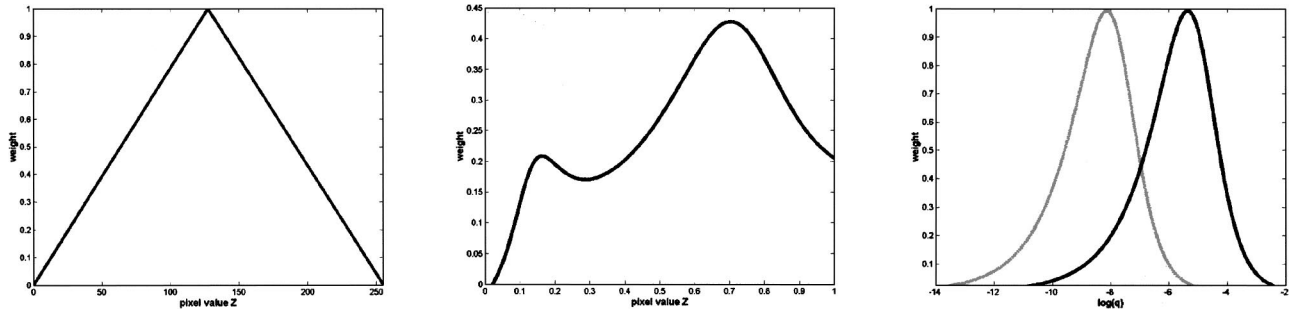

Fig. 3 From left to right, the weighting functions used to build the radiance maps as described, respectively, in Refs. 6, 7, and 8. The first and the last curve are plotted using a logarithmic scale. This representation is consistent with that used in the original research. The last curve shows the derivatives of the effective response functions just for two images (where the second received 16 times more light than the first). All curves refer to the sequence of Fig. 1.

Like in the previous model, it has been derived as an objective function to be minimized:

$$O=\sum_{j=1}^{N}\sum_{i=1}^{P}\left[\sum_{k=0}^{K} c_k Z_{i,j}^{k}-R_{j,j+1}\sum_{k=0}^{K} c_k Z_{i,j+1}^{k}\right]^2, \tag{12}$$

where N is the number of images and P the number of pixels. The scale is fixed by imposing $g(1)=1$ and the ranges for both I and Z values are rescaled in the range [0:1]. Since K is unknown, the evaluation of Eq. (12) is repeated for different values of K until the error is minimized. For this reason an upper bound for K must be chosen. An estimation of Eq. (12) is obtained using only a subset of the available pixel data by applying some selection criteria (like in the previous model). To merge the N images into one radiance map, the pixel values Z of every image are mapped onto the radiance values I by using the recovered function and the scaling factors e. A useful feature consists of the possibility of automatically estimating the ratios of the input images by using just raw initial estimates. This is useful when the exact values are unknown. The new ratio estimates are obtained starting from initial user-supplied ratio values and minimizing Eq. (12) by using Eq. (10). The process is repeated until the error induced by Eq. (12) meets a convergence criterion. However, we have found this estimate to be successfull only if at least five or six images are used. The obtained values are then merged using a weighting function. For this purpose, instead of using a *hat* or *Gaussian* function, we suggest taking the ratio of f and its first order derivative f', because the more a signal is sensitive to changes, the more it can be trusted. Figure 3 (middle) shows the certainty function related to the previously computed curve illustrated in Fig. 2 (middle), referring to the image sequences of Fig. 1. The map construction process is thus similar to the technique described in Ref. 6, differing only in the used weighting function. In the case of color images, after recovering separated functions for each channel, we propose a calibration step, since estimated radiances are expressed in relative units (with any physical meaning), where the scalings are unknown. Calibration is performed by estimating three scaling factors K_r, K_g, and K_b, such that the following color constraint holds for all neutral color points and for all channels c in the input images (e.g., $c=r,\ g,\ b$), thus preserving relative color balancing:

$$\frac{I_c K_c}{I_r K_r+I_g K_g+I_b K_b}=\frac{Z_c}{Z_r+Z_g+Z_b}. \tag{13}$$

A color Z is referred to as neutral when

$$|Z_c-\text{mean}(Z_r,Z_g,Z_b)|/\text{mean}(Z_r,Z_g,Z_b)<T, \tag{14}$$

where T is a suitable threshold value. This kind of simple balancing could also be used for the other described techniques.

