

Why I Want a Gradient Camera

Jack Tumblin, Amit Agrawal, Ramesh Raskar

TR2005-083 June 2005

Abstract

We propose a camera that measures static gradients instead of static intensities. Quantizing sensed intensity differences between adjacent pixel values permits an ordinary A/D converter to measure detailed high contrast (HDR) scenes. We measure alternating 'cliques' of sensors (small groups) that locally determine their own best exposure and reconstruct the image using a Poisson solver. This intrinsically differential design suppresses common-mode noise, hides and smoothes quantization, and can correct for its own saturated sensors. Simulations demonstrate these capabilities in side-by-side comparisons.

IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Why I want a Gradient Camera

Jack Tumblin
Northwestern University
jet@cs.northwestern.edu

Amit Agrawal
University of Maryland
aagrwal@umd.edu

Ramesh Raskar
MERL
raskar@merl.com

Abstract

We propose a camera that measures static gradients instead of static intensities. Quantizing sensed intensity differences between adjacent pixel values permits an ordinary A/D converter to measure detailed high contrast (HDR) scenes. We measure alternating ‘cliques’ of sensors (small groups) that locally determine their own best exposure, and reconstruct the image using a Poisson solver. This intrinsically differential design suppresses common-mode noise, hides and smoothes quantization, and can correct for its own saturated sensors. Simulations demonstrate these capabilities in side-by-side comparisons.

1 Introduction

Conventional digital cameras imitate the film cameras they intend to replace, and imitate their limitations as well. Can you think of any scene in which the picture from a digital camera is qualitatively better than film? Both share grainy-looking noise, but film noise has no grid-like structure. Both can suffer over- and under-exposure, but film’s asymptotic response compresses shadows and highlights gracefully, while an A/D converter imposes abrupt intensity limits. Noise obscures subtle intensity changes in both cameras, but film does not discard these changes in a quantized, piecewise-constant approximation. What can we do to make digital images better than film instead of worse?

We advocate a digital still camera that measures static gradients of log-intensity to better capture the visual appearance of stationary scenes. Instead of direct pixel intensity measurements as output, we propose a camera that selectively measures only the differences between adjacent pixel pairs. The sensor’s output is not directly displayable, but must be reconstructed using a Poisson solver as shown in Figure 5. However, the solver offers further advantages in noise hiding and error correction. This approach bypasses several limitations that digital cameras share with photographic film, and can provide new capabilities such as:

Fast, Sensitive, Wide-ranging Response: high contrast

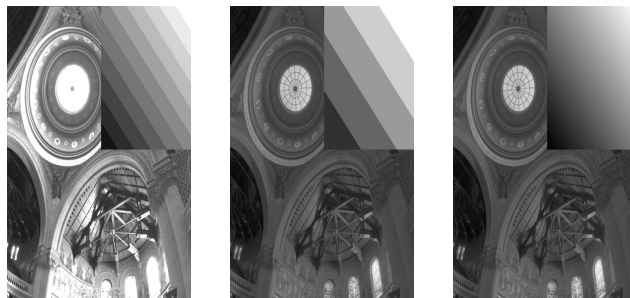


Figure 1. A log-gradient camera captures both large and small scene contrasts well: (a) an 8-bit intensity camera loses large contrasts ($> 10^{+5} : 1$) at A/D limits, but small contrasts (inset: $1.33 : 1$ ramp, simulated) are finely quantized (11 levels); (b) an 8-bit log-intensity camera captures large contrasts well, but small contrasts are coarsely quantized (5 levels); (c) an 8 bit log-gradient camera preserves both, hiding errors smoothly everywhere.

or ‘high dynamic range’ (HDR) scenes photographed without saturation or a long sequence of exposures, yet without loss of visible details. By measuring only the difference between adjacent pixels, we can reduce the A/D dynamic range for finer quantization, yet recover images whose aggregate dynamic range is many times larger.

Hidden Quantization, Smoother Noise: Quantization errors in gradients are subtle; as few as 3-4 bits appear equivalent to 8 bits of intensity quantization, whose errors cause spurious step-like edges. Poisson solvers reconstruct gradient noise as a low-resolution ‘cloudy’ error that may be less likely to mask visually important edges.

Correctable Saturation: sensor saturation is not correctable in conventional cameras, but errors from saturated gradient sensors are removable by correcting for the curl of the sensed gradient field.

In addition, several gradient-based image manipulation techniques can be provided as in-camera effects, such as eye-blink removal and composited focus [1].

1.1 Previous Work

Many researchers have sought replacements to conventional intensity-sensing cameras, particularly for com-

puter vision applications. For 15 years, ‘smart sensors’ have augmented photo-sensors with local processing for tasks such as edge detection, motion sensing and tracking. Mead’s *silicon retina* and *adaptive retina* [15] chips were among the first to mimic vertebrate retina computations, and inspired many later efforts [16]. For example, in Mitsubishi’s *artificial retina* [9] each photodetector’s sensitivity was controllably modulated by others nearby to avoid saturation and aid in fast edge detection. Brajovic et al. [2] developed a camera that creates reflectance images by removing estimated illumination to reduce dynamic range. This and other ‘smart sensors’ are intended for low-level vision tasks. However, these cameras usually did not create conventional images.

