An efficient multiple exposure image fusion in JPEG domain

Hebbalaguppe, Ramya.; Kakarala, Ramakrishna.

2012

Hebbalaguppe, R., & Kakarala, R. (2012). An efficient multiple exposure image fusion in JPEG domain. Proceedings of SPIE - Digital Photography VIII, 82990H.

https://hdl.handle.net/10356/99101

https://doi.org/10.1117/12.907899

An efficient multiple exposure image fusion in JPEG Domain

Ramya Hebbalaguppe[a] and Ramakrishna Kakarala[b]

[a] Center for Digital Video Processing, CLARITY: Centre for Sensor Web Technologies, Dublin City University, Ireland;

[b] School of Computer Engineering, Nanyang Technological University, Singapore

ABSTRACT

In this paper, we describe a method to fuse multiple images taken with varying exposure times in the JPEG domain. The proposed algorithm finds its application in HDR image acquisition and image stabilization for hand-held devices like mobile phones, music players with cameras, digital cameras etc. Image acquisition at low light typically results in blurry and noisy images for hand-held camera's. Altering camera settings like ISO sensitivity, exposure times and aperture for low light image capture results in noise amplification, motion blur and reduction of depth-of-field respectively. The purpose of fusing multiple exposures is to combine the sharp details of the shorter exposure images with high signal-to-noise-ratio (SNR) of the longer exposure images.[1]

The algorithm requires only a single pass over all images, making it efficient. It comprises of - sigmoidal boosting[2] of shorter exposed images, image fusion, artifact removal and saturation detection. Algorithm does not need more memory than a single JPEG macro block to be kept in memory making it feasible to be implemented as the part of a digital cameras hardware image processing engine. The Artifact removal step reuses the JPEG's built-in frequency analysis and hence benefits from the considerable optimization and design experience that is available for JPEG.

Keywords: Image Fusion, JPEG, HDR Imaging, Image Stabilization, Computational Photography, artifact removal

1. INTRODUCTION

Advances in optics and digital computing have greatly aided progress in the field of computational photography. The relative ease with which high dynamic range (HDR) images of the highest quality may be generated has made it accessible to amateur photographers. Due to the growing trend towards production of high resolution images, which also reveal details in both dark and bright areas, several algorithms have been proposed to generate HDR images by computing a weighted average of a stack of low dynamic range images. Image fusion may be performed either in the spatial domain, by combining the relevant regions of two or more images, or in a transform domain such as the wavelet or Fourier transform, by combining the relevant transform components. We describe in this paper an efficient method for fusing images during the JPEG (JPEG refers to the 1992 image compression standard by the Joint Photographic Experts Group.) file writing operation, which, as we describe below, is essentially a transform domain combination. JPEG is the preferred compression method for the vast majority of digital cameras in use today, which means that the necessary calculations, including the discrete cosine transform (DCT), run-length coding and entropy coding,[3] are performed by embedded software and hardware blocks residing within a camera's onboard image processor. Therefore, an image fusion algorithm that reuses the calculations which are part of JPEG stands to benefit from the considerable optimization and design experience that is available for JPEG.

The need for image fusion arises in many applications, including digital image stabilization and high dynamic range (HDR) capture. In stabilization, the motion blur that occurs in a long exposure-time image is corrected by fusing it with a second, short exposure-time image of the same scene.[4] The benefit of fusion is that the higher signal-to-noise ratio (SNR) of the long exposure image is combined with the sharp details of short exposure image, giving digitally the stabilization that is normally available only through opto-mechanical means. In HDR photography, a set of images with varying exposure times or ISO settings[5] are fused to capture a wide range of scene luminance, which would otherwise result in saturated regions or dark features of interest in a single exposure.

Further author information: (Send correspondence to.)
Ramya Sugnana Murthy Hebbalaguppe : E-mail: ramya.hebbalaguppe2@mail.dcu.ie
Ramakrishna Kakarala: E-mail: ramakrishna@ntu.edu.sg, Telephone: (+65)6790 5925

Digital Photography VIII, edited by Sebastiano Battiato, Brian G. Rodricks, Nitin Sampat, Francisco H. Imai, Feng Xiao, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8299, 82990H · © 2012 SPIE-IS&T · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.907899

The published research in image fusion places greater emphasis on image quality than on complexity or suitability for embedded systems. We review representative works briefly. To avoid "ghosting" artifacts when combining images containing moving objects E. Khan et al,[6] employ a weighting function, based on the probability of a pixel belonging to a moving object, to attenuate the contribution of moving pixels. The algorithm requires multiple iterations to converge. Lu et al[7] also present an iterative method for HDR image fusion, which relies on deconvolution to correct motion blur. Work done by Reinhard et al[8] use the variance across different exposures to determine the likelihood of a pixel resulting in ghosting; however due to slight misalignments and errors in camera calibration, the variance at pixel level is often noisy. There have been techniques to enhance the dynamic range of monochromatic and color images and videos by using optical flow estimation as a means of per-pixel registration by Bogoni[9] and Kang et al[10] . Jacobs et al[11] propose a local entropy-based method between different low dynamic range images to detect motion in a sequence. O. Gallo et al[12] select a reference image from the stack and detect regions from each of the other exposures that do not cause artifacts when combined with the reference image. The resultant HDR image is basically the collage of the patches which do not cause ghosting. M. Tico and K. Pulli[4] and M. Tico et al[13] describe a method for fusing a pair of short and long exposure images that works in the wavelet domain, which is useful for either image stabilization or HDR capture.