2.3 *HDR Image Recovery Using a Parametric Response Curve*

Another technique has been presented in Refs. 8 and 9. We propose approximating the camera response function by using a simple parametric function f that relates pixel values Z to the received light, called photoquantity q. Different parametrically formulated versions of Eq. (15) are proposed. The most flexible model is the following:

$$f(q)=\frac{1}{(1+\exp\{-[A\log(q)]\})^{C}}, \tag{15}$$

where the unknown quantities are the parameters A and C, and pixel values are supposed to range [0:1]. Parameters estimation can be performed by using the following observation: having two different images of the same scene, acquired with different exposure times, where the exposure ratio between them is k, it is possible to write the following statement:

$$g[f(q)]=f(kq), \tag{16}$$

where g is a new unknown function expressing how a pixel value of the first image becomes a new pixel value in the second one. This relationship can be easily observed by building up a *cross-histogram* of the two images: a 256 $\times$256 plot where an entry at position (x,y) means that the

gray level x of the first image becomes the gray level y in the second image. Using Eq. (15) we note that Eq. (16) can be rewritten as:

$$g[f(q)]=\frac{f(q)\cdot k^{AC}}{[\sqrt[C]{f}(q)\cdot(k^A-1)+1]^C}, \tag{17}$$

where k represents the exposure ratio between two images. By using the *cross-histogram*, the pair values $\{f(q),g[f(q)]\}$ are known, thus A and C can be retrieved by means of nonlinear regression across well populated pairs (for example, using the *Levenberg-Marquardt* method). Once again the radiance map construction involves a weighted average. From each image, the photoquantities q can be estimated by inverting the response function and scaling the ratios k_i. This is, to some extent, equivalent to the previous algorithms. The adopted weighting functions are the derivatives of the response functions shifted by their ratios k_i (their meaning is intuitive, since the higher the slope of the derivative for a fixed value, the higher its reliability). In this case the weight is computed starting from the logarithms of the *q quantities* (however, as noted in Ref. 8 the weight could also be expressed in terms of pixel values) and not from the pixel values. Figure 2 (right) illustrates the shifted response curves, and Fig. 3 (right) shows their derivatives for two images of the sequence of Fig. 1, where one image was 16 times brighter than the other. Curves are plotted using a logarithmic scale. Even if this technique is a sort of simplification, it provides an easily computable solution of the problem, avoiding the complex system solving required by the previous two methods. This can be useful because even if, in theory, the response functions should be estimated just once for a specific system, it could sometimes be necessary to repeat the process for every image sequence if particular image-dependant operations have been performed.

2.4 *Storage of Radiance Maps*

The previously discussed techniques are based on the same principles. Results in building radiance maps are very similar, expecially for the two more accurate techniques. In terms of recovered dynamic range, some differences can sometimes be observed, mainly due to the specific software implementation. For the sequence of Fig. 1, the recovered range is somewhere near 1:6000. Furthermore, the pixel values obtained after the radiance map recovery are usually expressed as floating point values; to store these numbers without losing any information, a floating point pixel map could be used. But this is an expensive solution, since 32*3 bits for each pixel of an RGB image are required (assuming 32-bit floating values). That is why a good memory-saving image format is described in Ref. 10, where color triplets are represented using a mantissa/exponent format requiring only 32 bits (8 for the common exponent+24 for the channels mantissas). Radiance image files can also be useful to simulate a kind of virtual camera, in the sense that, after having saved the maps, the acquisition process can be simulated by using the recovered response curves. Thus, one can imagine an application where the user is allowed to rephotograph the scene whenever he wants by using the most satisfying exposure. Some industries are starting to move in this direction.[11] In the next paragraph an example of this is used and shown.

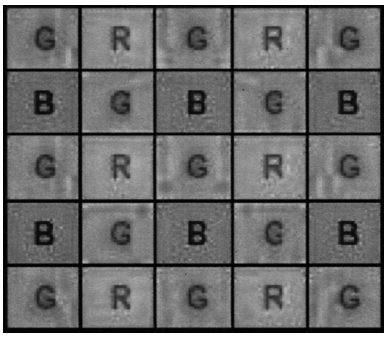

Fig. 4 The Bayer pattern.

