

A Versatile HDR Video Production System

Michael D. Tocci^{1,2}

Chris Kiser^{1,2,3}

Nora Tocci¹

Pradeep Sen^{2,3}

¹Contrast Optical Design & Engineering, Inc.

²University of New Mexico

³Advanced Graphics Lab

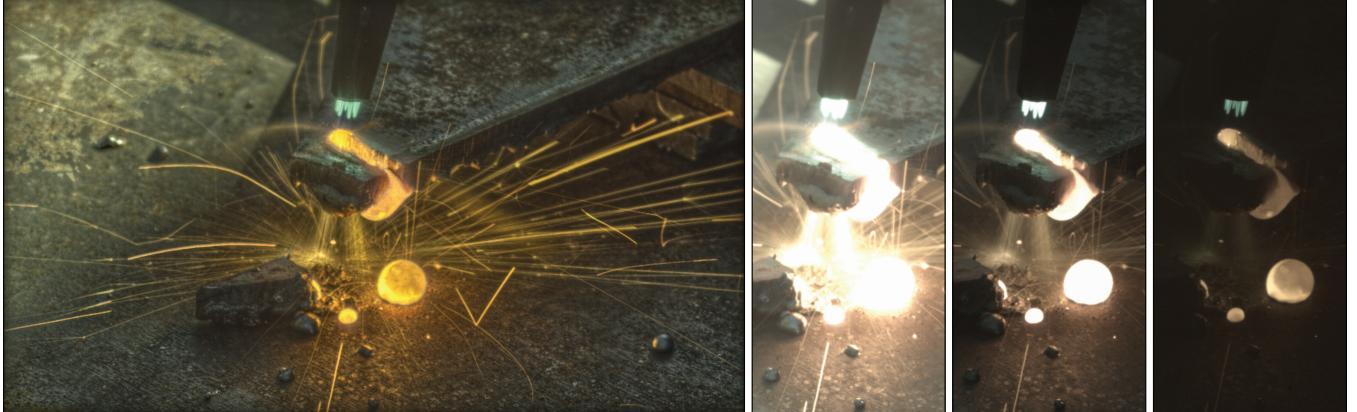


Figure 1: HDR image acquired with our proposed system. On the left we show the final image acquired with our camera and merged with the proposed algorithm. The inset photos show the individual LDR images from the high, medium, and low-exposure sensors, respectively.

Abstract

Although High Dynamic Range (HDR) imaging has been the subject of significant research over the past fifteen years, the goal of acquiring cinema-quality HDR images of fast-moving scenes using available components has not yet been achieved. In this work, we present an optical architecture for HDR imaging that allows simultaneous capture of high, medium, and low-exposure images on three sensors at high fidelity with efficient use of the available light. We also present an HDR merging algorithm to complement this architecture, which avoids undesired artifacts when there is a large exposure difference between the images. We implemented a prototype high-definition HDR-video system and we present still frames from the acquired HDR video, tonemapped with various techniques.

CR Categories: I.4.1 [Image Processing and Computer Vision]: Digitization and Image capture—Radiometry

Keywords: HDR video, merging HDR images

Links: DL PDF

1 Introduction

The extension of the dynamic range of digital images has been the subject of significant research in both academia and industry. Despite all this previous work, however, there are currently no readily-implemented solutions for capturing high-quality HDR video of fast-moving scenes. In this paper, we describe an end-to-end system for capturing HDR video with high pixel fidelity, using a light-efficient optical architecture that fits into a single hand-held unit.

Our proposed system is simple, uses only off-the-shelf technology, and is flexible in terms of the sensors that are used. Specifically, our HDR optical architecture: (1) captures optically-aligned, multiple-exposure images simultaneously that do not need image manipulation to account for motion, (2) extends the dynamic range of available image sensors (by over 7 photographic stops in our current prototype), (3) is inexpensive to implement, (4) utilizes a single, standard camera lens, and (5) efficiently uses the light from the lens.

To complement our system, we also propose a novel HDR image-merging algorithm that: (1) combines images separated by more than 3 stops in exposure, (2) spatially blends pre-demosaiced pixel data to reduce unwanted artifacts, (3) produces HDR images that are radiometrically correct, and (4) uses the highest-fidelity (lowest quantized-noise) pixel data available. We demonstrate a working prototype and present images and video acquired with this system.

2 Previous Work

2.1 HDR Acquisition systems

The process of capturing HDR images has been the focus of work by dozens of researchers and hundreds of artists and photographers. As a result, there are many published papers and patents describing methods and systems for capturing HDR images. Because of space limits, we focus only on the principal technologies currently available for HDR video and refer interested readers to texts on the subject (e.g., [Myszkowski et al. 2008]) for more information.

The simplest approach for HDR imaging involves taking a series of images with different exposure times (e.g., [Mann and Picard 1995;Debevec and Malik 1997]). Although this method works well for static scenes, it is not well-suited for video because of the different moments in time and exposure lengths for each photograph, which result in varying amounts of motion blur and other time-related effects. Nevertheless, researchers have extended this approach to video, by capturing frames with alternating bright and dark exposures [Ginosar et al. 1992; Kang et al. 2003] or using a rolling shutter with varying exposures [Unger and Gustavson 2007; Krymski 2008]. These approaches require image manipulation to register the images, which also introduces artifacts.

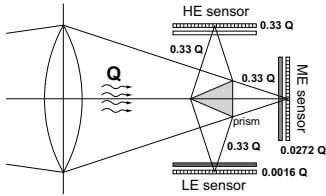


Figure 2: A traditional beamsplitting HDR optical system. Here a beamsplitting prism breaks up the light into three parts, one for each sensor fitted with different filters. Designs that use absorptive filters like this one make inefficient use of light.

Other researchers have proposed new camera sensors for HDR imaging. Some approaches place an array of neutral-density filters over the individual pixels of the sensor with varying amounts of absorption (e.g., [Nayar and Mitsunaga 2000]), which can require complex demosaicing algorithms. These approaches are also wasteful of light entering the camera. If the sensor has a filter pattern with four differently-exposed pixels (one of the four fully exposed), then only $\frac{1}{4}$ pixels would receive the full exposure level from the scene.

Other proposed HDR sensors have a unique response to light, either by adapting their sensitivity (e.g., [Nayar and Branzoi 2003]), measuring the pixel saturation time (e.g., [Brajovic and Kanade 1996]), or having a logarithmic response to mimic the human eye (e.g., [Seger et al. 1999]). The primary problem with all of these approaches is that they require the production of a new type of camera sensor. Although commercial-scale production of these sensors may someday be realized, they are currently expensive to manufacture, rendering these methods unusable by most researchers today. Our proposed architecture, on the other hand, performs HDR imaging independent of the sensor used, which makes it realizable using today’s technology and allows us to adopt better sensor technologies (with low-light level response, faster framerates, wider spectral response, etc.), as they are developed in the future.