All of the previously mentioned methods requires at least the following: an entire image to be kept in memory, two or more passes over each pixel, and significant computation (such as a wavelet transform) that is not required for other parts of the camera's normal image processing pipeline. While it is still possible to implement some of those methods in a mobile device, as demonstrated by Gelfand et al,[1] there is no published work describing a minimally complex fusion algorithm suitable for low-cost embedded cameras. The purpose of this paper is describe such an algorithm; unlike previous work, we focus on minimizing computation while providing reasonable image quality. In our previous work, R. Kalarala and R. Hebbalaguppe,[14] we demonstrate fusion of a pair of images while this paper introduces a method to fuse more than two images. Our algorithm is efficient in both computation and memory: it requires only a single pass over each pixel, and requires at most a single JPEG macroblock from the short-exposure image to be kept in memory. It reuses the DCT calculations already performed for JPEG encoding for a novel purpose: motion artifact detection. To further reduce complexity, the brightness transfer function between images is approximated by a lookup table implementing a parametric sigmoidal function.

The relevant background material for this paper include the operations of a digital camera's embedded image processing "pipeline" as described by J. Nakamura,[15] as well as the details of JPEG image compression and file storage as described by W. L. Pennebaker and J. L. Mitchell.[16]

Figure 1 summarizes the steps involved in our proposed method of image fusion. We describe the steps briefly in section 2, and elaborate on their implementation in the following sections.

2. OUR PROPOSED ALGORITHM

As mentioned in the introduction, our algorithm uses three images and fuses them in the JPEG domain. We assume that the images are taken in immediate succession to minimize the need for registration, and that the exposure ratios between them are known prior to the second image being taken. That assumption is reasonable for cameras operating in exposure bracketing mode, in which the ratios are set beforehand; typically the ratios are powers of two, but the values may be programmable. We assume that the exposures vary only in exposure time, while aperture and ISO setting are held constant. We use the well-known logarithmic exposure value (EV) notation of describing ratios of exposures. In that notation, $\mathrm{EV}(0)$ represents the exposure time determined by the camera's autoexposure routine, and relative to that time, $\mathrm{EV}(\Delta)$ represents an exposure time that is 2^{Δ} larger. For example $\mathrm{EV}(+1)$ is twice the exposure time of $\mathrm{EV}(0)$, and $\mathrm{EV}(-1)$ is one-half the exposure time of $\mathrm{EV}(0)$.

It is well-known that the human visual system is more sensitive to details in the luminance channel, which in the JPEG literature is denoted as Y in the $\mathrm{YC}_b\mathrm{C}_r$ color system. JPEG takes advantage of this by subsampling the two chrominance channels, denoted C_b, C_r, usually by a factor of two in each direction.

In our algorithm, short, long and normal exposure images are used, and we require that the long exposure image be taken prior to the shorter exposure images. The long exposure image is processed as normal on the digital camera, converted to JPEG, and then written to a file in secondary storage such as flash memory. Our algorithm makes no changes to the processing of the long exposure image. The short and normal exposure images, which is taken subsequently, is boosted in brightness by using a lookup table (LUT) modelling a sigmoidal function that is described in Section 3.This step may be

