2013 IEEE International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing

A ROBUST IMAGE REGISTRATION ALGORITHM FOR HDR IMAGING ON FEATURE PHONE PLATFORMS

Qiwei Xie,[1,2] *An Jiang,*[3] *Qian Long,* [1] *Seiichi Mita*[1] *and Xi Chen*[2]

[1]Toyota Technological Institute, Nagoya, Aichi, 468-8511, Japan
[2]Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[3]Mobile Phone Department, Nokia China Investment Ltd, 100176, China

ABSTRACT

A robust image registration method has been applied to High Dynamic Range (HDR) imaging on feature phone platforms successfully. The proposed algorithm obtains better exposure fusion effect under limited computing resources of feature phone platforms, which have only 12M RAM and 400MHz CPU available.

Index Terms— Image registration, Feature phone, High Dynamic Range, Exposure fusion

1. INTRODUCTION

HDR is a set of methods used in imaging and photography, to allow a greater dynamic range between the lightest and darkest areas of an image. HDR images can represent more accurately the range of intensity levels found in real scenes, from direct sunlight to faint starlight. HDR compensates for the loss of detail in bright or dark areas by taking multiple pictures at different exposure levels and intelligently combining them together to construct a single HDR image.

However, all HDR imaging algorithms are suffered from misalignments of the input images. Some of them assume that different exposure images are captured with a camera mounted on a tripod and thus well spatial aligned. Actually, most mobile device users would prefer to take photos by hand regardless of the time and location, and it becomes an essential requirement that registering series input images before image fusion in commercial HDR solution.

The problem of image registration was intensively studied during last years [1, 2] but not for registration of images of different exposures. One solution that addresses exactly the problem of capture of HDR photographs is proposed in [3], which uses the black and white-saturation areas in the input images to estimate the direction and the quantity of their displacement. Translational shifts are applied to one or both of the input images in order to achieve their spatial alignment. Another solution is presented in [4, 5]. 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.

In [6] 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 is designed for matching two video sequences and is not tested on still photographs.

There are a few techniques which compute camera response function based on misaligned photographs [7, 8]. However, these methods are not meant to create HDR images. The problem of removing ghosting artifacts in a multi-exposure sequence of photographs is also investigated [5, 9] but proposed algorithms do not take into consideration a compensation of camera movements.

In [10] A. Tomaszewska and R. Mantiuk present a fully automatic method for eliminating misalignments between a series of hand-held photographs taken at different exposures. The key component of the technique is the SIFT method [11] that is employed to search for feature-points in consecutive images. But using SIFT method is time and memory consuming, and it can not be applied to some lower end device platform, such as mobile phone.

A simple but effective image registration algorithm is presented in this paper as a plug-in to HDR imaging algorithms intended for implementation on feature phone platform, which has only 12M RAM and 400MHz CPU available. It deals with a couple of input images, one overexposed image and one underexposed image, and focuses on translational motion, ignores others of rotational, scale movements, considering that most images taken by hand can be well aligned using only translational shift.

2. ALGORITHM

In this paper, a simple but effective image registration algorithm is proposed. Firstly, the 'Double-Saturated' (DS) im-

978-0-7695-5088-6/13 $31.00 © 2013 IEEE
DOI 10.1109/HPCC.and.EUC.2013.315

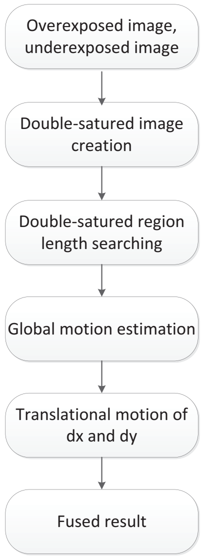

Fig. 1. Flowchart of the proposed process.

ages are generated based on the white-saturated image from overexposed image and the black-saturated images from underexposed image, using "AND" operation; Secondly, the lengths of DS region in horizontal and vertical directions are searched using several scan lines respectively; Finally, the translational shift is estimated from the histogram of DS region lengths, and the length with the highest frequency is likely to be the right global motion. The process of the proposed algorithms is described in Figure 1.