In approaches similar to our own, the light in the camera is split with a pyramid-shaped mirror or refracting prism and redirected toward a set of sensors fitted with absorptive filters to produce images with different exposures (e.g., Harvey [1998], Aggarwal and Ahuja [2001; 2004], and Wang et al. [2005]). The designs of these previous systems all suffer from parallax error, due to the fact that the image-forming beam is split into spatially-distinct subsections; each individual sensor “looks” through the camera lens from a slightly different angle. As shown in recent work on handheld plenoptic cameras (e.g., [Ng et al. 2005]), this provides each of the sensors with slightly different information, which significantly affects the imaging of scenes close to the camera.

These previous spatial-beamsplitting methods are also wasteful of light: the absorptive filters used to achieve the dynamic range allow only a fraction of the incoming light to the sensors. If Q Watts of radiative power enters the aperture of the camera, the three-way system shown in Fig. 2 (configured for the same dynamic range as ours) allows only $0.3622Q$ Watts to the sensors, wasting almost $\frac{2}{3}$ of the available light. As Aggarwal and Ahuja [2004] point out, it is possible to vary the amount of light to each sensor by moving the beamsplitting prism away from the optical axis instead of using filters. This effectively changes the size and shape of the aperture stop for each sensor, exacerbating the problem of each sensor getting different views of the scene. Furthermore, these shifted-optical-axis spatial-beamsplitting methods are not easily integrated with standard camera lenses, and require either custom lens manufacture or lens modification to work correctly.

Another option is to split the incoming light with beamsplitters prior to the lens. For example, McGuire et al. [2007] present

a design tool to create efficient beamsplitting trees with separate lenses for each sensor, and show examples of HDR imaging. This same concept was demonstrated by Soviet Montage Productions [Cole and Safai 2010], which can be implemented on a 3D filming rig with an intraocular distance of zero by using a beamsplitter to provide different light transmission to two identical lenses, one for each sensor. The two lenses must be perfectly matched, however, and zoom-, focus-, and iris-tracking can be difficult to maintain between them. In addition, putting the beamsplitter in front of the camera lens places a limit on the field of view. Finally, it is unclear how such a system could be developed into a single, hand-held unit. Our system places the beamsplitter behind a single camera lens, so it does not suffer from these limitations.

Finally, there are early prototype HDR systems in industry eventually intended for commercial use, such as the SpheronVR HDRv camera [SpheronVR 2011]. However, their method for achieving HDR capture has not been published. While all the systems we mention in this section are capable of producing HDR video, to date no method for producing high-quality HDR video has been demonstrated that is robust and yet simple enough to be readily introduced to a wide commercial audience or implemented in a modern optics laboratory. The goal of this work is to present such a system.

2.2 Algorithms for merging HDR images

A common method for merging multiple LDR images into a single composite HDR image is the one of Debevec and Malik [1997], which first solves for the camera response curve that translates pixel values to the log of scene irradiance and then blends irradiances from the images together. During the merging process, the algorithm combines values from every exposure by weighting each contribution by a triangle filter that falls off as the pixel value approaches cutoff or saturation and peaks in the middle. The idea is to give more weight to pixels in the “working range” of the camera, and less to the ones near the extrema of the camera’s operating range. As we describe in Sec. 4, however, this approach can suffer from undesirable artifacts when applied to widely-separated LDR images due to the blending between exposures.

Following the work of Debevec and Malik, other researchers have proposed different weighting functions for merging differently-exposed LDR images to reduce noise and improve the result (e.g., [Mitsunaga and Nayar 1999; Robertson et al. 2003; Kao 2008; Granados et al. 2010]). These approaches typically work on each pixel of the final HDR image independently and use only the information contained within the respective pixel in each of the LDR images. Unlike these approaches, we propose to use additional information available in the neighborhood of a pixel to reduce the noise in our final irradiance estimate. Finally, others have presented algorithms for fusing the LDR images together without explicitly creating an HDR image first (e.g., [Agarwala et al. 2004; Mertens et al. 2008]). These methods do not produce a true, radiometrically-correct HDR image, so the results cannot be incorporated into an HDR production workflow.

3 Efficient Optical Architecture for HDR Video

Our optical architecture is based on beamsplitters located between the camera lens and the sensors, which have been used in previous HDR camera designs as well as in 3-sensor color-splitting cameras [Kell and Sziklai 1951]¹. Like these previous methods, our optical system uses a set of partially-reflecting surfaces to split the light from a single photographic lens so that it is focused onto three

¹The optical splitting prisms used in 3CCD cameras employ dichroic filters to split the incoming light from a camera lens into red, green, and blue portions so that each color is imaged onto its own image sensor.

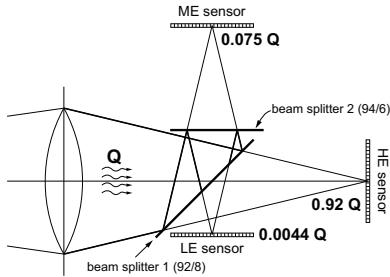


Figure 3: Illustration of our optical architecture. We also use beamsplitters between the lens and sensors, but the key difference is that we re-use the optical path to improve our light efficiency. In the end, 99.96% of light entering the aperture arrives at the sensors. Light efficiency is important in all imaging applications.

imaging sensors simultaneously. In our architecture, however, the light is directed back through one of the beamsplitters a second time, and the three sub-images are not split into red, green, and blue but instead are optically identical except for their light levels. This design, shown in Fig. 3, allows us to capture HDR images using most of the light entering the camera. We use the terms “high,” “medium,” and “low” exposure (HE, ME, LE, respectively) to refer to the sensors based on the amount of light each one receives.

The optical splitting system in our current implementation uses two uncoated, 2-micron thick plastic beamsplitters which rely on Fresnel reflections at air/plastic interfaces so their actual transmittance/reflectance (T/R) values are a function of angle. In our arrangement, the first beamsplitter is at a 45° angle and has an approximate T/R ratio of 92/8, which means that 92% of the light from the camera lens is transmitted through the first beamsplitter and focused directly onto the high-exposure (HE) sensor². This beamsplitter reflects 8% of the light from the lens upwards, as shown in Fig. 3, toward the second uncoated beamsplitter, which has the same optical properties as the first but is positioned at a 90° angle to the light path and has an approximate T/R ratio of 94/6.

Of the 8% of the total light that is reflected upwards, 94% (or 7.52% of the total light) is transmitted through the second beamsplitter and focused onto the medium-exposure (ME) sensor. The other 6% of this upward-reflected light (or 0.48% of the total light) is reflected back down by the second beamsplitter toward the first one (which is again at 45°), through which 92% (or 0.44% of the total light) is transmitted and focused onto the low-exposure (LE) sensor. With this arrangement, the HE, ME and LE sensors capture images with 92%, 7.52%, and 0.44% of the total light gathered by the camera lens, respectively. Therefore, the HE and ME exposures are separated by $12.2 \times$ (3.61 stops) and the ME and LE are separated by $17.0 \times$ (4.09 stops), which means that this configuration is designed to extend the dynamic range of the sensor by 7.7 stops.

This beamsplitter arrangement makes our design light efficient: a negligible 0.04% of the total light gathered by the lens is wasted. It also allows all three sensors to “see” the same scene, so all three images are optically identical except for their light levels. Of course, the ME image has undergone an odd number of reflections and so it is flipped left-right compared to the other images, but this is fixed easily in software. The three sensors are gen-locked to capture perfectly synchronized video frames with identical exposure times.

