Computer Vision and Image Understanding 193 (2020) 102929

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Multi-exposure photomontage with hand-held cameras☆

Ru Li, Shuaicheng Liu, Guanghui Liu *, Tiecheng Sun, Jishun Guo

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

ARTICLE INFO

Communicated by Nikos Paragios

MSC:
41A05
41A10
65D05
65D17

Keywords:
Multi-exposure image fusion
MRF
Rough registration

ABSTRACT

The paper studies the image fusion from multiple images taken by hand-held cameras with different exposures. Existing methods often generate unsatisfactory results, such as blurring/ghosting artifacts due to the problematic handling of camera motions, dynamic contents, and inappropriately fusion of local regions (e.g., over or under exposed). In addition, they often require a high-quality image registration, which is hard to achieve in scenarios with large depth variations and dynamic textures, and is also time-consuming. In this paper, we propose to enable a rough registration by a single homography and combine the inputs seamlessly to hide any possible misalignment. Specifically, the method first uses a Markov Random Field (MRF) energy for the labeling of all pixels, which assigns different labels to different aligned input images. During the labeling, it chooses well-exposed regions and skips moving objects at the same time. Then, the proposed method combines a Laplacian image according to the labels and constructs the fusion result by solving the Poisson equation. Furthermore, it adds some internal constraints when solving the Poisson equation for balancing and improving fusion results. We present various challenging examples, including static/dynamic, indoor/outdoor and daytime/nighttime scenes, to demonstrate the effectiveness and practicability of the proposed method.

1. Introduction

High-dynamic-range (HDR) imaging techniques have been increasingly used in consumer electronics, road traffic monitoring, and other industrial, security, or military applications (Darmont, 2012). However, digital cameras often fail to capture the irradiance range that visible to human eyes. It is thus quite significant to explore effective HDR synthesis methods or detailed low dynamic range (LDR) synthesis methods. HDR synthesis methods focus on generating HDR images directly and their results are always tone mapped to LDR images which preserve details better than any of its single exposure counterpart (Debevec and Malik, 1997; Reinhard et al., 2010; Sen et al., 2012; Kalantari, 2017; Wu et al., 2018; Yan et al., 2019). Detailed LDR synthesis methods directly synthesize the result from multi-exposure images (Burt, 1984; Burt and Kolczynski, 1993; Mertens et al., 2007; Wang et al., 2018; Ma et al., 2019). Our method belongs to the detailed LDR synthesis category.

Although the multi-exposure fusion (MEF) approaches have been studied extensively, there are still some drawbacks. For instance, many existing methods have employed some kinds of merging techniques, which assume that multiple exposure images are accurately aligned (Li and Kang, 2012; Li et al., 2013; Paul et al., 2016). Thus, any misalignment due to either camera motions or dynamic contents will lead to the so-called ghosting/blurring artifacts. In the meantime, a Laplacian pyramid reconstruction scheme for image fusion was proposed in Burt and Adelson (1983), which has been widely adopted in many subsequent works (Burt and Kolczynski, 1993; Mertens et al., 2007; Shen et al., 2014). However, the method (Mertens et al., 2007) also requires the inputs to be strictly aligned. For each pixel location, every aligned candidate pixel in the stack contributes to the final pixel value. Thus, if there are any misaligned regions, the fused results would suffer from the ghosting or blurring artifacts. Fig. 1 shows two examples, where the input images are aligned before fusion, but the scenes contain dynamic textures or objects (tree leaves in the left example and moving persons in the right example). The fused results by Mertens et al. (2007) suffer from the blurry (Fig. 1 left) and the ghosting (Fig. 1 right).

Later on, some de-ghosting methods are proposed to handle aforementioned problems (Tursun et al., 2015). Firstly, some methods based on energy optimization are introduced to maintain image consistency or distinguish different parameters (Jinno and Okuda, 2008; Granados et al., 2013). Secondly, some flow-based methods realize registration with pixel-level accuracy and are effective for aligning moving objects between two images (Kang et al., 2003; Zimmer et al., 2011; Kalantari, 2017). Thirdly, some patch-based methods (Sen et al., 2012; Hu et al., 2013) are proposed to reconstruct the input images by patch-based

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.cviu.2020.102929.

* Corresponding author.
E-mail address: guanghuiliu@uestc.edu.cn (G. Liu).

https://doi.org/10.1016/j.cviu.2020.102929
Received 23 July 2019; Received in revised form 4 February 2020; Accepted 5 February 2020
Available online 7 February 2020
1077-3142/

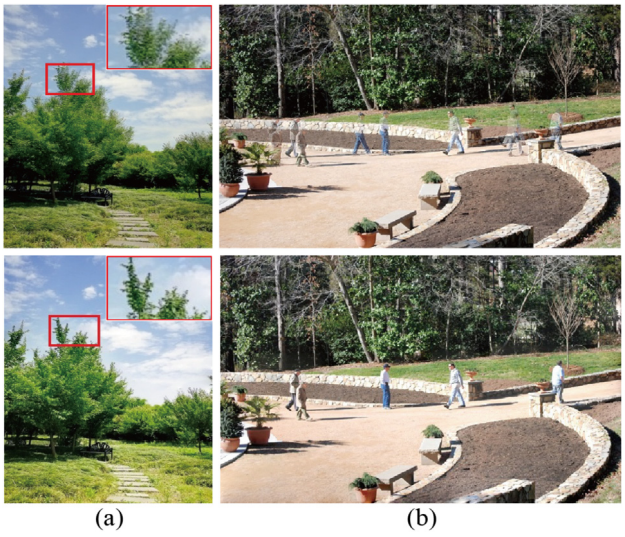

Fig. 1. Comparisons with Mertens et al. (2007). The inputs of left results are taken by us and the right inputs are Hu et al.'s scenes (Hu et al., 2013). Top images in (a), (b) are results of (Mertens et al., 2007). Our results are shown at the bottom. The comparisons indicate that our method can effectively solve blur and dynamic objects.

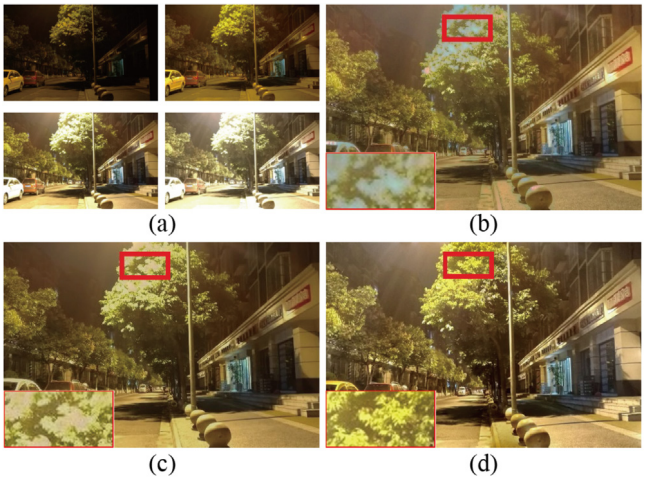

Fig. 2. Comparison of our method with two patch-based method: (Hu et al., 2013; Sen et al., 2012). (a) Source image sequence. (b) Result of Hu et al. (2013). (c) Result of Sen et al. (2012). (d) Our result.

synthesis according to one selected reference image, to form a fully registered image stack. The full alignment means that the reconstruction compensates both the camera and the scene motions. The synthesized candidates are then sent to the fusion framework. However, the patch-based reconstruction is not always robust in complicated situations, especially when encountered with dynamic textures (e.g., fountains, waterfalls, tree leaves in the wind), or structured regions. Fig. 2 shows such an example, where two patch-based methods generate blurry results in tree crown regions.,

High quality fully registration is challenging. For one thing, it is difficult to achieve a high-quality alignment under different appearances (Cui et al., 2017). For another, large foreground (Zhang et al., 2016) and near-range objects (Liu et al., 2016) would complicate the alignments, and scenes with large depth variations cannot be registered by a single homography, or by more sophisticated models (Lin et al., 2017). Besides, non-parametric approaches such as optical flows tend to generate errors at discontinuous depth boundaries (Kalantari, 2017) and the patch based reconstruction is also prone to produce errors as shown in Fig. 2.

To pursue a robust solution, the proposed method abandons the requirement of full alignment and replaces it by a rough registration with a single homography. As such, the photomontage idea proposed by Agarwala et al. (2004) is applied to compose the multi-exposure images that have been aligned roughly. However, our setting is different from Agarwala et al. (2004) in two aspects. First, Agarwala et al. generate composites interactively, which combines parts of a set of photographs into a single composite picture. Users select preferred image regions (e.g., a region containing a smiling face) at different pictures. In contrast, our solution is fully automatic because we combine image parts according to their exposure qualities. Second, the combined photos of Agarwala et al. (2004) were captured by a static tripod, whereas our inputs are captured by hand-held cameras. In our implementation, the method does not require the perfect registration, as long as it finds good seams to hide the misalignment.

