GHOST-FREE FUSION OF MULTI-EXPOSURE IMAGES IN THE GLOBAL GRADIENT REGION UNDER PATCH ALIGNMENT

Yulei Wang[1,2], Man Liu[1], Xi Chen[1], and Enyu Zhao[1]

[1] Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian 116026, China;
[2] State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xian 710071, China

ABSTRACT

High dynamic range (HDR) technology is one of the most widely used ways to improve image quality, and fusion of a series of low dynamic range (LDR) images is the main measure to obtain a HDR image. However, because moving objects are often found in a series of LDR images, the fused HDR images produce ghostly shapes. In order to eliminate ghosts, this paper proposes a ghost-free multi-exposure fusion method. Firstly, aligning the moving object in the input multi-exposure sequence images with the moving object in the reference image, and the aligned sequence images are obtained. In order to consider assigning more weight to pixels in the better exposure area, two weighting functions are defined. One is to measure pixel values relative to the overall brightness and adjacent exposure images, and the other one is to reflect pixel values within a range that has a larger global gradient relative to other exposures. Based on these two weighting functions, the low exposure sequence images aligned in the Laplacian pyramid are finally fused. Through experimental comparison, the obtained image has no ghost, good visual effect, and rich details.

***Index Terms*—**dynamic range image, ghost free, patch alignment, global gradient, Laplace pyramid

1. INTRODUCTION

With the development of image equipment and digital image processing technology, real scene reproduction has become the most urgent need of human beings. The real scene can show a wide dynamic range, and the human eyes can adapt to the real scene with a wide dynamic range, but the common image acquisition equipment and display equipment cannot capture and show the scene with a high dynamic range. Therefore, High Dynamic Range (HDR) image processing provides a new way to reproduce real scenes, and has become a research hotspot. HDR image acquisition mainly includes two methods. On one hand, HDR images can be directly obtained by hardware, which is expensive and difficult to popularize. On the other hand, the "HDR-like" image with rich details and extended dynamic range could be obtained by fusion of multiple LDR images with different exposure. Without camera curve calibration, HDR image reconstruction, tonal mapping and other steps, the fused images can reproduce the real scene on ordinary display devices [1].

However, when moving objects exist in the LDR images, such as moving cars, pedestrians on the road, branches in the wind, et al, obvious ghostly artifacts are bound to be generated in the fused image. In this paper, a new multi-exposure image fusion method is proposed, which can be applied to dynamic scene fusion and achieve ghost free fusion. First, the dynamic scene is converted to a "static scene" by aligning the moving areas in the multi-exposure sequence with the reference image according to the patch alignment formula. Then, a function is proposed which considers all multi-exposure images simultaneously to reflect the relative intensity and global gradient between images. On this basis, two kinds of weighting functions are designed and finally merged into the Laplacian pyramid to obtain the ghost free fusion image. Experiments show that the fused image obtained by this method has no ghost, rich details and good visual effect.

2. PROPOSE FUSION ALGORITHM

In this section the detailed flow diagram of the proposed algorithm is introduced in Fig.1. The proposed algorithm is divided into two steps. In the first step, in order to obtain the aligned LDR sequence images, a new multi-source bidirectional similarity measurement algorithm is proposed, which is optimized in the radiation domain, continuously iterated and selected for patches, and finally the "static" LDR sequence images of "motion-free region" are obtained. In the second step, two kinds of weighting functions are designed, one is based on pixel intensity and the other one is based on global gradient, and weighted average is then calculated to obtain the final weighting formula. Finally, the weighting formula is applied to the aligned sequence images in Laplace pyramid fusion to obtain the final fusion result and get the final HDR image.

