Motion Area based Exposure Fusion Algorithm for Ghost Removal in High Dynamic Range Video Generation

Shu-Yi HUANG†*, Qin LIU* , Hao WANG* and Takeshi IKENAGA†
*Software Institute, Nanjing University, Nanjing, China.
E-mail: {hsy10, qinliu, mf1332063}@software.nju.edu.cn Tel: +86 137-3918-4418
†Graduate School of Information, Production and Systems, Waseda University, Fukuoka, Japan.
E-mail: ikenaga@waseda.jp Tel: +81-93-692-5319

***Abstract*—High dynamic range (HDR) image is becoming prevalent for its superiority in recording details of the real-world scene. A conventional exposure fusion method can be utilized to generate a HDR image from a set of differently exposed low dynamic range (LDR) images according to their weight maps. However, this method causes ghosts in the HDR image if there are moving objects or the background changes in the LDR image sequence. Thus, conventional exposure fusion method is only suitable for HDR image generation, not for HDR video generation. A motion area based exposure fusion algorithm is therefore proposed to remove ghosts in HDR video generation. First, the motion areas are detected using block matching based motion estimation method. Then those areas are combined as an additional measurement to change values of original weight maps. The subjective and objective experimental results show ghost-free and highly detailed exposure fused image sequences from HDR videos.**

***Index Terms*—High Dynamic Range Video, Ghost Detection and Removal, Exposure Fusion, Motion Estimation**

I. Introduction

The dynamic range of a real world scene may reach four orders of magnitude, from shadows to sunlight, while the vast majority of CCD sensors people use today can only reproduce 256 levels per channel. Due to the limited dynamic range, auto-exposure cameras may correctly expose a region of interest such as a face, but fail to correctly expose the entire image and recover the whole dynamic range of the captured scene. Fortunately, high dynamic range (HDR) image technique can manifest both over-exposed and under-exposed regions simultaneously.

To extend the dynamic range of conventional cameras, the conventional works can be classified into two categories. Both of them achieve HDR image from couples of LDR images at different exposure levels. The first prevailing HDR process proposed by Debevec [1], both with and without the known response function of the camera, consists of three steps. First, estimate the camera response function (CRF). Second, recover the radiance map. Finally, apply tone mapping [2] to make the radiance map displayable on the commonly used LDR monitors. As an alternative, the other exposure fusion method proposed by Mertens [3] attempts to select the "good" pixels from the sequence and combines them into the final result, guided by quality measures like saturation and contrast. This method skips the typical HDR process, and no intermediate HDR image needs to be generated. Therefore, it avoids camera response curve calibration and is computationally efficient.

However, for both two approaches mentioned above, it is significant that the scene has to be completely static in order to avoid ghosts. Ghosts can be incurred by global camera motion such as hand shaking or moving objects in the input image sequence. In those cases, if LDR images are directly combined using exposure fusion method, the same object will appear at different locations in the resultant HDR image.

Thus, this paper proposes a method to generate a de-ghosting HDR video based on exposure fusion scheme. Given a stack of LDR frames, one is selected as a referential frame with the most useful information about the scene. Among other exposures in the stack, regions that cause ghosts are detected by using block matching based motion estimation algorithm. They are removed through changing weight map values when combined with the referential frame in the exposure fusion process. It then creates a better HDR frame, which removes the vast majority of ghosts in HDR videos.

Section 2 explains the details of our proposal. Subjective and objective evaluation results will be shown in Section 3. Finally, a conclusion is drawn in Section 4.

II. Our proposal

A comparison between this proposal and the original exposure fusion method is illustrated in Fig. 1. As two aligned camera solution will incur the angle difference of view, the inputs of this system are four adjacent frames taken by one single camera alternating with different exposure rates, in order like L1, H1, L2, H2 (L means low exposure rate, H means high exposure rate). The outputs of this system are two ghost-free HDR frames HDR1 and HDR2, generated from L1 and H1, L2 and H2. The whole framework consists of four steps:

1) ARPS3 Block Matching
2) Motion Area Registration
3) Motion Area based Weight Map Calculation
4) Fusion.

