High Dynamic Range Capture Techniques using Multiple Exposures and Optical Flare

The idea in this disclosure allows the capture of a much more dynamic range within only two exposures. This could apply to professional and consumer HDR video capture.

Abstract

High dynamic range imaging typically involves time sequential capture of multiple photographs. This cannot be easily applied to images with moving objects, especially if the motions are complex.

In this paper we provide a two step process to get around this. Firstly, we optically encode both the low dynamic range portion of the scene and highlight information into a low dynamic image that can be captured with a conventional image sensor. This is achieved using a cross screen or star filter.

Then we decode with software, both the low dynamic range image that can be captured with a conventional image and the highlight information. These two steps can then be combined to form an image of a higher dynamic range, than the regular sensor dynamic range.

Background

Camera sensors can capture a certain maximum number of photons before they start to saturate and no longer register additional light. Although it is impossible to increase the saturation point by increasing the capacity of the sensor electron, producing large sensors is excessively expensive and reduces sensor resolution. Such sensors are also hard to justify for general imaging applications because on average, only a small portion of a scene contains very bright spots and thus needs high capacity sensors.

The human visual system has developed a clever mechanism to cope with highly saturated scene regions, such as highlights or light sources. Like camera sensors, the photoreceptors in the human retina are also prone to saturate. However, the visual system is able to infer higher brightness of those saturated regions from glare, which is produced by the light that is scattered in the ocular fluid and spread over the retina. The glare surrounding bright areas boosts their perceived brightness, giving additional information to the brain that this part of the scene is much more than the photoreceptor saturation point. (See fig 1, Chap 1, pg 289. 'Glare Encoding of High Dynamic Image Ranges.')

Previous Trials

Previously we disclosed the idea of bracketing two exposures from a 3D camera to capture HDR, using an ND filter or differing exposures. However, the approach only allows for the capture of two exposures which must have some overlap. This limits the dynamic range that can be captured. Many real-world scenarios contain much higher dynamic range then can be captured with this technique. This invention attempts to redress this problem.

We have produced a method to spatially disperse bright features or highlights in a scene that would have otherwise saturated a sensor. The trick is to use an optical filter which causes flare with a characteristic pattern. Spatially spreading the light limits the amount that the sensor is saturated, thus preserving all the information in the scene. That pattern can then be digitally removed from the resulting image to partially reconstruct the original scene dynamic range.

However in the paper, this approach was limited to a single exposure with the optical filter. This approach makes it very difficult to reconstruct the original scene with a high level of video quality, due to the complexity of the overlapping flare patterns. Flare that is not removed properly results in a hazy, low contrast scene, with additional remnants of the flare filter.

We have improved on this process by considering two simultaneous captures of a scene. The base capture is a normal exposure without any optical filters. The base capture is artefact-free, but contains saturated highlight areas and thus limited dynamic range. However, in this disclosure the second filtered capture includes the optical filter as described in the paper, and so contains the information from the saturated highlight areas spatially spread over the image. We use the information from these two exposures to reconstruct the original scene dynamic range.

The Solution

In this paper we propose to use a similar approach to improve camera dynamic range without resorting to custom sensors, multi sensor cameras, or time sequential imaging. Unlike the eye, we are not limited to specific optics. Instead, we can choose to modify the optical system in order to increase the information that is encoded for the saturated areas. Our goal can thus be more ambitious than simply to estimate the overall brightness of the saturated image regions. Specifically, we propose a computational photography approach comprised of the following steps:

Encoding: Details of bright image regions in a high dynamic range (HDR) image, such as highlights and directly visible light sources, which are encoded into specially shaped glare patterns optically added to the image.

Capture: The encoded image is captured using a standard image sensor. Bright regions in the captured image are saturated due to limited sensor dynamic range.

Decoding: In software we separate the glare pattern from the low dynamic range version of the image. The glare pattern can be used to infer the radiometric intensity distributions in the saturated image regions.

