HYBRID SYNTHESIS FOR EXPOSURE FUSION FROM HAND-HELD CAMERA INPUTS

Ru Li, Shuaicheng Liu, Guanghui Liu, Bing Zeng

School of Information and Communication Engineering
University of Electronic Science and Technology of China

ABSTRACT

The paper proposes a hybrid synthesis method for multi-exposure image fusion taken by hand-held cameras. Motions either due to the shaky cameras or caused by dynamic scenes should be compensated before any content fusion. The misalignment will cause blurring/ghosting artifacts in the fused result. The proposed method can deal with such motions and maintain the exposure information of each input effectively. In particular, the proposed method first applies optical flow for a coarse registration, which performs well with complex non-rigid motion but produces deformations at regions with missing correspondences. To correct such error registration, we segment images into superpixels and identify problematic alignments based on each superpixel, which is further aligned by PatchMatch. After that, the proposed method obtains a fully aligned image stack which facilitates a high-quality fusion that is free from blurring/ghosting artifacts. We compare our method with existing fusion algorithms on various challenging examples, including the static/dynamic, the indoor/outdoor and the daytime/nighttime scenes. Experiment results demonstrate the effectiveness and robustness.

***Index Terms*—** Multi-exposure fusion, optical flow, patch match

1. INTRODUCTION

Dynamic ranges of natural scenes are much wider than those captured by commercial imaging products. Digital cameras often fail to capture the irradiance range that visible to the human eyes. High dynamic range (HDR) imaging techniques have attracted considerable interests due to they can overcome such limitation. HDR imaging has been increasingly used in consumer electronics, industrial and security [1]. Directly capturing and displaying an HDR image is expensive. A relatively cheap way is to capture a stack of different exposure images and then merge them together. There are two main categories to conduct the synthesis. One is to reconstruct an HDR image through camera response function (CRF) and then apply the tone mapping for the display [2, 3]. The other category, multi-exposure fusion (MEF), can directly synthesize a low dynamic range (LDR) image from several different exposure images that is more informative and detailed than any input. Our method belongs to the second category.

Since its first introduction in 1980's [4], image blending algorithms evolve quickly. They produce high-quality results when input sequence is captured by static cameras mounted on a tripod under static scenes. Mertens *et al.* [5] combined contrast, saturation and exposedness information to generate weight maps and applied pyramid reconstruction to fuse multi-exposure image sequence. The method is robust and works well for static inputs. However, the weight maps are often too noisy and, sometimes, may yield artifacts such as unnatural transitions and detail losses. Later, some modified exposure fusion methods [6, 7, 8, 9] were proposed to improve the performance of fusion image by filters [6, 7] or through gradient reconstructions [8, 9].

Above typical MEF algorithms require input exposures to be perfectly aligned. Otherwise, any motion, either due to dynamic scenes or hand-shakes, will cause blurring/ghosting artifacts. In particular, the method [5] needs the inputs with strict alignment because every candidate pixel in the stack contributes to the final pixel value. If there are any misaligned regions, the fusion results would suffer from artifacts. As shown in Fig. 1, the fused result of [5] suffers from severe ghosting. The performance of fusion is highly dependent on the accuracy of motion estimation. Therefore, exploring efficient motion compensation strategies is essential. According to Tursun *et al.* [10], existing motion compensation methods can be divided into several categories: the moving objects removal [11], the moving objects selection [12, 13], the optical flow based registration [14, 15, 16] and the patch-based registration [17, 18, 19, 20]. Their core idea is to detect moving objects, then the dynamic areas are excluded or assigned with small weights when synthesizing inputs.

In this paper, we propose a hybrid synthesis fusion method for hand-held camera inputs with dynamic contents. It first applies optical flow [22] to align inputs to the reference. Optical flow methods can align images with complex motions, but produce deformations in the regions with no correspondences. As a result, we segment images into superpixels [23] and use PatchMatch [24] to correct error superpixels which are identified by computing flow motion variances. After the registration of optical flow and the compensation of PatchMatch, a fully registered image stack is delivered. Finally, we fuse registered images by [5].