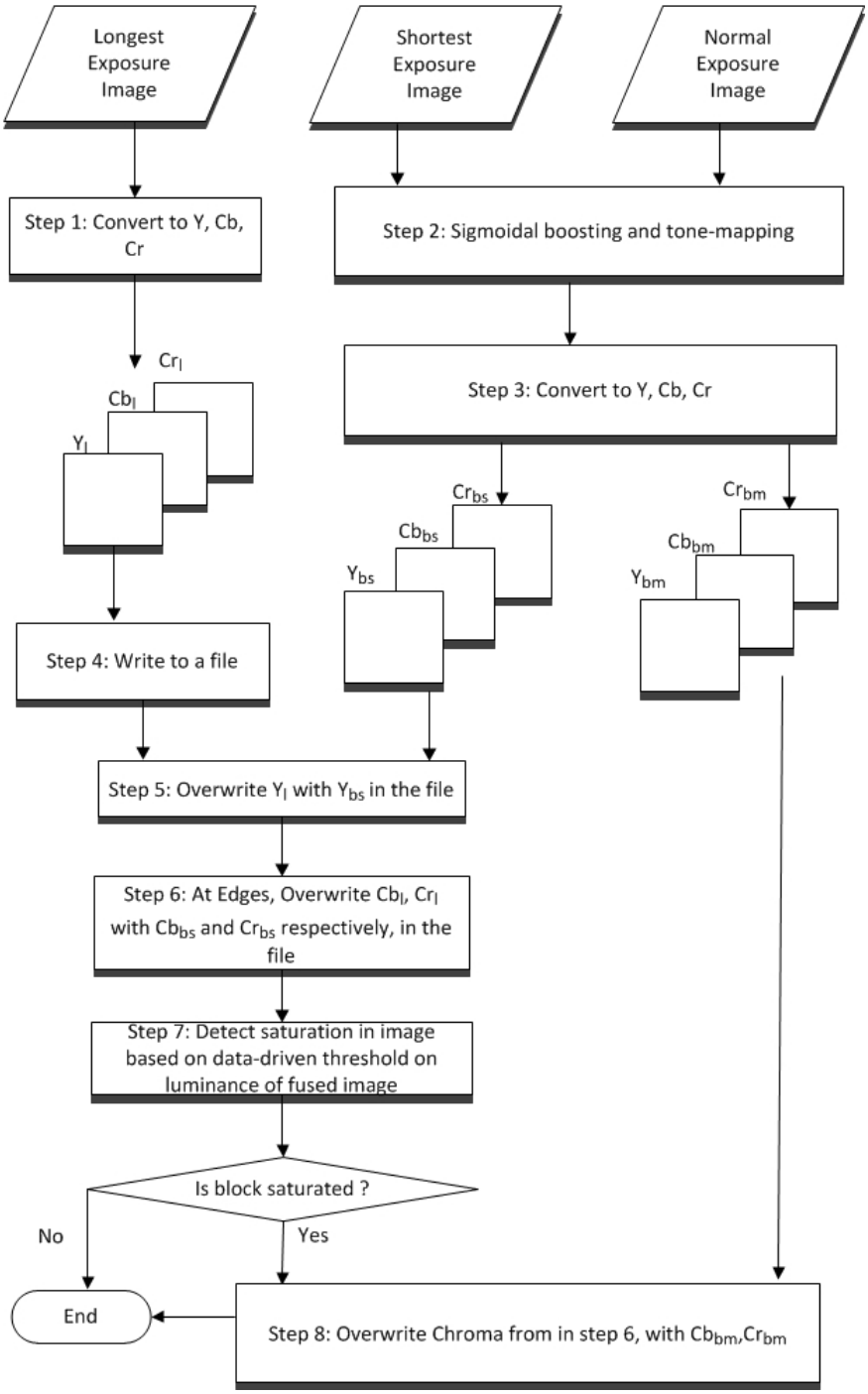


Figure 1. The proposed methodology for image fusion in the JPEG domain is shown. As explained in the text, the algorithm requires only a single pass on long , mid(normal) and short exposed images and requires in principle no more than a JPEG macro-block to be kept in main memory.

combined with tonemapping, which will be discussed later. The boosting process requires only one LUT transformation per pixel. Boosting is applied to all pixels in the image, and the boosted image is converted to JPEG as usual. However, instead of writing a separate JPEG file for the short-exposure image, we modify the file write operation to selectively overwrite parts of the long exposure's JPEG file with relevant portions of the short-exposure JPEG. Since the JPEG file interchange format (JFIF), which is the standard file format for storing JPEG images, places the luminance (Y) channel in separate 8×8 blocks of DCT values from the corresponding chrominance $(\mathrm{C}_b, \mathrm{C}_r)$ blocks, it is relatively easy to calculate the addresses of the two types of blocks within the file, and to overwrite selectively. If portions of the long-exposure JPEG file are replaced with suitable portions of the short-exposure JPEG, then the result, upon decompression and display, is a fused image. Hence, the JPEG file format serves as a medium for image fusion, and the JPEG decoder performs the actual fusion prior to display.

We take advantage of the JPEG file format to fuse the three images as follows. Specifically, we overwrite the Y blocks of the long-exposure JPEG file with the corresponding Y blocks of the *boosted* short exposure image. Hence the luminance of the fused image, in our algorithm, is exactly that of the boosted short exposure image. However, simply combining the luminance of one image (short exposure) with chrominance of a longest exposed image , which is taken earlier in time, would clearly lead to artifacts due to motion between the frames. Normally, motion is compensated by means of registration as in M. Tico and K. Pulli,[4] which uses for example, cross-correlation, to shift the second image in order to align with the first prior to fusion. In our algorithm, which is designed for minimal computation, we do not use image registration. Instead, we make use of the fact that motion artifacts are most noticeable at edges, and that the human visual system is less sensitive to motion artifacts in the chrominance domain than in the luminance domain. Our method of correcting for motion artifacts selects the 8×8 chrominance $(\mathrm{C}_b, \mathrm{C}_r)$ blocks of the long-exposure image where the corresponding luminance (Y) block contains significant high frequency information representing edges or texture, and replaces them with the corresponding $(\mathrm{C}_b, \mathrm{C}_r)$ blocks of the boosted short-exposure image. The method of detecting blocks to be replaced is described in Section 4. Image saturation is detected in the case of day light scenes. As shown in Figure 1, saturation is detected on fused image (containing luminance from short exposed image and chrominance from long exposed image), block by block basis based on a data-driven threshold as explained in sections below. If the image is saturated, chrominance from the mid(normal)-exposed image replaces the chrominance in the fused image.