2.1. DS image creation

An overexposed image tends to be white-saturated in bright area (Figure2.(a)), and underexposed image tends to be black-saturated in dark area on the contrary (Figure2.(b)). Ideally, if the overexposed and underexposed image is well aligned, the white-saturated region and the black-saturated region will be complementary, and no DS area will occur. Only the spatial global motion between the capture of these two input images will cause DS, which implies input image misalignment. The binary map of white-saturated image and black-saturated image is created by applying a simple threshold on the luminance component of the input images. In the overexposed image, if one pixel's luminance level is higher than percent of the Maximum Possible Luminance (MPL), then it's likely to be white-saturated, the corresponding position in the white-saturated binary map is set to be '1'. Otherwise, the corresponding position is set to be '0' (Figure2.(c)). This can be expressed in formula as:

$$B_{over}\left(i,j\right)=\left\{\begin{array}{ll}1 & \text{if } Y_{over}(i,j)>p\cdot MPL\\ 0 & else\end{array}\right. , \quad (1)$$

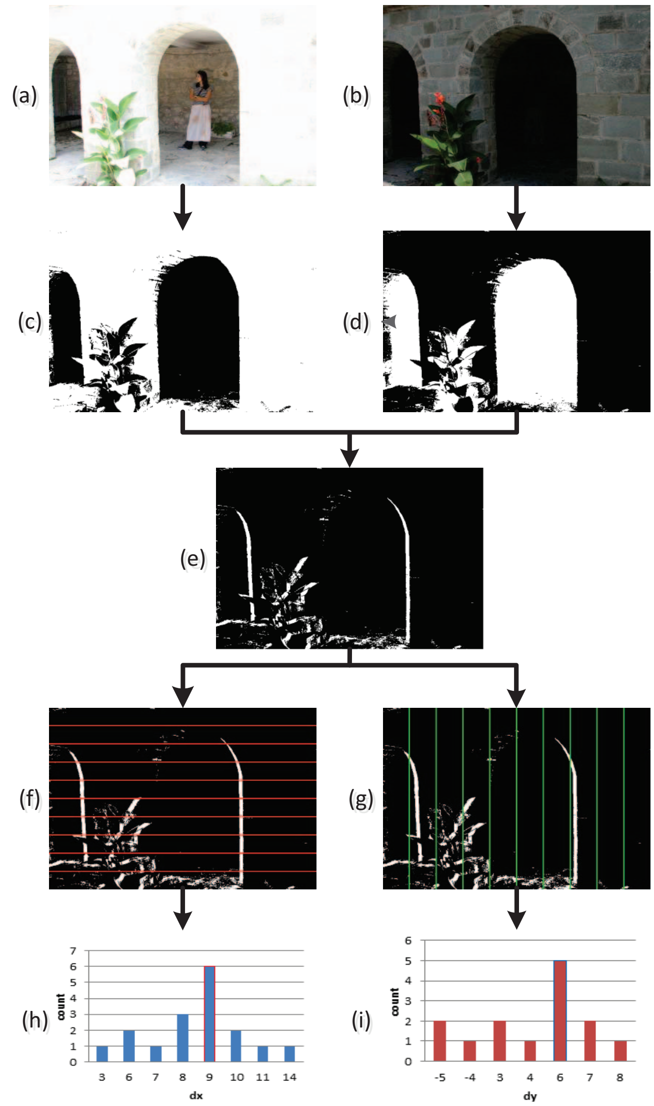

Fig. 2. Motion detection process of the proposed algorithm; (a) Overexposed image; (b) Underexposed image; (c) White-saturated image of (a); (d) Black-saturated image of (b); (e) double-saturated image (f)-(g) Horizontal / Vertical scan line in the DS image, which is "AND" operation result of (c) and (d); (h)-(i) Histogram of horizontal / vertical lengths of DS region, and the estimated movement is $dx = 9$, $dy = 6$.

where Y_{over} is the luminance level of the input overexposed image, and B_{over} is it's corresponding white-saturated binary map. Meanwhile, if one pixel's luminance level in the underexposed image is lower than $1-p$ percent of the MPL, then it's likely to be black-saturated, the corresponding pixel in the black-saturated binary map is set to be '1'. Otherwise, the corresponding position is set to be '0' (Figure2.(d)).

$$B_{under}(i,j) = \begin{cases} 1 & \text{if } Y_{under}(i,j) < p \cdot MPL \\ 0 & else \end{cases}, \quad (2)$$

where Y_{under} is the luminance level of the input underexposed image and B_{under} is it's corresponding black-saturated binary map. After that, we applying "And" operation pixel by pixel between the white-saturated and the black-saturated binary map, and then DS image (Figure2.(e)) can be created as:

$$DS = B_{over} \odot B_{under}, \quad (3)$$

where $\odot$ represents the element-wise multiplication operator.