More recently, high-speed CMOS sensors enabled **multiple capture single image** (MCSI) methods [26]. These cameras locally merge a rapid sequence of image measurements into one output image. For example, *Ranger* cameras from Integrated Vision Products [10] coupled pixels directly to a SIMD processor array, and user microcode operated on all pixels simultaneously [11]. The *Sony ID-Cam* sensed high-frequency optical codes from strobing LEDs, and used MCSI to reject ambient light [14]. The clique-adaptive gradient camera we propose fits the MCSI classification, but the sensor interactions are simple, we use very few capture steps (e.g. 2 or 4), and we require only modest on-chip processing power.

Reconstructing an image from its gradients by **Poisson solvers** has already proven useful for tone mapping [6], shadow removal [7] and novel forms of image editing [5, 21]. Noting that humans underestimate large gradients, Fat-tal et al. [6] devised a tone mapping scheme that compressed them. Earlier, Elder [5] explored image editing and encoding by oriented variable-sharpness edges. Tappen [24] saw that image gradient histograms peaked strongly at zero, and devised a Poisson solver-based method for super-resolution (‘de-mosaicing’) for interleaved color sensors.

Many ingenious **high dynamic range** (HDR) photography methods merge multiple mutually-aligned images with different exposure settings [13, 12, 4], or have varied or self-adjusting gain. For example, the Smal [23] camera’s AutoBrite self-adapting method reduced out-of-range contrasts before measurement, while Dalstar [3] and others extended A/D measurement abilities and reduced noise. A novel asynchronous binary camera measured wide-ranging intensity by variable pulse rate [27]. Nayar et al. [17] has proposed a suite of HDR techniques that included spatially-varying exposures and adaptive pixel attenuation, and micro-mirror arrays to re-aim and modulate incident light on each pixel sensor Nayar [18]. **Logarithmic intensity cameras** also avoid saturation well [8], but their increased quantization error and noise can hide small contrasts.

Unlike these specialized cameras, we propose a general purpose camera suitable for applications from family photographs to robotic vision to tracking and surveillance. It offers reduced quantization error, differential on-chip signalling and increased sensitivity to small contrasts in ordinary photos, but retains these abilities for HDR photography and includes HDR error correction.

2 Measurement Methods

Intensity Cameras: Film emulsions and most existing digital cameras measure static intensities best. They capture the approximate time-average of scene intensities by:

$$I_d(m, n) = (kI_s(m, n))^\gamma, \quad (1)$$

where $0.0 \leq I_d \leq 1.0$ is the normalized display value at pixel (m, n) ; I_s is the sensed light energy at the pixel; k is ‘exposure’ (e.g. gain, light sensitivity or ‘film speed’) and γ is the contrast sensitivity. Typically $\gamma \cong 1$ for CCDs, and $\gamma < 1$ will compress contrasts just as $\gamma > 1$ exaggerates them. Writing (1) in logarithmic units where differences directly correspond to contrast ratio

$$I_{ld} = \log(I_d) = \gamma(\log(I_s) + \log(k)) \quad (2)$$

reveals that γ is a scale factor for contrast, and exposure k is a simple offset in log units. Conventional cameras keep k and γ uniform, because pixel-to-pixel variations in k and γ have strong effects on the appearance of the captured image. A/D conversion of I_d makes integer pixel values for display and printing.

Most digital cameras are quasi-linear. The digitized I_d values are intended to be directly proportional to scene intensity I_s , but most include some contrast compression (e.g. $\gamma = 0.45$) to compensate for the contrast exaggeration of most computer displays (e.g. $\gamma = 2.2$). Specifying k by the I_s that causes $I_d = 1.0$ (e.g. ‘display white’) avoids complications from limited display device contrast.

The γ and A/D resolution set an upper bound on the contrast-capturing abilities of a quasi-linear camera. With 2^b uniform quantization levels for I_d , fixed k and fixed γ , the largest ratios of scene intensities the camera can capture is $C_{max} = I_s^{max}/I_s^{min} = 2^{-b/\gamma}$, i.e. $(\log(1) - \log(2^b)) = \gamma(\log(I_{smax}) - \log(I_{smin}))$. Just as with film, many photographically interesting scenes contain contrasts that are far too large for most A/D converters. Users must choose to lose visible scene features either to glaring white or to featureless black.