In principle, the above procedure does not require any part of either the long exposure or the shorter exposure images, to be stored in RAM during fusion. However, we assume that JPEG encoder contains a macroblock of information from 16×16 pixel region, which is divided into four 8×8 blocks for luminance Y, and for subsampled chrominance, a total of two 8×8 blocks for the C_b, C_r channels. Furthermore, note that the processing of the long exposure image is not altered, and that each pixel of the short-exposure image is processed only once, through the LUT boosting operation. Our principal modification is therefore to the JPEG encoder, and more importantly, to the file save operation. The next sections describe the steps in the algorithm in more detail.

3. IMAGE BOOSTING USING A SIGMOIDAL FUNCTION

When images of varying exposure are to be fused, the luminance of the shorter exposure images is boosted to match that of the longer one by estimating a compensating function that matches the camera's response. Following M. Tico and K. Pulli,[4] we call the compensating function the brightness transfer function (BTF). The BTF may be estimated by plotting the pixel values of longer exposed image against the corresponding luminance values in the shorter exposed image, and applying basic curve fitting as shown in Reinhard et al.[8] Another option is to calculate the BTF by smoothing the mean of pixel values for the longer exposure for the same pixel value in the shorter exposure, as proposed in M. Tico and K. Pulli.[4] Both methods require two passes over the shorter exposure image, the first to estimate the BTF and the second to boost the image accordingly. However, in our experiments, we found that the BTF varied primarily as a function of the EV difference between the images, which suggests a simpler way to estimate the BTF as we now describe.

Specifically, we found by testing a variety of cameras and EV differences that the BTF can be fit using a parametric sigmoidal curve. The advantage of using such a curve is that it requires only a single pass over both images. The sigmoidal function, proposed for tone-mapping in X. Zhang et al,[2] has the form

$$f(x; M, \alpha, \beta) = \frac{M}{1 + \beta e^{-\alpha x}}. \tag{1}$$

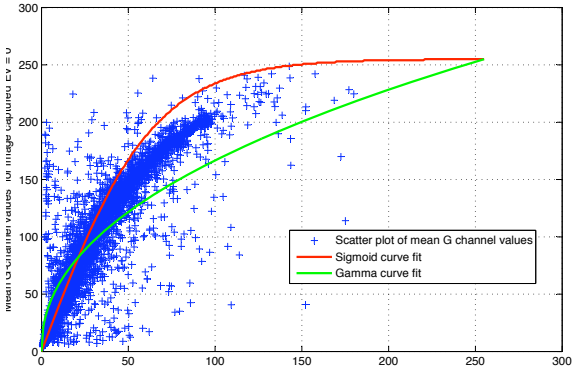


Figure 2. Sigmoid curve fit for a scatter plot of green channel local mean values (over 8x8 blocks) for an EV(-2) image against corresponding values for the EV(0) image in a short-long exposure pair for a Canon 400D camera. For comparison, the gamma curve for $\gamma = 2.2$ is shown.

As $x \to \infty$, this monotonically increasing function reaches towards M where M is the maximum pixel value, ranging from 255 to 4095 depending on the bit-depth of the pixel data. The slope of the function is controlled by α and its location by β. To best match our experimental results, we set $\beta = 1$, and define the scaled function

$$f_s(x; M, \alpha) = 2f(x; M, \alpha, 1) - M. \tag{2}$$

Note that $f_s(0) = 0$, and that $f_s'(0) = M\alpha/2$, so that the slope is proportional to α. We found the best match to our experimental results to be obtained by setting

$$\alpha = \frac{2(\Delta\text{EV} + 1)}{M} \tag{3}$$

Here ΔEV is the difference in exposure values between the long and short exposures. Note that using (3), we have the initial slope at the origin as $f_s'(0) = (\Delta\text{EV} + 1)$, and that f_s has an asymptotic value of M as $x \to \infty$.

In Figure 2, we show f_s plotted on the same axes as a scatter plot for the mean green channel values (mean computed over 8x8 blocks) of the two different exposure images (EV(0) and EV(-2)). The sigmoid function shows a good fit and works similarly to the BTF as computed in M. Tico and K. Pulli,[4] but with much lower memory requirement and computational time. Because the curve is parametric, and the ΔEV is known prior to taking the long-short exposure pair, the value of function may easily be computed to populate a LUT prior to the short exposure image. As the pixel values of the short exposure are read out, they are boosted by replacing their values with the corresponding values in the LUT.

It is worth comparing the sigmoidal curve to other parametric functions which are used in brightness compensation. Figure 2 shows for $\gamma = 2.2$ the gamma curve given the equation

$$V_{out} = V_{in}^{\gamma} \tag{4}$$

The gamma curve rises too steeply for low values, and levels off too quickly for high values, resulting in an under fitting of the BTF.