²Since T/R is also dependent on the wavelength of the light, we calculate T/R values for the full visible spectrum and integrate over the R, G, and B filter spectra in the Bayer pattern to arrive at separate T/R values for each color channel for use in our design and implementation. To simplify the discussion in this paper, however, we simply state a single average value of transmittance.

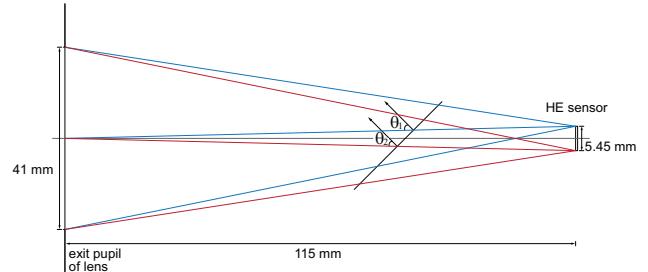


Figure 4: Scale drawing of optical path of first sensor. Since uncoated beamsplitters like ours can vary in transmittance as a function of angle, we examine the exact geometrical configuration of our system at f/2.8 to determine the range of transmittance values across our sensor. In this case, $\theta_1 = 46.4^\circ$ and $\theta_2 = 43.7^\circ$, which result in transmittance values of 91.85% and 92.38%, respectively.

3.1 Analysis of Optical System

Because the exact transmission/reflection properties of our beamsplitters vary with angle, we examine how these might vary over the area of the sensor by simulating the proposed optical architecture in ZEMAX [2011]. To calculate the range of transmittance values as a function of angle, we examine the largest angular variation possible on the pellicle beamsplitter. Approaches that place the beamsplitters outside the lens, such as the optical trees of McGuire et al. [2007], can have a large range of incident angles which results in significant variation in transmission over the field of view. Unlike these approaches, our system’s internal beamsplitters receive light in a much smaller range of field angles because of the geometrical configuration of the system, shown to scale in Fig. 4.

In our case, the top-left and bottom-right corner points on the sensor have chief-ray angles at the pellicle of 46.4° and 43.7° , a difference of 2.7° . At f/2.8, each of these two points receives a $\pm 10^\circ$ cone of rays from the lens, shown in blue and red (these cones are constant in angle over the entire sensor). We calculate the transmittance of the beamsplitter by integrating over this cone of rays using a ZEMAX simulation with 1 million random rays on a 2-micron thick, uncoated plastic pellicle beamsplitter, which yields a transmittance of 91.85% for the top-left and 92.38% for the bottom-right points, a difference of about 0.5%, and close to the 92% value we used in our design calculations. Therefore variation in transmittance across the sensor is not a major issue in our system.

Polarization of the incident light might affect the transmission properties of the beamsplitter as well. Although our simulations were all done with unpolarized light, it is possible to encounter linearly polarized light in outdoor scenes (e.g., from glancing reflections off water), which may change the exposure difference between sensors. However, in practice we did not see such polarization effects in the scenes we captured. We note that all of these effects may be reduced or eliminated by using a thin-film coating on the beamsplitter. This thin-film coating could be designed to have more constant transmission properties over the range of angles in the system or to reduce polarization effects. An examination of different beamsplitter coatings to address these factors is a topic for future work.

Advantages of the proposed optical splitting system are that its cost and complexity are relatively low, and it is compatible with standard camera lenses. The compact light path allows integration into a single hand-held unit, something difficult to do with designs that place the beamsplitters outside the lens [McGuire et al. 2007; Cole and Safai 2010]. The optical architecture is also flexible in terms of the kind of sensor used. The use of low-cost sensors, for example, could allow the design to be integrated into consumer electronics and bring HDR video to a wide audience.

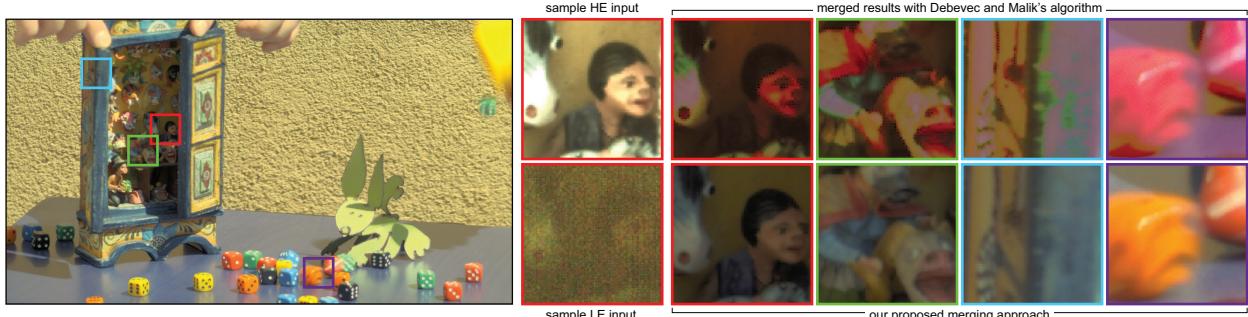


Figure 5: The problem with traditional merging algorithms. When applied to widely-separated LDR images, Debevec and Malik’s merging algorithm can produce artifacts in the tone mapped results, as seen in the top row of insets. Our proposed merging algorithm (left image and bottom insets) uses exclusively the data with the finest quantization available (from the brightest-exposed sensor) and blends in data from the next darker sensor when necessary. We show the input HE and LE images for a sample inset, with the LE image brightness scaled by $3\times$ to show the noise. The HDR image has been tonemapped with FDR Tools using the same settings for both merged images.

4 A New Algorithm for Merging HDR Images

In addition to the optical architecture, we also propose a novel algorithm for merging together images acquired with our system. In order to automatically create HDR images from the widely-separated component LDR images without artifacts, we found it necessary to depart from the standard methods for merging HDR images.

4.1 Limitations of current approaches

Most previous algorithms for merging HDR images from a set of LDR images with different exposures typically do so after demosaicing the LDR images and merge data pixel-by-pixel without taking neighboring pixel information into account. For this discussion, we focus on the original merging algorithm proposed by Debevec and Malik [1997]. Although this algorithm works well when the LDR images have a small exposure separation, we found the quality of the final HDR image degrades when the LDR images are separated by more than 3 photographic stops. To capture the widest dynamic range possible with the smallest number of camera sensors, it is necessary to position the LDR images further apart in exposure in our system than with traditional HDR acquisition methods. Debevec and Malik’s merging algorithm can yield undesired artifacts when applied to our camera’s data because of quantization and noise effects, as shown in Fig. 5.

This problem is exacerbated when we apply certain tone mapping operators (TMO’s), such as Fattal et al.’s local-contrast operator [2002]. These TMO’s amplify small gradient differences in the image to make them visible when the dynamic range is compressed, amplifying merging artifacts as well. Of course, when the LDR images are very close in exposure, these artifacts are considerably reduced and methods like Debevec and Malik’s work quite well. However, as HDR video systems with a small set of sensors become more widely used, this issue will become more important.