2.5 *Some Notes for Radiance Map Recovery on Digitally Acquired Images*

The raw output provided by a CCD image sensor represents an image where neither gamma correction nor other algorithms have been applied yet. In this case, camera response is almost linear and the aforementioned simple parametric model fits very well. We use some images acquired with CMOS sensors to perform experiments. Digital still cameras use CFA(color filter array) sensors to capture images. A CFA is typically made up of an array of 2×2 elements tiled across the sensor array (see Fig. 4). The number of green pixels is greater than the number of red and blue pixels because of the higher sensitivity of the human eye to green light. This filter is known as the Bayer pattern,[12] and the final color picture is obtained by using a color interpolation algorithm that joins together the information provided by the differently colored adjacent pixels. In our experiments, we use only two images for building radiance maps and then contrast reduction is applied. The main goal of such experiments is the possibility of using the described techniques to obtain a ''nice looking'' picture under difficult light conditions (back or front lighting), using as few pictures as possible. The expression ''nice looking' picture'' emphasizes the fact that we are not really interested in producing physically correct radiance maps and renderings, but images that exhibit less washed out details (in the sense of saturation and noise in dark areas). We believe that for consumer DSC users, this would be the most appealing feature: the possibility of capturing images without worrying too much about lighting conditions. Because of the presence of the color filter array, the radiance map recovery is performed after color interpolation by using a STMicroelectronics patented algorithm, treating the output images as simple RGB data. Figure 5 (left) shows two images captured using a CMOS sensor, where the second one received eight times more light than the first, allowing the coverage of both high- and low-light details of the scene. The map was built by using the simple parametric formulas described before. Figure 5 (right) shows the cross-histogram of the two images and how well the estimated parameters interpolate the sample data. It is worth mentioning that CMOS sensors convey a lot of noise; thus the real cross-histogram between the two images exhibits a lot of outliers

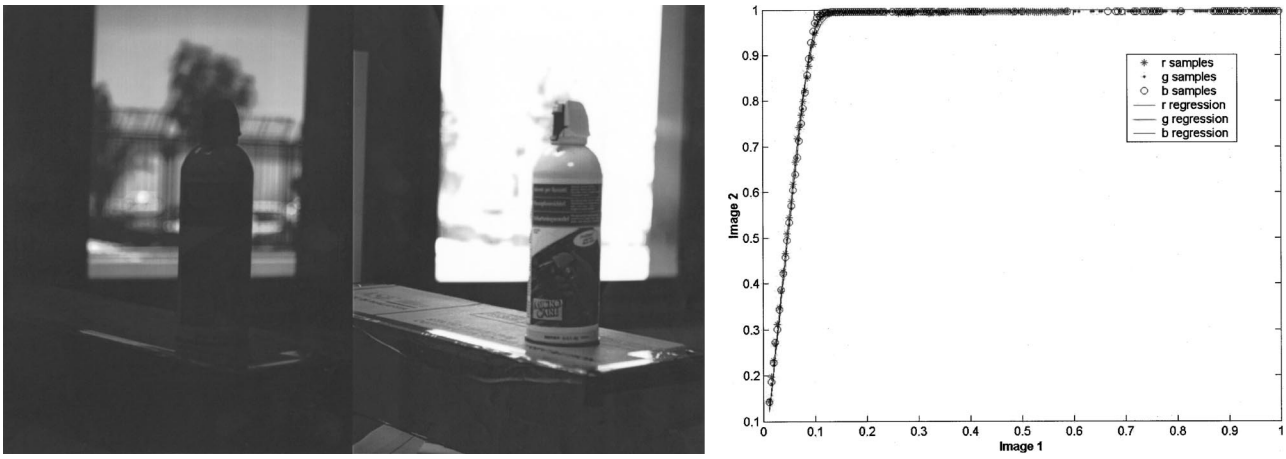

Fig. 5 Two images captured using a CMOS sensor with a relative exposure ratio of 8.8 (left) and cross-histograms with their interpolating curves (right).