Log responding cameras measure the logarithm of intensity (e.g. $I_{ld} = \log(I_d) = \gamma(\log(I_s) + \log(k))$). They use quantization steps of equal contrast instead of equal intensity and record higher-contrast scenes, but only by enlarging their quantization steps. If quantization step size

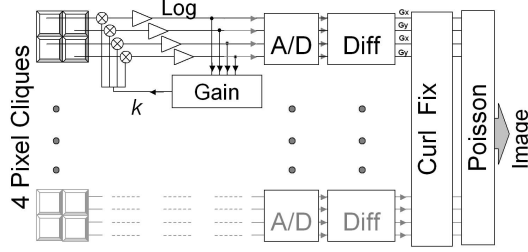


Figure 2. Log-gradient camera overview: intensity sensors organized into 4-pixel cliques share the same self-adjusting gain setting k , and send $\log(I_d)$ signals to A/D converter. Subtraction removes common-mode noise, and a linear ‘curl fix’ solver corrects saturated gradient values or ‘dead’ pixels, and a Poisson solver finds output values from gradients.

is less than about 1-2%, then Fechner’s Law [19] ensures that the camera will capture any visually detectable intensity differences, but coarser steps may lose visible scene details. Unfortunately, an 8-bit log-intensity camera with 1% quantization steps can’t even measure a 13:1 contrast range ($1.01^{255} = 12.6$). The $10^{+5} : 1$ contrast of Figure 1 requires quantization steps 4.6% or worse, and the 8-bit log intensity camera will lose many subtle but visible details to quantization. The log-gradient camera we propose (Figure 2) avoids this tradeoff; we get fine quantization *and* high contrasts.

2.1 Logarithmic Gradient Measurement

Rather than fitting contrast quantization to whole image, the clique-adaptive camera we propose finely quantizes only the differences between adjacent pixels. Even very limited contrasts (e.g. 8-bit, 1% step, $< 13 : 1$) between pixels are sufficient to describe any image, including even the most extreme HDR scenes, because a) the lens’ spatial impulse response (PSF) limits log-intensity gradient magnitude; b) large contrasts often span many pixels, and c) we can correct any out-of-range measurements that are surrounded by correctly-measured neighbors (see Section 3).

Differences between adjacent pixel values form a discrete approximation of the image gradients ∇I_d . Specifically, for a pixel at integer locations (m, n) we define $lg_x(m, n)$ and $lg_y(m, n)$ as the log-intensity forward differences:

$$(lg_x(m, n), lg_y(m, n)) \cong \nabla I_d(m, n) \quad (3)$$

Although we could have used forward differences of intensity rather than log intensity, I_d measurement is simpler because exposure value k disappears. Using (2):

$$\begin{aligned} lg_x(m, n) &= \log(I_d(m+1, n)) - \log(I_d(m, n)) \\ &= \gamma(\log(I_s(m+1, n)) - \log(I_s(m, n))) \\ lg_y(m, n) &= \gamma(\log(I_s(m, n+1)) - \log(I_s(m, n))) \end{aligned}$$

In this form, each gradient estimate lg_x and lg_y is computed from two locally-adjusted intensity detectors. As long as both detectors use the same k value, k has no effect; pairs of sensors can locally and independently regulate themselves to best avoid sensor saturation. While earlier, neurally inspired sensors also used self-regulated gain $k(m, n)$, it varied at each pixel, irretrievably discarding low-frequency image content. Gradient sensors that share k avoid this loss.

Unfortunately, shared k values pose a conundrum. To measure any forward difference accurately, two pixels must find a shared k value that will avoid saturation for both I_d values. More formally, our A/D converter can only measure values between I_d^{min} and I_d^{max} , and the shared k value acts as a shared offset chosen to fit both pixel values within the A/D input range. However, each sensor m, n is part of four separate forward differences to the right, left, top and bottom. Choosing k for one sensor pair means the other three pairs connected to m, n must also share that k . These pairs, in turn, must share the same k with their neighbors, and by induction the entire camera is forced to share a single k value. Without k variation we cannot measure HDR scenes, yet we cannot vary k if we measure all forward differences simultaneously.

2.2 Measurement Cliques

The ‘clique-adaptive’ design we propose solves the conundrum by splitting forward difference measurements into two or more disjoint sets, each measured in turn. For example, suppose we measured only the horizontal forward differences $lg_x(m, n)$ for all pixels with even-numbered m . No other forward difference shares the same pixel, so each pixel pair can now choose its own best k value independently. An undirected graph describes it well: make a graph node for each pixel (m, n) , and draw an edge to connect the pixels in each forward-difference measurement. The graph partitions all pixels into cliques (fully-connected subgraphs) and each clique ‘adapts’ to its own best k value. Three more sets of cliques can complete the image measurements. An obvious solution makes a second clique set from $lg_x(m, n)$ for all odd-numbered m , and a third and fourth clique set from $lg_y(m, n)$ for all even- and odd- numbered n respectively.

