MOTION-BLUR-FREE EXPOSURE FUSION

Marius Tico, Natasha Gelfand, Kari Pulli

Nokia Research Center, Palo Alto, CA, USA

ABSTRACT

We present a novel approach to HDR (high-dynamic-range) image fusion that copes with image blur degradation often present in long-exposed images. The proposed approach can deal with both camera and object motion blur in a computationally efficient manner suitable for implementation on a mobile device. The main idea is to exploit the differences between the image degradations that affect images captured for HDR fusion. Short-exposed images are mainly affected by sensor noise and less affected by motion blur, whereas longer exposed images are less noisy but potentially blurry due to motion during their exposure. Our approach consists of two steps. First we calculate an HDR representation of the scene by applying a typical HDR fusion approach. This representation could be blurry if some of the input images with long exposure time are blurry. We then fuse the HDR result with a photometrically modified version of the image with the shortest exposure time, which allows us to retain the sharpness of the short-exposed image and the noise-free characteristics of the HDR-fused image. The method does not assume an invariant blur PSF over any of the input images, and hence it can solve both for local and global blur, due to object or camera motion, respectively. We demonstrate the algorithm through a series of experiments and comparisons on natural images.

Index Terms— high dynamic range, motion blur, photometric calibration, multi-resolution fusion, exposure fusion

1. INTRODUCTION

Cameras have a limited dynamic range, much smaller than that of most scenes, or even the human eye. This means low-dynamic-range (LDR) images captured by cameras may lose detail in the very dark or bright areas of the scene. Several hardware and software techniques have been developed in order to capture the entire visual information present in a high-dynamic-range (HDR) scene [1]. The most common approach consists of fusing a set of differently exposed LDR images, that capture different regions of the dynamic range, into a single HDR representation of the scene radiance-map. The input LDR images can be captured by bracketing the exposure time: short-exposed LDR images capture the details in the bright areas, whereas long-exposed LDR images provide the visual information in the darker scene areas. Earlier approaches, such as Debevec and Malik [2], use two steps. First they build an HDR representation of the scene radiance by fusing multiple LDR images. Then the HDR radiance-map is tone-mapped back to an LDR representation to fit the dynamic range of the display or printer [1]. Another category of techniques consists of directly providing a tone-mapped LDR representation of the scene without an intermediate step of radiance-map creation. For instance, the "Exposure Fusion" (EF) approach by Mertens et al. [3] provides directly a tone-mapped LDR image from a bracketed exposure sequence, by applying a multi-resolution image fusion technique.

The main limitation of the HDR techniques based on fusing a bracketed exposure sequence consists of their inability to deal with image blur caused by either camera shake or objects moving during the long-exposure captures. Consequently, most HDR fusion methods have limited applicability in mobile consumer photography.

Some HDR fusion techniques can cope with ghosting artifacts due to objects motion [4, 5]. However, these approaches assume that all input images are sharp and only the objects in the scene may have changed their positions between different images. Consequently, these approaches cannot cope with motion blur inside individual frames due either to camera or object motion. The only HDR image fusion approach aiming to correct the motion blur artifacts was recently introduced by Lu et al. [6]. They use iterative multi-frame de-convolution that assumes a spatially invariant model for the blur, and cannot cope with blur caused by moving objects, or with ghosting artifacts. Moreover, due to the use of de-convolution in an iterative framework, the approach has a high computational complexity that may limit its applicability in mobile devices.

In this paper we build on noisy and blurred image fusion [7] and develop an algorithm for blur-free HDR imaging. The proposed method consists of two steps. First, the input LDR images are fused to create a pair of "blurred" and "noisy" HDR image representations. Second, the blurred and noisy images are fused to preserve the desired visual qualities of both, i.e., low noise from the blurred image, and edge sharpness from the noisy image. In our experiments we use the EF technique [3] for HDR image fusion.

We start by briefly reviewing the exposure fusion algorithm (Sec. 2). We then present our approach (Sec. 3) and experimental evaluation (Sec. 4), and finally conclude and summarize the paper.

2. EXPOSURE FUSION

We first briefly review the main principles of the exposure fusion technique [3]. The approach comprises two main operations: pixel weight calculation and multi-resolution image fusion.