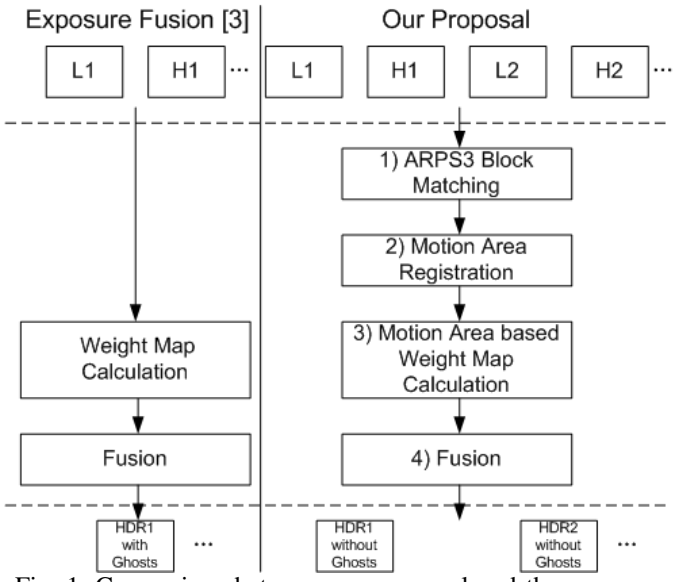

Fig. 1. Comparison between our proposal and the exposure fusion method[3]

The "ARPS3 Block Matching" step aims at obtaining motion vectors of "current frames", which are L1 and H1 in Fig. 1. The second step "Motion Area Registration" detects and labels moving objects. This step utilizes motion vectors calculated in the first step to estimate motion areas of each four frames in several, and then combines motion areas of two adjacent frames as a whole. For example, motion areas of frame L1 and H1 are combined as a whole and motion areas of L2 and H2 are combined as one another. Then those combined motion areas which result in ghosts are registered on two bitmaps. In the "Motion Area based Weight Map Calculation" step, the two bitmaps are used to reset values of four original weight maps. Finally, in step 4), L1 (L2) and H1 (H2) are fused to produce one ghost-free HDR frame. Fig. 2 is a graphic expression of the generation progress. It uses the 302nd to 305th frames in video *Bride* as an example.

A. ARPS3 Block Matching

This step aims at obtaining motion areas of four input frames. In order to remove ghosts, moving objects should be detected first. Block matching method is better suited for detecting moving regions as it has low complexity, making it suitable for practical applications. The goal of a block matching algorithm is to find a motion vector that describes the transformation from a block in a "current frame" i to its best matched block in a "reference frame" j. Various kinds of block matching algorithms are compared in the experiments. Considering both the processing time and the quality of composed HDR image, ARPS3 [4] is selected in this paper. Those images are gray-scaled before executing ARPS3 block matching. As images taken at different exposure levels focus on totally different parts of the scene, such as colorful sky in low exposure images and clear portrait in high exposure images, the corresponding current frame and its reference

