A Novel Approach for Image Fusion in JPEG Domain

Venkata Santhosh kumar A.[1], Ch. Ramesh[2]
*(M.Tech Student, computer science department, AITAM, Tekkali, India
** (Professor, AITAM, Tekkali, India

ABSTRACT: *Image fusion is a procedure of combining relevant information from more than one image into a single image. The resulting image will have more information than any of the input images. Fusion is a popular approach for producing an image without under- and over-exposed areas is to capture several input images with varying exposure settings, and later combine them into a single high-quality image. This technique is used in various disparate fields such as remote sensing, robotics and medical applications. In this paper, we proposed a method to fuse multiple images taken with varying exposure times. The key component of the method is a single pass sigmoidal booster is applied on the shorter exposed images implemented as LUT, unlike other methods which require two or more passes. We used this technique to capture high dynamic range images from a set of photographs taken at different exposures, where misalignments can produce blurring and artifact and prevent achieving high quality HDR images. We present our implementation of the technique and the results of tests made for variety of photographs.*

Keywords: *Artifact removal, HDR capture, High Dynamic Range Imaging, Image Stabilization, Image Fusion*

I. INTRODUCTION

Due to tremendous growth towards the production of high dynamic range (HDR) images. Modern image processing and graphics software becomes HDR enabled. Also HDR digital photography replaces low dynamic range (LDR) technologies. HDR photographs have better quality and easier to process in a digital darkroom, when compared to LDR. Unfortunately, HDR cameras are very expensive and not available for average users.[1] On the other hand, taking HDR photographs seems to be legitimate and crucial. In the near future LDR images may become almost obsolete due to the progress in LCD technology and it will not be easy to display LDR image correctly. LDR photographs will look pale and not interesting on HDR LCD displays. The multi-exposure HDR capture technique seems to be a good alternative to HDR cameras, which is used to create an HDR image from photographs taken with a conventional LDR camera. The technique uses differently exposed photographs to recover the response function of a camera. From the response function, the algorithm creates an HDR image whose pixel values are proportional to the true radiance value of a scene.[2] Because this technique requires multiple input photographs, there is a high likelihood of misalignments between pixels in the sequence of exposures due to moving of a hand-held camera or dynamic object in a scene. It is crucial that misalignments between input photographs should be removed before fusing an HDR image.

Image fusion is a procedure of combining relevant information from more than one image into a single image. The resulting image will have more information than any of the input images. Fusion is a popular approach for producing an image without under- and over-exposed areas is to capture several input images with varying exposure settings, and later combine them into a single high-quality image. The need for image fusion arises in many applications, including digital image stabilization and HDR image capture. In stabilization, the motion blur that occurs in a long exposure-time image is corrected by fusing it with a short exposure-time image of the same scene.[3] 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 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.[4] In the proposed technique short, long and normal exposure images are used. The long exposure image should 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. The short and normal exposure images which are taken consecutively is boosted in brightness by using a lookup table (LUT) modelling a sigmoidal function combined with tonemapping.[5] 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 (Cb,Cr) 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.

The rest of the paper is organized as follows. Section 2 deals with Related work, Section 3 discuss on Proposed Algorithm, Section 4 on Image fusion, Section 5 on **S**igmoidal function, Section 6 discuss about Artifact removal, Section 7 brief Image saturation detection, Section 8 discuss about Results and Section 9 concludes the paper.

II. RELATED WORK

In recent years significant progress has been made in the development of algorithms that allow to capture HDR images using low dynamic range sensors (standard LDR cameras) [20, 21, 22, 23].These algorithms retrieve high dynamic range information from a sequence of photographs. The authors suggest using tripod to avoid camera movement and they do not address the problem of eliminating misalignments. The problem of image alignment and matching was intensively studied during last year's [24, 25] but not for registration of images of different exposures. The only solution that addressed exactly the problem of capture of HDR photographs was proposed in [26, 27]. The technique employs conversion of input photographs into percentile threshold bitmaps. The bitmaps are analyzed and then aligned horizontally and vertically using shift and difference operations over each image.

Kang et al. [28] described a technique for creating high dynamic range video from a sequence of altering light and dark exposures. A part of the technique is a HDR stitching process, which includes global and local alignment to compensate for pixel motion. The stitching process can be also used to compensate for camera movement when creating an HDR still from a series of bracketed still photographs. However, the presented technique seems to be suitable for video where there are no large differences between consecutive frames.