Pixel weight calculation: The importance of each pixel of an input image can be weighted based on criteria such as exposure, local contrast, color purity, etc., [3, 8]. For instance, by analyzing pixel exposure, one would assign less weight to pixels that are either over-exposed or under-exposed and larger weight to pixels whose intensities are close to the middle of the intensity range. Another criteria could be the level of color saturation. By combining all these criteria one can build a weight map for each input image, where the weight corresponding to a pixel reflects the importance assigned to that pixel [3].

Multi-resolution fusion: The fusion process described in [3] is carried on in the transform domain, as inspired by Burt and Adelson [9]. Thus, each of the K input images I_k, $k \in \{1, 2, \ldots, K\}$, is decomposed into a Laplacian pyramid $\mathcal{L}(I_k)$, and the corresponding weight map W_k is decomposed into a Gaussian pyramid $\mathcal{G}(W_k)$. Next the Laplacian pyramid of the output image $\mathcal{L}(I)$ is determined

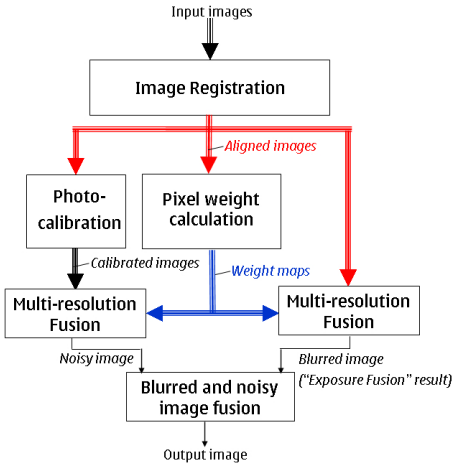

Fig. 1. The block diagram of the proposed method.

as the weighted average of the input Laplacian pyramids, where the weights are given by the Gaussian pyramids of the weight maps. Formally, this operation can be expressed as $\mathcal{L}(I) = \sum_k \mathcal{L}(I_k)\mathcal{G}(M_k)$, where the weighted sum is calculated for each transform coefficient independently. Finally the fused output image can be reconstructed from its Laplacian pyramid by inverse transform.

3. THE PROPOSED ALGORITHM

Figure 1 shows the diagram of the proposed algorithm. The input images are first registered. Next, the aligned images are submitted to three processes: photo-calibration, pixel weight computation, and multi-resolution fusion. In the photo-calibration step, the shortest-exposed image is successively calibrated with respect to all other input images, resulting in a set of increasingly brighter but noisier versions of the darkest image. The pixel weight calculation and the multi-resolution fusion operations are the same as in the exposure fusion techniques described above. In contrast to exposure fusion technique, we use the weight maps not only to fuse the original images, but also to fuse the photometrically calibrated images. At the end of these two multi-resolution fusion processes we obtain a noisy image and a potentially blurry image that are finally submitted to a blurred and noisy fusion operation.

3.1. Image registration

In the registration step, one image is chosen as reference (often the middle-exposed image), and all the other images are aligned with respect to it. The registration parameters that align two images are first estimated at coarse resolution, and are then progressively refined by matching corners at finer resolutions. At every resolution level we identify the best match for each reference corner feature inside the other image. Next, spurious corner correspondences due to noise, or moving objects in the scene, are eliminated by RANSAC based on an assumed motion model between the two images (we used affine motion model). Due to their different exposures, the input images exhibit significantly different brightness levels, therefore we use normalized cross-correlation as the similarity metric for corner matching.

3.2. Photometric calibration

Since the shortest-exposed image is usually the least affected by motion blur, we use this image as a sharpness reference to improve the result of the EF algorithm. In order to combine the visual information between the two images, we first need to photometrically calibrate one image with respect to another to match their brightness levels and color balances.

In this work we estimate a discrete brightness transfer function (BTF) between two images by analyzing a weighted version of their joint histogram. Given two images I_1 and I_2, the weighted joint histogram between them ($\mathbf{C}$) is

$$\mathbf{C}(i,j) = \sum_{\mathbf{x}} w(i)w(j)\delta[I_1(\mathbf{x}) - i]\delta[I_2(\mathbf{x}) - j] \qquad (1)$$

where δ stands for Dirac delta function; $\mathbf{x}$ denotes image pixel coordinates; $\mathbf{C}$ is an $L \times L$ matrix (usually $L = 256$); i and j correspond to brightness levels in I_1 and I_2 respectively; and the weight function w emphasizes the brightness levels that are closer to the middle of the range $[0, L]$. For instance, in this work we used a Gaussian weight function w of mean $L/2$ and standard deviation $0.4L$.