Fig. 6 shows the underlying cause for these artifacts by presenting the range of scene illumination measured by the three sensors in our architecture. For illustration, we simplify the system with 4-bit sensors (as opposed to the 12-bit sensors in our implementation), which measure only 16 unique brightness values and we separate the sensors by only 1 stop (a factor of 2) in exposure. Since CMOS sensors exhibit an approximately linear relationship between incident exposure and their output value, we graph the values from the three sensors as a linear function of incident irradiance instead of the traditional logarithmic scale.

In this figure, we see that the LE sensor measures the scene irradiance more coarsely than the other two sensors. For example, the HE sensor may measure 4 different pixel values in a gradient be-

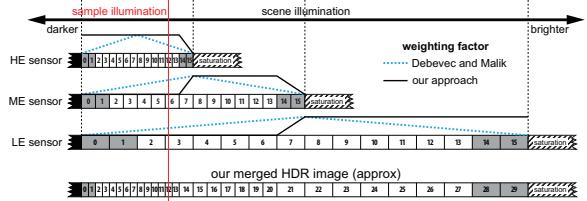


Figure 6: Quantization problems when merging LDR images. This figure shows the source of artifacts when merging our LDR images (simplified here to 4-bit sensors) with algorithms that always use data from all three sensors with simple weighting functions, such as that of Debevec and Malik. In this case, data from each sensor is weighted with a triangle function as shown, so there are non-zero contributions from the LE sensor at low brightness values (like the sample illumination level indicated), even though the data from the LE sensor is quantized more coarsely than that of the HE sensor. Our approach, in contrast, uses data from the higher-exposure sensor as much as possible and blends in data from the next darker sensor when near saturation.

fore the LE sensor records a single increment. In addition, there is always some small amount of noise in the pixel values, and an error of ± 1 in the LE sensor spans a 12 value range in the HE sensor for this example. Although Debevec and Malik’s algorithm blends these values together, we propose instead to use pixel values from only the longest-exposure sensor (which is less noisy) whenever possible, and blend in the next darker exposure when pixels approach saturation. More importantly, our algorithm not only examines individual pixels when merging the LDR images, but also takes into account neighboring pixels that might provide additional information to help in the denoising process.

4.2 Our proposed HDR-merging algorithm

While not a new idea, one key aspect of our image-merging algorithm is to use pixel data exclusively from the brightest, most well-exposed sensor possible. Therefore, we use pixels from the HE image as much as possible, and only use pixels in the ME image if the HE pixel is close to saturation. If the corresponding ME pixel is below the saturation level, it is multiplied by a factor that adjusts it in relation to the HE pixel based on the camera’s response curve (Fig. 7), given that the ME pixel receives $12.2\times$ less irradiance than the HE pixel. If the corresponding ME pixel is above the saturation level, then we simply apply a similar process to the same pixel in the low-exposure LE image.

When merging data between sensors, it may seem sufficient to follow a naïve “winner take all” approach and exclusively use the val-

ues from the HE sensor until they become saturated and then simply switch to the next sensor [JAI 2009]. We found that this does not work well in practice because it results in banding artifacts where transitions occur. Instead, we propose to gracefully transition from one sensor to the next by spatially blending pixel values between the two sensors. To do this, our algorithm scans a region around the pixel being evaluated. If any neighboring pixels in this region are saturated, then the pixel under consideration may be subject to pixel crosstalk or leakage, and the algorithm will estimate a value for the pixel based on its neighbors as described below.

Algorithm details – In our approach, HDR-merging is performed *prior* to demosaicing the individual Bayer color filter array images, because we found that demosaicing can corrupt colors in saturated regions (also noted by Ajdin et al. [2008]). For example, a bright orange section of a scene might have red pixels that are saturated while the green and blue pixels are not. If the image is demosaiced before being merged into HDR, the demosaiced orange color will be computed from saturated red-pixel data and non-saturated green/blue-pixel data. As a result, the hue of the orange section will be incorrectly reproduced. The only way to avoid these artifacts is to perform HDR-merging prior to demosaicing.

Since the images are merged prior to the demosaicing step, our algorithm works with pixel values instead of irradiance. To produce a radiometrically-correct HDR image, we must correctly match the irradiance levels of the HE, ME, and LE sensors using the appropriate beamsplitter transmittance values for each pixel color, since these change slightly as a function of wavelength. Although we use different values to match each of the color channels, for simplicity we explain the process with average values. We consider converting a pixel value through the camera response curve, where the resulting irradiance is adjusted by the exposure level ratio (average of $12.2 \times$ for HE/ME), and this new irradiance value is converted back through the camera response curve to a new pixel value. Fig. 7 shows this 3-step process graphically.

This conversion process may next be done for all HE pixel values (from 1 through 4096), to arrive at a pixel-ratio curve, which gives the scaling factor for converting each ME pixel’s value to the corresponding pixel value on the HE sensor for the same irradiance (Fig. 8). In practice, separate pixel-ratio curves are calculated for each color (R,G,B) in the Bayer pattern. When comparing pixel values between HE and ME images (or between ME and LE images), we use the pixel-ratio curves as lookup tables (LUTs) to convert HE pixel values less than 4096 into ME pixel values, or vice versa. When the HE pixel values are saturated, we simply extend the pixel-ratio curve using the last value obtained there (approximately 8).

The camera response curve can be measured with the method ofDebevec and Malik [1997], by taking a set of bracketed exposures and solving for a monotonically-increasing function that relates exposure to pixel value (to within a scale constant in the linear domain). Fig. 7 shows the curve computed from the raw camera data, although a curve computed from a linear best-fit could also be used. In our case, we can factor out the exposure time (since it is constant for all three images) and produce a curve that maps pixel value directly to scene irradiance. If we call this function $f(x)$, where x is our pixel value, then the three-step process described earlier can be reversed to map ME pixel values to HE pixel values, written as $g_{ME \rightarrow HE}(x) = f^{-1}(12.2f(x))$. The function $g_{ME \rightarrow HE}(x)$ is used to blend pixel values between the ME and HE sensors, and a similar function $g_{LE \rightarrow HE}(x)$ is used to blend between the LE and the HE sensors. Once the irradiance levels of the three images have been matched, we are ready to begin the merging process.