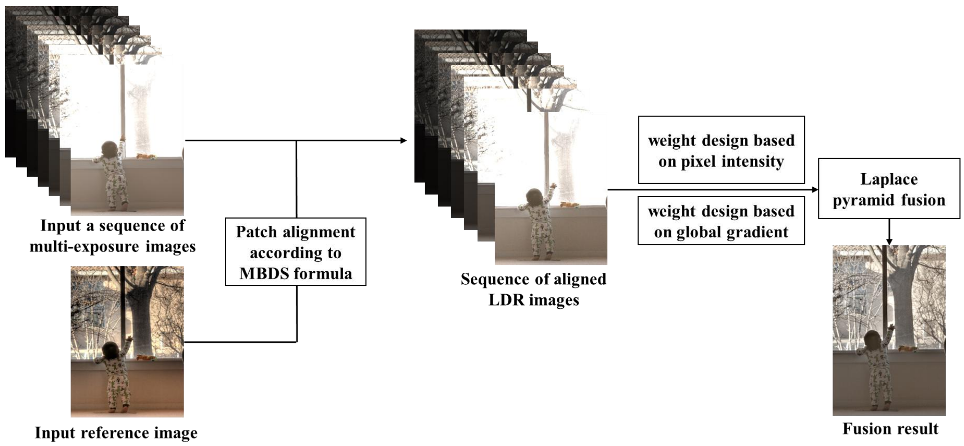


Fig.1. Graphical representation of the proposed image fusion algorithm

2.1 New method for eliminating ghosts

A new energy minimization formula is proposed to deal with multiple exposure sequences, and align the moving region in LDR with the patch method to eliminate the moving region. Enter a sequence of multi-exposure images, and a reference image $L_1, L_2, L_3 \cdots L_N$, L_{ref} . The proposed algorithm is designed to align the moving regions in the input LDR images with the moving regions in the L_{ref} to produce a set of "static" sequences. If the LDR sources are in JPEG or some other non-linear format, we first convert them into a linear space (range 0 to 1) using the appropriate camera response curve which is assumed to be known or can be estimated using established techniques. Our algorithm is improved based on the bidirectional similarity measure (BDS) proposed by Simakov et al in 2008 [2].

$$E_{BDS} = (L_{ref}|L_1, L_2 \cdots L_N) = \frac{1}{N}\sum_{P\in L_1,L_2\cdots L_N} \min d(P,Q) + \frac{1}{T}\sum_{Q\in L_{ref}} \min d(Q,P) \qquad (1)$$

In the above equation, T is the number of patches in the reference image, P and Q are patches in $L_1, L_2, L_3 \cdots L_N$ and L_{ref} respectively. $d()$ is similar to the $\mathcal{L}_2$ distance measure.

This function is the process of patch alignment, in which all facets of a multi-exposure sequence are found in the reference image (integrity) and vice versa (coherence). In some cases, when “aligned” with the reference image, what should be visible in the exposure may be obscured in L_k , so bidirectional similarity does not introduce missing information. To do this, we extend the BDS functionality to uses information from all other exposures, because the missing content may be visible (and well exposed) in one of these other images. This gives us a new measure of multi-source bidirectional similarity (or, more precisely, dissimilarity) for our applications.

A metric based on new multisource bidirectional similarity (MBDS) is proposed to measure this similarity concretely [3]. Minimizing MBDS implies that for every patch of pixels in L_k there should be a comparable patch in L_{ref} and for every patch in L_{ref} there is a comparable patch in L_k , across multiple scales.

$$E_{MBDS} = (L_{ref} \mid L_1, L_2, \cdots L_N) = \frac{1}{N}\sum_{k=1}^{N}\sum_{P\in L_1,L_2\cdots L_N} w_k(P) \min\nolimits_{Q\in L_{ref}} d(P,Q) + \frac{1}{T}\sum_{Q\in L_{ref}} \min\nolimits_{P\in L_1,L_2\cdots L_N} d(Q,P) \qquad (2)$$

We add a $w_k(P)$ term to the equation, meaning that when calculating integrity based on exposure, the source patch is weighted, which helps us ignore overexposed or underexposed patches and give priority to well-exposed source patches in multiple source images. We've normalized all of these weights.

The bidirectional nature of similarity measure is very important in our application. Using only consistency without regard to completeness can lead to duplication and other untrusted artifacts. Our algorithm focuses on the conditions around the edges of the ill-exposed area. If there is ambiguity at the edges, the integrity item algorithm can be used to fill the area by coherently copying the content from the correct part of the other images.