The brightness transfer function (BTF), assignes to each brightness level i from one image the most probable brightness level $j = BTF(i)$ from the other image. We start by constraining the BTF to pass through the highest entry (i_0, j_0) of $\mathbf{C}$, i.e., $BTF(i_0) = j_0$. Next, the remaining values of the BTF are calculated separately for $i > i_0$, and for $i < i_0$ by solving in each case the optimization problem $\max_{BTF} \sum_i \mathbf{C}(i, BTF(i))$. These two problems can be solved efficiently using Dynamic Programming and imposing the monotonically increasing condition on BTF function.

The approach of photo-calibration presented above does not work directly in calibrating the short-exposed image with respect to the EF result. This is because the EF result is no longer a uniformly exposed image, but it is rather a combination of differently exposed images. Consequently the optimal BTF function would no longer be spatially invariant, i.e., two pixels that have the same value in the shortest-exposed image may not necessarily map into the same value after calibration. In order to cope with this problem we build a photometrically calibrated version of the short-exposed image as a combination of multiple images. Denoting the shortest exposed image by I_1, and the remaining longer exposed images by $I_2, \ldots, I_K$, we first construct a set of photometrically calibrated versions of I_1 with respect to each one of the remaining images, i.e., $I_1^{(k)}$ the image obtained after calibrating I_1 with respect to I_k, for $k = 2, \ldots, K$. Next, we fuse the photometrically calibrated images $I_1^{(k)}$ using the multi-resolution fusion technique with the weight maps estimated in the exposure fusion process.

3.3. Blurry-Noisy image fusion

The blurry-noisy image fusion procedure is applied to combine the photometrically calibrated short-exposed image (I_n) and the result of the exposure fusion algorithm (I_b). The first one is affected by noise due to short exposure but it retains a sharp representation of the scene, whereas the result of the exposure fusion may be affected by camera and object motion blur. In order to retain the sharpness of I_n and the color representation of I_b we adopt different fusion procedures for the intensity and color components of the two images. For this we represent the two images in the YUV color space (similar results have been obtained also by using the CIE-Lab color space).

Intensity fusion: The intensity channels of the two images are fused in the wavelet domain extending our previous approach [7]. Thus,

Fig. 2. Example with object motion blur. Upper row: the input images. Bottom row: the exposure fusion result, the photometrically calibrated shortest exposure (i.e., noisy image), and our final result.

considering an orthogonal wavelet decomposition of the two intensity channels $\tilde{Y}_b$ and $\tilde{Y}_n$, and denoting their difference by $\tilde{D} = \tilde{Y}_b - \tilde{Y}_n$, we calculate the wavelet decomposition of the combined intensity channel $\tilde{Y}$ as a weighted average

$$\tilde{Y}(s,o,\mathbf{x}) = \tilde{Y}_n(s,o,\mathbf{x}) + w_{s,o,\mathbf{x}}\tilde{D}(s,o,\mathbf{x}), \qquad (2)$$

where the subscripts s and o stand for scale and orientation of a certain wavelet sub-band, inside which a coefficient is specified by its coordinates $\mathbf{x} = (x, y)$, and $w_{s,o,\mathbf{x}}$ is a positive weight (≤ 1).

As shown in [7] the MMSE (Minimum Mean Squared Error) estimate of the weight is given by $w_{s,o,\mathbf{x}} = \sigma_n^2(\mathbf{x})/E[\tilde{D}^2(s,o,\mathbf{x})]$, where $E[\cdot]$ stands for the expectation operator, and $\sigma_n^2(\mathbf{x})$ stands for the noise variance in the calibrated image at pixel $\mathbf{x}$.

We extend this approach by iteratively improving the weight $w_{s,o,\mathbf{x}}$ used in the previous estimate. Dropping the indices to simplify the notation, we can rewrite the weight as $w = \sigma_n^2/(E[(\tilde{Y}_b - \tilde{Y}^{(k)})^2] + \sigma_n^2)$, where $\tilde{Y}^{(k)}$ denotes the estimate of the output wavelet coefficients in the k^{th} iteration. In order to reduce the computational load we restrict the number of iterations to 5, as additional iterations do not seem to bring significant improvement.