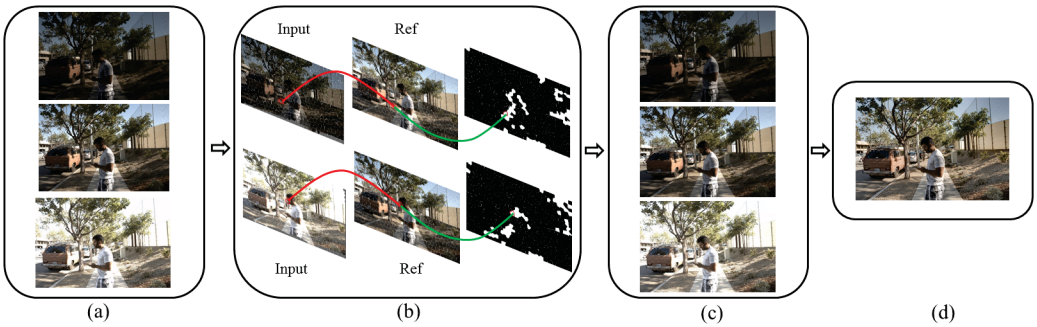

Fig. 2. The pipeline of our method. (a) Input images with different exposures. (b) Optical flow results and error regions caused by error flows. (c) PatchMatch results. (d) Fusion results.

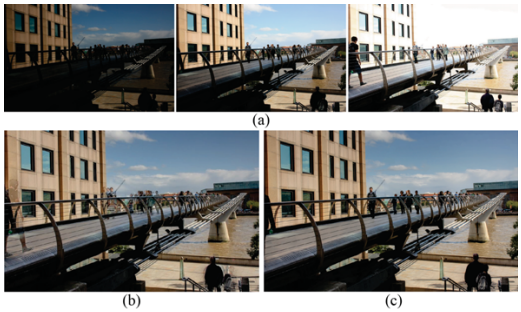

Fig. 1. The ghosting artifacts of MEF method. (a) Input images origin from Pece [21]. (b) Result of [5]. (c) Our result.

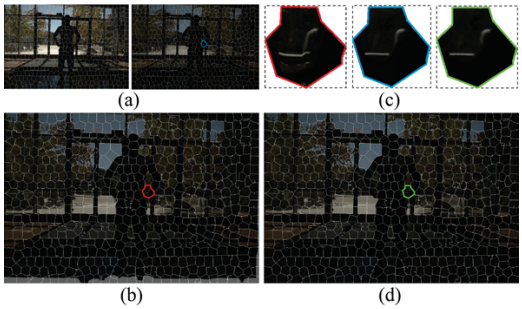

Fig. 3. Results of PatchMatch on Kalantari *et al.*'s scene [16]. (a) Reference image (left) and its exposure change result (right). (b) An optical flow result. (c) Some superpixel results. (d) PatchMatch result.

2. OUR METHOD

Figure 2 exhibits the hybrid fusion pipeline. Fig. 2(a) displays input image sequence. 'Input' images in Fig. 2(b) are optical flow alignment results. Fig. 2(b) shows the process of detecting and handling error regions caused by optical flow. The proposed method segments the reference image using superpixels and calculates flow variance within each of them. The corresponding superpixel is regarded as an error region once the variance exceeds a threshold. PatchMatch is utilized to further align error regions. Fig. 2(c) shows the results generated by PatchMatch correction. The combination of the two align methods achieves a fast speed as well as correctly aligns the challenging areas, such as regions containing occlusions and dynamic textures. Fig. 2(d) is the final fusion result.

2.1. Optical flow alignment

Given a sequence of LDR images with different exposures, the proposed method first aligns the low and high exposure images to the medium exposure one (reference) using optical flow. Kroeger *et al.*'s optical flow method [22] is chosen as coarse registration strategy. However, optical flow methods will produce deformations in the regions with discontinuous depth, which are caused by occlusions that either due to the parallax or owing to the dynamic contents. Locating these error regions and further align them are essential.

Instead of locating errors for every pixel, the proposed method segments images into superpixels, and detects errors for each superpixel. Dividing images into superpixels can not only maintain image continuity but also reduce the complexity of subsequent image processing tasks. SLIC segmentation [23] is applied to generate superpixels, which adapts a k-means clustering to efficiently generate superpixels. The proposed method first obtains the segmentation mask of the reference image by SLIC and then applies the mask to segment other aligned input images. As such, all input images share the same segmentation. When detecting error regions, flow motion variance is calculated for each superpixel. Error optical flow registration often possesses high motion variances. Therefore, applying a threshold is adapted to distinguish correct superpixels and problematic superpixels. If the variance exceeds an empirical threshold T_{flow}, the corresponding re-

gion is labeled as error region.

