GHOST-FREE DUAL-EXPOSURE HDR FOR DYNAMIC SCENES

Fahd Bouzaraa[‡] *Ibrahim Halfaoui*[‡] *Onay Urfalioglu*[†]

[‡] Technische Universität München
[†] Huawei Technologies Co. Ltd. - European Research Center

ABSTRACT

High Dynamic Range Imaging (HDRI) and Exposure Fusion (EF) are methods of choice to computationally extend the dynamic range of images depicting real world scenes. Unfortunately, those methods are still prone to certain artifacts. Among others, the so-called *Ghost Effect* is the most critical HDR limitation when it comes to dealing with motion (camera or scene motion) in input Low Dynamic Range (LDR) images. This problem becomes more challenging when the input LDR image stack contains only a couple of images with large color differences, which is the case in the mobile phone domain. To address this issue, a de-ghosting step is required to preserve the quality of the final HDR images. In this paper, we propose a robust de-ghosting approach based on the detection and the elimination of the motion induced effects on the final HDR images. The proposed method performs efficiently in all cases even on scenarios where only two differently exposed images with large illumination variations are available as input. Compared to the state-of-the-art, our results exhibit significant visual improvement and artifact reduction. Furthermore, our approach has low computational cost and complexity, which enables an efficient implementation especially for mobile phone-related applications.

Index Terms— High Dynamic Range Imaging; Exposure Fusion; De-ghosting; Motion Detection; Image Enhancement

1. INTRODUCTION

High Dynamic Range Imaging (HDRI) is an image processing task centered around combining differently-exposed LDRs of the same scene, into a final image with a greater dynamic range and more details. The merging process can be done using the inverse Camera Response Function (CRF) [1, 2, 3] or by means of Exposure Fusion [4]. In this case, the input LDRs are merged using weighting maps which evaluate the saturation, exposedness and contrast of the LDRs. These techniques are based on the assumption that the input LDRs are aligned (static scene). However, real world scenes are mostly dynamic and contain moving objects. This results in *Ghost Effects*, where objects appear in several locations in the final image. One way to deal with motion-related artifacts is by realigning the input LDRs to a chosen reference image. This can be done using several motion estimation techniques such as Optical Flow, Block Matching or PatchMatch [5, 6, 7, 8, 9, 10]. These methods depend strongly on the accuracy of the motion estimation step, which is a computationally expensive step.

On the other hand, straightforward de-ghosting techniques are interesting for applications where low-complexity is a priority. This set of methods is based on the modification of the weighting maps used to generate the final HDR image, in order to reduce the effect of the motion areas. In [11], Khan et al. proposed a technique based on the estimation of the likelihood that a pixel belongs to a moving object. Similarly, Galo et al. suggest in [12] to detect motion areas using the photometric relations between the input LDRs. Next, they propose a patch-wise merging operation, where patches belonging to motion regions are discarded. Likewise, Jaehyun et al. propose in [13] a method based on the photometric assumption, but relies on *Zero-Normalized Cross Correlation* (ZNCC) metric for the merging process. Alternatively, Pece et al. propose in [14] to detect motion using *Median Threshold Bitmap (MTB)*. The generated motion maps are refined using morphological transformations (erosion, dilation) and later used for the merging procedure. These methods perform generally well when the input stack offers a large number of differently exposed LDRs, however they depend strongly on the motion detection step, where assumptions such as the photometric relation do not always hold. In addition, these methods generally fail in case of 2 input LDRs with large illumination difference.

Recently, An et al. propose in [15] a new method for HDR de-ghosting which considers motion detection as a binary labeling task. the underlying idea is to model the distribution of the ZNCC images of each input LDR with respect to the reference image as a mixture of Gaussian functions which parameters are to be estimated. Despite the high quality of the de-ghosting operation, the generation of the ZNCC maps is computationally expensive. In addition, the motion detection and de-ghosting results are limited in case of few input LDRs.