The proposed method consists of some specific components as follows. It selects sub-image regions from different roughly aligned exposures by an MRF labeling and combines them seamlessly in the gradient domain. In this way, each pixel value belongs to a single image such that it is possible to maintain details well and handle blurring effectively. Moreover, it considers the dynamic identification and exposure selection in the MRF optimization simultaneously. The selected regions are not only well-exposed but are free from the interferences of dynamic objects/textures. Overall, the main contributions are:

(1) The proposed method relaxes the conditions of inputs. Conventional image alignment algorithms always fail to align inputs from hand-held cameras with large shaking. The proposed method abandons the requirement of full registrations, which can handle various complicated inputs and generate high-quality fusion results.

(2) The proposed method introduces the dynamic exclusion technique to handle moving objects. An energy optimization is first applied to detect moving objects and then a mask is generated to identify the dynamic pixels of each input, which reflects the probabilities of pixels being static or dynamic. The final results are free from ghosting with proper exposure.

(3) We propose to add some internal constraints to lighten underexposed regions.

(4) We conduct comprehensive comparisons to demonstrate the effectiveness of our method, including objective assessment, visual comparison, complexity comparison and subjective evaluation.

2. Related works

HDR images can be constructed by either directly capturing from special hardware (Nayar and Mitsunaga, 2000; Tocci et al., 2011), or synthesizing from multiple low dynamic range (LDR) images at different exposure levels using camera response function (CRF) (Mitsunaga and Nayar, 1999; Grossberg and Nayar, 2003), and then applying tone mapping (Fattal et al., 2002; Rana et al., 2018) to display (Mann and Picard, 1995; Debevec and Malik, 1997). MEF methods have become the most frequently used methods to generate fusion results (HDR outputs or detailed LDR outputs) for their low-cost and availabilities, which can be divided into two categories.

Methods that are suitable for static inputs. There are various MEF algorithms to achieve image fusion with fully static inputs. First, several edge-preserving filter-based methods have been proposed, including guided filter (Li et al., 2013; He et al., 2013), recursive filter (Li and Kang, 2012) and bilateral filter (Tomasi and Manduchi, 1998). Second, gradient reconstruction is also widely adopted in image fusion. The reconstructed image is usually obtained from the manipulated gradient by solving the Poisson equation (Pérez et al., 2003). Some variations from Pérez et al. (2003) are then designed to achieve an acceleration (Jia et al., 2006) or perform the image completion (Shen et al., 2007). We construct the Laplace image by the approach (Levin et al., 2004) in which we add some additional constraints for the lightening of some local image regions. As such, the proposed method

realizes natural transitions while maintaining image details. Third, Ma et al. proposed an optimization method, which optimizes an objective quality measure to improve the performance of typical MEF methods (Ma et al., 2017). Then, Some deep learning methods are proposed (Prabhakar et al., 2017; Ma et al., 2019) to achieve better fusion performance and improve the speed.

Methods based on motion detection or registration. Aforementioned MEF methods are just suitable for static scenes. To extend the scope of application, many algorithms approach the de-ghosting problem from different perspectives, providing solutions that range from rudimentary heuristics to advanced computer vision techniques (Tursun et al., 2015). Dynamic objects compensation should be properly processed, otherwise the fused results are easily ruined by the ghosting or blurring, whose core idea is to detect moving objects first and then exclude the dynamic areas or assign dynamic areas with small weights in synthesis process. The motion-based methods include global exposure registration (Tomaszewska and Mantiuk, 2007), moving objects removal (Zhang and Cham, 2010), moving objects selection (Wang and Tu, 2013; Lee et al., 2014; Li and Zhang, 2018) and moving objects registration (Sen et al., 2012; Hu et al., 2013; Kalantari, 2017). Eden et al. introduced a two-step graph-cut approach to detect and handle dynamic objects (Eden et al., 2006). Sen et al. proposed a patch-based energy minimization approach that integrates alignment and HDR reconstruction in a joint optimization (Sen et al., 2012). Hu et al. optimized image alignment based on brightness and gradient consistencies on the transformed domain (Hu et al., 2013). Granados et al. modeled the noise distribution of color values and incorporated it into MRF to produce HDR images (Granados et al., 2013). Kalantari et al. used optical flow to align the input images to the reference image, then employed a convolutional neural network to obtain the HDR image (Kalantari, 2017). Wu et al. introduced a non-flow-based deep framework for handling scenes with large-scale foreground motions (Wu et al., 2018). Yan et al. introduced an attention-guided network to obtain ghost-free results (Yan et al., 2019). In this work, the method identifies the dynamic areas by MRF labeling. In particular, it combines the region selection and dynamic detection into a unified optimization such that the selected regions are both free from ghosting and well-exposed.

3. Method

The input images are captured by hand-held cameras with varying exposures. The first step is to align them for motion compensation. By default, the image with median exposure is picked as the target, to which the other images are aligned. Slight misaligned errors could be tolerated in our implementation. We choose the Features from Accelerated Segment Test (FAST) (Rosten and Drummond, 2006) for the feature detection and track them by the Kanade–Lucas–Tomasi (KLT) (Shi and Tomasi, 1994). Specifically, we apply grid-based FAST feature detection (Guo et al., 2016) to prune the insufficient features of rich feature regions and balance flat regions with zero gradients. As such, features are more robust against the luminance differences (dark regions in under-exposed images and light regions in over-exposed regions cannot be ignored).

Fig. 3 shows the system pipeline after the alignment. Without loss of generality, we take four input images as an example. Fig. 3(a) shows the aligned input image sequence. Fig. 3(b) displays corresponding weight maps calculated by method (Mertens et al., 2007), which are then applied to produce a label map (Fig. 3(c)) through MRF energy minimization. The Laplace values are collected at each pixel from different images according to the label map to yield a Laplace image. By solving the Poisson equation properly, The final result is shown in Fig. 3(d).

3.1. Region selection

The following energy function is optimized for the labeling:

$$E(X) = \sum_{i \in \upsilon} E_1(x_i) + \lambda' \sum_{i \in \upsilon} E_2(x_i) + \lambda'' \sum_{(i,j) \in \varepsilon} E_3(x_i, x_j) \tag{1}$$

where each candidate image corresponds to a label and x_i is the label of the pixel i. $E_1(x_i)$ and $E_2(x_i)$ are data terms, in which $E_1(x_i)$ is the likelihood energy representing the exposure qualities. $E_2(x_i)$ encodes the dynamic information. $E_3(x_i, x_j)$ is the smoothness term that encourages the label similarities between neighboring pixels. υ is the set of all pixels and ε is the set of adjacent pixels. λ' and λ'' balance the terms. The energy can be minimized efficiently using graph-cut (Boykov et al., 2001).

3.1.1. Exposure weights

$E_1(x_i)$ represents the exposure qualities. It consists of three parts, the contrast, the saturation and the exposedness (Mertens et al., 2007). **The Contrast** evaluates differences in luminance or color that makes an object distinguishable. It is calculated by applying the Laplace convolution kernel to the grayscale version of each input image. **The Saturation** is determined by a combination of light intensity and how much it is distributed across the spectrum of different wavelengths. The saturation measure is defined as the saturation deviation within the R, G and B channels of each pixel. **The Exposedness** is defined as how light or dark an image will appear, revealing how well a pixel is exposed. A Gauss curve function is applied: $exp(-\frac{(i-0.5)^2}{2\sigma^2})$ (we set $\sigma = 0.2$ in our implementation), which evaluates an intensity based on how close it is to 0.5. It can overcome shortcomings of under-exposed (intensity is near to 0) and over-exposed (intensity is near to 1).

Mertens et al. combine the three measures equally to form the weight maps, and then merge the images according to the weight maps (Mertens et al., 2007). Our method utilizes these weight maps as the probability for selecting image regions. $E_1(x_i)$ is defined as:

$$E_1(x_i = label) = \frac{1}{W_{label}(i) + eps} \tag{2}$$

where $label$ corresponds to the image labels; eps is set to 0.001 to avoid $W_{label}(i) = 0$; W is the combined weight maps which are normalized between $(0, 1)$:

$$W_k = (\lambda_1 C_k) \cdot (\lambda_2 S_k) \cdot (\lambda_3 E_k) \tag{3}$$

where k represents the kth image; C_k, S_k, and E_k refer to the weights of contrast, saturation, and exposedness, respectively; "$\cdot$" is the Hadamard product; λ_1, λ_2 and λ_3 are three alterable parameters that modulate the influence of the weights. Generally, λ_1, λ_2, and λ_3 are all set to 1. Eq. (2) aims to select the largest weight value of pixel i between weight maps.

3.1.2. Dynamic exclusion

The dynamic areas need to be located so as to be excluded. To achieve this, we refer to the approach in Zhang et al. (2015). It applies an energy optimization to detect dynamic objects. After the detection, the dynamic pixels of each input are identified by a mask, which is then fed into $E_2(x_i)$ in Eq. (1). The energy function reflects the probabilities of pixels being static or dynamic.