Thinking about integrity helps ensure that appropriate content is extracted from the reference image by discouraging unnecessary patches. Of course, in cases where the motion area appears in another source image, integrity can mistakenly extract stray patches and appear in the final result. Therefore, in order to improve this phenomenon, we optimize it through iteration and selection. Spurious objects are usually suppressed by averaging during iteration and do not appear in the final result.

2.2 Iteration and selection

In order to realize MBDS measurement, the open algorithm of Barnes et al. was used to speed up iteration and selection by Patch-Match algorithm, and modification was made to process multiple sources of MBDS.

First, a mapping function is used to convert a linear multi-exposure image sequence into a parameter range.

$$h(L_k) = (L_k)^{(1/\gamma)} \times exposure(k), k = 1,2,3 \cdots N \tag{3}$$

where $exposure(k)$ represents the exposure rate of k exposure and reference exposure, assuming that the reference exposure is the unit radiation degree. γ is the coefficient of the gamma curve in the camera response function, $\gamma = 2.2$. Eq.3 shows map LDR image L_k to the linear HDR radiance domain. For each source image, our method uses the current image L_k as of that level as MBDS target input and runs the intensive search step multiple times over all the adjusted source exposure $g^k(L_q)$.

$$g^k(L_q) = clip(((h(L_k))/exposure(k))^{(1/\gamma)}), q = 1,2,\cdots N \tag{4}$$

$g^k(L_q)$ is the approximate inverse of $h(L_k)$, but not exact because of the clipping process that occurs when capturing an LDR image. The bidirectional search produces two nearest neighbor domains (NNF) for each source exposure q: one for coherence and one for completeness. Note that the integrity search is masked, which means that the search is done only in the well-exposed parts of each source $g^k(L_q)$ .This effectively implements the $w_k(P)$ term in the equation. For each pixel in the final coherent NNF, the algorithm selects the pixel in the NNF stack that causes the minimum $\mathcal{L}_2$ distance, which deals with the minimum terms of all the sources in the equation. This results in an NNF integrity item for each exposure level q, and an NNF coherence item (with an add-on identifying the source).

For selection, the patches of coherent NNF is summed in the standard manner, using the patch from the appropriate exposure per pixel. On the other hand, for the integrity of the NNF algorithm, our algorithm uses each NNF to sum up the patches for each adjusted exposure and then averages them together. The final result can be generated by adding these two terms and dividing by the appropriate weight, which gives us our new $h(L_k)$. Repeat this procedure for all N sources.

2.3 Design weighting functions

Through the dynamic scene patch-based after high dynamic range image reconstruction algorithm, obtained with the reference image registration more exposure LDR image sequence, we design two kinds of weighting function: weighting function design based on pixel intensity and global gradient, based on the weight of these two kinds of weighting function to get the final type, used for the final fusion.

a) Weight design based on the pixel intensity

It is easy to find that different exposure graphs have different brightness. In order to emphasize the bright area in low exposure and the dark area in high exposure, we design a weight formula of brightness. Mertens et al. [4] proposed a weight formula related to brightness in 2009, but this weight still has some defects in capturing different brightness. Therefore, we control the weight to the overall brightness relative to the image. Specifically, dark areas are given greater weight when the overall image is brighter (longer exposure) and vice versa. The specific weight formula is shown in Eq.5

$$W_{1,n}(x,y) = \exp\left(-\frac{\left(I_n(x,y)-(1-m_n)\right)^2}{2\sigma_n^2}\right) \tag{5}$$

$I_n(x,y)$ represents the value of the n-th pixel. The average pixel intensity of the n-th image is expressed as m_n . The average pixel intensity of the n-th image is expressed as m_n . When $I_n(x,y)$ is close to $1-m_n$, a larger weight should be assigned.