We have experimented with a number of specific optical encodings to implement this general principle. Some obvious candidates are regular lens glare and defocus blur to spread out energy from saturated image regions to other pixels. However, to provide enough information of the highlight regions for detailed reconstruction, energy spread must be significantly larger than standard lens flare. Likewise, a defocus blur implementation would have to use very large blur radii on the order of dozens of pixels. For such large blur, even the most recent deconvolution algorithms in combination with coded apertures fail to reconstruct high quality images.

In this paper, we therefore focus on the optical encoding that we found most successful, a glare pattern that scatters light in a fixed set of discrete directions. Such patterns are produced by inexpensive photographic cross-screen filters (also known as star filters), which are mounted in front of a camera lens. The scattering pattern of these filters is most salient for very bright scene features, since the star filters concentrate most energy in a Dirac peak rather than the glare rays. Star filters spread the light in discrete directions, and therefore one dimensional techniques can be applied instead of more expensive and less stable 2D techniques.

These properties let us estimate the amount of light spread from bright image features into several discrete directions, and then reconstruct clipped pixels using a tomographic reconstruction technique.

Related work

Multi-exposure HDR capture: Blending multiple exposures is the most accurate method for acquiring high dynamic range images with conventional cameras. However this approach is limited by ghosting and misalignment problems, which are still largely unsolved for difficult cases such as moving tree leaves or waves on the water. There are ways of obtaining multiple simultaneous exposures and to design sensors that directly support multi-exposure capture, but such cameras and sensors are not currently widely available.

LDR to HDR enhancement: Reconstructing an HDR image from a single exposure with clipped values is a challenging problem that yields only approximate solutions. Several techniques have been developed however; these are merely heuristics that are used to plausibly guess content that has ultimately not been captured.

Clipped signal restoration: For band limited 1D signals, reconstruction algorithms have been proposed for situations where the number of clipped samples is low, or where a statistical model of an undistorted signal is known. However, neither of these approaches can be trivially extended to images because natural image statistics are too weak to restore detailed texture in clipped regions. Therefore only special cases have been successfully solved in the image domain, for example images where only a subset of the colour channels is clipped, or noisy images with pixel values just above the clipping threshold.

Deconvolution: A large body of recent work has focused on the development of new deconvolution algorithms, as well as special, frequency-preserving convolution kernels for both motion blur and depth-of-field blur. In principle, both motion blur and depth-of-field blur could be used to spread energy of bright pixels in a fashion similar to what we propose in this paper. However, a sufficiently large energy spread can only be achieved with very large blur kernels. In our experiments, we found that even the combination of state-of-the-art deconvolution methods with special kernel shapes fails to recover a high quality, sharp image for these large radii. This is consistent with recently published results. Another problem with using convolution methods is that most recent deconvolution algorithms cannot reconstruct clipped pixels.

Our approach to using a cross-screen filter avoids these problems, since the filter produces a collection of 1D streaks that can be detected and removed reliably, while encoding enough information of the saturated regions to allow for detailed reconstruction of clipped pixel values.

Glare removal: Over the years, a number of approaches have been proposed for removing lens glare. Since we rely on strong glare for obtaining information about clipped image regions, the methods that optically suppress glare are not applicable in our setting. On the other hand, deconvolution methods that remove the glare after the fact suffer from the same shortcomings as the other deconvolution methods discussed above.

Image formation model

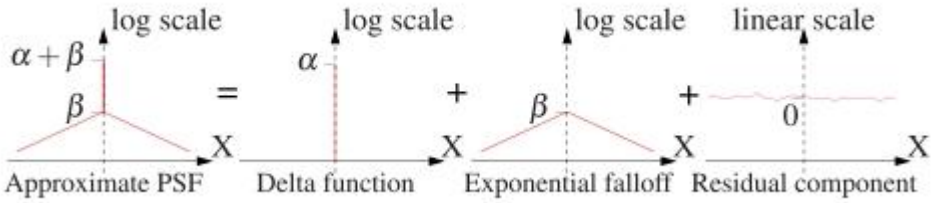
In the following, we outline the image formation process for cameras with a cross-screen filter before we go into the details of our approach.