In order to detect dynamic areas, one reference image is selected and the rest of the images are compared with the reference. The exposure of the reference image is mapped to the exposures of different inputs by intensity mapping function (IMF) (Grossberg and Nayar, 2003). IMF is capable of mapping between intensity values of any two exposures and is robust to scene and camera motions. Fig. 4(a) shows the input image sequence where the second image is defined as the reference. The remapped reference images are displayed in Fig. 4(b). For each input and its corresponding exposure-adjusted reference image pair, the dynamic mask is calculated by comparing the differences between the two images using an energy optimization. There are two terms: a data term that compares the intensity differences and

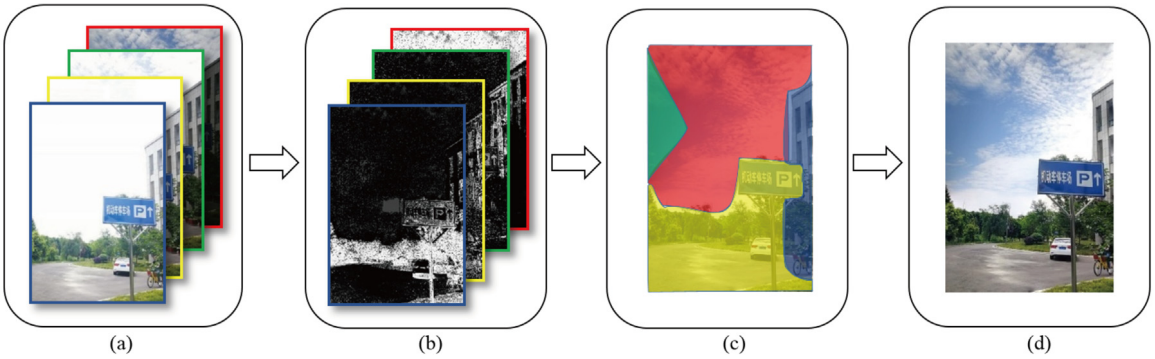

Fig. 3. The pipeline of our method. (a) Aligned input images. (b) Weights maps. (c) Final labels obtained by weight maps, with different color representing labels of different input images. (d) The fusion result.

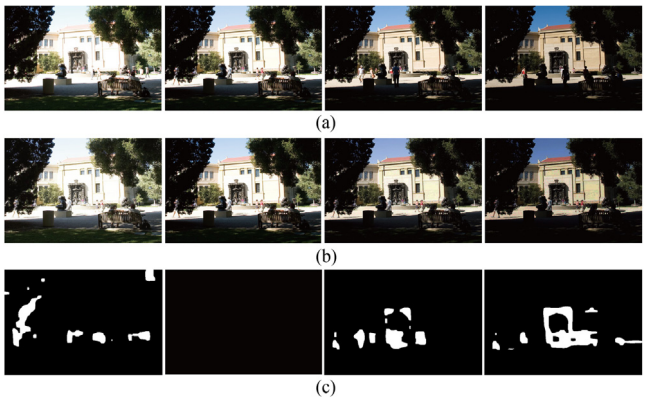

Fig. 4. Moving objects detection results. (a) Input image sequence by courtesy of Gallo et al. (2009). (b) Adjusted reference image sequence. (c) Detection results with white regions denoting moving areas. The second image is selected as reference image, so the values of second mask are all 0.

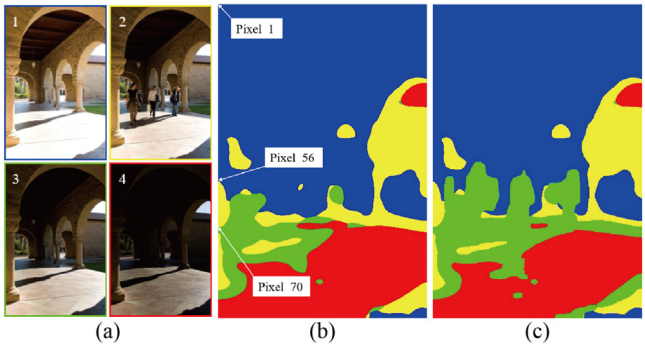

Fig. 5. (a) Aligned input images by courtesy of Gallo et al. (2009) where the third image is selected as the reference image. (b) Labels without dynamic term. (c) Final labels with dynamic term. Notably, if we want to keep the persons, the second image is selected as the reference.

a smoothness term that enforces smooth transitions of neighboring pixels. Two labels are optimized, which yields binary masks M, with 0 indicating the static pixels and 1 indicating the dynamic pixels (Fig. 4(c)).

$$M(i) = \begin{cases} 0 & i \in static\ areas \\ 1 & i \in dynamic\ areas \end{cases} \tag{4}$$

Skipping moving objects is necessary after obtaining the masks of dynamic objects. To this end, $E_2(x_i)$ is defined as:

$$E_2(x_i = label) = \begin{cases} \infty, & M(i) = 1 \\ 0, & M(i) = 0 \end{cases}, \tag{5}$$

When a pixel is static, $E_2(x_i)$ does not introduce any penalties. Otherwise, if a pixel i of one input is detected as dynamic pixels, we set $E_2 = \infty$ to exclude dynamic objects completely. Final detection results are shown in Fig. 4(c). It is obvious that the detection results are probably correspond to dynamic objects.

3.1.3. Spatial smoothness

$E_3(x_i, x_j)$ is the smoothness term, which is a function of the color gradient between two nodes i and j. Similar to Li et al. (2004), E_3 is defined as follows:

$$E_3(x_i, x_j) = |x_i - x_j| \cdot g(C_{ij}), \tag{6}$$

where $g(C_{ij}) = \frac{1}{1+C_{ij}}$ and $C_{ij} = \|C(i) - C(j)\|^2$. $C(i)$ represents color information:

$$C(i) = sqrt([R(i)]^2 + [B(i)]^2 + [G(i)]^2) \tag{7}$$

where R, G and B are three channels of input image. Clearly, C_{ij} is the square of $L2$-norm of the RGB color difference between pixel i and j. Therefore, when two pixels have large differences, $g(C_{ij})$ is near to 0. In a word, $E_3(x_i, x_j)$ is a penalty term when adjacent terms are assigned with different labels.

Fig. 5 demonstrates the results of our labeling. Fig. 5(a) displays four input images where the third image is selected as the reference. Fig. 5(b) and (c) show the result labels of without and with the dynamic detection. Fig. 5(b) is obtained by removing E_2 from Eq. (1), whose labels are purely based on the quality of exposures. In Fig. 5(c), the persons in the second image can be excluded if dynamic detection is enabled.

3.2. Constraints

The image Laplace values are collected according to the labels to form a Laplace image. The next step is to move back to the image domain by solving the Poisson equation, which leads to the solving of a linear sparse system: $Ax = b$, where A is a sparse matrix consisting of 0, −1, and 4, while b consists of boundary constraints and Laplace values.

3.2.1. Boundary constraints

The absolute scale is important when solving the Poisson equation, which controls the overall brightness of the fused result. Inappropriate processing methods would lead to some darker or brighter regions than natural scenes with undesirable detail lost. The proposed method solves the potential problem by adding the boundary constraints. Specifically, when we recover the color information, the proposed method selects the pixels at the image boundary according to the values of the corresponding labels. For example, in left border of Fig. 5(b), the boundary

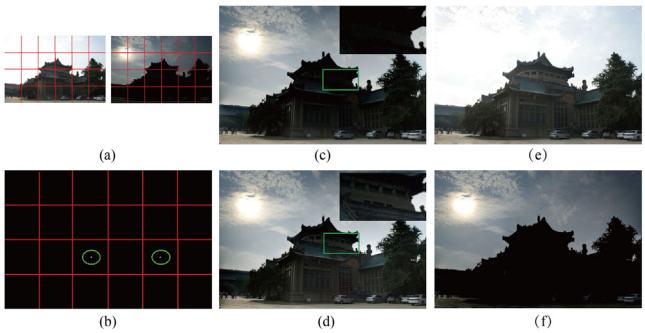

Fig. 6. The effect of constraints. (a) Input images by courtesy of Hu et al. (2013). (b) The location of internal constraints. (c) Without internal constraints. (d) With internal constraints. (e) and (f) are results generated by inappropriate boundary constraints. Please zoom in for a clearer observation.

constraint information of pixel 1 to pixel 56 origin from Image-1 and the constraints of pixel 57 to pixel 70 origin from Image-2.

However, the method does not exactly obey the labels, but employ a vote strategy. If one label dominates (over 70%), the entire boundary constraints will come from that label. In Fig. 5(b), the constraints at the top boundary come from Image-1 while at the left and right boundary, the constraints obey the labels. At the bottom, the red labels account for a large proportion, therefore, all constraints of bottom boundary come from the dominated forth image (which is dominating).

Fig. 6 displays an example, among which Fig. 6(a) shows two input images. Fig. 6(e) and (f) exhibit the over and under exposed fusion results by using the boundary constraints only from the first and the second input image, respectively. It is clear that boundary constraints are essential to determine a proper brightness.

3.2.2. Internal constraints

Apart from the boundary constraints, the method further adds some internal constraints, if necessary, to lighten some under-exposed regions. The internal constraints are realized by solving the Poisson equation twice, which needs to detect the dark regions firstly. If the preliminary fusion result exists dark regions, some internal constraints are placed in these regions and solve the Poisson equation again for the final result. Specifically, the original images (Fig. 6(a)) and preliminary result (Fig. 6(c)) are divided into some regular grids. For each grid, we sum the intensities of all pixels and calculate the difference V_{differ} between the input and preliminary fusion result:

$$V_{\text{differ}} = \sum_{i=1}^{h}\sum_{j=1}^{w} I_{ij} - \sum_{i=1}^{h}\sum_{j=1}^{w} R_{ij} \tag{8}$$

where h and w are the height and width of the grid; I_{ij} represents the input value at pixel (i, j); R_{ij} is the preliminary result value at pixel (i, j).