2.2. PatchMatch alignment

PatchMatch [24] algorithm is adopted to correct error registrations. The algorithm offers substantial performance improvements for the randomized patch-correspondence searching. The proposed method abandons running PatchMatch over the whole reference image by substitution of finding nearest neighbor field (NNF) in a small region. As shown in Fig. 3(c), the proposed method first gets the height h and width w of one superpixel and then runs PatchMatch in a region with height $2 * h$ and width $2 * w$ around the superpixel. Generally, running PatchMatch on the selected region can recover error registrations perfectly. However, PatchMatch method tends to maintain color and structural similarity between input and reference. Directly applying PatchMatch to align error regions eliminates the problem of dynamic objects but changes the exposure information of inputs.

The proposed method invokes IMF algorithm [25] to adjust the exposure of reference image to input image before utilizing PatchMatch. IMF is capable of mapping between intensity values of any two exposures. The combination of IMF and PatchMatch reconstruction excludes dynamic regions and maintains the exposure of inputs simultaneously.

Fig. 3(a) shows the reference image R (left) and its IMF adjusted exposure image L (right). Fig. 3(b) is a warped result of optical flow. Fig. 3(c) highlights some superpixels: left superpixel (red border) originates from Fig. 3(b), which suffers distortions (the door frame shows obvious bend); middle superpixel (blue border) is the corresponding region from L, which is regarded as the reference image when employing PatchMatch; right superpixel (green border) is PatchMatch result which is free from distortions. Fig. 3(d) is the final PatchMatch result without problematic superpixels.

2.3. Implementation details

The proposed approach is summarized in Algorithm 1. Two parameters are needed for clarification: number of superpixels N_s and the threshold T_{flow}. The value of N_s has a large impact on processing time. For an image with size 1500×1000, we set N_s around 580 empirically. Generally, it ranges from 560 to 600. The threshold T_{flow} takes different values when handling images with different exposures. It is set to 1.5 and 3.5 separately when dealing with under-exposure and over-exposure images.

3. EXPERIMENTS

Both objective evaluation and visual comparisons are conducted to verify the performance of the proposed method. In objective assessment section, we compare image quality metric values and calculation complexities of the proposed method against several state-of-the-art techniques which bear some resemblances with our method. In visual comparisons, apart from the comparisons with above methods of dynamic scenes, we also select several representative MEF algorithms for static scene comparisons. The combating MEF methods are chosen to cover a diversity of types.

Algorithm 1 Hybrid synthesis algorithm

Require: Source image sequence $\{S_k\} = \{S_k|1 \leqslant k \leqslant K\}$
1: Select the reference S_r and segment it into N_s superpixels
2: Generate $K-1$ latent image $\{L_k\} = \{L_k|1 \leqslant k \leqslant K, k \neq r\}$ using IMF
3: **for** each input image $S_{k(k\neq r)}$ **do**
4: Align the input to reference image using optical flow
5: Calculate flow motion variance for each superpixel
6: Detect errors according to T_{flow}
7: Further align error regions to L_k using PatchMatch
8: **end for**
9: Fuse aligned image sequence $\{\hat{S_k}\} = \{\hat{S_k}\,|1 \leqslant k \leqslant K\}$
Ensure: Fusion result $\hat{S}$

Table 2. Average execution time in seconds on 12 source image sequence of size $1500\times1000\times5$.

Alg	[16]	[19]	[17]	[20]	Our
Time(s)	68±5.1	381±38	252±20	16±0.7	12±1

3.1. Objective assessment

3.1.1. Objective evaluation values

The metric Q_S [29] is selected as evaluation criteria:

$$Q_S = \frac{1}{|W|}\sum_{w\in W}[\lambda(w)Q_0(a,f|w)+(1-\lambda(w))Q_0(b,f|w)] \quad (1)$$

where a and b are input images; f is fused image; the local quality $Q_0(a,b|w)$ is computed for the values $a(i,j)$ and $b(i,j)$ where pixels (i,j) lie in the sliding window w; W is the family of all windows. $\lambda(w)$ is a local weight between 0 and 1 indicating the relative importance of images a compared to image b. The quality metric evaluates how much of the salient information contained in each of the input images has been transferred into the fused image. It ranges from 0 to 1 with higher value indicating better performance. Table 1 displays the comparison results of our method with [17], [19] and [20]. As can be seen, the proposed method provides results with better Q_S values in most of the cases.