A cross-screen filter is a transparent photographic filter with parallel scratch marks or grooves on its surface (Figure 2(a), pg 291, Rouf, Mantiuk, et al). When mounted in front of a camera lens, the grooves disperse and diffract the light, creating a star shaped glare – linear streaks (Figure 3, pg 291), in a number of discrete directions. This glare is very faint and hence star shaped glare patterns are usually noticeable only around very bright areas. A captured image g, can be expressed as a result of applying a light transport operator H, describing the glare to the latent image f, and then clipping the result to the maximum sensor value. (Equation 1, pg 291, Rouf, Mantiuk, et al):

$$g(\mathbf{x}) = \min(1, \sum_{\mathbf{y}} f(\mathbf{y}) H(\mathbf{x}, \mathbf{y}) + n).$$

Here, x and y refer to two dimensional image coordinates, and n represents noise. For simplicity, we ignore noise n in the rest of the derivation and discuss its influence on results in the supplemental material. H can be modelled as a combination of following components:

- a Dirac peak representing the light that does not hit one of the scratches on the cross-screen filter,
- a glare function *K* which has been empirically found to be both shift- and depth-invariant, and
- a zero-mean residual waviness in glare, r, that is *not* shift-invariant, but several orders of magnitude weaker in intensity.



The kernel can be approximated by a sum of a dirac delta function and an exponential fall off. The residual component accounts for a shift-variant wavelength-dependent response.

Thus:

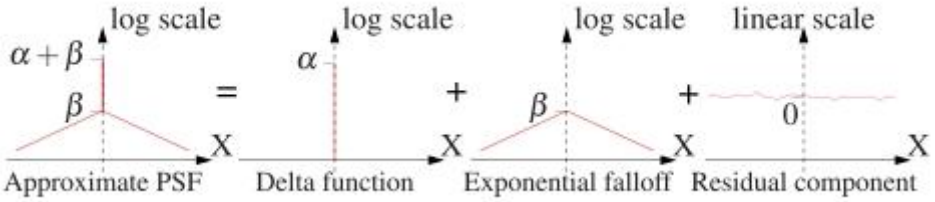
$$H(\mathbf{x}, \mathbf{y}) = \alpha\, \delta(\mathbf{y} - \mathbf{x}) + \beta\, K(\mathbf{y} - \mathbf{x}) + \gamma \rho(\mathbf{x}, \mathbf{y}). \quad (2)$$

For the filters we used, $\alpha \approx 1$, $\beta \approx 10^{-4}$ and $\gamma \approx 10^{-7}$. The glare function K is itself composed of 1D streaks,

$$K(\mathbf{x} - \mathbf{y}) = \begin{cases} \sum_{i=1}^{p/2} k_i(\mathbf{u}_i \cdot (\mathbf{x} - \mathbf{y})) & \text{when } \mathbf{v}_i \cdot (\mathbf{x} - \mathbf{y}) = 0 \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

Note that the parameters a, b, g and m can be measured for each cross-screen filter by capturing an (almost) point light source and measuring these statistics. In our experiments, we have observed that these quantities are independent of focal depth and position. The scene dependent residual waviness function r is primarily a function of the (unknown) spectral composition of the scattered light. Although this function is shift-variant, it too only distributes energy along radial-lines, like K.

(Figure 2 c, pg 291), shows cross-sections along glare streaks of 2D PSFs for several cross-screen filters we obtained. These measurements show that an exponential fall off model fits the overall shape of the glare quite well. In our application, this exponential model is sufficient for glare estimation with sufficient precision for saturated pixel reconstruction. The high-frequency variations captured in rare however, are important for removing glare from low dynamic range portion of the image. The overall image formation model is then given as (Equation 4, pg 292, Rouf, Mantiuk, et al):



In summary, our image formation model consists of a Dirac part and a combination of p/2 1D functions describing both an exponential fall off and a residual waviness. In the following, we can therefore consider the glare removal problem as a set of independent 1D problems.

