Multi-exposure Imaging on Mobile Devices

Natasha Gelfand
Nokia Research Center
natasha.gelfand@nokia.com

Andrew Adams
Stanford University
abadams@stanford.edu

Sung Hee Park
Stanford University
shpark7@stanford.edu

Kari Pulli
Nokia Research Center
kari.pulli@nokia.com

ABSTRACT

Many natural scenes have a dynamic range that is larger than the dynamic range of a camera's image sensor. A popular approach to producing an image without under- and over-exposed areas is to capture several input images with varying exposure settings, and later merge them into a single high-quality result using offline image processing software. In this paper, we describe a system for creating images of high-dynamic-range (HDR) scenes that operates entirely on a mobile camera. Our system consists of an automatic HDR metering algorithm that determines which exposures to capture, a video-rate viewfinder preview algorithm that allows the user to verify the dynamic range that will be recorded, and a light-weight image merging algorithm that computes a high-quality result directly on the camera. By using our system, a photographer can capture, view, and share images of HDR scenes directly on the camera, without using offline image processing software.

Categories and Subject Descriptors

I.4.1 [**Image Processing and Computer Vision**]: Digitization and Image Capture—*computational photography*

General Terms

Algorithms, Human Factors, Performance

1. INTRODUCTION

The contrast of many natural scenes that is visible to a human observer cannot be captured in a single photograph due to the limited dynamic range of the sensors found in modern digital cameras. Many amateur photographers are familiar with the frustration of trying to capture a portrait of a person in front of a beautiful sunset, or an image of a dark indoor scene which contains a bright window. In such cases, shadow and highlight areas of the image suffer from objectionable under-exposure and over-saturation artifacts.

A common approach to capturing such high-dynamic-range (HDR) scenes is to capture multiple differently exposed images and combine them together into a single high-dynamic-range result. Popular image editing programs such as Adobe Photoshop now include algorithms for creating high-dynamic-range images, and they are widely used by experienced photographers. This offline approach of creating merged images, however, is cumbersome for the average photographer.

While professional photographers know how to meter the highlight and shadow areas of the scene, estimate its dynamic range, and calculate the number of exposures needed, the best most amateurs can do is to use the "auto-exposure bracket" function on the camera. This does not always result in capturing the best set of exposures. If the images are merged offline, the user has no feedback on whether the correct set of images was captured, or the feedback comes too late to capture a new set with different parameter values.

To address this problem, a few digital cameras, such as Sony A550 and Pentax K-7, offer a multi-exposure mode, in which several images are captured, combined, and displayed to improve quality in the shadow and highlight regions. Both cameras use the standard "auto-expose and bracket" approach to metering and always capture the same number of photos (two for Sony and three for Pentax), regardless of the actual dynamic range of the scene. While these cameras solve one of the problems of consumer HDR imaging: being able to combine the images directly in-camera, the user still doesn't get a real-time preview of the total range of brightness that will be recorded. Such preview can be useful to help the user to correctly frame the scene, or even allow more experienced users to directly adjust parameters of the images to be captured.

This paper presents an automatic system for viewfinding and capture of high-dynamic-range scenes. Our application offers the following features:

- An automatic metering algorithm determines the exposure settings that capture the shadows and highlights, and the number of images that are required.
- A real-time viewfinder preview of all parts of the scene is available for framing the shot and examining the range of brightness that will be rendered in the final merged image.
- The merged image is created directly on the camera and presented for review. We adapt an existing image

MM'10, October 25–29, 2010, Firenze, Italy.

fusion algorithm to use less processing time and memory to make it feasible to process full-resolution images in-camera.

- The final result is provided as a regular (low-dynamic-range) image.

Our system is implemented on the Nokia N900 mobile phone, which we use as a programmable camera. The application behaves like a regular camera, therefore the average user does not need to know the details of our metering and merging algorithms to produce visually pleasing images. To our knowledge, this is the first application that deals with HDR scenes during both viewfinding and high-resolution capture.

2. HDR METERING AND VIEWFINDING

The first part of our application is the HDR viewfinding and metering mode that assists the photographer in composing the scene and capturing the input photos. Unlike the well-studied problems of HDR image creation and tonemapping, the problem of which images best capture the radiance of the scene has received comparatively less attention. With a few exceptions, image capture for HDR either uses the "auto-expose and bracket" mode available on many cameras, or the brute-force approach of capturing a range of exposures at pre-determined exposure increments until no clipped pixels remain.