We select the desired input I according to labels in the grid. If the grid contains only one label, the corresponding input is chosen. Otherwise, if a grid cell contains multiple labels, we calculate the numbers of each label and choose the dominated label. Then, the difference is compared with a threshold T_{cons} to determine whether the grid needs an internal constraint:

$$\begin{cases} V_{\text{differ}} > T_{cons} & \text{yes} \\ V_{\text{differ}} < T_{cons} & \text{no} \end{cases} \tag{9}$$

If the difference V_{differ} exceeds the empirical threshold, a constraint point is added to the middle position of that grid. After the detection, internal constraint values are attached to b and matrix A is adjusted accordingly, then Poisson equation is solved on the second time. Fig. 6(c) and (d) demonstrate the results of without and with internal constraints.

Algorithm 1 Multi-exposure photomontage method

Require: Source image sequence $\{S_k\} = \{S_k | 1 \le k \le K\}$
1: Align the inputs and select the reference S_r
2: Generate $K-1$ latent image $\{L_k\} = \{L_k | 1 \le k \le K, k \ne r\}$ using IMF
3: Obtain Laplace label map by Eq. (1)
4: Reconstruct preliminary result by solving Poisson equation
5: **for** each input image S_k **do**
6: Calculate the difference V_{differ} between input and result of each grid
7: **end for**
8: Identify which grids need constraints according to T_{cons}
9: Solve Poisson equation again by adding internal constraints
Ensure: Reconstructed result $\hat{S}$

3.3. Implementation details

Algorithm 1 summarizes our approach. The proposed method has five parameters, including (1) two balance parameters: λ' and λ'', (2) the height h and the width w of each grid, and (3) the threshold T_{cons}. Here, λ' influences the accuracy of detecting dynamic regions and λ'' determines the continuity of Laplace label map. We set $\lambda' = 3$ and $\lambda'' = 5$ in our implementation. Smaller λ' cannot detect dynamic objects absolutely, which may lead to blurring or ghosting, whereas larger λ' brings large seams to Poisson blending. Inappropriate λ'' causes problems to Laplace label map and further affects final fusion result because unsuitable λ'' cannot balance the data term and smoothness term. When dividing inputs and preliminary result into regular grids, we set the height h and the width w of each grid to 100. Empirically we have found that a value of 100 works well for many types of inputs. Smaller values would increase computational complexity, while larger values would not be good in lightening dark regions. Then, for a grid with size 100×100 and pixel values between 0 to 255, we set $T_{cons} = 8000$. It could be considered as 80×100, among which 80 means the average pixel difference between the input and preliminary fusion result and 100 means the number of darken pixels in preliminary fusion result. Generally, the threshold ranges from 7000 to 9000.

4. Experiments

We assemble a comprehensive dataset of 135 groups of multi-exposure image sequences from previous publications, Internet and our own captures, ranging from daytime–nighttime, static–dynamic, and outdoor–indoor. Based on the dataset, we conduct comprehensive experiments to verify the performance of our method, including objective assessment, visual comparison, complexity comparison and subjective evaluation. Several methods that are just suitable for static inputs (Li et al., 2013; Li and Kang, 2012; Paul et al., 2016; Mertens et al., 2007; Ma et al., 2019) and several de-ghosting methods (Sen et al., 2012; Hu et al., 2013; Kalantari, 2017; Wu et al., 2018) are selected to be compared with the proposed method. In order to make a fair comparison, we collect the codes from the authors of the above algorithms to generate their results with default settings.

4.1. Objective assessment

The main goal of image fusion is to integrate complementary information from multiple sources so that the fused images are more suitable for the purpose of human visual perception and computer processing. Three popular image fusion metrics: Q_{MI} (Hossny et al., 2008), Q_{NCIE} (Wang et al., 2008) and HDR-VDP (Mantiuk et al., 2011) are adopted to evaluate the performance objectively, which could estimate how much information is obtained from the input images.

Table 1
Information about input image sequences.

Source sequences	Size	Origin
TestChart1	1500 × 1000 × 5	EmpaMT[a]
Revel	1500 × 1000 × 5	EmpaMT
Knossos6	1500 × 1000 × 5	EmpaMT
Room	640 × 247 × 5	Banterle et al. (2011)
Lake1	1500 × 1000 × 5	EmpaMT
Forth4	1500 × 1000 × 5	EmpaMT
Garage	348 × 222 × 5	Li et al. (2013)
Lamp	512 × 384 × 5	Martin Ĉadík[b]
Drink	800 × 600 × 3	Our capture
Market	800 × 600 × 5	Our capture
HsLake	1500 × 1000 × 5	EmpaMT
Street	1024 × 682 × 5	Tursun et al. (2016)
Museum2	1024 × 682 × 5	Tursun et al. (2016)
Puppets	1024 × 812 × 5	Gallo et al. (2009)
Plants	1024 × 682 × 5	Tursun et al. (2016)
Toy	1024 × 682 × 5	Tursun et al. (2016)
Agia	1500 × 1000 × 5	EmpaMT
Forrest	1024 × 683 × 4	Gallo et al. (2009)
Book	800 × 600 × 3	Our capture
Cafe	800 × 600 × 5	Our capture
Train19	1500 × 1000 × 3	Kalantari (2017)
Train5	×1000 × 3	Kalantari (2017)
Colo	968 × 648 × 3	Hu et al. (2013)
Dome	968 × 648 × 3	Hu et al. (2013)
Duke	968 × 648 × 3	Hu et al. (2013)
Garden	968 × 648 × 3	Hu et al. (2013)
Happy	968 × 648 × 3	Hu et al. (2013)
Lady	968 × 648 × 3	Hu et al. (2013)
Lift	1500 × 1000 × 3	Our capture
Show	1500 × 1000 × 3	Our capture

[a] http://empamedia.ethz.ch/hdrdatabase/index.php.
[b] http://cadik.posvete.cz/tmo.

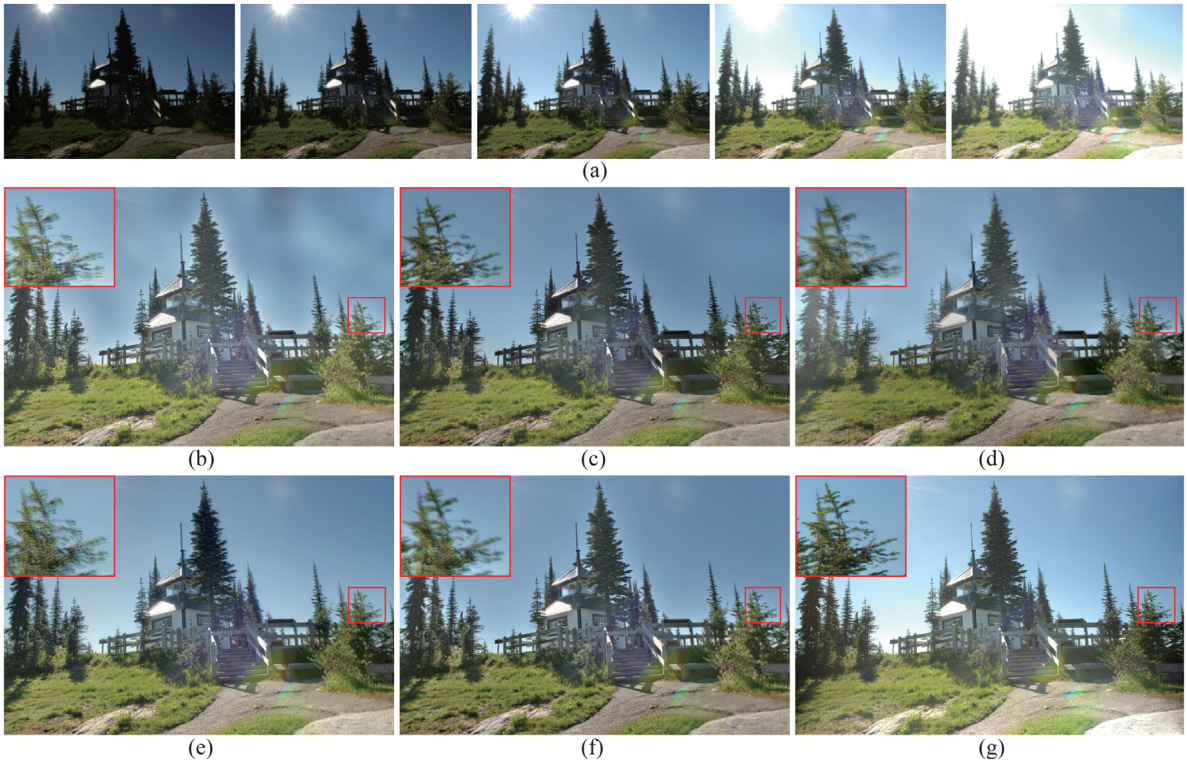

Fig. 7. Comparisons of the proposed method with several MEF methods. (a) Source image sequence by courtesy of EmpaMT dataset. (b) Result of Li et al. (2013). (c) Result of Li and Kang (2012). (d) Result of Paul et al. (2016). (e) Result of Mertens et al. (2007). (f) Result of Ma et al. (2019). (g) Our result.