Decoding Method

We now describe our proposed method for decoding both the low dynamic range image and the highlight details from a glare photograph taken with a cross-screen filter. Considering the light transport (Equation 4, see above), we can see that it is not possible to directly solve for the glare-free latent image due to sensor saturation. Instead, we split the problem by separately considering the saturated and the unsaturated pixels in the observed image g. We define gu to be the unsaturated pixels of g, with the values of all saturated pixels set to 0. We also define gs = g – gu to be a mask that is 1 for saturated pixels, and 0 for unsaturated ones. Similarly we define fs=f . gs and fu = f . (1 – gs). Finally, we define rs (ru) as only that part of the residual from equation 4, which is due to scattering of light from saturated (unsaturated) pixels.

With these definitions, we can rewrite the unsaturated component of Equation 4 as follows, (Equations 5, 6, Chap 4, pg 292):

$$g_{\mathbb{U}} = \alpha\,(f_{\mathbb{U}} + f_{\mathbb{S}}) + \beta\,K * (f_{\mathbb{U}} + f_{\mathbb{S}}) + \gamma(r_{\mathbb{U}} + r_{\mathbb{S}}) \qquad (5)$$

$$= \alpha\, f_{\mathbb{U}} + (\beta\,K * f_{\mathbb{U}} + \gamma r_{\mathbb{U}}) + (\beta\,K * f_{\mathbb{S}} + \gamma r_{\mathbb{S}}), \qquad (6)$$

As a result, we can now obtain the latent image by estimating and removing several kinds of glare.

Glare generated by unsaturated pixels that affects other unsaturated pixels — (First bracketed term of Equation 6, see above). This type of glare is fairly weak and does not contain high spatial frequencies. We can further simplify this term, since rU is so small as to be negligible.

Glare generated by saturated pixels that affects unsaturated pixels can be estimated and removed through the use of image priors (second bracketed term in equation 6, see above). The estimated glare also provides information about the saturated regions from which it emerges, and can therefore be used to reconstruct spatial detail within those regions.

Glare that contributes to already saturated pixels—either originating from unsaturated or saturated pixels — is not measured in the captured image and therefore does not need to be modelled. While in essence we do perform a 2D deconvolution, to make the solution possible and robust we decompose it into an 'easy' 2D deconvolution (a series of 1D problems), and finally a tomographic reconstruction. The supplemental material contains further discussion about the relationship to deconvolution.

Glare due to unsaturated pixels

Because the Dirac peak dominates the PSF of the cross screen filters, the glare due to unsaturated pixels is very weak. As mentioned above, we can further simplify the situation by neglecting the shift-variant residual rU, which is several orders of magnitude weaker than the shift-invariant part of the PSF. With these observations, we can remove the glare due to unsaturated pixels using a deconvolution approach similar to equations 7, and 8 :

$$g'(\mathbf{x}) = g(\mathbf{x}) - \beta(K * f_{\mathbb{U}})(\mathbf{x}) \quad \text{for } \mathbf{x} \in \mathbb{U} \qquad (7)$$

$$= \left(\sum_{t=0}^{\infty} \left(-\frac{\beta}{\alpha} K \right)^{t} * g \right)(\mathbf{x}) \quad \text{from (6)}, \qquad (8)$$

Where g' is the image with the unsaturated pixel glare removed, and the operator ·t denotes t-times convolution.

Glare due to saturated pixels

The next step is to estimate and remove glare due to saturated pixels. This glare component will also be used for reconstructing saturated pixel values in section 4.3. As mentioned in Section 3, we can factor this step into a number of 1D problems along directions **u***i*, where **u***i*,**v***i* form a coordinate frame aligned with the *i*th glare ray (Figure 5, pg 293). In the following, we consider each glare direction separately, and thus omit the i subscript for notational convenience.