For the parameter of p, some papers appoint it to be 80%[3], but it is not an optimized value. Notice that the white and black saturated areas are complementary in perfectly aligned input images and misalignment of the input images will not be too large, we proposed an iteration method to find the optimal p. That means an ergodic process can be executed from $p = 70\%$ to $p = 90\%$, the value of p corresponding to the most complementary case will be the optimized value. The optimization procedure can be written in formula as follows:

$$argmin_p |1 - \frac{SUM(B_{over}, p) + SUM(B_{under}, p)}{Size}|, \quad (4)$$

where SUM is the number of '1' of B_{over} or B_{under} with consideration to p, and $Size$ is the number of pixels of the input image.

2.2. DS region lengths search

The DS region lengths are searched using scan line method along the horizontal and the vertical direction of the DS image respectively. The horizontal scan line is located at 10%, 20%, $\cdots$, 90% position of the image height (Figure2.(f)) and the vertical scan line is located at the same percentages of the image width (Figure2.(g)). These scan lines cover the most significant part of the input images. Generally, the lengths of the DS area should not be too large or too small. Very small lengths usually occur at the edge of the image, or caused by noise; while very large lengths would be result of either huge exposure value difference or misalignment in inconsistent direction.

To each horizontal length of the DS region (call it as dx), the behavior of the transition tendencies is analyzed. If the white-saturated map changes from 1 to 0, and the black-saturated map vice versa, this mean that the overexposed image should be shifted to the right for pixels, and the displacement is recorded as dx. Otherwise, if the white-saturated map changes from 0 to 1, and the black-saturated map vice versa, the overexposed image should be shifted to the left for d pixels, and the displacement is recorded as $-dx$. The DS region lengths in the vertical direction (call it dy) are detected in the same way, and all the searching procedure in both directions need to be performed only once. If the transition tendency is in other form of behavior, the corresponding length of DS region will not be recorded.

2.3. Global motion estimation

Among variety of dx and dy, we need to select the correct dx and dy as the displacement between the input images. After all of the DS region lengths in both directions are recorded, the horizontal global motion is estimated based on the statistics histogram of dx (Figure2.(h)) and the vertical global motion is estimated based on the statistics histogram of dy (Figure1.(i)). The value with the highest frequency is likely to be the right global motion. Then image registration is performed according to the estimated global motion before HDR imaging.

The proposed algorithm is somewhat similar to that of [3], with three major differences as follows: Firstly, we use an optimized value of p to create DS image, but [3] only use a fixed value of 80%, and this will result in some uncertainties due to exposure value difference of different input image pairs; Secondly, the scan lines are used to search DS region lengths, and the search procedure only need to be performed once on each direction. Then the proposed algorithm more robust than that of [3], because the DS search and image shift procedure are repeated iteratively until the total amount of the DS pixels drops below $1/8$ of the initial; Finally, in [3] the global motion is the sum of the shift amounts from all of the iterations, while this paper realizes the estimation based on statistics histogram with more robust result compared with [3].

3. ALGORITHM IMPLEMENTATION

Though widely used in smart phone, HDR has not yet been used in feature phones, due to its computational complexity and high memory consumption. For example, there is only 12M RAM available on some feature phone platforms, and the main frequency is only about 400MHZ.

By use of the patented technology named "random access" [12, 13, 14], we can do image processing work of HDR block by block, and memory consumption can be large reduced by this way. A region decoder object in $ScaldoSDK$, named $CTRegionDec$, is responsible for decoding pixel data from a generic pixel source. This pixel data can be any rectangular subset of the original image, so a decoder supporting this interface can be used to ”random access” pixels of the image.

The main component of Scalado's image management is $Scalado^{TM} RAJPEG$. $Scalado^{TM} RAJPEG$ is a JPEG codec technology that reduces memory requirements and CPU load in an ingenious way.

JPEG is a complex compression format, which normally requires a complete decoding of an image in order to view or process it. $Scalado^{TM} RAJPEG$ technology allows instant processing of any region of the image in any scale while keeping the images in the compressed form, enabling viewing and processing of any region of a JPEG image, without the need to decode a single bit of unnecessary information from the image.