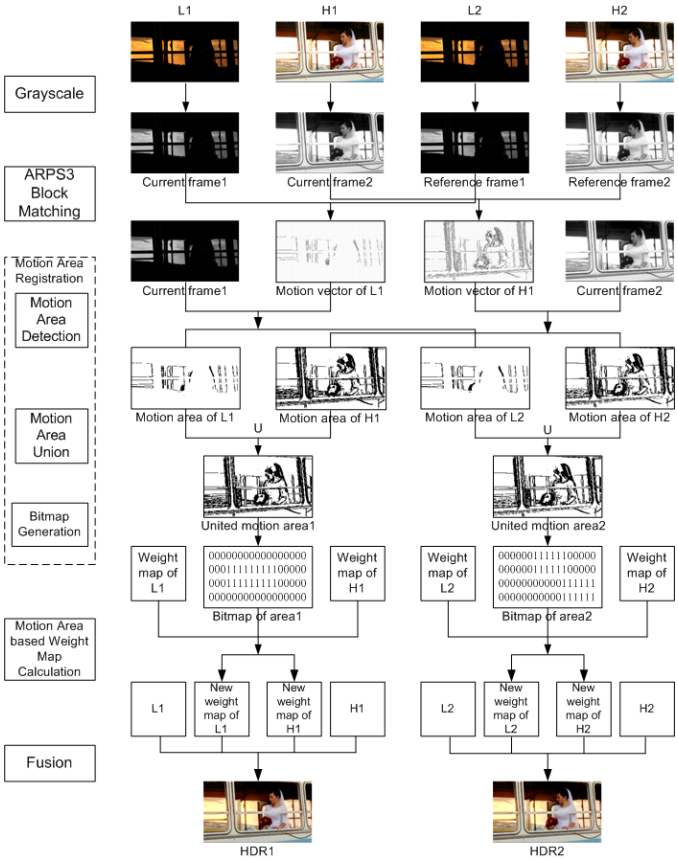

Fig. 2. Main flow of our proposal

frame must be at the same exposure level. For example, if L1 is selected as the current frame, then L2 rather than H1 should be selected as its reference frame. It makes no sense to execute block matching between different exposure frames. Similarly, if H1 is selected as the current frame then H2 should be its reference frame. Two steps are executed in this part: 1) Static Block Prediction, 2) Motion Vector Calculation. In the original ARPS3 method, macro-block with the size 16×16 and zero-motion threshold with value 512 are employed to perform motion estimation. Considering there is much camera noise in our image sequences, this proposal sets the value of zero-motion threshold as 2048 and uses block size 16×16.

Sum up "Block Matching" steps in our proposal as follows:

$$\begin{aligned} & if \quad SAD \quad < \quad 2048 \\ & \qquad MV = (0,0); \\ & else \\ & \qquad search \quad to \quad find \quad MV. \end{aligned} \tag{1}$$

B. Motion Area Registration

This part is to detect moving areas of four input frames and generate two 0-1 bitmaps in which 0 represents non-motion pixel and 1 represents motion pixel. Every bitmap reflects total moving areas of two adjacent frames such as L1 and H1, or L2 and H2. Two steps have to be executed in this part: 1) Motion Area Estimation, 2) Bitmap Generation.

1) Motion Area Estimation.

In this sub-step, our proposal calculates motion areas of current frames and reference frames separately.

i Motion Areas of Current Frames

During ARPS3 block matching, it is reasonable to consider the block of which motion vector doesn't equal (0,0) as moving area, otherwise, the reverse. Define k_{th} (k starts from 1) frame in the **current frame sequence** (Exp. L1H1L3H3L5H5$\cdots$) as C^k, define motion areas of C^k as CMA^k. Macro block at position (i,j) in C^k is defined as $CBlock^k_{i,j}$. Relatively, motion vector of $CBlock^k_{i,j}$ is defined as $MV^k_{i,j}$. Then, the acquisition of CMA^k can be formulized like (2):

$$\begin{aligned} CMA^k &= \bigcup CBlock^k_{i,j} \\ where \quad & MV^k_{i,j} \neq (0,0). \end{aligned} \tag{2}$$

ii Motion Areas of Reference Frames

For a block at location (i,j) in the current frame C^k, (i,j) + $MV^k_{i,j}$ represents the location of the best matched block in its reference frame, in other words, the new position of the moving object. Those best matched blocks compose motion areas in the reference frame. Define l_{th} frame in the **reference frame sequence** (Exp. L2H2L4H4L6H6$\cdots$) as R^l, define motion area of R^l as RMA^l, Macro block at position (p,q) in R^l is defined as $RBlock^l_{p,q}$. Then, RMA^l can be obtained as following (3):

$$\begin{aligned} RMA^l &= \bigcup RBlock^l_{p,q} \\ where \quad & p = i + MV^l_{i,j_x}, \\ & q = j + MV^l_{i,j_y} \\ and \quad & MV^l_{i,j} \neq (0,0). \end{aligned} \tag{3}$$