b) Weight design based on the global gradient

We observe the histogram of multi-exposure image sequence and find that the low-exposure image is at the pixel value of 0, while the high-exposure image is one pixel value. Therefore, bright areas usually have high contrast, that is, pixel values have large gradient values. According to the change of gradient value and considering the global effect, we design a global gradient weight formula.

$$W_{2,n}(x,y) = \frac{Grad_n\left(I_n(x,y)\right)^{-1}}{\sum_{n=1}^{N} Grad_n\left(I_n(x,y)\right)^{-1} + \varepsilon} \tag{6}$$

Where ε is a very small positive value to prevent the denominator from being 0, and $Grad_n\left(I_n(x,y)\right)$ denotes the gradient of the cumulative histogram of intensity $I_n(x,y)$.

c) Fusion of the weights

The final weight is shown in Eq.7:

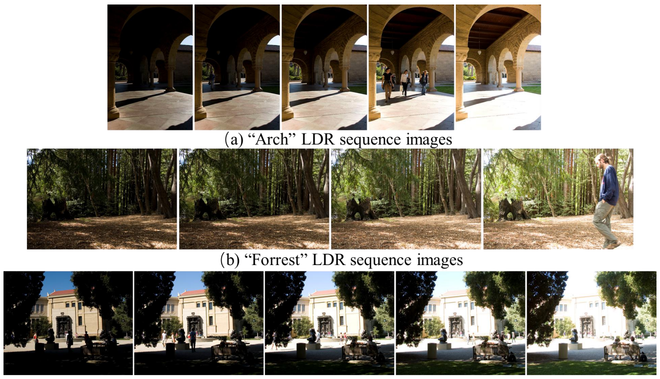
$$W_n(x,y) = \frac{W_{1,n}(x,y) \times W_{2,n}(x,y)}{\sum_{n=1}^{N} W_{1,n}(x,y) \times W_{2,n}(x,y) + \varepsilon} \tag{7}$$

After the weight formula is obtained, considering the phenomenon that the weight value is noisy and not smooth, we apply the equation in the Laplacian pyramid fusion, multi-resolution image processing, and get the final result.

3. EXPERIMENTAL RESULTS

In order to prove the efficiency of the proposed algorithm, it is compared with three algorithms, ZNCC algorithm proposed in [5], SIFT-GF algorithm proposed in [6] and FSPD algorithm proposed in [7], respectively. Three sets of exposure sequences are used in the experiments, which are ‘Arch’, ‘Garden’ and ‘Forrest’, shown in Fig.2 (a)-(c). The fusion results by four algorithms are shown in Fig.3.

It is difficult to distinguish the performance of four algorithms only by visual effects in Fig.3, and image quality assessment is needed for quantitative comparison. Blind Image Quality Index (BIQI) image quality assessment algorithm proposed by Moorthy is firstly used for comparison [8], shown in Table I. The BIQI score usually has a value between 0 and 100 (0 indicates the best and 100 indicates the worst). It can be seen from Table I that the BIQI score of the proposed algorithm (marked as Our) is the smallest among four, indicating that the proposed algorithm has the best performance for all three data sets.

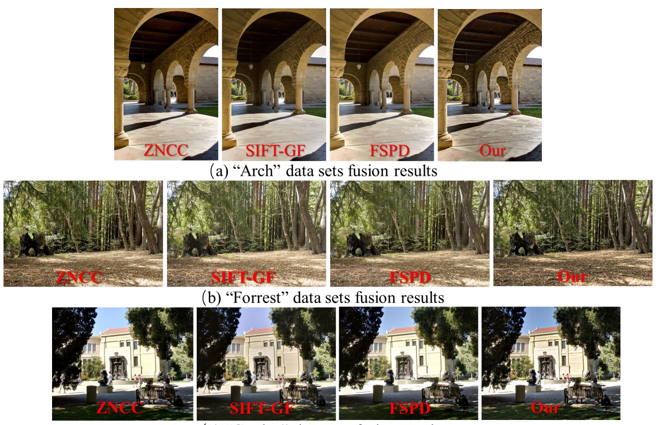


(a) "Arch" LDR sequence images

(b) "Forrest" LDR sequence images

(c) "Garden" LDR sequence images

Fig.2. Three sets of exposure sequences



(a) "Arch" data sets fusion results

(b) "Forrest" data sets fusion results

(c) "Garden" data sets fusion results

Fig.3. Fusion results of the four algorithms ZNCC, SIFT-GF, FSPD and OUR using three data sets