To explain our merging approach, we assume two registered LDR images (one high-exposure image I_{HE} and a second medium-

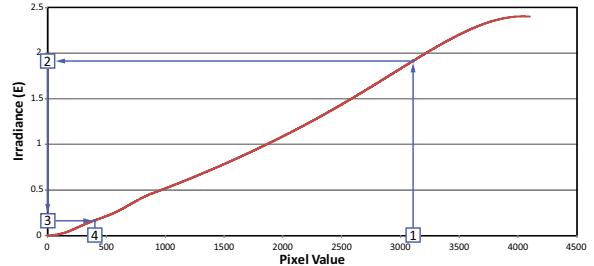


Figure 7: Camera response curve. This curve shows how the camera converts scene irradiance into pixel values. To compute what the ME pixel value should be for a given HE value, the HE pixel value (1) is first converted to a scene irradiance (2), which is next divided by our HE/ME attenuation ratio of 12.2. This new irradiance value (3) is converted through the camera response curve into the expected ME pixel value (4). Although this graph is approximately linear, it is not perfectly so because it is computed from the raw data, without significant smoothing or applying a linear fit.

exposure image I_{ME}) that are to be merged into an HDR image I_{HDR} . Our approach starts with the information in the high-exposure image I_{HE} and then combines in data from the next darker-exposure image I_{ME} , as needed. To reduce the transition artifacts described earlier, our algorithm works on each pixel location (x, y) by looking at the information from the surrounding $(2k + 1) \times (2k + 1)$ pixel neighborhood, denoted as $N(x, y)$. In our implementation, we used a 5×5 pixel neighborhood ($k = 2$), and we define a pixel to be *saturated* if its value is greater than 90% of the maximum pixel value (4096, for our sensor). We now specify the algorithm for each of the following 4 states for the pixel and its neighborhood:

Case 1: The pixel under consideration $I_{HE}(x, y)$ is **not saturated** and $N_{HE}(x, y)$ has **no** saturated pixels, so the pixel value is used as-is: $I_{HDR}(x, y) = I_{HE}(x, y)$.

Case 2: $I_{HE}(x, y)$ is **not saturated**, but $N_{HE}(x, y)$ has **1 or more** saturated pixels. Although most HDR merging algorithms would use this pixel as-is, we find that this calls into question the actual value of our pixel due to proximity effects (e.g., leakage or pixel cross-talk on the sensor). We therefore blend between the pixel value at $I_{HE}(x, y)$ and the one at the next darker-exposure $I_{ME}(x, y)$ depending on the amount of saturation present in the neighborhood. This is done in three steps:

1. Let U be the set of unsaturated pixels in neighborhood $N_{HE}(x, y)$, where $|U|$ is the number of unsaturated pixels.
2. Let $|N_{HE}(x, y)|$ be the number of pixels in neighborhood $N_{HE}(x, y)$. An interpolation coefficient α can be computed as $\alpha = |U|/|N_{HE}|$, which represents the fraction of unsaturated pixels in the neighborhood.
3. The output pixel is then given by the blended value: $I_{HDR}(x, y) = \alpha I_{HE}(x, y) + (1 - \alpha) g_{ME \rightarrow HE}(I_{ME}(x, y))$, where we use the pixel-ratio LUT to map the ME value into the HE range. This blends the measurement at the higher exposure $I_{HE}(x, y)$ with $I_{ME}(x, y)$ based on the number of saturated pixels in the neighborhood $N_{HE}(x, y)$.

Case 3: $I_{HE}(x, y)$ is **saturated** but $N_{HE}(x, y)$ has **1 or more** non-saturated pixels, which can be used to better estimate a value for $I_{HE}(x, y)$. We calculate the ratios of pixel values in the ME image between the unsaturated pixels in the neighborhood and the center pixel, and use this map of ME ratios to estimate the actual value of the saturated pixel under consideration. This is done in four steps:

1. As in Case 2, compute U , $|U|$, and the coefficient α

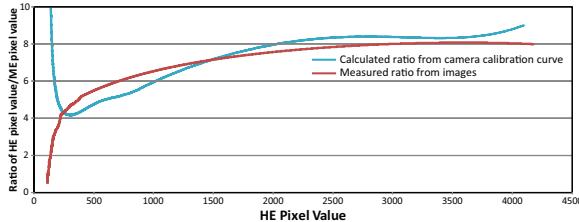


Figure 8: Pixel ratio curve. The blue curve is obtained when the 3-step process (Fig. 7) is applied to all red pixel values in the HE image using the camera response curve, and the ratio between HE and ME pixel values is calculated. The red curve is measured experimentally by taking the ratio of a large set of red HE and ME pixels in actual images captured by our system. The two match reasonably, validating our approach. This red curve is used by the algorithm as a look-up table (LUT) of scale factors to convert red ME pixel values to blend them smoothly with red HE pixel values. Similar curves were computed for the green and blue channels.

2. Compute a ratio map R of the ratios between the center pixel and each pixel of the neighborhood from the ME image: $R(x, y)_i = I_{ME}(x, y)/N_{ME}(x, y)_i$, for all pixels i in the ME neighborhood $N_{ME}(x, y)$.
3. Now compute an estimated $\hat{I}_{HE}(x, y)$ for the saturated pixel by scaling the unsaturated pixel values in the neighborhood N_{HE} with the ratios computed in step 2:

$$\hat{I}_{HE}(x, y) = \frac{1}{|U|} \sum_{i \in U} R_i N_{HE}(x, y)_i.$$
4. Finally, blend estimated $\hat{I}_{HE}(x, y)$ with $I_{ME}(x, y)$ using an equation similar to that of Case 2, step 3:

$$I_{HDR}(x, y) = \alpha \hat{I}_{HE}(x, y) + (1 - \alpha) g_{ME \rightarrow HE}(I_{ME}(x, y)).$$

Case 4: $I_{HE}(x, y)$ is **saturated** and **all** pixels in $N_{HE}(x, y)$ are saturated, so we do not have any valid information from the high-exposure image. In this case, we simply use the ME image and set $I_{HDR}(x, y) = I_{ME}(x, y)$.

This is our algorithm for combining data from two sensors. When there are three LDR images, the process above is simply repeated in a second iteration, substituting I_{HDR} for I_{HE} and I_{LE} for I_{ME} . In this manner, we merge data from the higher exposures as we work our way toward the lowest exposure, and only use data from lower exposures when the higher-exposure data is at or near saturation. We found that this algorithm reduces artifacts in the merged image considerably, as shown in Fig. 5.

The output of our algorithm is an HDR image that can be demosaiced and converted from pixel values to irradiance using a camera response curve similar to that of Fig. 7 accounting for all 3 color channels. The final HDR full-color image may then be tonemapped with commercial software packages (e.g., FDRTools, HDR Expose, Photomatix, etc.), or as described in the literature (e.g., [Smith et al. 2006]).

5 System Implementation

We implemented a prototype of our design using three Silicon Imaging SI-1920HD high-end cinema CMOS sensors mounted inside a custom camera body. Each 12-bit sensor was connected to a Dalsa/Coreco x64-CL iPro frame grabber in a PC. These sensors have 1920×1080 pixels (5 microns square) with a standard Bayer color filter array, and can measure a dynamic range of around 10 stops (excluding noise). The sensors were aligned by aiming the camera at small pinhole light sources, locking down the HE sensor and then adjusting setscrews to align the ME and LE sensors. In our prototype, the final image registration was accurate to less than

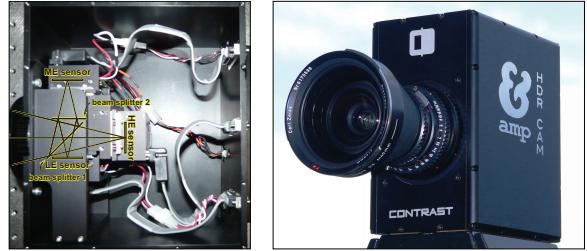


Figure 9: Completed camera prototype. (left) The optical path of our completed implementation. (right) The finished prototype used to acquire the video images in this paper.