In [29] Sand and Teller present a global and local matching algorithm, which is robust to changes in exposure of photographs. The key idea behind this technique is to identify which parts of the image can be matched without being confused by parts that are difficult to match. Such assumption seems to be not valid for images with large differences in exposures, where there is usually not enough information for correct matching. The technique was designed for matching two video sequences and was not tested on still photographs.

Recently, Cerman and Hlavac [30] presented an alignment method based on unconstrained nonlinear optimization. In this method, each image is linearized using the estimated camera response function and multiplied by the exposure ratio. Then, a normalized difference summed across all corresponding pixels is used to estimate misalignments. The method can compensate global rotation and horizontal and vertical shifts. There are a few techniques which compute camera response function based on misaligned photographs [31]. However, these methods are not meant to create HDR images. The problem of removing ghosting artifacts in a multi-exposure sequence of photographs was also investigated [32] but proposed algorithms do not take into consideration a compensation of camera movements.

III. PROPOSED ALGORITHM

The proposed technique 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, EV(0) represents the exposure time determined by the camera's auto exposure routine, and relative to that time, EV(D) represents an exposure time that is 2^{Δ} larger.[7] For example EV(+1) is twice the exposure time of EV(0), and EV(1) is one-half the exposure time of 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 Y,Cb,Cr color system. JPEG takes advantage of this by subsampling the two chrominance channels, denoted Cb, Cr usually by a factor of two in each direction.

Proposed Algorithm:

Step 1: convert the longest exposure image into Y_l,Cb_l,Cr_l

Step 2: Apply Sigmoidal boosting and tone-mapping for Shortest and normal exposure images.

Step 3: convert the shortest, normal exposure images into Y_s,Cb_s,Cr_s and Y_n,Cb_n,Cr_n respectively.

Step 4: write the file into secondary storage device.

Step 5: Overwrite Y_l with Yb_s in the file

Step 6: At Edges, Overwrite Cb_l, Cr_l with Cb_s and Cr_s in the file.

Step 7: Detect saturation in image based on data-driven threshold on luminance of fused image.

Step 8: check whether block is saturated or not. If it is YES Overwrite Chroma in step 6, with Cb_n,Cr_n otherwise stop the process

In the proposed technique for correcting motion artifacts selects the 8*8 chrominance (Cb,Cr) 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 (Cb,Cr) blocks of the boosted short-exposure image. Image saturation is detected in the case of day light scenes. 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-exposed also called as normal exposed image replaces the chrominance in the fused image.[11] [10]

IV. IMAGE FUSION

In modern digital cameras we can preview the captured photo immediately, after capture. Hence we perform the image merging directly on the camera. The proposed 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. So to compensate for camera motion using the image alignment algorithm which is discussed 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 an input images, exposure fusion computes a scalar-valued weight map for each image, and performs a weighted blend of the inputs to obtain the final HRD image which is described in Figure 1.

Let (I_1.....I_n) be n images captured by the camera. The weight for each pixel (i, j) of image k is computed as:

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

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:

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

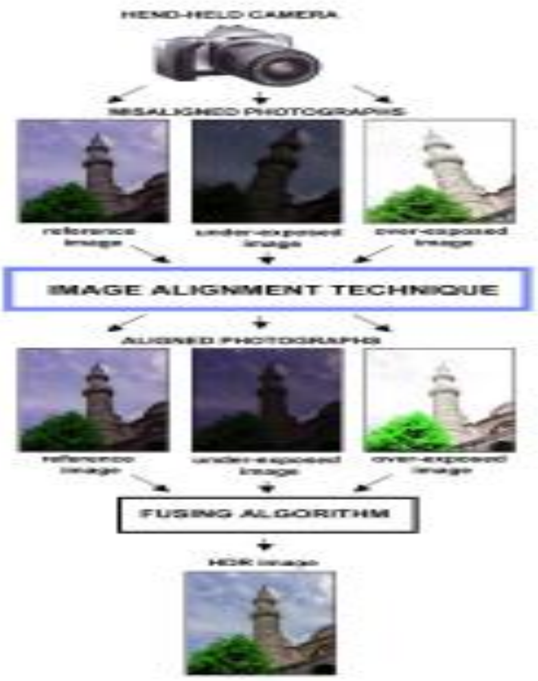

Figure 1: Image Fusion Technique

V. SIGMODIAL 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. M. Tico and K. Pulli proposed a compensating function the brightness transfer function (BTF).[12] The BTF is estimated by plotting the pixel values of longer exposed image against the corresponding luminance values in the shorter exposed image, applying basic curve fitting and by smoothing the mean of pixel values for the longer exposure for the same pixel value in the shorter exposure. Both methods require two passes over the shorter exposure image, the first to estimate the BTF and the second to boost the image accordingly.[8]