from the regular plot shown in Fig. 5. In fact, in order to get rid of this noise, the histogram was regularized by discarding not well-populated entries and by pixel averaging. If a pixel p_1 in the first image corresponds to many pixels $p_{21}, p_{22} \ldots, p_{2n}$ in the second picture, then they are all collapsed in a single value P_2. To test, even if roughly, the accuracy of the recovered map, this was virtually photographed into one of the input images using Eq. (15). Figure 6 shows the map photographed in the second of the input images and the difference between the original and the replica (properly scaled for visibility purposes). A better alternative for radiance map construction consists of using the Bayer pattern data directly; this would prevent a lot of overhead computation. In this approach, the response curves are recovered from the three channels; an extended range Bayer pattern is generated and its range is compressed and quantized to 8 bits before color interpolation.

3 Contrast Reduction

It has been shown how it is possible to combine multiple exposures when creating high dynamic range radiance maps. The next step consists of mapping the computed radiance values, expressed in floating point precision, and have arbitrary dynamic range in the input range (typically [0:255]) of a common display device. Problems arise from the fact that usually CRT displays are capable of represent-

Fig. 6 A virtually photographed image (left) and the scaled differences between the original (right).

Fig. 7 Two images captured using a CMOS sensor with respectively short and long exposure times. The exposure ratio was not known but roughly estimated to be near 3.5.

ing a limited dynamic range (usually around 50, in terms of luminance values) which, furthermore, need to be quantized into a limited number of possible gray values. Various solutions have been proposed in the literature, some of them taking into account very sophisticated features of human perception to produce *ad hoc* solutions. Some of these methods, such as those described in Refs. 13 and 14 deal with data expressed in absolute units [so that the response of the human visual system (HVS) can be taken into account], which are usually not available to the common user. Thus, these methods cannot be applied to images rendered in fictitious raw units without any physical meaning. Generally speaking, the contrast reduction can be stated as the following quantization problem:

$$Q[\mathrm{Val}] = \lfloor (N-1)\cdot F(\mathrm{Val}) + 0.5 \rfloor \quad \text{with}$$

$$F:[\mathrm{LoVal}:\mathrm{HiVal}] \rightarrow [0:1], \tag{18}$$

where N is the final quantization level (typically $N=256$) and [LoVal:HiVal] is the input dynamic range. The simplest solution to this problem is to use a *linear mapping* such that the maximum value is mapped to 1:

$$F(\mathrm{Val}) = \frac{\mathrm{Val}}{\mathrm{HiVal}}. \tag{19}$$

Of course this kind of mapping is useless when the image has a really wide dynamic range: the resulting image will usually appear too dark. *Gamma-corrected linear mapping* is another widely used solution in computer graphics, and attempts also to compensate for nonlinear responses of display devices:

$$F(\mathrm{Val}) = \left(\frac{\mathrm{Val}}{\mathrm{HiVal}}\right)^{1/q} \quad q \in [1:3]. \tag{20}$$

Other simple methods are *exponentation* and *logarithmic mappings. Exponentation mapping* can be expressed as:

$$F(\mathrm{Val}) = \left(\frac{\mathrm{LoVal}}{\mathrm{HiVal}}\right)^{p/q} \quad p \in [0:1] \;\; q \in [1:3]. \tag{21}$$

A form of *logarithmic mapping* can be found in Ref. 15:

$$F(\mathrm{Val}) = \left[\frac{\log(1+p\cdot \mathrm{Val})}{\log(1+p\cdot \mathrm{HiVal})}\right]^{1/q} \quad p \in [0,\infty) \;\; q \in [1:3]. \tag{22}$$

The main problem with formulas such as Eqs. (21) and (22) concerns an appropriate selection of the involved parameters. To overcome this problem, a mapping that produces similar results, providing a way to automatically set the parameters, has been proposed in Ref. 16. The mathematical formulation of this operator can be expressed as:

$$F(\mathrm{Val}) = \frac{p\cdot \mathrm{Val}}{p\cdot \mathrm{Val} - \mathrm{Val} + \mathrm{HiVal}} \quad p \in [1,\infty). \tag{23}$$