Table I. BIQI quality evaluation index

	ZNCC	SIFT-GF	FSPD	OURS
Arch	26.3588	23.4733	23.5338	**19.5374**
Garden	20.0731	31.6562	29.3509	**14.0210**
Forrest	46.0214	57.7348	53.3395	**42.6511**

Apart from BIQI score, some reference evaluation algorithms are selected for quality evaluation, such as Visual Saliency-based Index (VSI), Feature Similarity (FSIM), and Universal Quality Index (UQI). We use the "Arch" data set for comparison in this experiment, shown in Table II. It is also proved that the proposed algorithm has the highest score among all three quality evaluation indexes, indicating the effectiveness of the proposed algorithm among four algorithms.

Table II: Three quality evaluation indexes for "Arch" fusion results

	ZNCC	SIFT-GF	FSPD	OURS
VSI	0.8441	0.8912	0.8889	**0.8951**
FSIM	0.5646	0.6538	0.6527	**0.6668**
UQI	0.4362	0.3923	0.4355	**0.5051**

4. CONCLUSIONS

In this paper, a patch image fusion algorithm is proposed and applied to multiple-exposure of the dynamic scenarios. Firstly, patch block algorithm is used to align the motion areas in each source image with the reference image, obtaining a set of alignment "static" scenes. Then, two kinds of weighting function are applied, one is based on the pixel intensity and the other one is based on the global gradient. Based on these two kinds of weighting functions, the final weight is obtained by combining two weight functions with the Laplace pyramid. Through experimental analysis, the fusion image obtained by the proposed algorithm is saturated in color, rich in details, good in visual effect, and eliminates the influence of ghost, with best image quality evaluation indexes of BIQI, VSI, FSIM, and UQI.

5. ACKNOWLEDGEMENTS

This work is supported by the National Nature Science Foundation of China (61801075, 41801231), China Postdoctoral Science Foundation (2020M670723), Open Research Funds of State Key Laboratory of Integrated Services Networks, Xidian University (ISN20-15) and the Fundamental Research Funds for the Central Universities (3132019341).

REFERENCES

[1] Reinhard E, Heidrich W, Debevec P, et al. High dynamic range imaging: acquisition, display, and image-based lighting[M]. Morgan Kaufmann, 2010.

[2] Simakov D, Caspi Y, Shechtman E, et al. Summarizing visual data using bidirectional similarity[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008: 1-8.

[3] Sen P, Kalantari N K, Yaesoubi M, et al. Robust patch-based hdr reconstruction of dynamic scenes[J]. ACM Trans. Graph., 2012, 31(6): 203:1-203:11.

[4] Mertens T, Kautz J, Van Reeth F. Exposure fusion: A simple and practical alternative to high dynamic range photography[C]//Computer graphics forum. Oxford, UK: Blackwell Publishing Ltd, 2009, 28(1): 161-171.

[5] An J, Ha S, Cho, N. Probabilistic motion pixel detection for the reduction of ghost artifacts in high dynamic range images from multiple exposures[J]. EURASIP Journal on Image and Video Processing 2014 :1-15.

[6] Hayat N, Imran M. Ghost-Free Multi Exposure Image Fusion Technique Using Dense SIFT Descriptor and Guided Filter[J]. Journal of Visual Communication and Image Representation, 2019, 62: 295–308.

[7] Li H, Ma K, Yong H, et al. Fast Multi-Scale Structural Patch Decomposition for Multi-Exposure Image Fusion[J]. IEEE Transactions on Image Processing, 2020, 29: 5805-5816.

[8] Moorthy A K, Bovik A C. A two-step framework for constructing blind image quality indices[J]. IEEE Signal processing letters, 2010, 17(5): 513-516.