Our approach advances upon the previously mentioned methods by proposing a low-complexity yet very accurate algorithm for motion detection and subsequent merging using EF. The proposed approach allows for greater color difference (different exposures) between the input LDRs. Furthermore,

the quality of the final HDR image does not depend on the number of inputs, unlike the previously mentioned methods, which makes the proposed method suitable to computationally limited devices.

2. OUR APPROACH

In this paper we propose a method for performing de-ghosting based on motion detection and modification of the weighting maps of the exposure fusion algorithm. In the following, we describe the steps of the de-ghosting for the case of 2 input LDRs I_b (bright image) and I_d (dark image). The proposed approach can be easily extended to more input LDRs. The main contributions of our approach:

- A low-complexity and accurate motion detection algorithm which deals with different types of motion (scene or camera motion) as well as large exposure differences. The generated motion maps indicate the location of motion related regions with respect to the chosen reference image.

2.1. Motion Detection and Motion Map Refinement

The underlying idea of the proposed motion detection algorithm is to explore the difference image between I_b and I_d. However, I_b and I_d present large color difference which needs to be reduced. This is done using *Histogram Matching* (HM) [16]. HM is a low-complexity algorithm which matches the *Cumulative Distribution Function* (CDF) of a source image to the CDF of a target image with the desired color distribution. In our case, we designate I_d as the source image, since it contains generally less saturated areas then I_b. This results in image I_h, which has the same color properties as I_b while having the same content of I_d. This step does not influence the choice of the reference image, as explained later in the exposure fusion stage.

Next, in case the input images contain camera-related motion (translation and rotation), we propose an optional low-complexity step for global motion registration. This step is based on the computation of a *Homography* matrix H [17]. To this end, we use *SURF* [18] to detect and extract features in I_b and I_h. Several approaches for features matching can be used, such as *RANSAC* [19]. Finally, I_b is warped to the view of I_h using the computed matrix H, resulting in an image I_w. Understandably, if no camera motion is detected, I_w is a copy of I_b.

Once the color as well as content (camera motion) differences are reduced, we compute the difference image I_{diff} on down-sampled versions of I_w and I_h. The down-sampling step reduces HM-related noise and artifacts, as we use a *Gaussian Pyramid* which acts as a low-pass filter. Additionally, the down-sampling decreases the computational cost of the algorithm. Empirically, it is sufficient to down-sample to 1 or 2 levels. The difference image is computed according to:

$$I_{diff}(i,j) = |D(I_h(i,j)) - D(I_w(i,j))|, \quad (1)$$

where D represents the down-sampling operator and (i,j) pixel coordinates. The ensuing detection step relies strongly on the information provided by the difference image. Typically, 2 types of difference values in I_{diff} can be identified:

- Difference values from motion related objects. These values are generally large and less frequent.
- Difference values originating from the inherent color difference between I_h and I_w. These values are typically smaller and more frequent.

The goal of the next steps is to accurately distinguish between the previously mentioned difference values. To this end, we process I_{diff} using the *logistic function*:

$$\hat{I}_{diff}(i,j) = \frac{1}{1 + k_1 \mathrm{e}^{-k_2(I_{diff}(i,j)-0.5)}}, \quad (2)$$

where k_1 and k_2 are control parameters. Empirically, these parameters are respectively set to 0.09 and 12. The logistic function allows for manipulating the contrast of the difference image, so that large difference values corresponding to motion pixels are enhanced in comparison to small difference values. This is important for the subsequent detection step, since it allows for a better classification of the difference values, as shown in fig. 1. The logistic function-based manipulation allows for an accurate approximation of the ground truth motion map, in comparison to the case where no manipulation is done. The ground truth motion map of the shown image pair in fig. 1 was drawn manually. The processing using the logistic function is performed for all 3 *RGB* color channels separately.