Bilcu et al. [2] propose an algorithm for automatic bracketing that takes three photos: one whose exposure is metered using the standard camera auto-exposure algorithm, and two more whose exposure time is adjusted so that the under-exposed and over-saturated pixels found in the first image are well exposed. Hasinoff et al. [4] address the problem of computing the exposures and gains of images so that the noise in the final merged result is reduced. Kang et al.'s [5] automatic metering algorithm captures two images, one exposed for the highlights and one for the shadows, for the purpose of creating a high-dynamic-range video. This is similar to our approach, however, while their final video is created offline, we display a merged preview of the two frames in the camera's viewfinder. This allows the photographer to preview the full dynamic range that will be recorded in the final capture, assisting in composing the photograph. We are not aware of any other camera that provides such a real-time preview mode for HDR.

The metering and viewfinding component of our system operates as follows. We use the FCam programmable camera API [1] for the N900 to continuously stream two viewfinder-resolution (640×480) images, alternating between a short exposure to capture the highlights and a long exposure to capture the shadows. Every time an image is received from the sensor, we also receive its associated brightness histogram, which is computed automatically by the hardware ISP unit (image signal processor) in the N900. For each short exposure, we require that fewer than 10% of the pixels in the image are bright (have values above 239). If there are too many bright pixels, we decrease the exposure time for the subsequent short exposures. If there are too few (less than 1%) bright pixels, we increase the exposure time or gain. Similarly, for long exposure we require that fewer than 10% of pixels have values less than 16, otherwise we increase the exposure time or gain for the subsequent long exposures. In addition, we constrain the long exposure to be at most 6 stops above the short exposure, in order to avoid artifacts during the creation of the full-resolution photograph. When the long and short exposures stabilize, the metering is complete. Figures 1(a) and (b) show the short and long exposure frames from a viewfinder sequence.

While regular cameras directly display the captured image stream on the viewfinder, doing so in our case would result in distracting flickering as the long and short exposures alternate. In our system, the user can switch between viewing the long exposure, the short exposure, and the merged preview of the two, which is the default. To compute the viewfinder preview we first experimented with the method for HDR video described in [5]. While we discovered that it is possible to implement this method to run in real-time on the camera, the quality of the result is highly sensitive to the tone-mapping curve that is used to convert from HDR images to the 8-bit display. A better result was obtained by simply computing the viewfinder preview as the per-pixel average of the most recent pair of long and the short exposure frames. This does not require converting to HDR and then back to LDR and is extremely fast. Notice also that since the method is very fast, we do not need to do any tracking or alignment of the images. We stream the viewfinder at 30 frames per second, which makes the artifacts caused by misalignment due to handshake small enough that they are not visible on the low-resolution viewfinder preview. Figure 1(c) shows the merged result as the photographer sees it in the camera's viewfinder. While the colors are somewhat washed out, it is possible to see details within both the highlight and the shadow regions of the scene. Figure 1(d) also shows a single frame whose exposure setting is at the geometric mean of the long and short exposure. The higher contrast of this single middle frame causes loss of detail due to under-exposure and over-saturation.

3. HIGH-RESOLUTION CAPTURE

The second component of our on-camera HDR system is the high-resolution image capture and merging. We use the exposure values for the long and short exposures computed during viewfinding to determine which images need to be captured. In order to minimize capture and processing time, we do not capture the same pre-determined number of images for every scene (as existing cameras with HDR mode do), but instead use the total number of stops between the long and short exposures to determine how many images need to be captured.

The total exposure of each image is given by its exposure time and analog gain. During metering, we try to keep gain as low as possible to minimize noise, only raising it when the exposure time becomes so long that the resulting images may suffer from motion blur. Let (T_L, G_L) be the exposure time and gain parameters of the long exposure that captures the shadows and (T_S, G_S) be the parameters of the short exposure that captures the highlights. The total exposure is given as $E_L = T_L \cdot G_L$ and $E_S = T_S \cdot G_S$. The metering algorithm described in the previous section constrains the difference between the two exposures to be at most 6 stops, that is $\frac{E_L}{E_S} <= 2^6 = 64$. The reason for this limitation is that our camera hardware has enough memory to store and merge at most three high-resolution (5-megapixel) images, therefore 6 stops allows us to capture 3 images with 3-stop intervals. Using more than three stops between consecutive images makes image registration and merging difficult due

(a) (b) (c) (d)

Figure 1: Viewfinder metering and preview. (a) and (b) show short and long exposures respectively. (c) shows the merged result as displayed to the user at video frame-rate and (d) shows the single exposure whose parameters are the geometric mean of (a) and (b). Notice that our viewfinder captures more detail in both the shadow and highlight regions than any of the single exposures.

to too much difference between the inputs.