For K inputs, the metric Q_{MI} is defined as:

$$Q_{MI} = \frac{1}{K}\sum_{i=1}^{K}\frac{MI(I_i, F)}{H(I_i) + H(F)} \tag{10}$$

where I_i $(i = 1, \ldots, K)$ are inputs, F is fusion result, H represents the marginal entropy of an image, MI is mutual information between

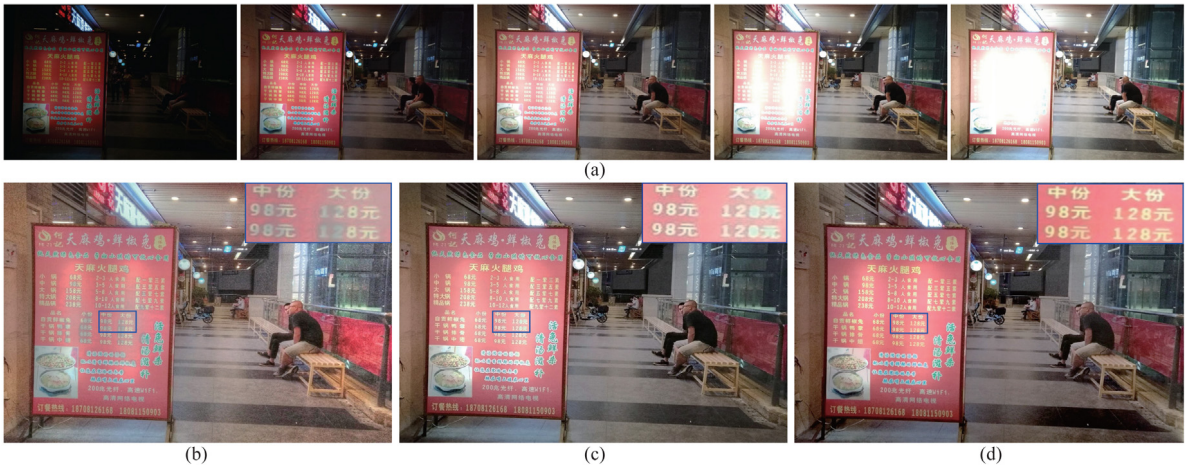

Fig. 8. Comparisons of the proposed method with Sen et al. (2012) and Hu et al. (2013). (a) Input image sequence. (b) Result of Sen et al. (2012). (c) Result of Hu et al. (2013). (d) Our result.

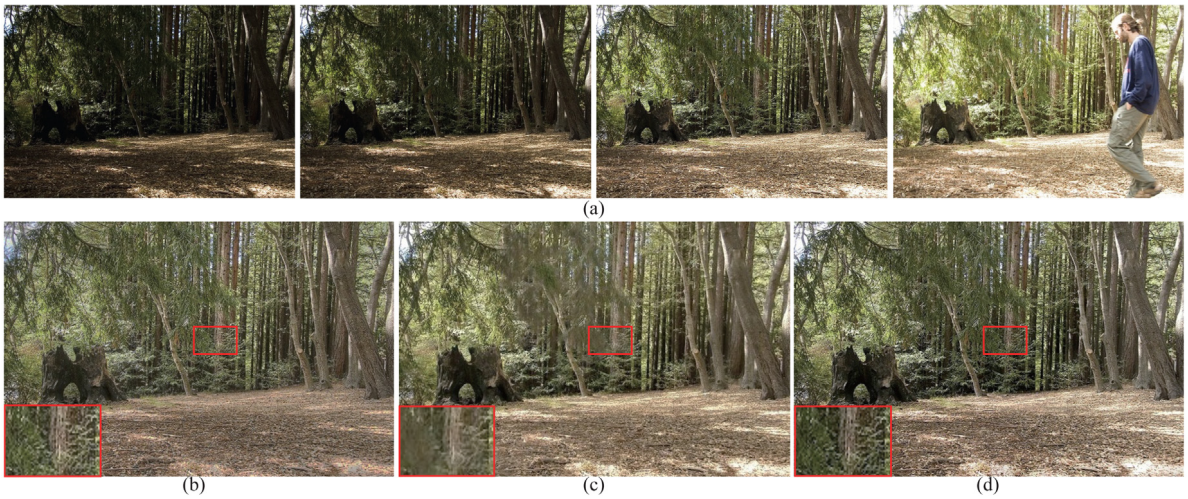

Fig. 9. Comparisons of the proposed method with Sen et al. (2012) and Hu et al. (2013). (a) Input image sequence by courtesy of Gallo et al. (2009). (b) Result of Sen et al. (2012). (c) Result of Hu et al. (2013). (d) Our result.

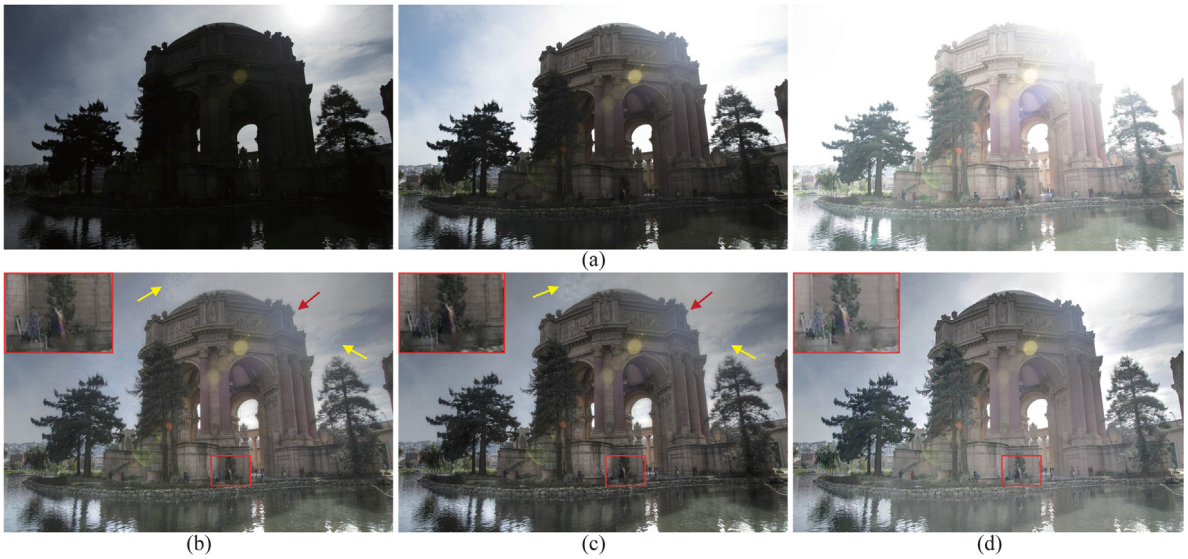

Fig. 10. Comparisons of the proposed method with Kalantari (2017) and Wu et al. (2018). (a) Input image sequence by courtesy of Hu et al. (2013). (b) Result of Kalantari (2017). (c) Result of Wu et al. (2018). (d) Our result. Please zoom in for a clearer observation.

Table 2
Quantitative comparisons of the proposed method with some representative image fusion methods. Q_{MI} scores range from 0 to 1 with higher values indicating better quality. Q_{NCIE} and $HDR-VDP$ range from 0 to 100 with higher values indicating better performance.