3.1.2. Calculation complexity

We conduct a complexity comparison which is shown in Table 2. Upon the same inputs, all experiments are conducted on a computer with i7 3.4 GHz CPU and 32G RAM. Our method takes approximately 12 seconds. Table 2 demonstrates that the proposed method is more computationally efficient.

Table 1. Performance comparisons with Sen *et al.* [17], Hu *et al.* [19] and Ma *et al.* [20].

	Pedestrian	Fountain	Noise	Cafe	Forth2	Falls	Sunset	Zurich	Zurich2	Train1	Capt1	Capt2	Avg
[17]	0.731	0.573	0.265	0.684	0.794	0.589	0.688	0.745	0.729	0.619	0.486	0.476	0.615
[19]	0.766	0.635	0.753	0.751	0.825	0.767	0.759	0.744	0.782	0.619	**0.499**	0.480	0.698
[20]	0.777	0.616	**0.783**	0.725	0.818	0.717	0.730	**0.771**	0.739	**0.639**	0.468	0.448	0.686
Our	**0.781**	**0.637**	0.769	**0.767**	**0.845**	**0.781**	**0.802**	0.754	**0.799**	0.620	0.490	**0.486**	**0.711**

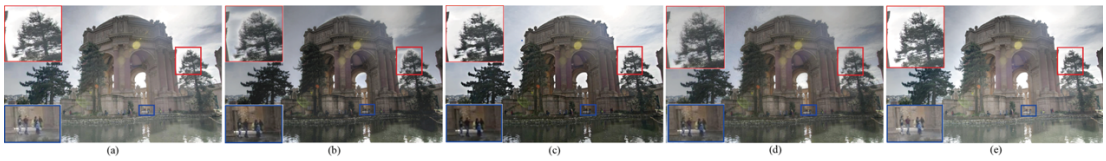

Fig. 4. Comparisons with other methods. (a) Result of [19]. (b) Result of [17]. (c) Result of [20]. (d) Result of [16]. (e) Our result. Please zoom in for a clearer observation.

Fig. 5. Comparisons with other MEF methods. (a) Result of [6]. (b) Result of [7]. (c) Result of [26]. (d) Result of [27]. (e) Result of [8]. (f) Result of [5]. (g) Result of [28]. (h) Our result.

3.2. Visual comparisons

Fig. 4 shows the comparisons with [19], [17], [20] and [16]. Fig. 4 (a) and (b) exhibite that patch-based methods can not deal with the structural regions well. Method [20] cannot recover some information perfectly such as the left tree in Fig. 4(c). The flow-based deep learning method [16] suffers from artifacts in tree leaves regions (Fig. 4(d)). Our method performs well and is free from such artifacts. Then we select several typical MEF algorithms to compare: [6], [7], [26], [27], [8], [5] and [28]. They require static inputs without shaky scenes and dynamic objects. Fig. 5 shows the comparison results. Other MEF methods can not deal with dynamic objects/texture properly so the tree leaves in the middle region suffer different degree of fuzziness (Fig. 5(a)-Fig. 5(g)). Our method can maintain the color and structural information simultaneously and generate natural fusion result.

4. CONCLUSION

We have presented a hybrid synthesis method for MEF. The method combines the strengths of flow-based methods and patch-based methods. Optical flow is applied for global registration which guarantees the computation efficiency. PatchMatch is used to align error regions, which can fully exclude moving objects. The method is evaluated both qualitatively and quantitatively to demonstrate the effectiveness.

5. ACKNOWLEDGEMENTS

This work was supported in part by National Foundation of China under Grants 61872067 and 61720106004, in part by Department of Science and Technology of Sichuan Province under Grant 2019YFH0016.

References

[1] A. Darmont, *High dynamic range imaging: sensors and architectures*. SPIE Washington, 2012.