We found that the sigmoidal function fit the BTF reasonably well for different makes of recently-manufactured digital SLR cameras including the Canon 400D, Pentax K7, Olympus E-450, Nikon D40 and point and shoot cameras like Sony DSC-W30, Casio EX-F1. Figure 3 shows scatter plots containing the experimental results. In each case, the scatter plots are produced by using a longer exposed image along the vertical axis and a shorter exposed image along horizontal axis, where $\Delta\text{EV} = 1, 2, 4$ were chosen as explained in the Figure 3.

Tone mapping may be applied after sigmoidal boosting for daylight scenes if there are large regions that appear to be nearly saturated in the boosted short exposed image. This situation occurs especially when ΔEV is large. In our daylight results, shown below, we applied a simple global tone map (the same across the image) to the short-exposure image, and also limited the value of α in (3) to be at most $6/M$. Note that it is possible to automatically distinguish between nighttime

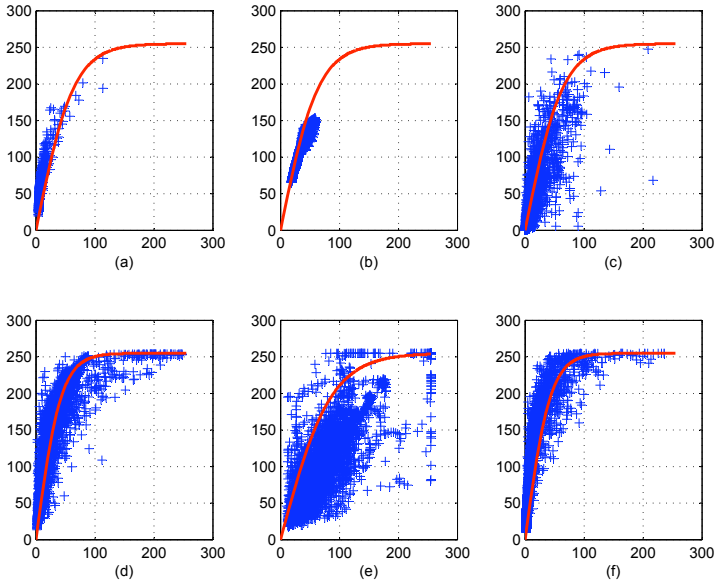


Figure 3. Results of sigmoidal boosting as applied to various camera makes: (a) , (b) and (c) are sigmoid curve fits for green channel local mean values for an EV(-2) image plotted against the EV(0) image for the Nikon D40, Pentax K-7 and Canon 400D cameras respectively; (d) sigmoid curve fit for green channel local mean values for an EV(-2) image plotted against the EV$(+2)$ image in a short-long exposure pair for Casio EX-F1 camera; (e) sigmoid curve fit for green channel local mean values for an EV(-1) image plotted against the EV(0) image in a short-long exposure pair for Olympus E-450 camera; (f) sigmoid curve for green channel local mean values for an EV(-2) image plotted against the EV$(+2)$ image in a short-long exposure pair for Sony DSC-W30 camera.

and daylight scenes by looking at the exposure time while the shot is framed in the viewfinder, or meter reading if the camera has one. The tone map is implemented as a LUT as with the sigmoid function. The LUT for the tone map may be combined with the LUT for the sigmoid for efficiency.

We illustrate our algorithm using an example of short and long exposure pair as in Figure 4 and using three images as shown in Figure 6. In Figure 4, we illustrate the steps involved in the proposed algorithm by applying them to hand-held images captured at two different exposures under low light. In this scenario, we are performing digital image stabilization by combining the sharp details of the short exposure with the high SNR of the long exposure. The captured images contain some artifacts: in the long exposure in Figure 4(a), note the motion blur artifacts around the lady on the left and the child on the right. In the next step, we combine the luminance Y of the boosted shorter exposure image (using the sigmoidal function (2) in the same JPEG file with the chrominance (C_b, C_r) of the longer exposure image. Note that the long exposure file has been written first, and we are overwriting its luminance components with those of the boosted short-exposure image. The resulting image after decompression is shown in Figure 4(c), but contains several artifacts, in particular around the lady's legs and the child's hands. To remove those artifacts, we detect blocks containing significant high frequency information using the method described below in Section 4, and replace the chrominance for those blocks in the long exposure with those of the boosted short exposure image. We do not need to store either image in main memory. We write the JPEG file for the long exposure first, and as we are writing the compressed information for each block of the boosted short exposure, we decide on a block basis whether to overwrite the chrominance of the long exposure with that of the short. Note that we always overwrite the luminance of the long exposure with that of the short exposure.

4. ARTIFACT REMOVAL

As noted in Section 2, the chrominance from longer exposure image, denoted by Cb_l, Cr_l is merged with the boosted luminance of shorter exposed image denoted by Y_{bs}. The merged image (Y_{bs}, Cb_l, Cb_l) contains motion artifacts like ghosting and color bleeding due to mismatch of luminance and chrominance values at the edges from images taken at two different times. Below, we propose a novel technique for detecting and removing those artifacts in the JPEG domain.

(a) Image captured at EV(0).