Using the processed difference image $\hat{I}_{diff}$, we compute a threshold T_c for each color channel c using an approach inspired from [20]. These thresholds enable us to distinguish between motion-related difference values and HM-related values. To this end, we compute the histogram of $\hat{I}_{diff}$. The desired threshold depends on the mean of the bins where an abrupt decrease of the number of pixels occurs:

$$T_c = \arg\max_{T_c^i} |N_p(T_c^i) - N_p(T_c^{i+1})|, i = 0, ..., B-2, \quad (3)$$

where $N_p(T_c^i)$ is the number of pixels around the bin center T_c^i of the bin number i out of B bins. The threshold T_c is equal to $\frac{T_c^{i+1} - T_c^i}{2}$. Accordingly, a pixel is marked as *motion-pixel* if at least 1 difference value in $\hat{I}_{diff}$ of a color channel c is larger than the corresponding threshold T_c. This results in an initial binary motion map M, which indicates the location of the motion-pixels.

Finally, we apply morphological operations on the full-size motion map M. These operations aim at removing possible detection noise (wrongly detected pixels) and enhance the

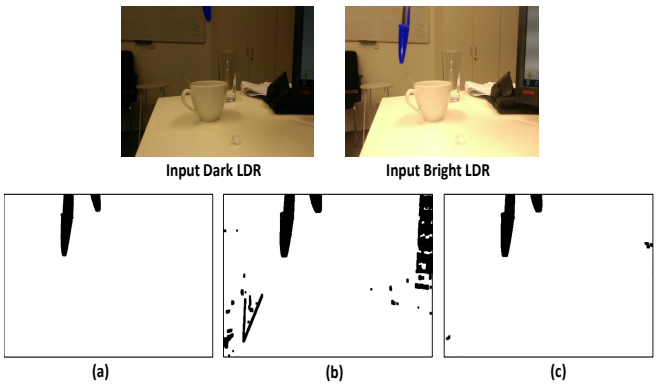

Fig. 1: Illustration of the visual impact of the **logistic function**-based manipulation of the difference image on the final motion map. **(a)** Ground truth motion map **(b)** Motion map **without** manipulation **(c)** Motion map **with** the proposed manipulation.

shape and filling of motion objects in the final motion map. The first operation starts with the counting of the number of motion pixels N_w inside a window of size $w \times w$ centered around each motion pixel in the motion map M. The parameter w is typically set to 3 or 5. The processing of the motion pixel under investigation is done as following:

- $N_w \leq \lfloor \frac{w^2}{2} \rfloor$: Probable motion noise. Motion pixel will be discarded from the motion map.
- $N_w > \lfloor \frac{w^2}{2} \rfloor$: Motion pixel is confirmed.

The second morphological operation marks areas in the immediate vicinity of motion-pixels as motion areas as well, using a similar window-based approach as previously described. This step enables to fill-up possible missing motion-pixels inside motion objects, and thus improves the shape of these objects in the final motion map $\hat{M}$. Likewise, the window is typically of size 3×3 or 5×5.

2.2. HDR De-ghosting

In this step, we propose to modify the exposure fusion algorithm, similar to [15], by including the gained motion map (or motion maps in case of more than 2 LDRs) into the weighting maps of the input images:

$$W_i(p) = (C_i(p))^{\omega_C} \times (S_i(p))^{\omega_S} \times (E_i(p))^{\omega_E} \times \hat{M}_i(p), \quad (4)$$

where $C_i(p)$ is the **contrast** map for image i at pixel p, $S_i(p)$ is the **saturation** map, $E_i(p)$ is the **exposedness** map. The parameters ω_C, ω_S and ω_E represent the corresponding power values. $\hat{M}_i(p)$ is the previously computed motion map of the image i. In the case of 2 input LDRs, the motion map corresponding to the designated reference image is composed of ones, as we assume that all pixels in the reference image are valid. The computed motion map $\hat{M}$ is assigned to the non-reference image and used according to the equation 4. In the case of more than 2 inputs, the weighting maps of the non-reference images are set to zero for motion related areas according to the equation 4, so that these pixels are excluded from the final HDR image.