$Scalado^{TM} RAJPEG$ reduces the amount of memory required by a factor of up to 20x, and increases speed for image manipulation by a factor of up to 10x. All operations can be performed in real-time on the phone's display until the user is satisfied and wants to save the complete image

4. RESULTS

In this paper, we limit the length of the DS region from 3 pixels to 10% of the image height or width according to its searching direction, and this means misalignment below 3 pixels and above 10% of the image height or width will not be detected.

The performance of the algorithm is presented in Figure 3-6. It's obviously that the image in the right column show great improvement over the image in the left column, and the phenomenon of ghost is also solved with clear details. Meanwhile, For the input images of Figure 3, the optimized value of p and the estimated global motion are: $p = 0.83, dx = 9, dy = 6$; for that of Figure 4, $p = 0.82, dx = 3, dy = 3$; for that of Figure 5, $p = 0.84, dx = 6, dy = 3$; for that of Figure 6, $p = 0.77, dx = -7, dy = -3$.

Remark: Notice that image registration would result in black bars on the sides of fused image, because after image registration there is no data available on the sides of shifted image. The problem can be easy solved by some other method like cropping or rending.

5. CONCLUSION

We propose a computationally efficient, simple and effective algorithm which incorporates Double-Saturated image, scanline searching and histogram method to detect camera motion from images with different exposures. Since no complicated method is involved, the proposed algorithm is computational efficient and intended for implementation on some lower devices such as feature phone platforms.

Fig. 3. Performance of the proposed algorithm; left top: Overexposed image; right top: Underexposed image; left down: HDR images composited from unaligned exposures; right down: Exposures aligned with the proposed algorithm yield superior HDR images.

Fig. 4. Performance of the proposed algorithm; left top: Overexposed image; right top: Underexposed image; left down: HDR images composited from unaligned exposures; right down: Exposures aligned with the proposed algorithm yield superior HDR images.

Fig. 5. Performance of the proposed algorithm; left top: Overexposed image; right top: Underexposed image; left down: HDR images composited from unaligned exposures; right down: Exposures aligned with the proposed algorithm yield superior HDR images.

Fig. 6. Performance of the proposed algorithm; left top: Overexposed image; right top: Underexposed image; left down: HDR images composited from unaligned exposures; right down: Exposures aligned with the proposed algorithm yield superior HDR images.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 61101219, 61201050, 61201375) and Anhui Provincial Natural Science Foundation (Grant No. 1208085ME65).

6. REFERENCES

[1] B. Zitova et al., "Image registration methods: a survey," *Image and Vision Computing.*, vol. 21, pp. 977–1000, Oct. 2003.

[2] L.G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, pp. 276–325, Dec. 1992.

[3] T. Kartalov et al., "A real time global motion compensation for multi-exposure imaging algorithms," *International Conference on Computer as a Tool (EUROCON).*, pp. 1–4, 2011.

[4] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from handheld exposures," *Journal of Graphics Tools.*, vol. 8, pp. 17–30, 2003.

[5] E. Reinhard et al., "High dynamic range imaging: Acquisition, display, and image-based lighting," *Morgan Kaufmann Publishers Inc.*, 2005.

[6] P. Sand et al., "Video matching," *ACM Trans. Graph.*, vol. 23, pp. 592–599, Jul. 2004.

[7] M.D. Grossberg et al., "Determining the camera response from images: What is knowable?," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1455–1467, Nov. 2003.

[8] S.J. Kim et al., "Radiometric self-alignment of image sequences," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 645–651, 2004.

[9] E.A. Khan et al., "Ghost removal in high dynamic range images," *IEEE International Conference on Image Processing.*, pp. 2005–2008, 2006.

[10] A. Tomaszewska et al., "Image registration for multi-exposure high dynamic range image acquisition," *Proc. of International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision.*, pp. 49–56, 2007.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision.*, vol. 60, pp. 91–110, Nov. 2004.

[12] Scalado AB., "Method for processing a digital image and image representation format (7831101)," Dec. 2008, United States.

[13] Scalado AB., "Method for processing a digital image and image representation format (7480418)," Nov. 2004, United States.

[14] Scalado AB., "Method for processing a digital image and image representation format (7720298)," Dec. 2008, United States.