Generalizing this approach leads to a wide variety of clique-adaptive sensor designs. For any design, begin by partitioning the detector grid into a first set of cliques—small disjoint groups of adjacent pixels. Each clique finds the k value to avoid saturation of any of its pixel members, and we only measure forward differences within a clique. In this way, each clique is similar to a tiny auto-exposure camera with only a few pixels. Finally, design one or more additional clique sets to ensure each forward difference in the image at least once.

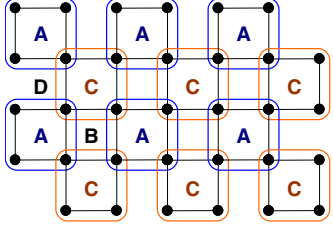


Figure 3. ‘Box clique’ photosensor groups: each 4-sensor clique adjusts to measure local intensities, alternating between the A and C clique sets.

Figure 3 shows the design we chose to simulate, as this checkerboard-like arrangement of 4-pixel cliques seems well suited for hardware implementation. Each clique is a square of 4 adjacent pixels, and two clique sets A, C measure all forward differences once. Each photo-detector selects between two local k values, yet an $M \times N$ pixel sensor must find only $\frac{MN}{2}$ separate k values. In Figure 2, we give an overview of the proposed log-gradient camera. Note that the A/D converter does not measure (l_{g_x}, l_{g_y}) directly, but instead measures I_{ld} for each clique member and then subtracts the results digitally. This may initially seem unwise because it doubles quantization noise, but this approach permits asynchronous measurement of clique members, and ensures that (l_{g_x}, l_{g_y}) and any diagonal links measured within a clique will have zero curl (see Section 3). It also permits the A/D converter to keep $I_{ld}^{min}, I_{ld}^{max}$ input limits fixed, and keep all measured analog signals positive-valued.

Measurement by cliques can also improve common-mode **noise rejection**. Many existing image sensor chips transfer intensity-indicating signals as an analog voltage, current or charge to A/D converters located away from the image-sensing area of the chip, and this long-distance transfer is susceptible to noise, cross-talk from nearby digital switching circuits and external EMI/RFI. Differential signalling improves noise immunity by sending a signal and its negative $(+S, -S)$ along two adjacent paths; unwanted ‘common mode’ noise N that invaded both pathways is removed at the receiver by subtraction: $(N + S) - (N - S) = 2S$. Sending each of the four clique member’s I_{ld} signals along adjacent paths can provide common mode noise rejection without new signal pathways, because unwanted signals are cancelled by the subtraction used to compute the clique’s four (l_{g_x}, l_{g_y}) values.

2.3 Reconstruction Methods

Reconstruction from gradients amounts to solving a Poisson equation [6]. Specifically, we wish to recover 2D log intensity I_{ld} whose gradients I_x and I_y are close to the sensed gradients l_{g_x} and l_{g_y} in least-squares sense. This

amounts to minimizing the following functional:

$$J(I) = \int \int (I_x - l_{g_x})^2 + (I_y - l_{g_y})^2 dx dy \quad (4)$$

The Euler-Lagrange equation to minimize J is

$$\frac{\partial J}{\partial I_{ld}} - \frac{d}{dx} \frac{\partial J}{\partial I_x} - \frac{d}{dy} \frac{\partial J}{\partial I_y} = 0 \quad (5)$$

which gives the Poisson equation

$$\nabla^2 I_{ld} = \frac{\partial}{\partial x} l_{g_x} + \frac{\partial}{\partial y} l_{g_y} \quad (6)$$

where $\nabla^2 I_{ld} = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$ is the Laplacian. I_{ld} can then be obtained from I_{ld} . For solving the Poisson equation, we use a sine transform based method [22].

Dirichlet boundary conditions are a natural choice for image reconstruction, and require absolute intensity values around the periphery of the image. Instead of direct measurement of an image that may include very high contrasts, we propose to encircle the entire sensor array with a 1D ring of *periphery* sensors that encode only the difference from their periphery neighbors. Sensors inside the periphery then measure their differences from these sensor’s values, rather than zero-valued Dirichlet boundaries. Image reconstruction then proceeds in two steps:

1. Determine boundary intensity values using the periphery sensors alone (a 1D Poisson problem),
2. Combine these boundary values with (l_{g_x}, l_{g_y}) to solve for all interior pixel values (a 2D Poisson problem).

The unknown offset in the reconstructed 1D signal is then equivalent to a global exposure setting. However, if our 1D periphery sensor solution contains large errors, then these errors will propagate into the 2D interior solution.

3 Sensor Error Corrections

An HDR scene will easily saturate detectors in a conventional camera, but the consequences are simple; the output image will contain featureless regions of white or black. The opposite is true for our proposed gradient camera: each clique’s shared, self-adjusting k value greatly reduces the chances of any detector saturation, but when saturation occurs the erroneous gradients can severely disrupt the reconstructed image. Fortunately, these errors are both detectable and correctable.