The most appealing aspect of this method, besides its computational speed, is the automatic selection of the parameter p. In fact, Schlick has made the assumption that what really changes on a visualization device, when several viewing parameters are modified, is the M value of the smallest dark gray level that can be distinguished from black. Thus Eq. (23) should map the smallest nonzero floating-point value (LoVal) of our raw image to this value M. This gives:

$$p = \frac{M\cdot \mathrm{HiVal} - M\cdot \mathrm{LoVal}}{N\cdot \mathrm{LoVal} - M\cdot \mathrm{LoVal}} \approx \frac{M\cdot \mathrm{HiVal}}{N\cdot \mathrm{LoVal}}. \tag{24}$$

The techniques described so far are very simple and do not rely on the specific image content. A more adaptive technique has been proposed in Ref. 17. The relative adaptive histogram adjustment is part of a bigger and really impressive work, which also incorporates models for simulating human visual system limitations such as *glare*, *color sensitivity*, and *visual acuity*. However, histogram adjustment is the first step of the algorithm, and also the only one that can be implemented using arbitrary unit-expressed data. The process starts by computing a small image, the *fovea* im-

Fig. 8 From left to right, linear gamma corrected, Schlik, histogram adjusted, Retinex compressed rendering for the radiance map recovered for the sequence of Fig. 1. The radiance map is built using the method described in Ref. 6.

age, where each pixel represents one degree of the visual area that will be used to determine adaptation levels at each point. After having converted all luminance values into brightness values (by simply taking their logarithm), an histogram is built, where values between minimum and maximum bounds $L_{w\,\min}$, $L_{w\,\max}$ (of the raw input image) are equally distributed on the *log* scale. The cumulative distribution function $P(b)$ (where b is a generic bin entry) is defined in the usual way as:

$$P(b)=\frac{\Sigma_{b_i<b} f(b_i)}{T}, \tag{25}$$

where T is the total number of samples and $f(b_i)$ is the frequency count for bin i. The derivative of this function can be expressed as:

$$\frac{dP(b)}{db}=\frac{f(b)}{T\Delta b}, \tag{26}$$

where Δb is the size of each bin. Applying a histogram equalization on the raw input, the output becomes an image where all brightness values have equal probability. The equalization formula, which always gives us a way to map luminance values to display values, can be expressed as:

$$B_d=\log(L_{d\,\min})+[\log(L_{d\,\max})-\log(L_{d\,\min})]\cdot P(B_w), \tag{27}$$

where B_d is the output display brightness, $L_{d\,\min}$ and $L_{d\,\max}$ are minimum and maximum display luminance values, and B_w is the incoming brightness value. The main problem with histogram equalization is that it easily exaggerates contrast in those regions that are highly populated. To prevent this, a ceiling procedure is applied on the histogram, such that the contrast produced by equalization will never be greater than those obtained using a linear mapping operator. The ceiling can be written as:

$$\frac{dL_d}{dL}\leqslant\frac{L_d}{L}. \tag{28}$$

And from Eqs. (26) and (27) it can be easily obtained:

$$f(b)<=\frac{T\Delta b}{\log(L_{d\,\max})-\log(L_{d\,\min})}. \tag{29}$$

When the frequency count of some bin exceeds Eq. (29), it is simply cut off, and the procedure is repeated until the number of the corrected bins is under some defined threshold. Figure 8 shows some renderings obtained using some of the described techniques applied on the map constructed by starting from the sequence of Fig. 1. Figures 9 and 10 show some mappings for the pair of Figs. 5 and 7, where it can be seen that both low- and high-light information is present, producing a complete rendering of the scene that

Fig. 9 From left to right, linear gamma corrected, Schlik, histogram adjusted, Retinex compressed rendering for the radiance map recovered (using the parametric model[8,9]) from the two images of Fig. 5.