		TestChart1	Revel	Knossos6	Room	Lake1	Forth4	Garage	Lamp	Drink	Market	Mean
Li et al. (2013)	Q_{MI}	0.605	0.417	0.379	0.221	0.392	0.438	0.503	0.637	0.482	0.314	0.4388
	Q_{NCIE}	72.231	66.183	67.774	66.119	65.936	69.022	66.158	68.825	63.911	62.406	66.8565
	HDR-VDP	24.227	30.818	30.267	30.099	23.227	27.728	27.936	29.994	31.487	26.398	28.218
Li and Kang (2012)	Q_{MI}	0.577	0.527	0.482	0.330	0.450	0.539	0.611	0.648	0.489	0.220	0.4873
	Q_{NCIE}	72.064	66.519	68.034	66.238	60.097	69.344	66.526	68.905	64.007	62.263	66.3997
	HDR-VDP	24.728	32.173	32.091	30.871	24.209	29.793	28.012	29.673	30.194	27.762	28.951
Paul et al. (2016)	Q_{MI}	0.695	0.518	0.492	0.624	0.629	0.635	0.542	0.492	0.311	0.280	0.5350
	Q_{NCIE}	72.584	66.631	68.119	67.028	66.822	**69.732**	66.314	62.363	63.743	62.299	66.5638
	HDR-VDP	27.564	**33.702**	**35.446**	33.200	26.736	29.662	29.623	29.085	34.273	26.139	30.543
Mertens et al. (2007)	Q_{MI}	0.600	0.514	**0.573**	0.463	0.504	**0.644**	0.607	**0.708**	**0.549**	0.489	0.5651
	Q_{NCIE}	72.224	66.612	**68.413**	66.628	66.259	69.616	66.594	69.105	64.502	62.900	67.2873
	HDR-VDP	25.904	31.266	31.788	31.540	23.481	29.824	28.311	30.032	35.293	**29.116**	29.655
Ma et al. (2019)	Q_{MI}	0.699	**0.584**	0.552	0.508	0.682	0.624	0.609	0.660	0.535	0.571	0.6024
	Q_{NCIE}	72.594	**66.761**	68.165	66.524	66.728	69.511	66.431	**70.024**	64.127	63.010	67.3875
	HDR-VDP	25.140	31.608	30.446	32.236	25.16	28.898	29.044	30.004	25.396	28.974	29.6879
Ours	Q_{MI}	**0.724**	0.575	0.510	**0.655**	**0.730**	0.597	**0.633**	0.636	0.491	**0.563**	**0.6114**
	Q_{NCIE}	**72.635**	66.743	68.096	**67.038**	**67.231**	69.435	**66.742**	68.834	**64.511**	**63.527**	**67.4792**
	HDR-VDP	**30.565**	32.221	32.098	**35.315**	**27.191**	**29.844**	**29.624**	**31.112**	**36.575**	29.115	**31.366**
		HsLake	Street	Museum2	Puppets	Plants	Toy	Agia	Forrest	Book	Cafe	Mean
Sen et al. (2012)	Q_{MI}	0.430	0.416	0.551	0.525	0.340	0.421	0.233	0.376	0.212	0.389	0.3893
	Q_{NCIE}	66.121	63.796	66.178	67.524	62.442	64.716	62.120	67.033	60.882	61.151	64.1963
	HDR-VDP	27.104	23.920	**24.887**	25.637	19.061	25.644	**22.577**	30.086	24.882	27.89	25.169
Hu et al. (2013)	Q_{MI}	**0.595**	0.462	0.646	0.678	0.408	**0.520**	0.200	**0.419**	0.240	0.460	0.4628
	Q_{NCIE}	66.715	64.008	66.437	67.933	62.483	**65.038**	62.100	**67.526**	61.225	61.844	64.5309
	HDR-VDP	29.841	23.873	23.514	25.530	19.509	**27.072**	21.589	30.052	24.67	26.643	25.227
Ours	Q_{MI}	0.592	**0.484**	**0.687**	**0.745**	**0.578**	0.503	**0.323**	0.379	**0.392**	**0.629**	**0.5312**
	Q_{NCIE}	**66.732**	**64.234**	**66.529**	**68.014**	**63.128**	64.909	**62.300**	67.458	**62.917**	**66.256**	**65.2477**
	HDR-VDP	**29.974**	**23.971**	23.116	**25.649**	**20.116**	26.983	21.777	**30.101**	**25.01**	**28.125**	**25.482**
		Train19	Train5	Colo	Dome	Duke	Garden	Happy	Lady	Lift	Show	Mean
Kalantari (2017)	Q_{MI}	0.622	0.495	0.588	0.786	0.678	0.262	0.351	0.704	0.388	0.494	0.5368
	Q_{NCIE}	76.230	75.534	75.418	76.334	75.653	75.113	75.208	75.970	70.524	69.437	74.5421
	HDR-VDP	27.199	24.574	25.445	26.979	25.733	22.953	25.514	26.461	26.445	25.574	25.6879
Wu et al. (2018)	Q_{MI}	**0.805**	0.614	0.790	**0.798**	**0.812**	0.505	0.455	0.812	0.501	0.542	0.6634
	Q_{NCIE}	76.374	75.653	75.652	**76.512**	75.701	**75.330**	75.290	76.063	**71.583**	68.123	74.6281
	HDR-VDP	30.779	27.212	**30.458**	25.634	27.967	**27.432**	23.775	24.472	26.811	26.547	27.1087
Ours	Q_{MI}	0.774	**0.751**	**0.800**	0.778	**0.828**	**0.510**	**0.578**	0.743	**0.535**	**0.575**	**0.6872**
	Q_{NCIE}	**76.472**	**75.751**	**75.664**	76.355	**75.924**	75.318	**75.660**	**76.180**	70.568	**70.609**	**74.8501**
	HDR-VDP	**31.547**	**27.724**	27.137	**27.707**	**29.644**	25.602	**27.019**	**28.716**	**26.900**	**27.537**	**27.9533**

two images. The quality metric Q_{MI} measures how well the original information from source images is preserved in the fused image.

The nonlinear correlation information entropy Q_{NCIE} (Wang et al., 2008; Liu et al., 2011), used as a nonlinear correlation measure of the concerned variables, is defined as:

$$Q_{NCIE} = 100 * (1 + \sum_{i=1}^{K} \frac{\lambda_i{}^R}{K} \log_b \frac{\lambda_i{}^R}{K}) \tag{11}$$

where b is determined by the intensity level, i.e., $b = 256$; R is the nonlinear correlation matrix of the concerned K variables; λ_i $(i = 1, \ldots, K)$, are the eigenvalues of R. NCIE owns strong suitability as a measure for the nonlinear type of correlation of multiple variables.

HDR-VDR computes visual difference based on human perception rather than mathematical differences (Mantiuk et al., 2011). However, HDR-VDP compares a pair of images and predicts the visibility and quality. To apply it for multi-exposure inputs, we first obtain the value between each input and fusion result $(hdr - vdp)_i$, and then get their average result, as described in Eq. (12):

$$HDR-VDP = \frac{1}{K} \sum_{i=1}^{K} (hdr - vdp)_i \tag{12}$$

Thirty image sequences are collected for the comparison (listed in Table 1). These sequences are divided for different categories because the compared methods are suitable for different inputs. The first 10 sequences are either captured by cameras mounted on a tripod or aligned before the fusion calculation. Our method is compared with several MEF methods that are suitable for static inputs (Li et al., 2013; Li and Kang, 2012; Paul et al., 2016; Mertens et al., 2007; Ma et al., 2019) on these 10 scenes and the comparison results are shown in Table 2 (from the second row to seventh row). As can be seen, the proposed method provides results with better Q_{MI}, Q_{NCIE} and HDR-VDP values in most of the cases. The middle 10 sequences are dynamic scenes with moving objects or camera motions. Two state-of-the-art de-ghosting methods (Sen et al., 2012; Hu et al., 2013) are compared on these sequences. They reconstruct the underlying image stacks by patch match oriented optimization. Table 2 (from eighth row to tenth row) shows that although these three methods get similar Q_{NCIE} scores in some cases, the proposed method owns superior performance overall. The last 10 sequences are applied for the comparison with the deep learning image fusion methods (Kalantari, 2017; Wu et al., 2018). They are separated from typical de-ghosting methods because they are just suitable for inputs with fixed images (three images). Deep learning methods have the advantage that they can exploit information extracted from training data to identify and compensate for image regions that do not meet the assumptions underlying the HDR process. However, they lack flexibility and robustness. Deep learning methods may also encounter uncertain problems caused by feature extraction and deficiency of respective field. Therefore, they may not get high metric values when computing the degree of information preservation from inputs to fusion results.

Note that some methods (Sen et al., 2012; Kalantari, 2017; Wu et al., 2018) generate HDR outputs and their final comparison results are obtained by tonemapping using Photomatix.[1]

4.2. Visual comparison

In this part, our method is first compared with Li et al. (2013), Li and Kang (2012), Paul et al. (2016), Mertens et al. (2007) and Ma et al. (2019). These five methods take the inputs without camera motion. We feed the inputs that are pre-aligned or captured by static cameras to their algorithms and generate the results for the comparison (Fig. 7). Some slight movements of the right tree in Fig. 7 lead to blur artifacts of other MEF methods. Although some compared methods have their own strategies to tackle with dynamic objects, such as median and recursive filters of Li and Kang (2012), they still produce unsatisfactory blur. Li et al. treated RGB channels separately, making it difficult to make proper use of color information (Li et al., 2013). As a result, it produces results with unnatural color (Fig. 7(b)). Paul et al. reconstructed image using Haar wavelet (Paul et al., 2016) which is easy to blur structural information (Fig. 7(d)). Mertens' method (Mertens et al., 2007) cannot avoid blur artifacts for its weight combination strategy (Fig. 7(e)). Ma et al. applied a context aggregation network to generate weight maps (Ma et al., 2019), which has some improvement but leads to similar artifacts with Mertens' method. The proposed method performs well with respect to the slight movements, which does not involve synthesis process and is likely to select areas continuously.

Figs. 8 and 9 display the comparisons with two de-ghosting methods: (Sen et al., 2012; Hu et al., 2013). The two patch-based approaches aim to reconstruct ghost regions in the output image by transferring information from inputs which are determined by patch matching. However, they cannot recover the structured regions properly. In Fig. 8, the change of exposure degrades the performance of their results. Especially, the gray artifacts in the 'text' regions of Fig. 8(b) are unexpected. Note that, Hu's method tends to suffer from color drifting in some cases, which occurs in computing generic intensity mapping function and then causes the radiance consistency measure to be ineffective. In Fig. 9, Hu's approach blur the tree leaves (middle part in (Fig. 9(b)). Moreover, the trunks in the right region of their result lose texture information severely. Sen's method performs better than Hu's method, however, their result exists slight blur in red arrow region (Fig. 9(b)). The errors in motion estimation are difficult to avoid in the presence of tiny random motions for patch-based method although they generally produce relatively good results. Our method abandons generating results from every input. It selects regions from single image which avoids blur artifacts effectively.