(b) Image captured at EV(-2).

(c) Combining chrominance of (a) and boosted luminance of (b).

(d) Replacement map.

(e) Final HDR image after artifact removal.

Figure 4. Results of the proposed algorithm on a pair of short and long exposure images of a basketball court from a hand-held camera in low light. The image in (a) is captured at EV(0) ,or normal exposure (note motion blur around lady's legs); (b) is captured at EV(-2), or deliberately underexposed by a factor of 4; (c) shows the result of fusing luminance from boosted short exposure image and chrominance from long exposed image; (d) magenta pixels indicate 8×8 blocks where the chrominance of long exposure image is replaced by the corresponding chrominance of the short-exposure (e) shows the final result after applying the replacement. Note that the ghosting artifacts observed in (c) around the lady's legs on the left and the child's hand on the right are removed; closeups of those regions are shown in the subsequent figure.

(a) Before artifact removal (b) After

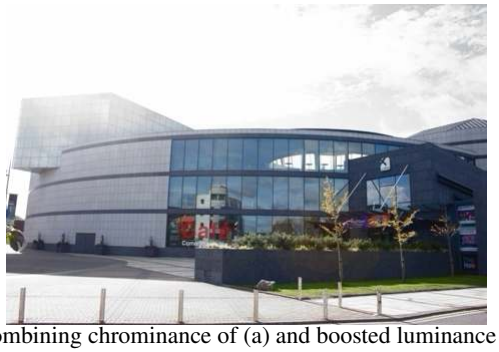
(c) Before artifact removal (d) After

Figure 5. Closeups of selected regions of the basketball court example to illustrate the effectiveness of ghosting artifact removal. Part (a) shows the ghosting artifacts visible as a blue discoloration on the lady's legs. Part (b) shows the result after artifact removal; notice that the skin tone is restored. Parts (c) and (d) show respectively another before and after example of artifact removal. Again, notice that the blue artifact that is noticeable on the baby's shirt and legs is removed.

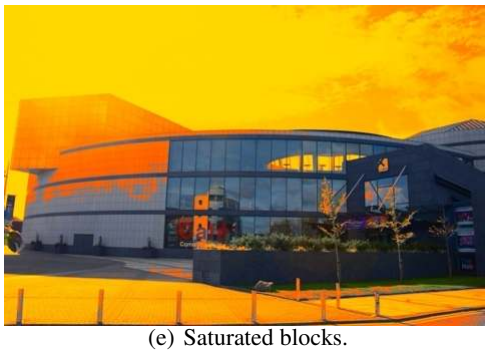
(a) Image captured at EV$(+2)$.

(b) Image captured at EV(0).

(c) Image captured at EV(-2).



(d) Combining chrominance of (a) and boosted luminance of (c).



(e) Saturated blocks.

(f) Three image fusion.

Figure 6. Results of the proposed algorithm on three images, of Helix theater located on Dublin City University,from a hand-held camera in day light. The image in (a) is captured at EV$(+2)$, or long exposure exposure (note the image saturation on the left portion of helix building has blown out the details); (b) is captured at EV(0), or normal exposed image; (c) is captured at EV(-2), or deliberately underexposed exposed image by a factor 4; (d) shows the result of fusing luminance from boosted short exposure image and chrominance from long exposed image with artifact removal; (e) Orange pixels indicate 8×8 blocks which are detected as saturated blocks as explained in image saturation detection section.(f) shows the final result after applying the replacement. Note that the image is neither underexposed nor overexposed revealing better colors on the building and sky compared to (d) which is the result of short and long exposure image fusion.

Our technique for artifact removal works on the shorter exposure image and takes advantage of JPEG's built-in frequency analysis using DCT, to perform texture or edge detection. That DCT coefficients may be used to detect 8×8 blocks containing strong edges is well-known; see W. L. Pennebaker and J. L. Mitchell[16] and R. Kakarala and R. Bagadi.[17] The JPEG compression algorithm computes the DCT of each 8×8 block, which are then quantized and subsequently compressed by using run-length encoding in a "zig-zag" scan. The End of Block (EOB) symbol, which occurs in every block as part of JPEG syntax, indicates the location of the last non-zero AC coefficient in the 64-coefficient zig-zag scan. Since our goal is minimal computation, we note that classification of blocks as smooth or edge blocks may be accomplished simply by noting whether the EOB signal occurs early or late in the scan. We apply the detection to the EOB location in the DCT coefficients of the luminance blocks Y_{bs} of the boosted short-exposure image. An empirically determined EOB location threshold of 15 (out of 64) is used for all experiments in this paper, and the quantization matrices used are the defaults described in W. L. Pennebaker and J. L. Mitchell.[16] If the EOB occurs after this threshold location, then the block is classified as an edge block for which the chrominance comes from the short exposure, otherwise it is considered a smooth block and the chrominance from the long exposure is used.