3.1. Curl Correction

Independently adapted cliques greatly reduce the chance that any pixel within them will saturate. Image statistics

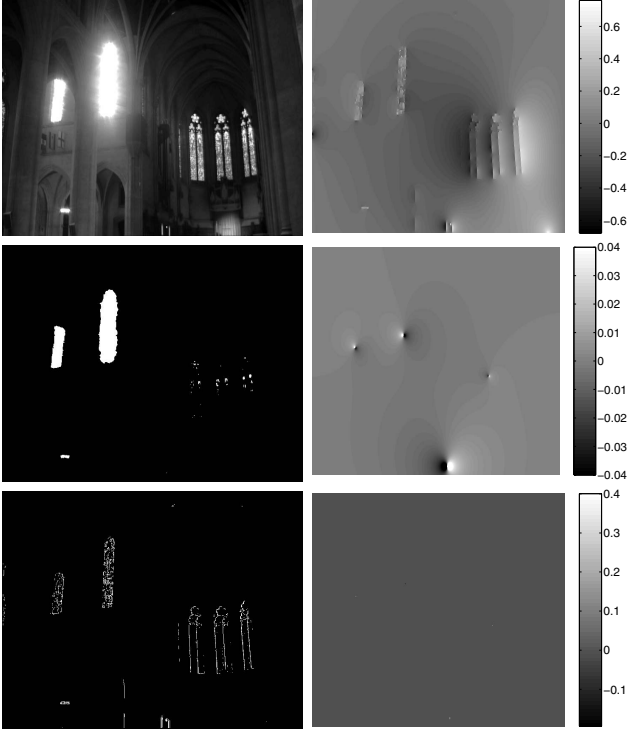


Figure 4. Sensor Error Correction: In the left column, an 8-bit intensity camera captures the Nave HDR scene [top], but loses 2.29% pixels to unmeasurable white; these pixels are marked white in [mid]. However, an 8-bit, 1% step log-gradient camera can measure all but 2788 or 0.41% of its gradients; these are marked white in [bott]. In the right column, [top] shows intensity error caused by reconstructing uncorrected gradients; the Poisson solver propagates these errors widely across the image. After curl correction, only 4 disjoint graphs remain, with a total of 28 unknown gradients, causing 4 small dimple-like errors in intensity [mid]. After disjoint graph correction [bot], error falls to 4 dots caused by underestimated offsets for the 4 graphs.

are also in our favor: images of natural scenes usually have power spectra that fall rapidly with spatial frequency. Relying on Pentland’s fractal analysis [20], Weiss [25] showed that derivative filters applied to such scenes produce sparse outputs with near-zero values almost everywhere. In addition, our ‘box clique’ design ensures that if a pixel sensor was saturated within one clique set, it may escape saturation when measured in another. None of these tendencies are guarantees, of course; as shown in Figure 4 occluded light sources and transparency can sometimes saturate pixels within cliques and corrupt their gradient measurements. We can detect corrupt gradients directly from saturated sensors, and then compute corrected values from the curl.

Image intensity, like any other 2D scalar function, defines a unique gradient vector field with zero curl, a conservative field where the integral of the gradient over any closed-loop path is zero. Any non-zero curl in the gradient

camera’s output always indicates an error in image sensing. These errors come either from unresponsive *dead pixels*, or from cliques whose contrast exceeds the A/D input range, forcing one or more pixels to out-of-range or ‘saturated’ values. Any gradient (lg_x, lg_y) made from a saturated pixel is incorrect, causing nonzero curl. On a discrete image, the curl C at pixel m, n is the sum of forward differences along the smallest closed path:

$$C(m, n) = (lg_x(m, n + 1) - lg_x(m, n)) - (lg_y(m + 1, n) - lg_y(m, n)) \quad (7)$$

and is the discrete equivalent of $lg_{xy} - lg_{yx}$. Nonzero curl at $C(m, n)$ is caused by one or more erroneous gradients along the closed square path $(lg_y(m, n), lg_x(m, n + 1), -lg_y(m + 1, n), -lg_x(m, n))$ that begins and ends at (m, n) . If the $lg_x(m, n)$ gradient is wrong, its error also causes a nonzero curl at $C(m, n - 1)$. Similarly, erroneous $lg_y(m, n)$ will also cause nonzero curl at $C(m - 1, n)$. To compute a corrected set of gradients we find the K pixels with nonzero curl, write a curl equation (7) for each one, and solve the system of equations to find replacement values for the erroneous gradients.