We select capture parameters for high-resolution images as follows:

- If $\frac{E_L}{E_S} < 2$, that is the long and short exposure are less than one stop apart, we capture only one image, with parameters $T_1 = T_S$ and $G_1 = G_S$.
- If $2 \leq \frac{E_L}{E_S} \leq 8$, that is the long exposure and short exposure are between one and three stops apart, we capture two images with the same parameters as the metered long and short exposures: $T_{1,2} = T_{L,S}$ and $G_{1,2} = G_{L,S}$.
- If $8 < \frac{E_L}{E_S} \leq 64$, we capture three images, where $(T_1, G_1) = (T_S, G_S)$, $(T_3, G_3) = (T_L, G_L)$ and $(T_2, G_2) = (\sqrt{T_S \cdot T_L}, \sqrt{G_S \cdot G_L})$.

Therefore, our application behaves like a regular camera for scenes with low dynamic range, and uses multiple images only when required.

4. IMAGE FUSION

One of the major attractions of modern digital cameras is the ability to preview the captured photo immediately after capture, therefore we perform the image merging directly on the camera. We do not wish to require that the photographer use a tripod, so our algorithm must be robust to camera motion during capture. While this was not an issue for viewfinding, where the low resolution and high frame rate make misalignment artifacts barely noticeable, such artifacts become objectionable in the high-resolution case. Therefore, we compensate for camera motion using the image alignment algorithm described in [9]

The input images are merged into the final result using an adaptation of the *exposure fusion* algorithm introduced by Mertens et al. [6]. Given a stack of input images (two or three in our case), exposure fusion computes a scalar-valued weight map for each image, and performs a weighted blend of the inputs to obtain the final result.

Let $(I_1 \ldots I_n)$ be the stack of n 5-megapixel input images captured by the camera. The weight for each pixel (i, j) of image k is computed as:

$$W_k(i,j) = \exp(-\frac{(I_k(i,j) - \mu \cdot 255)^2}{2(\sigma \cdot 255)^2}). \quad (1)$$

We set the parameters $\mu = 0.5, \sigma = 0.2$, and normalize the maps so that the sum of mask values for every pixel is 1.

The result consists of multiplying the images by their weight maps and blending:

$$R(i,j) = \sum_{k=1}^{n} W_k(i,j) \cdot I_k(i,j). \quad (2)$$

To produce a seamless result, the blending is computed using the multi-resolution algorithm of Burt and Adelson [3]. The blending computes a Laplacian pyramid of the inputs and a Gaussian pyramid of the maps, blends each layer separately, and then collapses the resulting Laplacian pyramid to obtain the result.

Exposure fusion is an attractive algorithm for producing HDR-like images on mobile devices since all computation is done using regular 8-bit images, without producing an intermediate high-dynamic-range image and then tonemapping it for display, as is done in most existing HDR approaches [7]. However, given the constrained memory of our target device, we make a further modification to the exposure fusion method which saves time, memory, and produces a higher-quality result.

The most time- and memory-consuming step of our algorithm is computing the weight maps and blending the images at the highest resolution of the image pyramid. It has been observed [8] that displaying an HDR image can be simulated by a low-resolution HDR base and a higher resolution LDR detail layer. Therefore, instead of applying exposure fusion to the full 5-megapixel inputs, we apply it to downscaled version of each image and then add the missing detail layer. Let L_i be the result of Gaussian filtering and subsampling each input image by a factor of k in width and height. Let U_i be the result of upsampling L_i back to its original resolution. The missing detail is given by $D_i = L_i - U_i$. We first compute a low-resolution merged image R_{low} by applying exposure fusion to $(L_1 \ldots L_n)$, and compute the final result as:

$$R(i,j) = R_{low}(i,j) + \alpha \cdot \max(D_1(i,j) \ldots D_n(i,j)) \quad (3)$$

that is, we add the maximum detail of all inputs to the low-resolution result. Setting α to a value greater than 1 has the effect of amplifying the detail layer, resulting in a sharpening of the resulting image. In our implementation, we set the subsampling factor $k = 2$ and the detail amplification $\alpha = 1.5$. The results of the fusion algorithm are shown in Figures 2 and 3. Table 1 shows the running time of our approach, and compares it with regular exposure fusion.