3. EXPERIMENTAL RESULTS

We tested the performance of our approach on several sequences, with different exposure ratios between input LDRs as well as different numbers of input images. As shown in fig. 2, our motion map yields the best approximation for the ground truth motion map, despite the large exposure difference between the input LDRs. The accuracy of the motion map is important for the ensuing de-ghosting step, since wrongly detected pixels will be exempted from the exposure fusing procedure. This results in the deterioration of the final HDR image, especially in the case of 2 input LDRs, where the final weighting map for motion-pixels is extracted only from the reference image. The improvement brought by accurate motion maps is shown in fig. 3, where 2 examples of HDR de-ghosting are shown. The first set (first row) depicts a scene with large color difference and only 2 inputs. In addition, the scene contains camera as well as scene motion. As shown in the figure, the final HDR is ghost-free, and the computed motion map accurately detects the motion-related areas. The second scene is composed of 4 LDRs with significant scene motion. Likewise, the resulting HDR is totally ghost-free with no artifacts or additional noise.

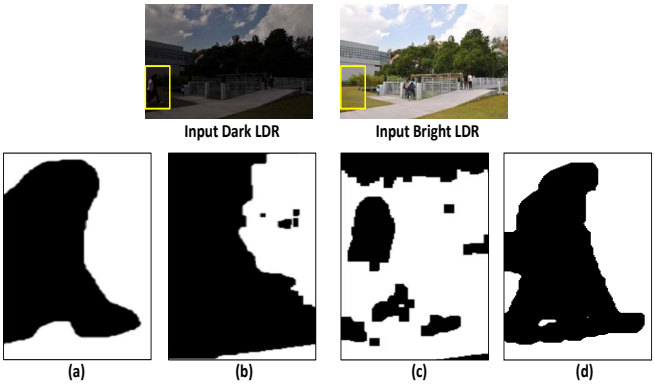

Fig. 2: Zoom-in on different motion maps computed using several approaches. **(a)** Ground truth map as presented by [15] **(b)** Motion map from [15] **(c)** Motion map from [14] **(d)** Our result. Images and maps (a), (b) and (c) are courtesy of [15].

In addition, fig. 4 contains a comparison of the results generated on a sequence provided by [10], with the results of [15]. As shown in the figure, our motion map accurately detects motion objects, with relatively small number of false positives. The motion map computed using the approach of [15] contains a high number of wrongly labeled pixels. This decreases the quality of the final HDR. The weighting maps in these areas are extracting information only from the