For macroblocks where an edge is detected in the Y_{bs} component, the chrominance from the boosted shorter exposure (Cb_{bs},Cr_{bs}) image is used to overwrite the corresponding values from the longer exposed image in the merged HDR image. In the standard $4:1:1$ color format, there are four 8×8 luminance (Y) blocks in each 16×16 macroblock, and one each of C_b, C_r block. Our algorithm overwrites the two chrominance blocks of the long image with those of the boosted short exposure if any of the four Y_{bs} blocks contains an edge. Since we replace one block by another, we do not require RAM memory for the algorithm beyond the storage of a JPEG macro-block.

Figure 5 shows details of the fused image of the basketball court set before, and after, artifact reduction. It can be seen that the method, though simple, is effective in removing ghosting artifacts.

The sensitivity of the JPEG-based edge detector is easily adjusted using the EOB threshold, which can be varied from 1 to 64. While the presence of noise complicates the selection of chrominance blocks, we show in the next section that the algorithm when using the threshold of 15 performs well in a variety of scenes taken at night with varying exposure times and ISO sensitivities.

5. IMAGE SATURATION DETECTION

We extend the algorithm described in R. Kakarala and R. Hebbalaguppe[14] to fuse more than two images, we detect saturation in the fused image by considering the luminance component alone. The fused image is obtained by fusing luminance from shortest exposed image and chrominance from the longest exposed image as explained in R. Kakarala and R. Hebbalaguppe.[14] We use a data-driven threshold which is the average of the difference between the maximum and minimum luminance values of the fused image.

$$Threshold_{sat} = \frac{\max Y_{fused} - \min Y_{fused}}{2} \tag{5}$$

Figure 6 demonstrates that our algorithm can fuse three images taken at different exposures from a hand-held camera in automatic bracketing mode. Saturation is detected block-by- block on the fused image. We notice that image saturation in fused image can be overcome replacing chrominance from in the fused image (Figure 6(d)) with the chrominance from the normal exposed image (Cb_{bm}, Cb_{bm}) as shown in Figure 6(f)

6. HDR RESULT

We have tested our algorithm on a variety of scenes, using hand-held cameras both in day light and low light. We used auto exposure bracketing in the continuous drive mode (bracketing the exposure time, all other settings such as ISO and the F number do not vary). The result in Figures (4, 6) show that the proposed algorithm is successful in producing sharp stabilized images with no noticeable artifacts. Due to space constraints , we have made other image sets available online*

We evaluate the performance of the algorithm not only by the images produced, but by using quantitative measures of signal-to-noise-ratio (SNR) and sharpness using effective bandwidth B. Fishbain and L. Yaroslavsky[18] and blur metric by Fishbain et al.[19]The latter is relatively simple and effective. These performance metrics are described in greater detail in R. Kakarala and R. Hebbalaguppe[14]

*https://picasaweb.google.com/100251433381871209447/SPIE.

7. COMPARISON WITH OTHER ALGORITHMS

Our method is single pass in nature and it is much simpler than the method of M. Tico and K. Pulli[4] and Tico et al[13] which uses two passes for BTF estimation and image boosting, in addition to implementing chrominance fusion in the wavelet domain. Our method also bypasses image registration which is used in M. Tico and K. Pulli[4] as the image stack is assumed to be captured in bracketing mode which ensures minimal motion between the two successive frames captured.

Given their higher complexity, the image quality provided by the algorithms of M. Tico and K. Pulli[4] and Tico et al[13] should be better than our simple method. There is clearly benefit to image quality in registration, deconvolution, and wavelet-based image fusion, steps that are used in those algorithms. Implementing any of those steps on a mobile device is not an easy task, and requires a powerful processor in order to produce the results without requiring a long wait on the part of the user. In contrast, our method requires only minimal computation: a LUT, detecting the position of the EOB symbol in each luminance block, and modifying the file write operation to selectively overwrite portions of the long exposure JPEG file.

The method of Lu et al[7] is iterative in nature, requiring 6-15 iterations in practice, and therefore clearly not capable of real-time application on digital cameras. The method of M. Tico and K. Pulli[4] uses a wavelet transform of both images, a significant overhead which cannot be reused for compression unlike our method. We have obtained from the authors of Lu et al[7] and Tico et al[13] the fusion results using their algorithms on the basketball court image set shown in this paper, and to enable visual comparison we have posted them online at the website mentioned in the previous section. The result of Lu et al[7] which takes 187 seconds to compute, shows considerable noise and ghosting artifacts. However, as pointed out in their paper, their method is not suited for images containing subject motion. The result of Tico et al,[13] which takes roughly 60 seconds for 10 megapixel images, is very good. The SNR improvement of the result from Tico et al[13] is 66 percent which is roughly the same as our method. Furthermore, it requires much more processing (registration, intensity equalization, wavelet transform) than our method. We have implemented our method on a laptop computer with a 1.6 GHz dual-core processor, and find that it takes 0.8 seconds for 2 megapixel images, much faster than either of the methods in comparison.