Fig. 10 shows the comparison with two deep learning methods (Kalantari et al., 2013; Wu et al., 2018). Kalantari et al. applied optical flow to align inputs first and then sent them to a conventional neural network to obtain fused results, which may produce artifacts due to two main reasons: misalignment of optical flow and the limitation of merging process. Wu et al. improved Kalantari et al.'s method and embedded the alignment into the network. The two methods adopt similar network architecture and produce unnatural transformations in sky region (yellow arrows). Moreover, they also generate results with undesirable artifacts around the building (red arrows). Our method can properly handle such problems and obtain results with more details and higher sensory experience (red box region).

4.3. Complexity comparison

Computing efficiency is also important for evaluating fusion performance. Methods that are just suitable for static inputs spend less time (less than 10 s) for whole image fusion process because they are not necessary to handle dynamic objects. We conduct a complexity comparison with two de-ghosting methods (Sen et al., 2012; Hu et al., 2013) and two deep learning methods (Kalantari, 2017; Wu et al., 2018), which is exhibited in Table 3. Upon the same inputs with size $1500 \times 1000 \times 3$, all experiments are conducted on a computer with i7 3.4 GHz CPU and 32G RAM. Results show that the proposed method is more computationally efficient. The methods of Sen et al. (2012) and Hu et al. (2013) take more time on patch match oriented optimization. The optical flow alignment in Kalantari et al. (2013) spends approximately $50s$ on average. Our optimized code is a little bit faster than Wu et al. (2018) and significantly faster than other methods.

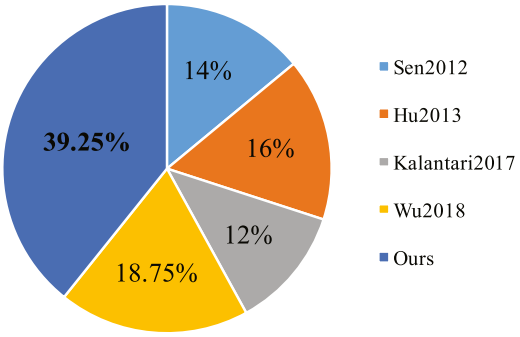

Fig. 11. The numbers are the performance of votes obtained by each method.

4.4. Subjective evaluation

We further conduct a user study (Abebe et al., 2018) to evaluate our method subjectively. We invited 20 viewers (12 male and 8 female subjects aged between 20 and 50) to evaluate the visual quality of different image fusion methods concerning the following two aspects:

- the maintenance of the image details
- the overall perception

The participants are allowed to zoom in and zoom out the images for better observation. To minimize the influence of fatigue effect, the length of a session is limited to a maximum of 30 min. The subjective testing environment was set up in a normal indoor office with an ordinary illumination level. In the evaluation, 20 groups of results are randomly selected from the dataset and every group involves the results of Sen et al. (2012), Hu et al. (2013), Kalantari (2017) and Wu et al. (2018) and our method. We do not subjectively compare our results with the MEF methods whose inputs are fully static because their results are obviously inferior to aforementioned methods' results when dealing with dynamic scenes. During the test, five fused results are shown to the viewers at the same time on a computer screen but in random spatial order. These images are displayed on an LCD monitor at a resolution of 1920×1080 pixels with Truecolor (32 bit) at 60 Hz. For each image set, the viewers are asked to give 5 integer scores that best reflect the perceptual quality of 5 fused image. The scores range from 1 to 10, where 1 denotes the worst quality and 10 is the best. Once all scores are entered for a given scene, the viewers save their decisions and proceeded to the next scene.

We first calculate the number of best scores of 400 groups (20 viewers $\times$ 20 groups of results) and display the results in Fig. 11. It is obvious that our results are favored by a majority of viewers (approximate 40%). Then, Fig. 12 shows the mean scores and corresponding errors of 2000 trials (20 viewers and each viewer saw 100 images). Our method has the highest mean score, while Wu et al. (2018) is the second best on average. Wu et al. (2018) is an optimized method based on Kalantari (2017). Two patch-based de-ghosting methods (Sen et al., 2012; Hu et al., 2013) do not have significant differences, among which the latter one performs a little better than the former one by properly dealing with large saturated regions.

[1] https://www.hdrsoft.com/index.html.

Table 3
Average execution time in seconds on test image sequences.

Methods	Sen et al. (2012)	Hu et al. (2013)	Kalantari (2017)	Wu et al. (2018)	Ours
Time (s)	252 ± 20	181 ± 38	68 ± 5.1	17 ± 3.8	11 ± 4.7

Table 4
One-way ANOVA results comparing the different methods.

Comparisons	F	p-value
Overall	$F(4, 1995) = 23.07$	$p < 0.001$
Ours vs. Sen et al. (2012)	$F(1, 798) = 57.42$	$p < 0.001$
Ours vs. Hu et al. (2013)	$F(1, 798) = 57.50$	$p < 0.001$
Ours vs. Kalantari (2017)	$F(1, 798) = 73.20$	$p < 0.001$
Ours vs. Wu et al. (2018)	$F(1, 798) = 26.80$	$p < 0.001$

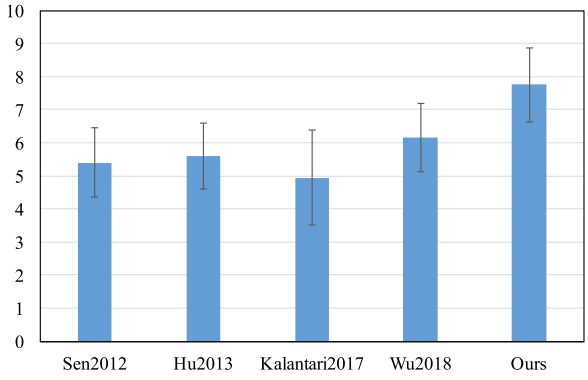

Fig. 12. Mean and standard deviation of subjective evaluation votes over 20 users.

Next, these 2000 scores are analyzed for significance test using one-way ANOVA (ANalysis Of VAriance), as shown in Table 4. Overall and individual comparisons of the proposed method with other four methods indicate that the subjective results are significant, whose p-values are smaller than 0.001. Except for the comparisons in Table 4, we analyze the significance between Sen et al. (2012), Hu et al. (2013), Kalantari (2017) and Wu et al. (2018), among which the p-value between Sen et al. (2012) and Hu et al. (2013) equals to 0.5174 ($F(1, 798) = 0.43$) and other p-values are smaller than 0.05. Sen et al. (2012) and Hu et al. (2013) also have similar mean and standard deviation in Fig. 12.

Last, we compute the coefficient of consistency (Kappa coefficient) between different viewers of one specific method. There are 190 combinations of any two viewers and we randomly select 30 groups to calculate their Kappa values. The total 150 Kappa results (5 methods × 30 groups of scores) range from 0.63 to 1 (average 0.69), which demonstrates that users agree with each other to a significant extent on the performance of any individual method.

5. Conclusion

We have presented a method for accurately fusing multi-exposure images captured by hand-held cameras. In image fusion, high-quality image registration is hard to achieve when scenes have large depth variations and dynamic textures. The proposed method does not require high-quality registration before fusion. It selects well-exposed regions and detects dynamic objects from roughly aligned images using MRF energy minimization. Then, the method finds good seams to hide misalignment when solving Poisson equation. Proper boundary constraints and internal constraints are added for desired brightness. It thus offers the prospect of more extensive applications of image fusion. We conduct comprehensive comparisons with several typical MEF methods to demonstrate its effectiveness.

CRediT authorship contribution statement

Ru Li: Conceptualization, Methodology, Software, Validation, Investigation, Writing - original draft, Visualization. **Shuaicheng Liu:** Formal analysis, Writing - review & editing, Project administration. **Guanghui Liu:** Writing - review & editing, Supervision. **Tiecheng Sun:** Writing - review & editing. **Jishun Guo:** Writing - review & editing.

Acknowledgements

This research was supported in part by National Natural Science Foundation of China (NSFC) under Grants: 61872067 and 61720106004, in part by Department of Science and Technology of Sichuan Province under Grant 2019YFH0016.