Fig. 10 From left to right, linear gamma corrected, Schlik, histogram adjusted, Retinex compressed rendering for the radiance map recovered (using the parametric model[8,9]) from the two images of Fig. 7.

would seldom be possible to acquire using a single shot. Other computationally more expensive solutions, aimed to reduce the contrast of radiance maps not suitable for real-time implementations, have been proposed in Refs. 18, 19, and 20. These techniques are not global remapping functions, as those described before. Instead, they act locally and differently depending on the spatial locations of values in the image. One of the first works stressing the need for local operators was Ref. 21. In Ref. 18 the main idea that we describe briefly consists of mimicking the way in which a skilled artist acts when representing a high-contrast scene: first he draws the main features and boundaries, then he adds all fine details. Low curvature image simplifiers (LCIS) are a set of differential equations that govern diffusion of fluids across solids and that are based on the work of Perona and Malik.[22] Each equation is driven by a different conductance threshold k to produce a set of different details of the same image. Large values for k produce a simplified image where only the main features are present, while small values produce more detailed images. Once the set of different detailed images has been generated, the mentioned artist-like approach is inverted by recovering select details from the input map and subtracting smoothed images from more detailed ones. At this point, the main features are strongly compressed by multiplying the logarithmically expressed input values with a scaling factor, while fine details are only slightly compressed (using bigger scaling factors). Eventually, all differently scaled details are summed together, exponentiated, and mapped to final display values using a linear or gamma corrected mapping to $[0{:}255]$. The main advantage of this technique is related to the possibility of exhibiting all details contained in a scene without losing contrast (as opposed to what usually happens using, for example, a logarithmic compression in the presence of very wide dynamic ranges), while drawbacks consist of the unnatural final aspect of the picture and high computation times (the anisotropic diffusion equations are applied iteratively on the input image). On the other hand, the correct choice of the various parameters involved in the algorithm is very difficult, leading easily to anomalous noise enhancement and *halo* artifacts. Figures 11 and 12 shows the detailed decomposition of a classical image obtained from Ref. 23 that exhibits an enormous dynamic range, and its LCIS rendering compared with a logarithmic compressed one. The last rendering technique that we mention here consists in taking advantage from the Retinex dynamic range compression feature. The multiscale Retinex mathematical formulation (we are considering the center/surround version of Refs. 19 and 20) is:

$$R_i(x,y)=\sum_{n=1}^{N} w_n\cdot\{\log[I_i(x,y)]-\log[F_n(x,y)*I_i(x,y)]\}, \tag{30}$$

where $I_i(x,y)$ is the i'th color channel,$*$ denotes convolu-

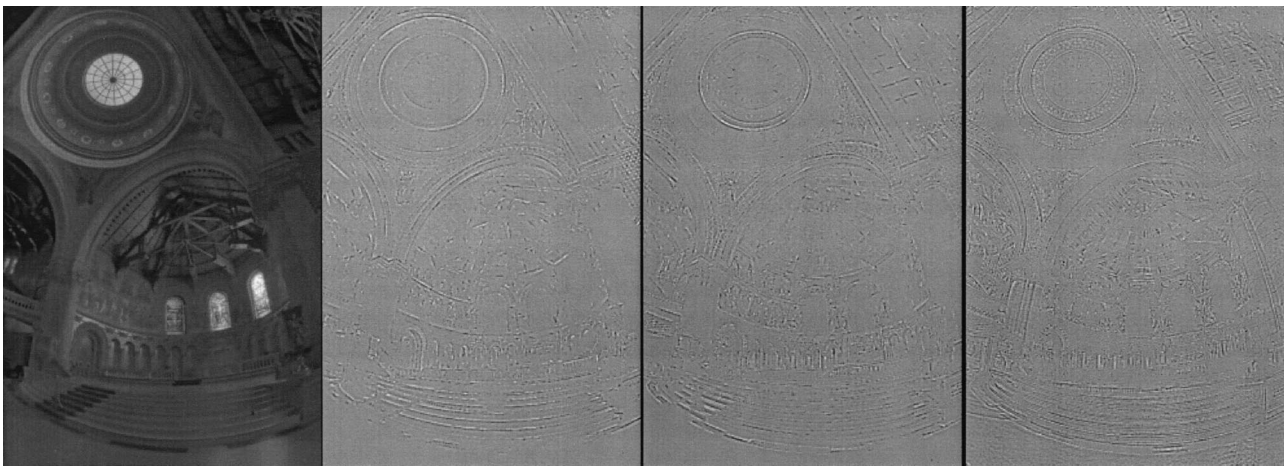

Fig. 11 Base (simplified) image and three detail levels of LCIS decomposition. Images have been properly scaled to achieve visibility and appear in gray scale, since LCIS is applied only on the Y (luminance) channel.