5 microns with a rotation error of less than 0.1° . Note that, in general, the required registration accuracy is modulo the 2×2 pixel array of the Bayer pattern, because while pixels of one color in a sensor must match with the same color in another, a shift of any number of 2×2 Bayer patterns can be handled in software. We discuss how registration tolerances affect our results in Sec. 7.

The camera body has a Hasselblad lens mount to allow the use of high-performance, interchangeable commercial lenses. For beam-splitters, our current prototype system employs uncoated pellicle beam-splitters, such as the ones sold by Edmund Optics [part number NT39-482]. Although pellicle beam-splitters are not ideal for commercial products (they can be delicate and difficult to clean and maintain), we used them because of their low cost and accessibility. However, we found them to be robust for our application, as we were able to use our system successfully in various locations and environmental conditions. A picture of the completed camera prototype, which cost less than US\$15k in parts, is shown in Fig. 9.

We calculated the camera’s response curve as described by Devebec and Malik [1997] with a set of bracketed exposures from the HE sensor (Fig. 7). We also measured the ratio between the HE and ME pixel values experimentally for each color channel and observed that it matched the curve predicted by our 3-step process (Fig. 8). This curve was implemented as a LUT to translate values from the ME to the HE sensor range for merging, and a similar process was used to obtain and implement the HE/LE ratio curve. The LE sensor values are also adjusted with a slight offset before being used in the merging algorithm to account for a small amount of stray light.

The workflow for processing the data coming from the camera is shown in Alg. 1. Most of the steps are self-explanatory and are typical for HDR systems. The non-uniform correction (NUC) step is applied to each image to correct for non-uniformities in the sensor. This is a 2-point correction, which involves taking a dark field image for offset and a uniformly-lit diffuse scene for gain. The HDR merging algorithm presented in Sec. 4 was implemented in MATLAB. The demosaicing step used the algorithm of Malvar et al. [2004], and we tonemapped the final HDR images using HDR processing tools such as HDR Expose and Photomatix.

6 Results

Once the system was built, we performed a radiometric calibration test to measure its dynamic range. To do this, we aimed the camera at a stepped array of neutral density filters, in steps of 1 stop each, and focused down the sun’s light to an integrating sphere behind the array to provide a uniform, bright light source. A single frame of image data was captured with all 3 sensors simultaneously at an exposure time of $1/30$ second. Fig. 10 shows the result of this experiment. In our current prototype, we are able to clearly measure a dynamic range of over 2^{17} to 1, equivalent to over 17 photographic stops or over 100dB. More importantly, we demonstrate that our optical architecture can increase the dynamic range of these partic-

Algorithm 1 HDR Processing Workflow

- 1: Read in HE, ME, and LE raw images from the camera
 - 2: Apply NUC (2 point correction) and fix registration shifts modulo Bayer pattern
 - 3: Merge HDR image using algorithm in Sec. 4
 - 4: Demosaic HDR image to produce full-color image
 - 5: White balancing, color correction, conversion to HDR format
 - 6: Tonemapping and filtering (saturation, brightness, sharpness)
 - 7: Build final video from individual frames
-

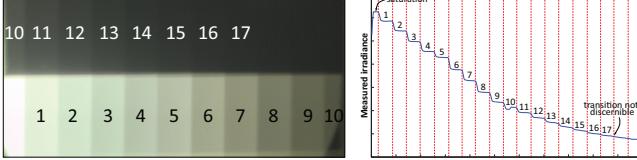


Figure 10: Dynamic range test. The dynamic range of our camera was tested by imaging a set of neutral density filters varying in transmittance by powers of 2. We focused the sun onto an integrating sphere behind the filters for uniform illumination. On the left is a tonemapped HDR image of the filters taken by our prototype. On the right is a log plot of the measured intensity across the HDR image (the kink in the curve at step 10 is due to the wrap-around in the image). The dynamic range of the camera can be measured by finding the last pair of filters where there is no noticeable difference in the measured intensity. This happens after filter 17, which means our prototype can capture over 17 stops of dynamic range.

ular camera sensors by 7 stops (7.7 was calculated from the design) using available optical and electronic components.

We also tested our proposed merging algorithm against that of Debevec and Malik [1997]. In every scene we tested, our approach was able to produce higher quality images than this algorithm because of the quantization artifacts it introduces with widely-separated exposure images (see Figs. 5 and 11). We found it useful to visualize the merging process by indicating which parts of the image were being selected from the different LDR images (Fig. 11).

Finally, Figs. 1, 11, and 13 show frames from video captured with our prototype and tonemapped in different ways. All videos were captured at 1920×1080 at 24 or 30 fps and show a wide variety of HDR scenes with complex movement and lens effects. The CUTTING TORCH scene in Fig. 1 shows an oxy-acetylene torch cutting a steel plate, tonemapped with Photomatix. The sparks are correctly motion-blurred because of the simultaneous capture on all sensors. We are also able to capture detail in the bright molten metal and the dark surroundings at the same time, suggesting that the proposed system could be used in industrial and manufacturing applications.

The BALLS & WATER scene in Fig. 11 shows rapidly-moving, colored objects all correctly motion-blurred in bright sunlight. The uniform-color spheres present a challenging gradient for the merging algorithm. The first two images of Fig. 13, SNOW DOG and WASH ME, were captured in the snow in direct sunlight. Filming snowy scenes in bright sunlight is notoriously difficult, especially with dark-colored subjects, yet our system captures detail in both the snow and the dark areas. The GO-CHIPS & FLAME scene shows an intense flame and a set of objects in direct sunlight and deep shadow. MELTING SNOW is a very challenging scene that is impossible to capture with conventional cameras, because it was taken from inside a dimly-lit room looking out at a bright outdoor scene. Our system can simultaneously capture both the wall decoration inside and the structure of the clouds outside which are several orders of magnitude brighter. This scene also features a focus-pull during the shot from the distant mountains to the foreground, which is difficult to do with systems that split the beam before the lens (e.g., [Cole and Safai 2010]).

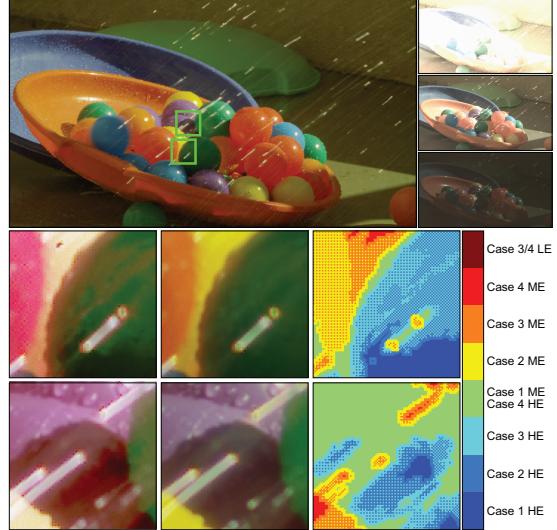


Figure 11: Another comparison of our merging algorithm. The top image shows the HDR image merged by our algorithm after tonemapping with Photomatix, with the 3 LDR inputs on the side. The bottom two rows show comparisons between Debevec and Malik’s algorithm (left) and ours (middle), along with a visualization showing how we select pixel values (right).