References

Abebe, M.A., Booth, A., Kervec, J., Pouli, T., Larabi, M.-C., 2018. Towards an automatic correction of over-exposure in photographs: Application to tone-mapping. Comput. Vis. Image Underst. 168, 3–20.
Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S., Colburn, A., Curless, B., Salesin, D., Cohen, M., 2004. Interactive digital photomontage. ACM Trans. Graph. 23 (3), 294–302.
Banterle, F., Artusi, A., Debattista, K., Chalmers, A., 2011. Advanced High Dynamic Range Imaging: Theory and Practice. AK Peters (CRC Press).
Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23 (11), 1222–1239.
Burt, P.J., 1984. The pyramid as a structure for efficient computation. In: Multiresolution Image Processing and Analysis. pp. 6–35.
Burt, P.J., Adelson, E.H., 1983. The Laplacian pyramid as a compact image code. IEEE Trans. Commun. 31 (4), 532–540.
Burt, P.J., Kolczynski, R.J., 1993. Enhanced image capture through fusion. In: Proc. ICCV. pp. 173–182.
Cui, Z., Wang, O., Tan, P., Wang, J., 2017. Time slice video synthesis by robust video alignment. ACM Trans. Graph. 36 (4), 1–10.
Darmont, A., 2012. High Dynamic Range Imaging: Sensors and Architectures. SPIE Washington.
Debevec, P.E., Malik, J., 1997. Recovering high dynamic range radiance maps from photographs. ACM Trans. Graph. 369–378.
Eden, A., Uyttendaele, M., Szeliski, R., 2006. Seamless image stitching of scenes with large motions and exposure differences. In: Proc. CVPR. pp. 2498–2505.
Fattal, R., Lischinski, D., Werman, M., 2002. Gradient domain high dynamic range compression. ACM Trans. Graph. 21 (3), 249–256.
Gallo, O., Gelfandz, N., Chen, W.-C., Tico, M., Pulli, K., 2009. Artifact-free high dynamic range imaging. In: Proc. ICCP. pp. 1–7.
Granados, M., Kim, K., Tompkin, J., Theobalt, C., 2013. Automatic noise modeling for ghost-free HDR reconstruction. ACM Trans. Graph. 32 (6), 201.
Grossberg, M.D., Nayar, S.K., 2003. Determining the camera response from images: What is knowable? IEEE Trans. Pattern Anal. Mach. Intell. 25 (11), 1455–1467.
Guo, H., Liu, S., He, T., Zhu, S., Zeng, B., Gabbouj, M., 2016. Joint video stitching and stabilization from moving cameras. IEEE Trans. Image Process. 25 (11), 5491–5503.
He, K., Sun, J., Tang, X., 2013. Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. 35 (6), 1397–1409.
Hossny, M., Nahavandi, S., Creighton, D., 2008. Comments on 'Information measure for performance of image fusion'. Electron. Lett. 44 (18), 1066–1067.
Hu, J., Gallo, O., Pulli, K., Sun, X., 2013. HDR deghosting: How to deal with saturation? In: Proc. CVPR. pp. 1163–1170.
Jia, J., Sun, J., Tang, C.-K., Shum, H.-Y., 2006. Drag-and-drop pasting. ACM Trans. Graph. 25 (3), 631–636.
Jinno, T., Okuda, M., 2008. Motion blur free HDR image acquisition using multiple exposures. In: Proc. ICIP. pp. 1304–1307.
Kalantari, N.K., 2017. Deep high dynamic range imaging of dynamic scenes. ACM Trans. Graph. 36 (4), 1–12.
Kalantari, N., Shechtman, E., Barnes, C., Darabi, S., Goldman, D.B., Sen, P., 2013. Patch-based high dynamic range video. ACM Trans. Graph. 32 (6), 202.
Kang, S.B., Uyttendaele, M., Winder, S., Szeliski, R., 2003. High dynamic range video. ACM Trans. Graph. 22 (3), 319–325.
Lee, C., Li, Y., Monga, V., 2014. Ghost-free high dynamic range imaging via rank minimization. IEEE Signal Process. Lett. 21 (9), 1045–1049.
Levin, A., Zomet, A., Peleg, S., Weiss, Y., 2004. Seamless image stitching in the gradient domain. In: Proc. ECCV. pp. 377–389.

Li, S., Kang, X., 2012. Fast multi-exposure image fusion with median filter and recursive filter. IEEE Trans. Consum. Electron. 58 (2), 626–632.
Li, S., Kang, X., Hu, J., 2013. Image fusion with guided filtering. IEEE Trans. Image Process. 22 (7), 2864–2875.
Li, Y., Sun, J., Tang, C.-K., Shum, H.-Y., 2004. Lazy snapping. ACM Trans. Graph. 23 (3), 303–308.
Li, H., Zhang, L., 2018. Multi-exposure fusion with CNN features. In: Proc. ICIP. 1723–1727.
Lin, K., Jiang, N., Liu, S., Cheong, L.F., Do, M., Lu, J., 2017. Direct photometric alignment by mesh deformation. In: Proc. CVPR. pp. 2701–2709.
Liu, Z., Blasch, E., Xue, Z., Zhao, J., Laganiere, R., Wu, W., 2011. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. IEEE Trans. Pattern Anal. Mach. Intell. 34 (1), 94–109.
Liu, S., Xu, B., Deng, C., Zhu, S., Zeng, B., Gabbouj, M., 2016. A hybrid approach for near-range video stabilization. IEEE Trans. Circuit Syst. Video Tech. 27, 1922–1933.
Ma, K., Duanmu, Z., Yeganeh, H., Wang, Z., 2017. Multi-exposure image fusion by optimizing a structural similarity index. IEEE Trans. Comput. Imaging 4 (1), 60–72.
Ma, K., Duanmu, Z., Zhu, H., Fang, Y., Wang, Z., 2019. Deep guided learning for fast multi-exposure image fusion. IEEE Trans. Image Process.
Mann, S., Picard, R.W., 1995. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In: Proceedings of IS and T. pp. 442–448.
Mantiuk, R., Kim, K.J., Rempel, A.G., Heidrich, W., 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans. Graph. 30 (4), 1–14.
Mertens, T., Kautz, J., Van Reeth, F., 2007. Exposure fusion. Comput. Graph. Forum 28 (1), 382–390.
Mitsunaga, T., Nayar, S.K., 1999. Radiometric self calibration. In: Proc. CVPR. pp. 374–380.
Nayar, S.K., Mitsunaga, T., 2000. High dynamic range imaging: Spatially varying pixel exposures. In: Proc. CVPR, Vol. 1. pp. 472–479.
Paul, S., Sevcenco, I.S., Agathoklis, P., 2016. Multi-exposure and multi-focus image fusion in gradient domain. J. Circuits Syst. Comput. 25 (10), 1650123.
Pérez, P., Gangnet, M., Blake, A., 2003. Poisson image editing. ACM Trans. Graphics. 22 (3), 313–318.
Prabhakar, K.R., Srikar, V.S., Babu, R.V., 2017. DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In: Proc. ICCV. pp. 4724–4732.
Rana, A., Valenzise, G., Dufaux, F., 2018. Learning-based tone mapping operator for efficient image matching. IEEE Trans. Multimedia 21 (1), 256–268.
Reinhard, E., Ward, G., Pattanaik, S., Debevec, P., 2010. High Dynamic Range Imaging : Acquisition, Display, and Image-Based Lighting. Princeton University Press.
Rosten, E., Drummond, T., 2006. Machine learning for high-speed corner detection. In: Proc. ECCV. pp. 430–443.
Sen, P., Kalantari, N.K., Yaesoubi, M., Darabi, S., Goldman, D.B., Shechtman, E., 2012. Robust patch-based hdr reconstruction of dynamic scenes.. ACM Trans. Graph. 31 (6), 1–11.
Shen, J., Jin, X., Zhou, C., Wang, C.C., 2007. Gradient based image completion by solving the Poisson equation. Comput. Graph. 31 (1), 119–126.
Shen, J., Zhao, Y., Yan, S., Li, X., et al., 2014. Exposure fusion using boosting Laplacian pyramid. IEEE Trans. Cybern. 44 (9), 1579–1590.
Shi, J., Tomasi, C., 1994. Good features to track. In: Proc. CVPR. pp. 593–600.
Tocci, M.D., Kiser, C., Tocci, N., Sen, P., 2011. A versatile HDR video production system. ACM Trans. Graph. 30 (4), 41.
Tomasi, C., Manduchi, R., 1998. Bilateral filtering for gray and color images. In: Proc. ICCV. pp. 839–846.
Tomaszewska, A., Mantiuk, R., 2007. Image registration for multiexposure high dynamic range image acquisition. In: International Conference on Computer Graphics, Visualization and Computer Vision.
Tursun, O.T., Akyüz, A.O., Erdem, A., Erdem, E., 2016. An objective deghosting quality metric for HDR images. Comput. Graph. Forum 35 (2), 139–152.
Tursun, O.T., Erdem, A., Erdem, E., 2015. The state of the art in HDR deghosting: A survey and evaluation. Comput. Graph. Forum 34 (2), 683–707.
Wang, Q., Chen, W., Wu, X., Li, Z., 2018. Detail preserving multi-scale exposure fusion. In: Proc. ICIP. pp. 1713–1717.
Wang, Q., Shen, Y., Jin, J., 2008. Performance evaluation of image fusion techniques. Image Fusion: Algorithms Appl. 19, 469–492.
Wang, C., Tu, C., 2013. An exposure fusion approach without ghost for dynamic scenes. In: International Congress on Image and Signal Processing. pp. 904–909.
Wu, S., Xu, J., Tai, Y.-W., Tang, C.-K., 2018. Deep high dynamic range imaging with large foreground motions. In: Proc. ECCV. pp. 117–132.
Yan, Q., Gong, D., Shi, Q., van den Hengel, A., Shen, C., Reid, I., Zhang, Y., 2019. Attention-guided network for ghost-free high dynamic range imaging. In: Proc. CVPR. pp. 1751–1760.
Zhang, W., Cham, W.K., 2010. Gradient-directed composition of multi-exposure images. In: Proc. CVPR. pp. 530–536.
Zhang, B., Liu, Q., Ikenaga, T., 2015. Ghost-free high dynamic range imaging via moving objects detection and extension. In: 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA. pp. 459–462.
Zhang, F.-L., Wu, X., Zhang, H.-T., Wang, J., Hu, S.-M., 2016. Robust background identification for dynamic video editing. ACM Trans. Graph. 35 (6), 197.
Zimmer, H., Bruhn, A., Weickert, J., 2011. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. Computer Graph. Forum 30 (2), 405–414.