Although we can merge the images in under a minute, we also provide a low-resolution preview immediately after cap-

(a) (b) (c) (d)

Figure 2: Three-exposure sequence. The metered dynamic range was 4 stops. (a) – (c) show the short, middle, and long exposures, (d) shows our combined result.

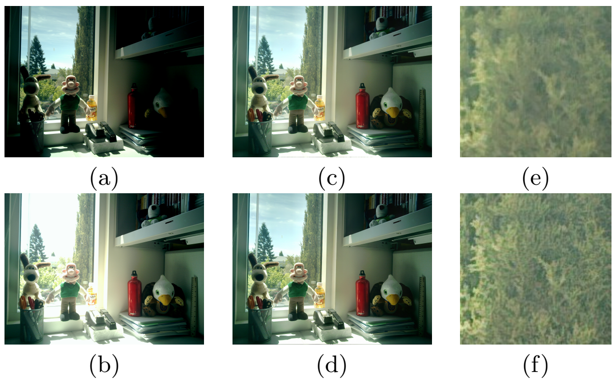

Figure 3: Two-exposure sequence. The metered dynamic range was 3 stops. (a) and (b) show the short and long exposures, (c) is the result of regular exposure fusion, (d) is the merged result using our base and detail algorithm, (e) shows a crop of the bushes computed by the regular algorithm, and (f) is a crop of our result. In addition to being faster, notice the sharper details of our approach.

Num. Images	Preview Total	Registration	Regular Fusion	Our Fusion	Highres Total
2	1.5	12.1	43.1	13.0	26.9
3	2.3	17.6	68.2	23.4	43.2

Table 1: Running time (in seconds) of our algorithm on two and three inputs. Preview resolution is 640×480, high resolution images are 2592×1968. Total time is the sum of registration, our fusion, and saving.

ture by subsampling the inputs to VGA resolution thumbnails and applying the regular exposure fusion algorithm to generate a quick preview. This way the user is able to see the result right after capture, while the full-resolution merge algorithm runs in the background as the user examines the preview or composes the next photo.

5. CONCLUSIONS AND FUTURE WORK

We have presented a system for metering, viewfinding, capture, and merging of images of high-dynamic-range scenes running on a programmable camera. The system provides a real-time HDR viewfinder, automatic HDR metering, and an image-merging algorithm that is fast enough to run directly on the camera, but effective enough to produce pleasing results.

Several areas are possible for future work. Given the hardware limitations of mobile devices, algorithm speed and memory optimizations are extremely important. Our current hard limit of at most three input images is dictated not by the algorithm we use, but by the amount of memory of the Nokia N900 cameraphone. While we can handle more images by using flash memory, the slow memory access on the N900 makes such approach unacceptably slow.

One promising approach to improve efficiency is to apply the image registration and merging algorithms directly to RAW inputs. We can also broaden the applicability of the method by incorporating an efficient de-ghosting algorithm, which would allow us to handle scenes with subject motion – the Achilles' heel of many algorithms that rely on merging multiple frames.

In the future, as the computing capabilities of digital cameras get better, and with the availability of programmable camera APIs such as FCam [1], we expect that multi-image algorithms implemented directly on the capture device will become increasingly popular.

6. REFERENCES

[1] A. Adams, E.-V. Talvala, S. H. Park, D. E. Jacobs, B. Adjin, N. Gelfand, J. Dolson, D. Vaquero, J. Baek, M. Tico, H. P. A. Lensch, W. Matusik, K. Pulli, M. Horowitz, and M. Levoy. The Frankencamera: An experimental platform for computational photography. *Proc. SIGGRAPH*, 2010.

[2] R. C. Bilcu, A. Burian, A. Knuutila, and M. Vehviläinen. High dynamic range imaging on mobile devices. *Proc. ICECS*, 2008.

[3] P. Burt and T. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communication*, 1983.

[4] S. Hasinoff, F. Durand, and W. T. Freeman. Optimal capture for high dynamic range photography. *Proc. CVPR*, 2010.

[5] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *Proc. SIGGRAPH*, 2003.

[6] T. Mertens, J. Kautz, and F. V. Reeth. Exposure fusion. *Pacific Graphics*, 2007.

[7] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High Dynamic Range Imaging.* Morgan Kaufmann.

[8] Brightside Technologies. High dynamic range display.

[9] M. Tico and K. Pulli. Image enhancement method via blur and noisy image fusion. *Proc. ICIP*, 2009.