iii Motion Area Combination

As moving regions in two adjacent differently exposed frames don't overlap entirely in the sequence frame, a union operation is executed in this sub-step. This step combines the motion areas of two adjacent frames which are going to be fused. Define the motion regions of the r_{th} frame in the **whole sequence** as MA^r, then we get the combined motion regions UMA^r of the r_{th} and the $\{r+1\}_{th}$ frames following (4):

$$UMA^r = MA^r \cup MA^{r+1} \tag{4}$$

2) Bitmap Generation

In this step, a 0-1 bitmap for registering moving pixels of two adjacent frames is generated. Define bitmap for the r_{th} and the $\{r+1\}_{th}$ frames in the **whole sequence** as $Bitmap^r$. Define pixel at position (i,j) in the r_{th} frame as $P^r_{i,j}$, bit value of $P^r_{i,j}$ in $Bitmap^r$ is $b^r_{i,j}$. Then, this proposal utilizes combined motion areas obtained in the previous step to generate the bitmap:

$$\begin{aligned} & if \quad P^r_{i,j} \in UMA^r \quad or \quad P^{r+1}_{i,j} \in UMA^r \\ & \qquad b^r_{i,j} = 1 \\ & else \\ & \qquad b^r_{i,j} = 0 \end{aligned} \tag{5}$$

C. Motion Area based Weight Map Calculation

This step is an improvement of the original weight map calculation from the exposure fusion method. First, weight maps of input frames are calculated following (6) [3]. For each pixel in one LDR image, the method combines information from different measures into a scalar weight map. With measures C, S and E, being contrast, saturation, and well-exposeness, and corresponding "weighting" exponents ω_C, ω_S, and ω_E. The subscript ij,k refers to pixel (i,j) in the k_{th} image.

$$W_{ij,k} = (C_{ij,k})^{\omega_C} \times (S_{ij,k})^{\omega_S} \times (E_{ij,k})^{\omega_E} \tag{6}$$

Define the weight map of the r_{th} frame in **whole frame sequence** as W^r and weight value of pixel at location (i,j) in W^r is $w^r_{i,j}$. As moving regions of two adjacent differently exposed frames don't overlap exactly, the same moving object will appear twice in the directly fused HDR image. Thus, for keeping the moving object and removing the ghosts, moving regions at a selected exposure level should be reserved and moving regions at other exposure levels should be removed. Our proposal achieves this result by setting weight values of motion pixels in the selected frame's weight map as 1 while setting weight values of motion pixels in the other frame's weight map as 0. Assume that moving objects in the $\{r+1\}_{th}$ frame are reserved. Then, this proposal utilizes the bitmap obtained in the previous step to refine weight maps according to (7).

$$\begin{aligned} & if \quad b^r_{i,j} = 1 \\ & \qquad w^r_{i,j} = 0, \quad w^{r+1}_{i,j} = 1; \\ & else \\ & \qquad w^r_{i,j} \quad and \quad w^{r+1}_{i,j} \quad remain \quad unchanged. \end{aligned} \tag{7}$$

D. Fusion

Finally, under the guidance of newly refined weight maps W^r and W^{r+1}, this proposal blends the r_{th} frame and the $\{r+1\}_{th}$ frame to get the ghost-free HDR frame. The method fuses differently exposed images using a Laplacian decomposition of the images and a Gaussian pyramid of the weight maps. Let the l_{th} level in a Laplacian pyramid decomposition of an image A be defined as $L\{A\}^l$, and $G\{B\}^l$ for a Gaussian pyramid of image B. Then, the method blends the coefficients in a similar fashion to (8) [3]. Finally, the pyramid $L\{R\}^l$ is collapsed to obtain resultant image. $\hat{W}_{ij,k}$ is the normalized value of the N weight maps such that they sum to one at each pixel (i,j) [3].

$$L\{R\}^l_{ij} = \sum_{k=1}^{N} G\{\hat{W}\}^l_{ij,k} L\{I\}^l_{ij,k} \tag{8}$$

TABLE I
ORIGINAL INPUT FRAME SEQUENCE

Bride(30fps)	Campus(50fps)	Tunnel(59fps)	Cup(25fps)