Image priors

Knowing both which pixels are saturated in the observed image, as well as the direction of the 1D glare rays, we can determine which image pixels exhibit a glare contribution from saturated pixels. In order to separate the latent image information from the glare in these pixels, we employ results from natural image statistics, specifically a sparse gradient prior. We model the distribution of gradients in the latent image using a Laplace distribution, which is the best approximation of the heavy-tailed distribution that still leads to a convex problem.

Glare rays cause the largest distortion of the image gradients in the direction *orthogonal* to the glare rays.

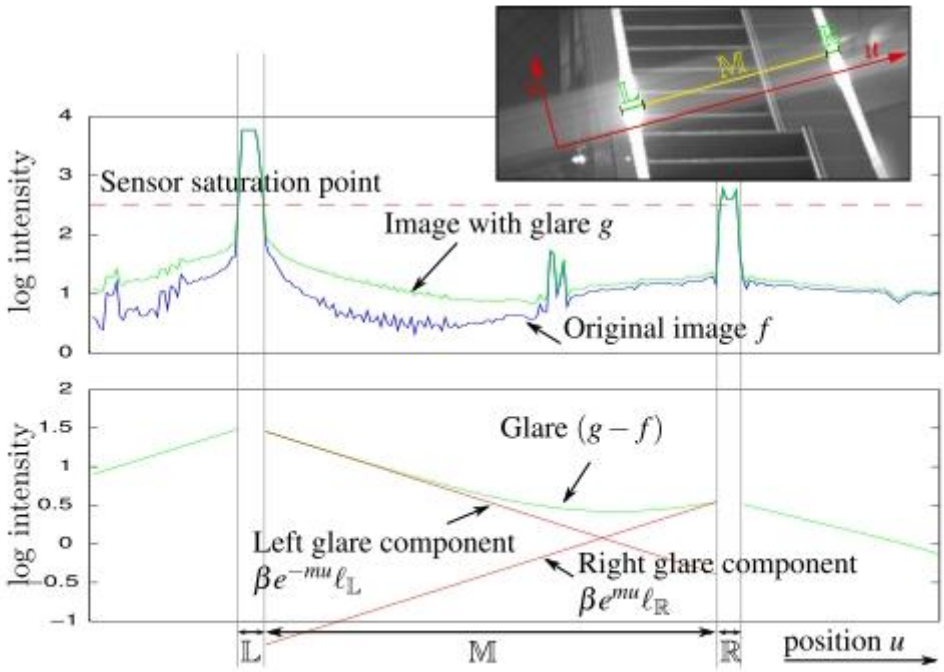
we obtain $\frac{\partial f}{\partial v} \sim \text{Laplace}(0, b)$.

According to the spares gradient prior, any deviations from a zero mean in the observed image g are attributed to glare. In the supplemental material we show that the Maximum Likelihood (ML) estimator for the mean of a Laplace distribution is obtained by minimizing the L1 norm. Therefore, we can solve for the intrinsic image as follows, (Equation 9, pg. 293, Rouf, Mantiuk, et al):

$$\arg\min_f \left\| \frac{\partial}{\partial v} \left(g' - \beta K * f_{\mathbb{S}} - \gamma r_{\mathbb{S}} \right) \right\|_1 + R, \qquad (9)$$

where R gives constraints on $r_{\mathbb{S}}$ (see supplemental material, Section 2.2 for details):

$$R = \left(\lambda_1 \left\| r_{\mathbb{S}} \right\|_2 + \lambda_2 \left\| \frac{\partial}{\partial v} r_{\mathbb{S}} \right\|_1 + \lambda_3 \left\| \frac{\partial}{\partial u} r_{\mathbb{S}} \right\|_1 \right). \qquad (10)$$



Optimization

To actually apply the image prior in the glare estimation, we consider a single continuous segment M of unsaturated pixels along a glare direction u. M is bounded by two sets of saturated pixels L and R on the left and on the right. (Figure 5, pg. 293).