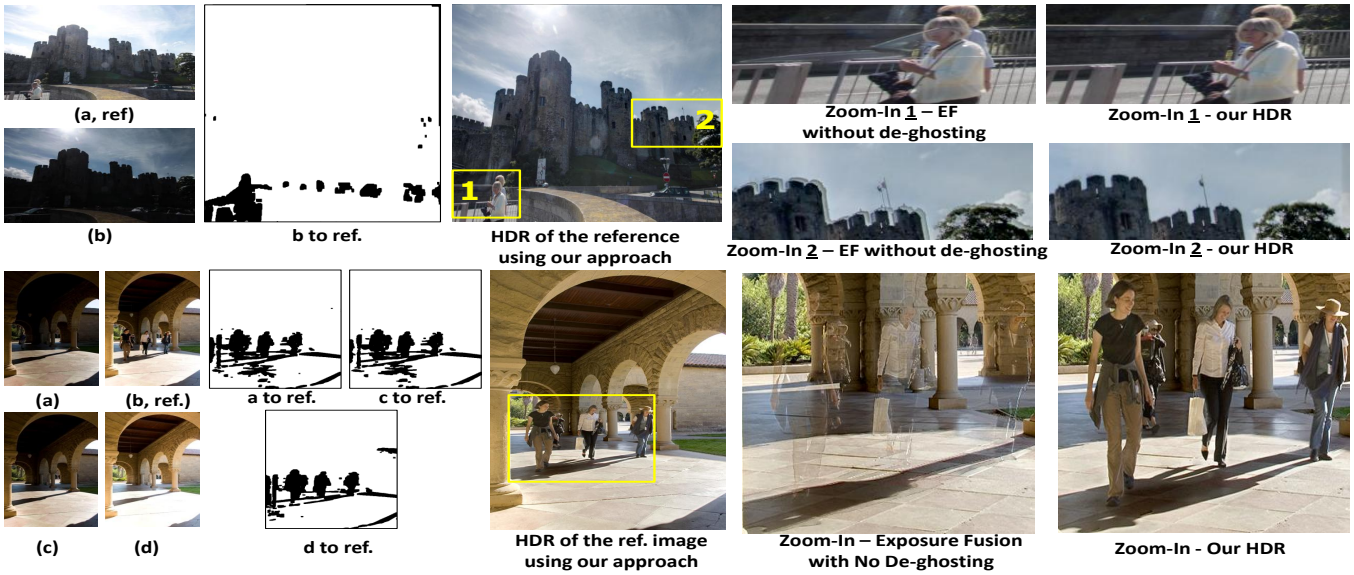

Fig. 3: Motion detection and HDR results from our approach on sequences from [21] (first row) and [12] (second row). The tested sequences present large color differences between the input LDRs as well as different motion types (camera and scene motion). The accurate binary maps improve the quality of the final HDR image. This is visible when we compare our results to the case where no de-ghosting is performed, as shown in the areas marked by the yellow squares.

reference image, which explains the observed blurriness. Besides, the approach [15] contains computationally expensive steps such as the computation of the **ZNCC** maps.

Furthermore, we compared the de-ghosting results of our approach to computationally expensive algorithms such as in [9]. As shown in fig. 5, our approach achieves similar to better de-ghosting results, while keeping a considerable low-complexity of the enabling algorithm.

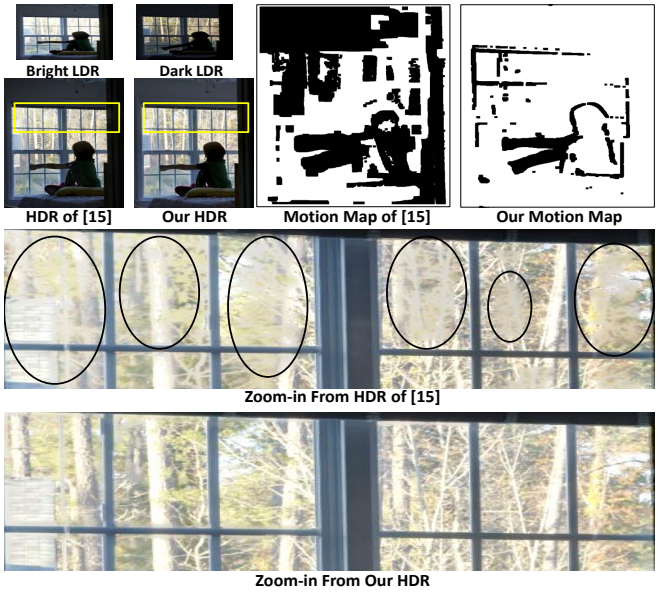

Fig. 4: Comparison of the results of our approach with the results of the method proposed in [15]. The motion map gained from our approach is more accurate, with less wrongly detect pixels. This improves the overall quality of the final image, as shown in the provided zoom-ins.

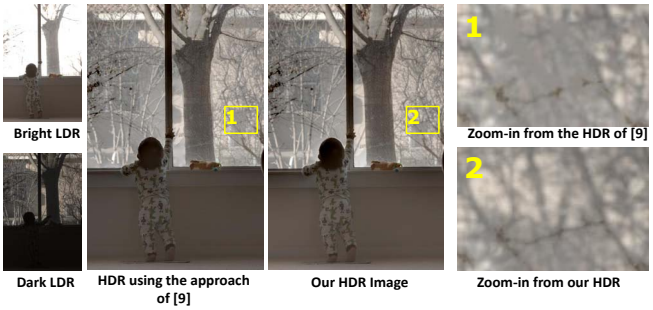

Fig. 5: Comparison of our HDR with the results of [9]. For this comparison, we aligned the input LDRs using the approach of [9] and generated the final HDR using EF. The comparison shows that even against high-complexity approach with motion correction steps, our approach provides comparable to better results, as shown in the zoom-ins.