[2] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM Trans. Graph.*, pp. 369–378, 1997.

[3] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, 2002.

[4] P. J. Burt, "The pyramid as a structure for efficient computation," in *Multiresolution image processing and analysis*, pp. 6–35, 1984.

[5] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion," *Computer Graphics Forum*, vol. 28, no. 1, pp. 382–390, 2007.

[6] S. Li and X. Kang, "Fast multi-exposure image fusion with median filter and recursive filter," *IEEE Trans. on Consumer Electronics*, vol. 58, no. 2, pp. 626–632, 2012.

[7] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. on Image Processing*, vol. 22, no. 7, pp. 2864–75, 2013.

[8] S. Paul, I. S. Sevcenco, and P. Agathoklis, "Multi-exposure and multi-focus image fusion in gradient domain," *Journal of Circuits Systems and Computers*, vol. 25, no. 10, p. 1650123, 2016.

[9] R. Li, X. He, S. Liu, G. Liu, and B. Zeng, "Photomontage for robust hdr imaging with hand-held cameras," in *Proc. ICIP*, pp. 1708–1712, 2018.

[10] O. T. Tursun, A. Erdem, and E. Erdem, "The state of the art in hdr deghosting: A survey and evaluation," *Computer Graphics Forum*, vol. 34, no. 2, pp. 683–707, 2015.

[11] W. Zhang and W. K. Cham, "Gradient-directed composition of multi-exposure images," in *Proc. CVPR*, pp. 530–536, 2010.

[12] C. Wang and C. Tu, "An exposure fusion approach without ghost for dynamic scenes," in *International Congress on Image and Signal Processing*, pp. 904–909, 2013.

[13] C. Lee, Y. Li, and V. Monga, "Ghost-free high dynamic range imaging via rank minimization," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1045–1049, 2014.

[14] L. Bogoni, "Extending dynamic range of monochrome and color images through fusion," in *International Conference on Pattern Recognition*, pp. 7–12, 2000.

[15] H. Zimmer, A. Bruhn, and J. Weickert, "Freehand hdr imaging of moving scenes with simultaneous resolution enhancement," *Computer Graphics Forum*, vol. 30, no. 2, pp. 405–414, 2011.

[16] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–12, 2017.

[17] P. Sen, E. Shechtman, E. Shechtman, E. Shechtman, E. Shechtman, and E. Shechtman, "Robust patch-based hdr reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–11, 2012.

[18] J. Hu, O. Gallo, and K. Pulli, "Exposure stacks of live scenes with hand-held cameras," in *Proc. ECCV*, pp. 499–512, 2012.

[19] J. Hu, O. Gallo, K. Pulli, and X. Sun, "Hdr deghosting: How to deal with saturation?," in *Proc. CVPR*, pp. 1163–1170, 2013.

[20] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, "Robust multi-exposure image fusion: A structural patch decomposition approach," *IEEE Trans. on Image Processing*, vol. 26, no. 5, pp. 2519–2532, 2017.

[21] F. Pece and J. Kautz, "Bitmap movement detection: Hdr for dynamic scenes," in *2010 Conference on Visual Media Production*, pp. 1–8, 2010.

[22] T. Kroeger, R. Timofte, D. Dai, and L. V. Gool, "Fast optical flow using dense inverse search," in *Proc. ECCV*, pp. 471–488, 2016.

[23] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. SuSstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012.

[24] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, p. 24, 2009.

[25] M. D. Grossberg and S. K. Nayar, "Determining the camera response from images: What is knowable?," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1455–1467, 2003.

[26] J. Shen, Y. Zhao, S. Yan, X. Li, *et al.*, "Exposure fusion using boosting laplacian pyramid," *IEEE Trans. on Cybernetics*, vol. 44, no. 9, pp. 1579–1590, 2014.

[27] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Trans. on Image Processing*, vol. 20, no. 12, pp. 3634–46, 2011.

[28] K. Ma, Z. Duanmu, H. Yeganeh, and Z. Wang, "Multi-exposure image fusion by optimizing a structural similarity index," *IEEE Trans. on Computational Imaging*, vol. 4, no. 1, pp. 60–72, 2018.

[29] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. ICIP*, pp. 98–173, 2003.