We can now use the exponential nature of the 1D glare streaks from Equation 3 (pg. 291), and expand the convolution operator in Equation 9 (pg. 293). Note that ℓL and ℓR have the same value for all unsaturated pixels u ∈ M, and therefore all pixels in M can be used to robustly estimate these two quantities. Also note that ℓL and ℓR represent the amount of energy present in the glare from the saturated pixels to the left and to the right of M. These quantities, which we refer to as line integrals, will be useful for reconstructing detail in the saturated regions.

Now we can reformulate the glare estimator in terms of line integrals ℓL and ℓR rather than saturated pixel values. From Equations 9, 11, and 12, we obtain, (see equation 13, pg. 293):

$$(k * f_{\mathbb{S}_v})(u,v) = \ell_{\mathbb{L}}\, e^{-mu} + \ell_{\mathbb{R}}\, e^{mu} \quad \text{for } u \in \mathbb{M}, \tag{11}$$

where

$$\ell_{\mathbb{L}} = \sum_{i \in \mathbb{L}} e^{mi} f(i,v), \;\; \ell_{\mathbb{R}} = \sum_{i \in \mathbb{R}} e^{-mi} f(i,v). \tag{12}$$

$$\underset{\ell_{\mathbb{L}},\ell_{\mathbb{R}},r}{\arg\min} \sum_{u \in \mathbb{M}} \left\| \frac{\partial}{\partial v} g'(u) - \beta \frac{\partial}{\partial v} \left(\ell_{\mathbb{L}}\, e^{-mu} + \ell_{\mathbb{R}}\, e^{mu} \right) - \gamma \frac{\partial}{\partial v} r_{\mathbb{S}}(u) \right\|_1 + R. \tag{13}$$

This equation allows us to efficiently optimize on each segment $\mathbb{M}$ independently. However, to solve for $\ell_{\mathbb{L}}$ and $\ell_{\mathbb{R}}$, the partial derivatives $\frac{\partial \ell_{\mathbb{L}}}{\partial v}$ and $\frac{\partial \ell_{\mathbb{R}}}{\partial v}$ must be found for all segments and then integrated. To solve the minimization problem efficiently, we use several EM iterations. We initially set $\gamma \frac{\partial r}{\partial v} = 0$. Since $\gamma \ll \beta$, this provides a reasonable initial estimate of exponential glare component, but enhances color artifacts when this monochromatic glare is removed. In the E-step, we solve for $\ell_{\mathbb{L}}$ and $\ell_{\mathbb{R}}$, and in the M-step we refine the estimate of r. Minimizing Equation 13 is sufficient to remove most of the glare (Figure 6).

(Figure 6, page 293)

Finally, we prepare line integral estimates for the energy contributed by individual, continuous regions of saturated pixels, which will be used in the next section. Each value ℓL and ℓR can contain contributions from multiple saturated segments on the left and right of M. However, isolating glare due to each saturated region is trivial since there are exactly as many line integrals as there are regions M along a glare line, and therefore the contributions for each region can be found with a simple linear system. For convenience, we shift the origin of (u, v) to the left most or rightmost pixel of each segment M to get isolated line integrals bℓL and bℓR.

Reconstruction of saturated pixels

So far we have decoded the values of the intrinsic image f for the previously unsaturated pixels, only the values of the saturated pixels are still unknown. However, glare removal procedure (Section 4.2 pg. 292), also yields line integrals along p discrete directions, as shown in Figure 7(a), (pg. 294). In the final step of the decoding procedure, we use this information to reconstruct the saturated region. To this end, we need to find saturated pixel values that can produce the line integrals matching the observations. This requires solving a standard tomographic reconstruction problem.