Begin with the left-hand-side of (7). Stack all the nonzero curl values C from the curl-afflicted pixels to form a $K \times 1$ vector \mathbf{C} , using lexicographical ordering of elements. The right-hand side of (7) becomes a sparse matrix A of constants times a vector \mathbf{x} . This vector contains all (lg_x, lg_y) from our collection of K (7) equations, but segregates the corrupted or unknown gradient values from the known, properly measured gradients: $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$. The upper part \mathbf{x}_1 is the $P \times 1$ vector of stacked lg_x and lg_y gradient measurements that we know are trustworthy; the lower part \mathbf{x}_2 is an $L \times 1$ vector holds the unknown (saturated) gradient measurements we wish to recover. The linear system of equations is:

$$A\mathbf{x} = A \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \mathbf{C} \quad (8)$$

where A is a $K \times (P + L)$ sparse matrix. Each row of A represents one of the K equations and will have only 4 non-zero values: two 1’s corresponding to $lg_x(m, n + 1)$ and $lg_y(m, n)$ and two -1 ’s corresponding to $lg_x(m, n)$ and $lg_y(m + 1, n)$. Partitioning A as $A^{K \times (P+L)} = \begin{bmatrix} A_1^{K \times P} & A_2^{K \times L} \end{bmatrix}$, lets us write

$$A_2\mathbf{x}_2 = c - A_1\mathbf{x}_1 \quad (9)$$

Thus $\mathbf{x}_2 = (A_2^T A_2)^{-1} A_2^T (c - A_1\mathbf{x}_1)$ and hence the saturated gradient measurements can be recovered, as demonstrated in Fig. 4.

This solution requires $\text{rank}(A_2) \geq L$, a condition easily met if erroneous gradients do not completely enclose an image feature. The next section describes a reasonable solution to the $\text{rank}(A_2) < L$ case as well.

3.2. Disjoint Graph Correction

The same formalism we used for cliques helps us describe gradient corrections for the $\text{rank}(A_2) < L$ case. Suppose we regard the entire image as a graph with pixel nodes and with forward-difference links. If we remove all links made from unreliable measurements that cause nonzero curl, then we may have partitioned the graph. If we did not, then $\text{rank}(A_2)$ will be equal to L , and curl correction can restore the missing links. However, if the graph separates into two or more sets of pixels, then the rank drops below L and curl correction fails.

For example, Figure 4 [left top] shows windows backlit by direct sunlight, causing extremely high gradients around the window edges. The [left bottom] image shows there are many invalid gradient estimates there, but not enough to completely enclose any window. After curl correction, only four tiny, very bright spots within the window remain, each forming its own disjoint graph [right middle]. We know that all gradients connecting the two graphs exceeded the A/D’s measuring ability, but we don’t know how much. Our solution is simple and *ad-hoc*: we choose just one broken link between the two disjoint graphs, and assign it a gradient value of one A/D quantizing level higher than the A/D converter’s maximum. Adding just one link to each disjoint graph reconnects it to the image. We then repeat the curl correction to construct values for all remaining unknown gradients. The Poisson solver reconstructs the zero-curl result whose error consists of an under-estimated offset for the formerly disjoint graphs [right bottom].

4 Applications and Results

4.1 HDR with Low-Contrast Details

Figure 5 shows simulation results for a log-gradient camera using the box-clique design of Section 2. We began with two copies of the floating-point HDR source image from [6], one for the A and C clique sets respectively. We simulated an 8-bit A/D with 1% quantization steps between $-\log(12.6) \leq I_{ld} \leq 0$, and whose midpoint is $mid = -\log(1.01^{128}) = -\log(3.57)$. For each clique of 4 log-intensity pixels, we chose the k value that ensures the average of all 4 pixels maps to the A/D midpoint mid , and then found an 8-bit value for each pixel in the clique. As the left side Figure 5 shows, both the A and C cliques produce values tightly clustered around 128, with a few edge-driven outliers. On the right, the simulated camera output I_{ld} faithfully reconstructs the original scene. Though difficult to illustrate on low-contrast printed paper, both Figures 1(c) and 5(c) show simulations of our proposed design reproduces high contrast (HDR) scenes quite well. Even scenes such as Figure 4 with directly visible light sources sharp-



Figure 5. Simulation of 4-pixel clique design on Synagogue HDR map. [top left] A image, [top right] C image, [bottom] reconstructed I_{ld} image

ened with in-focus occluders did not induce unrecoverable errors; in fact, scenes that generated disjoint graphs in our simulated camera were particularly difficult to find. Our results assumed very modest hardware consisting of an 8-bit A/D converter. For comparison, most consumer-grade digital cameras use 10, 12, or even 14-bit A/D converters, and typically provide a usable contrast range of $\cong 1000 : 1$. Figures 1(c) and 5(c) both depict scenes with contrasts $> 10^{+5}$, and would overwhelm these ordinary cameras.

4.2 Quantization Hiding

Even though the proposed gradient camera captures HDR images easily, its measurements and reconstructed output images usually have far less visible quantization error than a traditional intensity-measuring camera. Intensity cameras approximate images as functions with piecewise-constant values, where the A/D converter sets a fixed number of uniformly spaced levels. Gradient cameras approximate images as intensity functions with piecewise-linear values (piecewise-constant gradients) instead, with the number of describable gradients set by the A/D converter. In addition, the Poisson solver’s results do not always follow these strictly quantized gradients, but instead finds the image whose gradients best match the given values in the least squares sense.