The noise variance $\sigma_n^2(\mathbf{x})$ needed in the fusion process is estimated based on the digital gain applied at each pixel of Y_n with respect to the un-calibrated intensity channel Y_1 of the shortest-exposed image. Thus we have $\sigma_n(\mathbf{x}) = \sigma_1\,[Y_n(\mathbf{x})/(Y_1(\mathbf{x}) + \epsilon)]$, where ϵ is a small value used to avoid division by zero, and σ_1 is the noise level estimated in Y_1 as in Donoho and Johnstone [10].

Color fusion: Colors are captured better in the blurry image, except where they are affected either by significant blur or by ghosting artifacts caused by moving objects in the scene. Consequently we should take the colors from the blurry image in all pixels except the blurry image pixels. In this work we introduce an approach to identify the image areas affected by blur or ghosting artifacts based on analyzing the difference between the result of the intensity fusion $Y(\mathbf{x})$ and the intensity component of the blurry image $Y_b(\mathbf{x})$. In the non-blurred areas (i.e., apart of image edges and locally blurred objects) this difference is mainly caused by a residual noise that we model as zero-mean normally distributed with variance σ^2. Assuming that the image edges and locally blurred areas only occupy a fraction of the entire image, the sample median of the absolute difference between the two images is mainly determined by the residual noise. Consequently, in accordance to the theoretical relation between the standard deviation and the absolute value of a zero-mean normal distribution, we can estimate the residual noise standard deviation as $\sigma = 1.4826 \cdot \text{median}\{D(\mathbf{x})\}$, where $D(\mathbf{x}) = |Y(\mathbf{x}) - Y_b(\mathbf{x})|$.

Fig. 3. Example with ghosting artifacts in addition to blur caused by camera shake blur: exposure fusion (second row left), and our result (second row right). Both results have been obtained by fusing the three images (top).

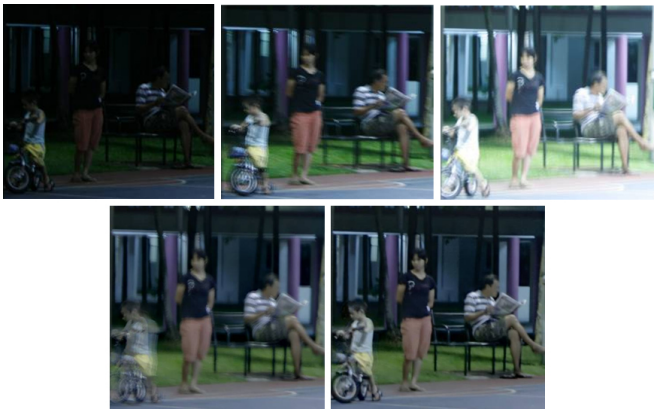

Fig. 4. Detail from input images (first row), exposure fusion result (second row left), and our result (second row right).

Next, we can classify a pixel $\mathbf{x}$ as a blurry (or ghosted) pixel if $D(\mathbf{x}) > \tau\sigma$, where τ is a constant that determines the extent of the confidence interval for the image difference in the non-blurred areas (in our experiments we used $\tau = 3$). In order to avoid artifacts due to few pixel misclassifications, we transform this decision into a fuzzy decision by assigning to each pixel a weight $w(\mathbf{x})$ that reflects its membership to the blurry pixel set. We use the weight function $w(\mathbf{x}) = 1/[1 + \exp(|D(\mathbf{x}) - \tau\sigma|/\sigma)]$. Finally, denoting by C_n a color channel from the noisy image (i.e., one of U_n, or V_n), and similarly by C_b a color channel from the blurred image, we calculate the corresponding color channel in the pixel $\mathbf{x}$ of the output image as $C(\mathbf{x}) = C_n(\mathbf{x}) + w(\mathbf{x})(C_b(\mathbf{x}) - C_n(\mathbf{x}))$.

4. EXPERIMENTAL RESULTS

We tested our approach on several natural images. One such example is shown in Fig. 2. We note that the second and third input images are highly affected by local blur, which is reflected also in the exposure fusion result shown on second row left. The photometrically calibrated short-exposed image which is the noisy version of our result (second row center) eliminates the blur but it is significantly affected by noise. Our final result is shown on second row right, and it clearly reduces both the blur and noise content in the

Fig. 5. Example of color selection weights. Dark areas correspond to blurry pixels and their colors are mainly taken from the noisy image.