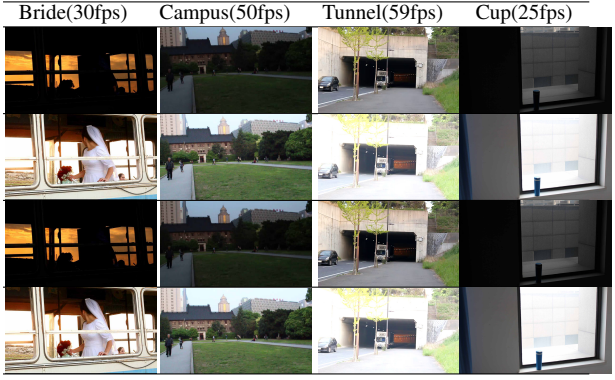

TABLE II
AVERAGE BIQI OF COMBINED HDR FRAMES

	Bride	Campus	Tunnel	Cup
Directly Fusion [3]	38.50	40.80	27.86	59.56
Pece [5]	40.67	38.36	30.89	46.38
our method	34.98	38.67	27.91	57.48

TABLE III
OVERVIEW COMPARISON FOR DIFFERENT METHODS

	Directly Fusion [3]	Pece [5]	Our Proposal
Bride			
Cam-pus			
Tunnel			
Cup			

III. Experimental Results

Experiments are run on Dell OptiPlex 790. CPU is Intel (R) Core (TM) i5-2400 and the RAM is 4GB. Video *Campus* and *Cup* as shown in the following Table I are recorded by Canon 650D, and video *Tunnel* is recorded by Canon 550D. The *Bride* video can be found on http://hdr.glogger.mobi/.

In the experiment, this proposal is compared with direct exposure fusion [3] method and the de-ghost method proposed by Pece [5]. For evaluating the entire quality of composed HDR frame, BIQI [6] values of generated HDR frames are calculated. The lower the value, the better the image quality will be. BIQI results in Table II are satisfactory. The overview and detailed of combined HDR frames are shown in Table III and IV respectively. Obviously, our proposal can remove the vast majority of ghosts.

TABLE IV
DETAILED COMPARISON FOR DIFFERENT METHODS

	Directly Fusion [3]	Pece [5]	Our Proposal
Bride			
Cam-pus			
Tun-nel			
Cup			

IV. Conclusion

Ghost problem is a critical problem in high dynamic range video generation. This paper proposes a ghost detection and removal method combining block matching based motion estimation method and exposure fusion methods. First, motion areas are detected using block matching based motion estimation method ARPS3. Then in exposure fusion process, the weights of motion areas which need to be reserved in the certain LDR image are set as 1, while the weights of motion areas which need to be removed in other LDR images are set as 0. Experiments show this proposal can remove the vast majority of ghosts in different video sequences.

Acknowledgment

The Project is supported by the Natural Science Foundation of Jiangsu, China(Grant No. BK20130588) and KAKENHI-Japan(26280016).

References

[1] Debevec, Paul E., and Jitendra Malik. “Recovering high dynamic range radiance maps from photographs.” ACM SIGGRAPH 2008 classes. ACM, 2008.

[2] Durand, Fredo, and Julie Dorsey. “Fast bilateral filtering for the display of high-dynamic-range images.” ACM Transactions on Graphics (TOG). Vol. 21. No. 3. ACM, 2002.

[3] Mertens, Tom, Jan Kautz, and Frank Van Reeth. “Exposure fusion.” Computer Graphics and Applications, 2007. PG’07. 15th Pacific Conference on. IEEE, 2007.

[4] Ma, Kai-Kuang, and Gang Qiu. “Unequal-arm adaptive rood pattern search for fast block-matching motion estimation in the JVT/H. 26L.” Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on. Vol. 1. IEEE, 2003.

[5] Pece, Fabrizio, and Jan Kautz. “Bitmap movement detection: HDR for dynamic scenes.” Visual Media Production (CVMP), 2010 Conference on. IEEE, 2010.

[6] Moorthy, Anush K., and Alan C. Bovik. “A two-step framework for constructing blind image quality indices.” Signal Processing Letters, IEEE 17.5 (2010): 513-516.