As shown in Fig 6, reconstruction even from coarsely quantized gradients is still quite accurate, because quantization induces discontinuities only in the intensities second

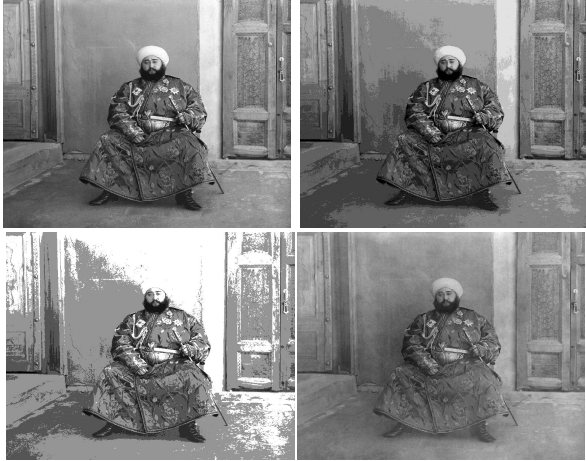


Figure 6. Gradient quantization is far less visible than intensity quantization [top row]: source image; source after 3-bit intensity quantization; bottom row: source after 3-bit log(intensity) quantization, source viewed by simulated 3-bit gradient camera output.

derivatives or higher. Figure 6 top row shows the original intensity image before and after uniform 3-bit intensity quantization, where the step-like discontinuities cause noticeable ‘contouring’ artifacts. If 3-bit quantization is applied to $\log(\text{intensity})$ for side-by-side comparison to the gradient camera, the contouring artifacts are even worse (lower left). However, a 3-bit gradient camera’s output (lower right) avoids these artifacts and is visually very similar to the original. While dramatic, this comparison is not entirely fair; the intensity camera measures only one value per pixel, but the gradient camera measures two (lg_x, lg_y). However, even a reduction to 2-bit quantization for (lg_x, lg_y) will enjoy the interpolation and smoothing provided by Poisson solvers and will approximate the original signal far more better than intensity quantization.

4.3 Motion Rejection

Several existing digital cameras now include stabilizers (e.g. Canon S1-is, Minolta A-2). These mechanisms help prevent image blurring caused by hand tremors during long time exposures. A gradient camera can also offer some blur prevention by combining multiple image capture with thresholds, as demonstrated in Figure 7.

The principle is quite simple. Any fixed but noisy camera viewing a stationary scene can reduce image noise by averaging a steadily growing set of new images. When the image at any pixel changes significantly, then the wisest choice is to stop averaging before the changes can cause blurring. Now suppose each pixel is allowed to choose its own stopping time. The result for a traditional intensity camera is only slightly different; a few pixels will continue averaging for slightly longer because they happen to follow

a constant-intensity path through the scene. However, the results are very different for a gradient-sensing camera. Instead of stopping when a pixel’s value changes, we stop averaging only when its gradient is different. Gradient sensors can ignore the uniform intensity changes caused by moving but smooth-shaded surfaces, averaging until it encounters a step-like or ridge-like discontinuity in intensity. In Figure 7 we simulate a motion-rejecting gradient camera where each gradient sensor stops averaging only when the incoming gradient differs from the average by more than the noise variance for new images.



Figure 7. Gradient cameras reject motion and reduce noise better than intensity cameras when used as a gated estimator. A gated estimator finds the average of a stream of input values, locks its value permanently if given an outlier. [Top Row] source scene; time-average of source image + Gaussian noise moved in a 5-pixel-wide circle in 15 frame-times. [middle row]: Gated intensity camera result; Gated gradient camera result. [bottom row]:time-to-lock (from 0 to 15) for intensity camera; time-to-lock for gradient camera. Both pixel-gated cameras avoided motion blur well, but intensity camera lock time was much shorter and the results show much more noise. Similar adjacent gradients are much more common than similar adjacent intensities. Please use zoom with your PDF browser.

Figure 7 shows promising results. On the first row, $+/- 5$ pixel (x, y) translation blurs the ordinary camera considerably. The middle image shows the intensity image result; as bottom row shows, most pixels stopped almost immediately (very dark), preserving sharpness but performing very little noise reduction. Back on the 1st row, the gradient camera result is both sharp and low noise, and the bottom row shows both a longer averaging time and more uniform distribution of this time across the image.

5 Summary and Conclusions

A gradient camera is similar to existing intensity cameras in electro-optical structure, but by measuring only the local changes in the image, we gain some significant advantages. It needs little or no exposure metering to capture high contrast scenes, hides effects of quantization well and distributes noise as low-frequency error rather than masking high frequencies.