MONKEY shows a stuffed animal with bright white fur in direct sunlight. Fine details in the fur and the dark bucket are accurately captured. Finally, the LIGHTING TORCH scene shows a focus pull from a bright background to a dimly-lit subject in the foreground. A bright torch flame is added, which provides flickering illumination to the subject’s otherwise darkened face. These results all demonstrate the quality of images that our prototype system can capture.

7 Discussion

Sensor alignment and merging – Our prototype was aligned so that mis-registration between images was less than 5 microns. We note that mis-registration could cause sub-pixel artifacts in regions of sharp detail when the data is merged into HDR prior to the demosaicing process. However, in general, our spatially-blending merging algorithm helps reduce these by using values from a single sensor as much as possible, and only blending in data from other sensors when it approaches saturation. This limits the regions where blending mis-registered values would cause a problem to those that are near saturated pixels and have sharp detail. Furthermore, although there can be artifacts from merging the data prior to demosaicing, the color artifacts caused by merging data post-demosaic (which happens in areas where only some color channels are saturated) are more serious because they can happen in large, smooth regions in the image. Overall, mis-registration artifacts can be mitigated by simply using sensors with larger pixels. Also, the same challenge of sensor alignment is addressed in 3CCD systems for full-resolution color. These commercial products are typically aligned to very tight tolerances (much smaller than a single pixel), and we expect production versions of our proposed system to benefit from similar commercial alignment techniques.

Merging algorithm and noise reduction – Although our algorithm relies on data from only a single sensor until it approaches saturation to reduce quantization artifacts and noise from darker sensors, it might be possible to leverage information from the other sensors appropriately for further noise reduction. Of course, if our images were spaced very closely together, simple approaches like that of Debevec and Malik would reduce noise. With our spacing, it might

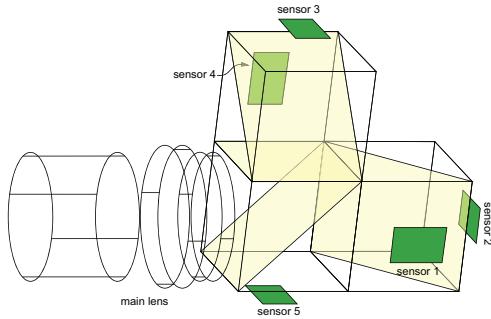


Figure 12: Extension of our optical architecture. Our optical architecture can be extended to more sensors for a better image quality or higher dynamic range. Here we show a design for a system with 5 sensors, where we essentially replace the HE and ME sensors with two other beamsplitters and a pair of sensors for each. The design was simulated with ZEMAX [2011].

be possible to combine our spatial-blending approach (which looks at a neighborhood of pixels, not just the pixel being merged) with merging algorithms that weigh other images based on their noise content (e.g., [Granados et al. 2010; Hasinoff et al. 2010]) to produce better results. This is an interesting subject for future work.

Flexible design – We present a simple blueprint for an HDR video imaging system using off-the-shelf components that can produce high-quality HDR video footage. The use of commercially-available sensors makes it easy to change them to extend the capabilities of the camera. Furthermore, the number of sensors can be extended with minor modifications to the optical path. Fig. 12 shows an optical design configured with 5 sensors, which would allow the camera to have an even higher dynamic range, or to produce higher quality images at the same dynamic range.

Toward better, low-cost consumer cameras – Although a system like ours could be used in high-end imaging applications such as feature film production, it could also impact the consumer camera market. First, the proposed architecture might enable high-quality, low-cost imaging because it allows three cheap sensors with limited dynamic range to capture images with dynamic range comparable to that of a single, high-end sensor. Second, HDR imaging would arguably benefit consumers more than professional cinematographers because the typical consumer does not have the lighting rigs or training needed to achieve HDR effects with LDR systems.

Limitations – Any HDR system (including the human eye) is ultimately limited in its ability to capture wide dynamic range by veiling glare in the optics [McCann and Rizzi 2007]. Our design is compatible with higher-performance lenses, which may help reduce veiling glare. Our optical architecture also provides the high-exposure sensor with only 92% of the light of the scene, which compares favorably with the 33% provided by previous internal beam-splitting systems, but is less light than would be captured with a single sensor and lens in a traditional camera.

Future work – In terms of the optical system, a study of thin-film coatings on the beamsplitters to reduce angle- or polarization-dependent effects (as discussed in Sec. 3.1) could be explored. The optical system would also benefit from improvements in the sensor alignment. In terms of the proposed merging algorithm, studies of noise optimization during the image merging process, in conjunction with our spatial-blending method, would make an interesting subject for future work. Finally, it would be useful to develop software packages for HDR cinematography for use with systems like the one proposed. These would give directors of photography the same kinds of tools in post-process that they are used to having at their disposal when lighting the set.

8 Conclusion

In conclusion, we have presented an optical architecture for high-dynamic range video that makes efficient use of the incoming light and produces cinema-quality images. We also proposed a novel HDR merging algorithm that minimizes quantization artifacts from the LDR data by using the highest-exposure data until saturation before spatially-blending in darker-exposure data. To test our design, we built a working prototype of the system and showed images from video acquired with the system.

Acknowledgment – We thank the anonymous reviewers for their helpful comments that improved the final version of the paper.

References

- AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM Trans. Graph.* 23 (August), 294–302.
- AGGARWAL, M., AND AHUJA, N. 2001. Split aperture imaging for high dynamic range. In *Proceedings of ICCV 2001*, 10 – 17.
- AGGARWAL, M., AND AHUJA, N. 2004. Split aperture imaging for high dynamic range. *International Journal of Computer Vision* 58, 7–17.
- AJDIN, B., HULLIN, M., FUCHS, C., SEIDEL, H.-P., AND LENSCHE, H. 2008. Demosaicing by smoothing along 1D features. In *Proceedings of CVPR 2008*, 1–8.
- BRAJOVIC, V., AND KANADE, T. 1996. A sorting image sensor: an example of massively parallel intensity-to-time processing for low-latency computational sensors. In *Proceedings of ICRA*, 1996, vol. 2, 1638–1643.
- COLE, A., AND SAFAI, M., 2010. Soviet Montage Productions. <http://www.sovietmontage.com/>.
- DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proceedings of ACM SIGGRAPH 1997*, 369–378.
- FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM Trans. Graph.* 21 (July), 249–256.
- GINOSAR, R., HILSENRATH, O., AND ZEEVI, Y., 1992. Wide dynamic range camera. United States Patent #5,144,442.
- GRANADOS, M., AJDIN, B., WAND, M., THEOBALT, C., SEIDEL, H.-P., AND LENSCHE, H. 2010. Optimal HDR reconstruction with linear digital cameras. In *Proceedings of CVPR 2010*.
- HARVEY, R., 1998. Optical beam splitter and electronic high speed camera incorporating such a beam splitter. United States Patent #5,734,507.
- HASINOFF, S. W., DURAND, F., AND FREEMAN, W. T. 2010. Noise-optimal capture for high dynamic range photography. In *Proceedings of CVPR 2010*.
- JAI. 2009. Using image fusion to capture high-dynamic range (HDR) scenes. Tech. Rep. TN-0903. JAI, Inc., Japan.
- KANG, S. B., UYTENDAELE, M., WINDER, S., AND SZELISKI, R. 2003. High dynamic range video. In *Proceedings of ACM SIGGRAPH 2003*, ACM, 319–325.
- KAO, W.-C. 2008. High dynamic range imaging by fusing multiple raw images and tone reproduction. *IEEE Transactions on Consumer Electronics* 54, 1 (Feb.), 10 – 15.