VI. ARTIFACT REMOVAL

The chrominance from longer exposure image, denoted by Cb_l, Cr_l is merged with the boosted luminance of shorter exposed image denoted by Yb_s. The merged image (Yb_s, Cb_l, Cr_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. Proposed 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 which is proposed by W. L. Pennebaker and J. L. Mitchell16 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. The 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. Apply the detection to the EOB location in the DCT coefficients of the luminance blocks Yb_s of the boosted short-exposure image.[13] 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.[14]

For macroblocks where an edge is detected in the Yb_s component, the chrominance from the boosted shorter exposure (Cb_s,Cr_s) 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 Cb, Cr block. Our algorithm overwrites the two chrominance blocks of the long image with those of the boosted short exposure if any of the four Yb_s 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 macroblock.[15] which is described in Figure 2, 3,4 and 5.

VII. IMAGE SATURATION DETECTION

The fused image is obtained by combining the luminance from shortest exposed image and chrominance from the longest exposed image which is described by R. Kakarala and R. Hebbalaguppe [14]. Data-driven threshold is used to calculate average of the difference between the maximum and minimum luminance values of the fused image. If the luminance value in the fused image is greater than the data driven threshold given then replace the chrominance of the fused image by the chrominance of the normal exposed image.[18,19]

VIII. RESULTS

We evaluate the performance of the algorithm by taking three images at various exposure times called as Longest, Normal and Shortest exposure time images and the technique efficiently fuses multiple images taken at various exposure times in the JPEG domain which has shown in Figure 6,7,8 and 9.

Figure 2: Before Artifact removal

Figure 3: After Artifact Removal

Figure 4: Before Artifact removal

Figure 5: After Artifact Removal

Figure 6: Longest Exposure image

Figure 7: Normal Exposure image

Figure 8: Shortest Exposure image

Figure 9: Final image after Image Fusion

IX. CONCLUSION

Image fusion is a procedure of combining relevant information from more than one image into a single image. The resulting image will have more information than any of the input images. Fusion is a popular approach for producing an image without under- and over-exposed areas is to capture several input images with varying exposure settings, and later combine them into a single high-quality image. The proposed technique efficiently fuses multiple images taken at various exposure times in the JPEG domain. The technique has categorized images into three types based on the exposure time, they are longest, normal and shortest exposure images. A single pass sigmoidal booster is applied on the shorter exposed images implemented as LUT. 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. Experimental Results shows our proposed technique has efficiently fused multiple images into single image with better quality HRD.

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 , 1–11. 10.1007/s11554-011-0231-8.

[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-Cr´et´e, 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).

[20] S. Mann and R. Picard. Being 'Undigital' with Digital Cameras: Extending Dynamic Range by Combining Differently Exposed Pictures. Proc.of IST's 48th Annual Conference, pages 442– 448, May 1995.

[21] P.E. Debevec and J.Malik. Recovering high dynamic range radiance maps from photographs. In Proceedings of ACM SIGGRAPH 1997,pages 369–378, 1997.

[22] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. pages 374–380, 1999.

[23] M. Robertson, S. Borman, and R. Stevenson. Dynamic range improvement through multiple exposures. In Proceedings of the 1999 International Conference on Image Processing (ICIP- 99), pages 159–163, Los Alamitos, CA, October 24–28 1999.

[24] Barbara Zitova and Jan Flusser. "Image registration methods: a survey. Image and Vision Computing", 21(11):977–1000, October 2003.

[25] Lisa Gottesfeld Brown. "A survey of image registration techniques. ACM Comput". Surv.,24(4):325–376, 1992.

[26] Greg Ward. Fast, "robust image registration for compositing high dynamic range photographs from hand-held exposures. journal of graphics tools", 8(2):17–30, 2003.