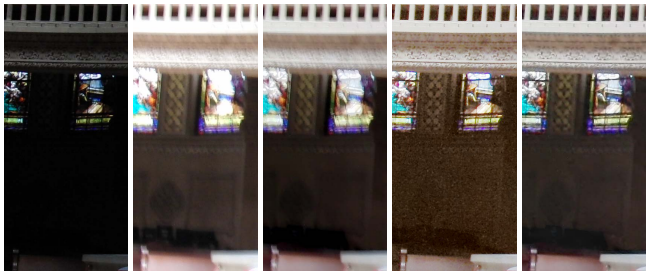

Fig. 6. Solving under-exposed areas in shortest exposed image. From left to right: two input images, the exposure fusion result, the photometrically calibrated shortest exposure (i.e., noisy image), and our result.

images.

A second example affected by both camera motion blur and ghosting artifacts is shown in Fig. 3. We note the effectiveness of our approach to reduce the blur artifacts present in the longer-exposed images, as well as the ghosting artifacts caused by people and objects moving during the image capture. A fragment from this result shown in Fig. 4 reveals that in addition to ghosting artifacts, long-exposed images are also affected by motion blur due to camera shake.

The proposed color fusion approach has been applied on both examples described before. The blurry image color weight maps for the two examples are shown in Fig. 5. We note that our approach detects the regions that are affected by blur or ghost artifacts and weights them less (darker in the image) than the other regions.

Fig. 6 exemplifies the way the algorithm behaves in areas where the shortest exposed image is significantly under-exposed. Thus, the under-exposed regions, which are lost in the noisy image, are reconstructed in the final result based on the visual information available in the blurry image. This illustrates the fact that the fusion between the two images is necessary since a single image de-noising method applied only onto the short-exposed image could not resolve the lost image areas.

Fig. 7. Comparison: method from [6] (left), and our result (right).

A comparison with the recently proposed HDR fusion approach of Lu et al. [6] is shown in Fig. 7. Theirs is the only method that provides a solution to the blur in HDR images, however it cannot cope with spatially varying blur or with ghosting artifacts. We ran 20 iterations of Lu et al. [6], and we tuned the parameters to provide the best visual result for this example. In this comparison we used the tone-mapping approach proposed in [11] in order to display the result of Lu et al. [6]. Our approach does not need tone-mapping, however, in order to facilitate the visual comparison, we applied the same tone mapping procedure to our initial image result (shown in Fig. 3). We note that besides solving the local motion blur and ghosting artifacts our approach provides a sharper and less noisy result.

5. CONCLUSIONS

We have introduced a method for capturing extended-dynamic-range images with a mobile camera. The main problem faced by such an application is the increased likelihood to have blur degradation in longer-exposed input images, due to either camera shake or objects moving in the scene. Our approach eliminates the blur and ghosting artifacts by fusing the input images in two stages. First, we create two versions of the final image: one noisy but unaffected by motion blur, and one potentially blurry but colorful and less affected by noise. In the second stage we fused the noisy and blurred images to preserve their desired qualities in the final picture. We tested our approach on several real images captured with different digital cameras.

6. REFERENCES

[1] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High dynamic range imaging: acquisition, display, and image-based lighting*, Morgan Kaufmann, 2006.

[2] P. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," in *SIGGRAPH*, 1997.

[3] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion," *Pacific Conf. on Computer Graphics and Applications*, pp. 382–390, Jan 2007.

[4] E. Khan, A. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *ICIP*, Oct 2006, pp. 2005 – 2008.

[5] O. Gallo, N. Gelfand, W-C. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *ICCP*, April 2009.

[6] P-Y. Lu, T-H. Huang, M-S. Wu, Y-T. Cheng, and Y-Y. Chuang, "High dynamic range image reconstruction from hand-held cameras," in *CVPR*, 2009, pp. 509–516.

[7] M. Tico and K. Pulli, "Image enhancement method via blur and noisy image fusion," in *ICIP*, Cairo, Egypt, 2009.

[8] M. Robertson, S. Borman, and R. Stevenson, "Dynamic range improvement through multiple exposures," in *ICIP*, Kobe, Japan, 1999, pp. 159–163.

[9] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, Jan 1983.

[10] D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[11] D. Jiang and Q. Guoping, "Fast tone mapping for high dynamic range images," in *ICPR*, 2004, vol. 2, pp. 847–850.