Fig. 12 LCIS rendering (left) and logarithmically compressed rendering. Fine texture details have been preserved for the left rendering. The radiance map was obtained from Ref. 23.

tion, and w_n and $F_n(x,y)$ are the weight of the n'th scale and the n'th Gaussian surround function, respectively. Each scale represents a different *Gaussian* function of different-sized space constant. The final RGB values are obtained by remapping the computed $R_i(x,y)$ values to [0:255] after doing some clipping on their histogram distribution. The main drawbacks of Retinex concern the wide kernels of the surround functions involved in the convolution (which require moving into the Fourier space) and the optimal choice of parameters (the number of scales, the dimensions of space constants, the optimal remapping of the final obtained values, etc.). Examples of Retinex renderings are shown in Figs. 8, 9, and 10, where parameters were manually tuned until images looked satisfactory. The rightmost image of Fig. 9 shows also the *color constancy* feature of Retinex. Recently, another very promising algorithm acting on similar principles of LCSI (but avoiding many of artifacts) has been described in Ref. 24 where image gradients of large magnitude are selectively compressed using an attenuating function. Solving a differential equation from the modified gradients generates the final image. Drawbacks consist, once again, of the high computational complexity required to solve the main differential equations and the computation of an attenuation function based on multiresolution decomposition.

4 Conclusions and Future Works

Various techniques dealing with expanding the dynamic range of images are described. All techniques rely on using differently exposed pictures of the same scene to enhance the usual 8 bit depth representation of pictures. The improvement is obtained by estimating (or using a preestimated) a response curve and then building a floating point radiance map that gives an approximate reconstruction of how the original scene looked before the acquisition phase. The map needs to be contrast reduced for viewing purposes on a common CRT display. Various possible requantization techniques are also described. Some of them are very simple and could be easily implemented in real-time environments, while others are computationally more expensive. Such techniques can be useful when trying to obtain a good picture under difficult light conditions using typical consumer DSCs. Further investigation is needed to adapt some of the presented techniques to work directly in the *Bayer pattern* domain; such an implementation allows the use of image data before color interpolation takes place, as opposed to what is done for the images of this article. In addition, if the input scene is critical, it is possible to automatically decide when and how, in terms of exposure settings, to acquire the frames. Automatic alignment of images should also be included (some sensor manufacturers[25] are starting to produce devices capable of acquiring in ''one shot'' both underexposed and overexposed pictures without any misalignment). The basic ideas behind these techniques introduce exciting possibilities for providing more range detail to the final user and a better approximation of the real scene that could also be saved and virtually photographed. Dynamic range extension using differently exposed pictures could also be achieved by using techniques based on different principles such as the use of *multiresolution fusion*. Interesting ideas have been proposed in Ref. 26.

Acknowledgments

The studio sequence used in this work is courtesy of M.A. Robertson, coauthor of another dynamic range extension technique based on similar principles of the described algorithms, not reported here due to lack of space. Details can be found in Ref. 27.