8. SUMMARY

An efficient method to fuse multiple images taken with varying exposure times in the JPEG domain is presented. The algorithm uses the spatial frequency analysis provided by the DCT within JPEG to combine the uniform regions of the longest-exposure image with the detailed regions of the short-exposure images, thereby reducing noise while providing sharp details. Advantages of the proposed method are great reduction in processing time, improved memory management, and efficient ghost removal in obtaining reasonably good quality HDR images.

Experiments show both quantitative and qualitative improvement over the short-long exposed images. Qualitatively, the fused image looks sharp with better colors. Quantitatively, the fused image shows improvement in SNR over the shortest exposed image and the sharpness (obtained by blur metric) over the longest exposed image.

To summarize our method, we use a single pass sigmoidal boosting on the shorter exposed images implemented as LUT, unlike other published methods which require two or more passes. Reuse of edge detection which is a part of JPEG for removal of artifacts further optimizes the algorithm. Lastly, the method requires no more than a single macro block to be kept in memory, because the image fusion is performed essentially in the JPEG file and rendered only on decoding the image.

ACKNOWLEDGMENTS

We thank Tz-Huan Huang of National Taiwan University and Dr. Marius Tico of Nokia Research for providing estimates of running times and fusion results from their algorithms. We gratefully acknowledge the help of D. Mao, N. Nagaraju, R. Bagadi, V. Premachandran for providing experimental image sets from various cameras. Many thanks to Prof. A.F. Smeaton, Prof. N. O'Connor and Ms J. Kuklyte for equipment and software to improvise the algorithm.

REFERENCES

[1] Gelfand, N., Adams, A., Park, S., and Pulli, K., "Multi-exposure imaging on mobile devices.," in [*ACM Multimedia*], 823–826 (2010).

[2] Zhang, X., Jones, R. W., Baharav, I., and Reid, D. M., "System and method for digital image tone mapping using an adaptive sigmoidal function based on perceptual preference guidelines," *US Patent 7,023,580* (2006).

[3] Wallace, G., "The jpeg still picture compression standard.," *IEEE Trans. Consumer Electronics* **38** (Feb. 1992).

[4] Tico, M. and Pulli, K., "Image enhancement method via blur and noisy image fusion.," in [*Int. Conf. on Image Processing (ICIP)*], 1521 – 1524 (2009).

[5] Hasinoff, S. W., Durand, F., and Freeman, W. T., "Noise-optimal capture for high dynamic range photography," in [*CVPR*], 553–560 (2010).

[6] Khan, E., Akyuz, A., and Reinhard, E., "Ghost removal in high dynamic range images," in [*IEEE Intl. Conf. Image Processing (ICIP)*], 530–533 (2006).

[7] Lu, P., Huang, T., Wu, M., Cheng, Y., and Chuang, Y., "High dynamic range image reconstruction from hand-held cameras.," in [*Computer Vision and Pattern Recognition (CVPR)*], 509 – 516 (2009).

[8] Reinhard, E., Ward, G., Pattanaik, S., and Debevec, P., [*High dynamic range imaging: Acquisition, display and image-based lighting*], Morgan Kaufmann Publishers (2005).

[9] Bogoni, L., "Extending dynamic range of monochrome and color images through fusion," in [*15th International Conference on Pattern Recognition (ICPR)*], **3**, 7–12, IEEE (2000).

[10] Kang, S., Uyttendaele, M., Winder, S., and Szeliski, R., "System and process for generating high dynamic range video." U.S. Patent 7,382,931 (Jun 2008).

[11] Jacobs, K., Loscos, C., and Ward, G., "Automatic high dynamic range image generation for dynamic scenes," in [*IEEE Computer Graphics and Applications*], **28**, 24–33, IEEE (Mar-Apr 2008).

[12] Gallo, O., Gelfand, N., Chen, W., Tico, M., and Pulli, K., "Artifact-free high dynamic range imaging," in [*IEEE Intl. Conf. Computational Photography (ICCP)*], 1–7 (April 2009).

[13] Tico, M., Gelfand, N., and Pulli, K., "Motion-blur free exposure fusion.," in [*Proc. Int. Conf. on Image Processing (ICIP)*], 1521–1524 (2010).

[14] Kakarala, R. and Hebbalaguppe, R., "A method for fusing a pair of images in the jpeg domain," *Journal of Real-Time Image Processing,Springer.* . Accepted for publication.

[15] Nakamura, J., [*Image sensors and signal processing for digital still cameras*], CRC Press (2005).

[16] Pennebaker, W. L. and Mitchell, J. L., [*The JPEG still image data compression standard*], Kluwer Academic Publishers (1993).

[17] Kakarala, R. and Bagadi, R., "A method for signalling block-adaptive quantization in baseline sequential jpeg," in [*IEEE TENCON*], 1–6 (2009).

[18] Fishbain, B., Yaroslavsky, L., Ideses, I., and Roffet-Crété, F., "No-reference method for image effective-bandwidth estimation," in [*SPIE - Image Quality and System Performance*], **6808** (2008).

[19] Fishbain, B., Ideses, I. A., Shabat, G., Salomon, B. G., and Yaroslavsky, L. P., "Superresolution in color videos acquired through turbulent media.," *Optics Letters* **34** (2009).