KELL, R., AND SZIKLAI, G., 1951. Simultaneous color television. United States Patent #2,560,351.

KRYMSKI, A., 2008. High dynamic range imager with a rolling shutter. United States Patent #7,397,509.

MALVAR, H., WEI HE, L., AND CUTLER, R. 2004. High-quality linear interpolation for demosaicing of Bayer-patterned color images. In *Proceedings of ICASSP 2004*, vol. 3, 485–488.

MANN, S., AND PICARD, R. W. 1995. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of Society for Imaging Science and Technology's 48th Annual Conference*, 442–448.

MCCANN, J., AND RIZZI, A. 2007. Veiling glare: the dynamic range limit of HDR images. In *Human Vision and Electronic Imaging XII*, SPIE.

MCGUIRE, M., MATUSIK, W., PFISTER, H., CHEN, B., HUGHES, J., AND NAYAR, S. 2007. Optical splitting trees for high-precision monocular imaging. *IEEE Computer Graphics and Applications* 27, 2 (march-april), 32–42.

MERTENS, T., KAUTZ, J., AND REETH, F. V. 2008. Exposure fusion: A simple and practical alternative to high dynamic range photography. *Computer Graphics Forum* 28, 1, 161–171.

MITSUNAGA, T., AND NAYAR, S. 1999. Radiometric self calibration. In *Proceedings of CVPR 1999*, vol. 1, 374–380.

MYSZKOWSKI, K., MANTIUK, R., AND KRAWCZYK, G. 2008. *High Dynamic Range Video*. Morgan & Claypool.

NAYAR, S., AND BRANZOI, V. 2003. Adaptive dynamic range imaging: optical control of pixel exposures over space and time. In *Proceedings of ICCV 2003*, 1168 – 1175.

NAYAR, S., AND MITSUNAGA, T. 2000. High dynamic range imaging: spatially varying pixel exposures. In *Proceedings of CVPR 2000*, 472 – 479.

NG, R., LEVOY, M., BRÉDIF, M., DUVAL, G., HOROWITZ, M., AND HANRAHAN, P. 2005. Light field photography with a hand-held plenoptic camera. Tech. Rep. CSTR 2005-02, Stanford University.

ROBERTSON, M. A., BORMAN, S., AND STEVENSON, R. L. 2003. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging* 12, 2 (april), 219–228.

SEGER, U., APEL, U., AND HÖFFLINGER, B. 1999. HDRC-Imagers for natural visual perception. In *Handbook of Computer Vision and Application*, B. Jähne, H. Haußecker, and P. Geißler, Eds., vol. 1. Academic Press, 223–235.

SMITH, K., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. Beyond tone mapping: Enhanced depiction of tone mapped HDR images. *Computer Graphics Forum* 25, 3, 427–438.

SIPHERONVR, 2011. <http://www.sipheron.com/>.

UNGER, J., AND GUSTAVSON, S. 2007. High-dynamic-range video for photometric measurement of illumination. SPIE, vol. 6501, 65010E.

WANG, H., RASKAR, R., AND AHUJA, N. 2005. High dynamic range video using split aperture camera. In *Proceedings of OMNIVIS 2005*.

ZEMAX, 2011. Optical design software. <http://www.zemax.com/>.

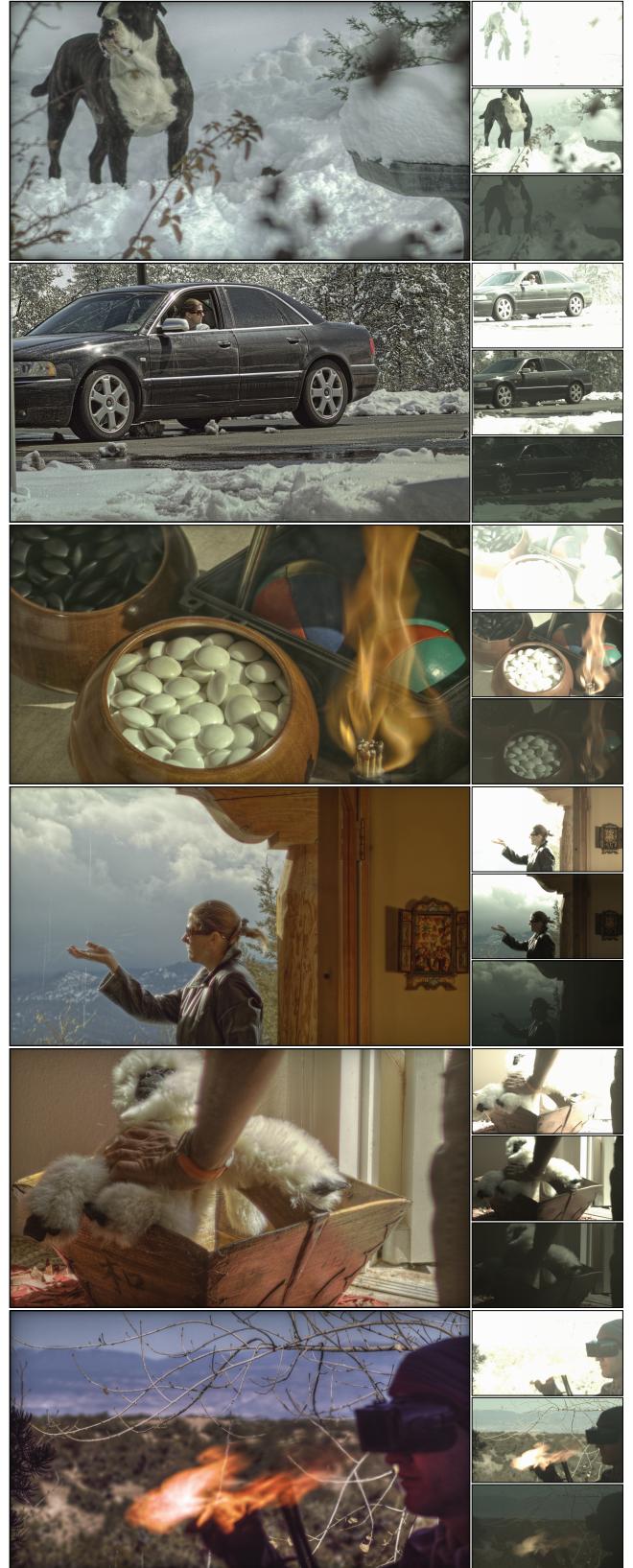


Figure 13: Images from videos captured with our camera. These are sample tonemapped images that demonstrate different imaging HDR effects. The LDR images captured by three sensors in our system are shown on the right. The scenes are, from top to bottom, SNOW DOG, WASH ME, GO CHIPS & FLAME, MELTING SNOW, MONKEY, and LIGHTING TORCH.