However, the gradient camera requires significant computation to construct a displayable image, and may require extensive modifications to existing sensor designs. As the Poisson solvers required for reconstruction are demanding tasks to execute at interactive rates, it may be difficult to implement a real-time digital viewfinder for this camera. Similarly, our current proposal involves taking two measurements at each pixel with different, locally adjusted gains. This two-step exposure method may limit extensions of gradient cameras to video, though a globally-adjusted two-step video system has been successfully implemented [12]. The extra on-chip circuitry required for cliques and adaptation may also reduce the size of the light sensor area. However, chip level CMOS processing makes per-pixel operations increasingly practical.

Our paper addressed only stationary cameras viewing stationary scenes, but several other directions look promising. We proposed only a luminance camera, but extensions to color suggest novel opportunities. As chrominance is usually directly attributable to reflectance rather than illumination, separate HDR capture for each color channel may be unnecessary, and instead chrominance might be sensed and encoded as a low-dynamic range adjunct to luminance. As our sense of color has far lower spatial resolution than luminance, there may be interesting possibilities to include color sensing as part of each clique. Also, given the variety of clique and sensor patterns possible, alternatives to rectangular grids deserve a thorough exploration. For example, a diagonal grid superimposed on an axis-aligned grid has been shown to improve convergence in some projectors.

Despite the challenges and limitations, gradient based sensing is worth further exploration. We believe digital cameras might substantially improve their own abilities to capture visually meaningful assessments of a scene by measuring its changes.

References

- [1] A. Agarwala *et al.* Interactive digital photomontage. *ACM Trans. on Graphics*, 23(3):294–302, Aug. 2004.
- [2] V. Brajovic. Brightness perception, dynamic range and noise: a unifying model for adaptive image sensors. *CVPR*, 2(22):189–196, 2004.
- [3] DalStar, Inc. <http://www.pinnaclevision.co.uk>, 2004.
- [4] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. *ACM SIGGRAPH*, pages 369–378, Aug. 1997.
- [5] J. H. Elder. Are edges incomplete? *International Journal of Computer Vision*, 34(2/3):97–122, 1999.
- [6] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Trans. on Graphics*, 21(3):249–256, 2002.
- [7] G. Finlayson, S. Hordley, and M. Drew. Removing shadows from images. *Proc. European Conf. Computer Vision*, 4:823–836, 2002.
- [8] Fraunhofer IMS. One Mpixel CMOS HDR array, 2001.
- [9] E. Funatsu *et al.* An artificial retina chip with a 256×256 array of N-MOS variable sensitivity photodetector cells. *Proc. SPIE. Machine Vision App., Arch., and Sys. Int. IV*, 2597:283–291, 1995.
- [10] Integrated Vision Products. <http://www.ivp.se>, 2004.
- [11] R. Johansson *et al.* A multi-resolution 100 GOPS 4 Gpixels/s programmable CMOS image sensor for machine vision. *IEEE Workshop on CCDs and Adv. Image Sensors*, 2003. <http://www.ek.isy.liu.se/leif/m12ccdaais.pdf>.
- [12] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Trans. on Graphics*, 22(3):319–325, July 2003.
- [13] S. Mann and R. Picard. On being undigital with digital cameras. *Proc. of IS&T 48th Annual Conference*, pages 422–428, 1995.
- [14] N. Matsushita *et al.* ID CAM: a smart camera for scene capturing and ID recognition. *ISMAR*, 2(1):227–236, 2003.
- [15] C. Mead. *Analog VLSI implementation of neural systems*, chapter Adaptive Retina, pages 239–246. Kluwer, 1989.
- [16] A. Moini. Vision chips or seeing silicon. <http://www.iee.et.tu-dresden.de/iee/eb/analog/papers/mirror/visionchips>, 1997.
- [17] S. K. Nayar and V. Branzoi. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. *ICCV*, pages 1168–1175, 2003.
- [18] S. K. Nayar, V. Branzoi, and T. Boult. Programmable imaging using a digital micromirror array. *CVPR*, 1:436–443, 2004.
- [19] S. Palmer. *Vision Science: Photons to Phenomenology*. Bradford Books, 1999.
- [20] A. P. Pentland. Fractal based description of natural scenes. *IEEE PAMI*, 6(6):661–674, 1984.
- [21] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *ACM Trans. on Graphics*, 22(3):313–318, July 2003.
- [22] W. H. Press *et al.* *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.
- [23] Smal Camera. <http://www.smalcamera.com/>, 2004.
- [24] M. F. Tappen, B. C. Russell, and W. T. Freeman. Exploiting the sparse derivative prior for super-resolution and image demosaicing. *3rd Intl. Workshop on Stats. and Computational Theories of Vision*, 2003.
- [25] Y. Weiss. Deriving intrinsic images from image sequences. *ICCV*, 2:68–75, 2001.
- [26] F. Xiao *et al.* Image analysis using modulated light sources. *Proc. SPIE Image Sensors*, 4306:22–30, 2001.
- [27] D. Yang *et al.* A 640x512 cmos image sensor with ultrawide dynamic range floating-point pixel-level adc. *IEEE Journal on Solid State Circuits*, 34:1821–1834, 1999.