Unlike the glare estimation, the tomographic reconstruction is inherently a 2D problem. We gather the estimated line integrals along all *p* directions in a linear system that describes the relationship between line integrals and saturated pixels *f* . We therefore use a one-index representation for all line integrals contributing to a given region: bℓ*i*. This relationship is then expressed as, (Equation 14, pg. 294):

$$\widehat{\ell}_i = \sum_j w_{ij} f_j, \qquad (14)$$

where the weight term wi j for line integral i and an unknown pixel j is the product of exponential fall off and a bilinear resampling weight ai j , as shown in Figure 7(b)(Equation 15, pg. 294):

$$w_{ij} = a_{ij} e^{-m|u_i - u_j|}. \qquad (15)$$

Here, ui is the reference location used while computing bℓi. The absolute value consolidates different signs for glare fall offs to the left and right. We solve this tomography problem using Simultaneous Iterative Reconstruction. We start with an initial guess, f (0) = 0. Then in each iteration t, the residual error in the current estimate of line integrals (Equation 16, pg. 294, Rouf, Mantiuk, et al):

$$\Delta\widehat{\ell}_i = \widehat{\ell}_i - \sum_j w_{ij} f_j^{(t)}, \qquad (16)$$

is back projected over the participating unknown pixels regardless of distance from the reference location, i.e., energy distribution is proportional to resampling weight (a) only (Equation 17, pg. 294):

$$f_j^{(t+1)} = f_j^{(t)} - \Delta\widehat{\ell}_i \frac{a_{ij}}{\sum_k a_{ik}}. \qquad (17)$$

Using a uniform distribution for the back projected residual independent of any falloff, is a standard procedure in tomography. One should think of this as a (weak) prior on the intensity distribution within the unknown region. We employ a simple two-scale approach which solves the problem for a low resolution image first. Since we know that actual values at the saturated pixels are larger than the saturation threshold for the camera, we enforce this simple constraint during back projection.

Results

(See Figure 8 pg. 296, 'Glare Encoding of High Dynamic Image Ranges'). The figure shows a number of examples of HDR images, decoded from single images captured as RAW images with a Canon 40D DSLR camera using 8- and 16-point cross screen filters, and Canon lenses ranging from 50mm to 100mm. In this figure the first two columns represent two exposures of the 12-bit input image, while the right two columns represent two virtual exposures of our reconstructions.

Saturated regions are reconstructed and glare produced by the filter is removed. For colour images, we run our algorithm separately and independently on each colour channel. Radial lens distortion was removed in a pre-processing step. Insets in the right column show ground-truth comparisons for some of the results, i.e. short exposure images taken without the filter, using the same camera and lens.

Note that the geometric and photometric alignment may not be perfect due to the changes in the acquisition setup. These results demonstrate a number of points:

Glare estimation: Accurate estimation of glare is necessary not only to correctly reconstruct saturated regions, but also to remove glare. Our sparse-gradient prior was robust enough to estimate glare both for a multitude of small light sources (Figure 8a, pg. 296), as well as relatively large saturated areas (Figure 8c, pg. 296). The main requirement for successful glare estimation is that saturated regions be both bright and large enough (i.e. sufficient cumulative energy) to produce glare above the camera noise level.

Highlight reconstruction: Given only 8–16 directional line integrals, tomographic reconstruction is a challenging task. Even so, the results demonstrate that our method can estimate the total energy of the saturated regions as well as the approximate values of the saturated pixels. This is in contrast to the previous single-image methods, which could achieve neither of these two goals. Our method can also easily distinguish between very bright light sources and diffuse surfaces that are just above the clipping level, thus making complicated classification methods for the LDR-to-HDR enhancement unnecessary. Figure 8(a), (pg. 296, Rouf, Mantiuk, et al) also demonstrates that the multi-exposure HDR can exhibit some artefacts due to alignment issues, particularly at the outline of the light sources. Ours being a single exposure method does not show any such artefacts.

Conclusion

The distinctive feature of our proposed single-image HDR capture method is that the information lost in clipped pixels is encoded in the remaining portions of an image. This approach is very different from existing HDR capture methods, which attempt to register HDR information within each pixel or a group of closely located pixels. Unlike the LDR to HDR methods that only enhance clipped pixels; the proposed method can restore a close approximation of their original values. Our method does all that without requiring specialized sensor or invasive camera modifications, as it needs only a cross-screen filter mounted on top of a lens. Our reconstruction method contains several technical contributions, including the use of natural image priors to separate encoded information (glare) from image content.