References

1. B. Long, *Complete Digital Photography*, Charles River Media, Hingman, MA (2001).
2. S. Battiato and M. Mancuso, ''An introduction to the digital still camera technology,'' *ST J. Syst. Res. Special Issue on Image Processing for Digital Still Camera*, **2**(2), 2–9 (Dec. 2001).
3. S. A. Bhukhanwala and T. V. Ramabadran, ''Automated global enhancement of digitized photographs,'' *IEEE Trans. Consum. Electron.* **40**, 1–10 (1994).
4. G. Messina, S. Battiato, M. Mancuso, and A. Buemi, ''Improving image resolution by adaptive back-projection correction techniques,'' *IEEE Trans. Consum. Electron.* **48**(3), 409–416 (Aug. 2002).
5. C. A. Poynton, *A Technical Introduction to Digital Video*, Wiley and Sons, New York (1992).
6. P. Debevec and J. Malik, ''Recovering high dynamic range radiance maps from photographs,'' *Proc. ACM SIGGRAPH*, pp. 369–378 (1997).
7. T. Mitsunaga and S. K. Nayar, ''Radiometric self calibration,'' *Proc. IEEE Conf. Computer Vis. Patt. Recog.* **1**, 374–380 (1999).
8. S. Mann, ''Comparametric equations with practical applications in quantigraphic image processing,'' *IEEE Trans. Image Process.* **9**, 1389–1406 (2000).
9. S. Mann and R. W. Picard, ''Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures,'' *Proc. IS&T 46th Annual Conf.*, pp. 422–428 (1995).
10. G. W. Larson, *Real Pixels, Graphic Gems II*, pp. 80–83, Academic Press, New York (1992).
11. Spheron VR, Digital Camera Solution, see http://www.spheron.com/ (2003).
12. B. E. Bayer, ''Color imaging array,'' U.S. Patent No. 3,971,065 (1976).
13. J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg, ''A model of visual adaption for realistic image synthesis,'' *Computer Graph. Proc. ACM SIGGRAPH*, pp. 249–258 (1996).
14. J. Tumblin and H. Rushmeier, ''Tone reproduction for realistic images,'' *IEEE Comput. Graphics Appl.* **13**, 42–48 (1993).
15. J. Blinn, ''Dirty pixels,'' *IEEE Comput. Graphics Appl.* **9**, 100–105 (1989).
16. C. Schlik, ''Quantization techniques for visualization of high dynamic range pictures,'' *5th Eurographics Workshop on Rendering*, pp. 7–18, (1994).
17. G. W. Larson, H. Rushmeier, and C. Piatko, ''A visibility matching tone reproduction operator for high dynamic range scenes,'' *IEEE Trans. Vis. Comput. Graph.* **3**, 291–306 (1997).
18. J. Tumblin and G. Turk, ''LCIS: A boundary hierarchy for detail-

preserving contrast reduction,'' *Proc. ACM SIGGRAPH*, pp. 83–90 (1999).
19. D. J. Jobson and Z. Rahman, ''Properties and performance of a center/surround Retinex,'' *IEEE Trans. Image Process.* **6**, 451–462 (1997).
20. D. J. Jobson and Z. Rahman, ''A multiscale Retinex for bridging the gap between color images and the human observation of scenes,'' *IEEE Trans. Image Process.* **6**, 965–976 (1997).
21. K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman, ''Spatially non uniform scaling functions for high contrast images,'' *Proc. Graph. Interface 93*, pp. 182–191 (1993).
22. P. Perona and J. Malik, ''Scale-space and edge detection using anisotropic diffusion,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 629–639 (1990).
23. P. Debevec home page, http://www.debevec.org (2003).
24. R. Fattal, D. Lischinski, and M. Werman, ''Gradient domain high dynamic range compression,'' *ACM Trans. Graph. (Proc. ACM SIGGRAPH 2002)* **2**(3) 249–256 (July 2002).
25. STMicroelectronics, Imaging Division, see http://www.edb.st.com (2003).
26. L. Bogoni and M. Hansen, ''Pattern-selective color image fusion,'' *Pattern Recogn.* **34**, 1515–1526 (2001).
27. M. A. Robertson, S. Borman, and R. L. Stevenson, ''Dynamic range improvement through multiple exposure,'' *Proc. IEEE Int. Conf. Image Process. (ICIP-99)*, Vol. 3, 159–163, Kobe (1999).

S. Battiato received his Italian degree in computer science in 1995 and PhD in applied mathematics and computer science in 1999, both at Catania University. Since November 1999 has been working at STMicroelectronics in the Advanced System Technology (AST) Catania Laboratory in the Imaging Group as a team leader. He is the author of several papers and patents.

A. Castorina received his degree in computer science in 2000 at the University of Catania doing a thesis about watermarking algorithms for digital images. Since September 2000 he has been working at STMicroelectronics in the AST Digital Still Camera Group as a system engineer. His current activities include exposure compensation and high dynamic range imaging.

M. Mancuso received the degree in electronics engineering from the University of Palermo in 1990. After a research period at the University of Palermo in the image processing field, he joined STMicroelectronics. At presents he work as Program Manager with the Advanced System Technology team in Catania, Italy. He is the author of several papers and patents.