4. CONCLUSION

In this paper we propose a novel de-ghosting approach based on the detection of camera and scene motion. The motion maps are constructed based on the manipulation of the difference image between the input LDR images. The high accuracy of the gained motion maps allows for a ghost-free final HDR. Aside from its ease of implementation and low-complexity, our approach obtains good quality even on extreme cases with only 2 or 3 input LDR images.

5. REFERENCES

[1] T. Mitsunaga and S. K. Nayar, "Radiometric Self Calibration," in *IEEE Computer Society Conference On*

Computer Vision and Pattern Recognition, vol. 1, 1999, pp. 374–380.

[2] P. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," in *Proceedings SIGGRAPH*, 1997, pp. 369–378.

[3] M. A. Robertson, S. Borman, and R. L. Stevenson, "Dynamic Range Improvement Through Multiple Exposures," in *International Conference on Image Processing*, vol. 3, 1999, pp. 159–163.

[4] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure Fusion," in *Pacific Graphics*, 2007, pp. 369–378.

[5] S. Kang, M. Uyttendaeke, S. Winder, and R. Szelisk, "High Dynamic Range Video," in *ACM SIGGRAPH*, 2003, pp. 319–325.

[6] H. Zimmer, A. Bruhn, and J. Weickert, "Freehand HDR Imaging of Scenes with Simultaneous Resolution Enhancement," in *EUROGRAPHICS*, 2011, pp. 319–325.

[7] C. Barnes, E. Schechtman, D. B. Goldman, and A. Finkelstein, "PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing," in *ACM Transactions on Graphics SIGGRAPH*, 2009.

[8] J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR Deghosting: How to deal with Saturation?" in *International Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1036–1170.

[9] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Schechtman, "Robust Patch-Based HDR Reconstruction of Dynamic Scenes," in *ACM Transactions on Graphics SIGGRAPH Asia*, 2012, pp. 203–214.

[10] D. Hafner, O. Demetz, and J. Weickert, "Simultaneous HDR and Optic Flow Computation," in *International Conference on Pattern Recognition*, 2014, pp. 2065–2070.

[11] E. A. Khan, A. O. Akyiiz, and R. E., "Ghost Removal in High Dynamic Range Images," in *IEEE International Conference on Images Processing*, 2006, pp. 2005–2008.

[12] O. Gallo, N. Gelfand, W. Chen, M. Tico, and K. Pulli, "Artifact-free High Dynamic Range Imaging," in *IEEE International Conference on Computational Photography*, 2009, pp. 1–7.

[13] J. An, S. H. Lee, J. G. kuk, and N. L. Cho, "A Multi-Exposure Image Fusion Algorithm Without Ghost Effect," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 1565 – 1568.

[14] F. Pece and J. Kautz, "Bitmap Movement Detection: HDR for Dynamic Scenes," in *Conference on Visual Media Production*, 2010, pp. 1–8.

[15] J. An, S. J. Ha, and N. I. Cho, "Probabilistic Motion Pixel Detection for the Reduction of Ghost Artifacts in High Dynamic Range Images from Multiple Exposures," in *EURASIP Journal on Image and Video Processing*, 2014, pp. 1–15.

[16] R. D. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. New Jersey: Prentice Hall, 2006.

[17] T. Vincent and R. Laganiere, "Detecting Planar Homographies in an Image Pair," in *Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis*, 2001, pp. 182–187.

[18] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded Up Robust Features," in *European Conference on Computer Vision*, 2006, pp. 404–417.

[19] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[20] F. Bouzaraa, O. Urfalioglu, and G. Cordara, "Dual-Exposure Image Registration for HDR Processing," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015, pp. 1553–1557.

[21] K. Karaduzovic-Hadziabdic, J. Hasic Telalovic, and R. Mantiuk, "Expert Evaluation of Deghosting Algorithms for Multi-Exposure High Dynamic Range Imaging," in *Second International Conference and SME Workshop on HDR imaging*, 2014, pp. 1–4.