

Robust Match Fusion Using Optimization

Xiameng Qin, Jianbing Shen, *Senior Member, IEEE*, Xiaoyang Mao,
Xuelong Li, *Fellow, IEEE*, and Yunde Jia

Abstract—In this paper, we present a novel patch-based match and fusion algorithm by taking account of moving scene in a multiple exposure image sequence using optimization. A uniform iterative approach is developed to match and find the corresponding patches in different exposure images, which are then fused in each iteration. Our approach does not need to align the input multiple exposure images before the fusion process. Considering that the pixel values are affected by various exposure time, we design a new patch-based energy function that will be optimized to improve the matching accuracy. An efficient patch-based exposure fusion approach using the random walker algorithm is developed to preserve the moving objects from the input multiple exposure images. To the best of our knowledge, our algorithm is the first patch-based exposure fusion work to preserve the moving objects of dynamic scenes that does not need the registration process of different exposure images. Experimental results of moving scenes demonstrate that our algorithm achieves visually pleasing fusion results without ghosting artifacts, while the results produced by the state-of-the-art exposure fusion and tone mapping algorithms exhibit different levels of ghosting artifacts.

Index Terms—Exposure fusion, moving scenes, optimization, patch-based match, random walker.

I. INTRODUCTION

A NATURE scene often has a high dynamic range (HDR) that exceeds the capability of cameras, which can only capture a small range of the radiance of the real world, and

a single image will often be insufficient to show the details of the full dynamic range in a scene. HDR images can capture a greater dynamic range of luminance than classical low dynamic range (LDR) images for natural scenes [12]. However, the dominant display devices have limited dynamic ranges, they can not directly display the HDR images. Therefore, the HDR images need to be compressed into the dynamic range of the conventional displays while preserving the details and color appearances of the original scene content, which is defined as the tone mapping or tone reproduction operations [1]. Many tone mapping approaches [1], [7], [12] have been proposed to resolve the problem of compressing the HDR images into the LDR images during the last decade.

Specialized image sensors have also been employed to capture HDR contents directly [27]. These hardware are designed in different manners, such as allocating different exposures for different parts of a scene [13], or displaying HDR images directly on HDR displays [9]. Since the dominant displays such as CRT and LCD monitors only have limited dynamic ranges, an HDR image is first reconstructed from a sequence of LDR images and then the tone mapping operator is used to compress the dynamic range for viewing on normal displays [15], [23], [36]. This workflow has some advantages such as no specialized hardware and various tone mapping operations can be performed on HDR images with desired appearance for different display purposes. However, this two-phase workflow is usually not as efficient as the process of directly fusing the multiple exposure images into a single LDR image [16], [21].

The task of exposure fusion is to obtain the full dynamic range of a scene by fusing multiple exposure images into a single high-quality image while preserving the details and color appearances. A variety of exposure fusion methods [16], [21], [33] have been developed in recent years. However, the most challenging problem is how to deal with the situations where there are moving objects in the scene, or the scene is moving. The ghosting artifacts will be produced when we directly fuse the multiple exposure images that contain moving objects or dynamic scenes with camera motion [16]. This ghosting phenomenon will be alleviated when we first align the multiple exposure images containing moving objects and then fuse these aligned images. However, it is still very difficult to accurately align the input images taken under different exposure settings, which will introduce the ghosting artifacts again when the accuracy of registration process is not enough. Although some methods [21], [35] can remove the moving objects using some certain assumptions such as the movement of objects should not be complex or the moving objects should

Manuscript received March 1, 2014; revised June 11, 2014 and August 1, 2014; accepted August 30, 2014. Date of publication September 16, 2014; date of current version July 15, 2015. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2013CB328805, in part by the Key Program of NSFC-Guangdong Union Foundation under Grant U1035004, in part by the National Natural Science Foundation of China under Grant 61272359 and Grant 61125106, in part by the Program for New Century Excellent Talents in University under Grant NCET-11-0789, in part by JSPS KAKENHI under Grant 26240015 and Grant 26560006, in part by the Key Research Program of the Chinese Academy of Sciences under Grant KGZD-EW-T03, in part by the Beijing Higher Education Young Elite Teacher Project, and in part by the Joint Building Program of Beijing Municipal Education Commission. This paper was recommended by Associate Editor J. Wang. (*Corresponding author: J. Shen.*)

X. Qin, J. Shen, and Y. Jia are with the Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China (e-mail: qxm0405@bit.edu.cn; shenjianbing@bit.edu.cn; jiayunde@bit.edu.cn).

X. Mao is with the Interdisciplinary Graduate School of Medical and Engineering, University of Yamanashi, Kofu 400-0016, Japan (e-mail: mao@yamanashi.ac.jp).

X. Li is with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China (e-mail: xuelong_li@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCYB.2014.2355140

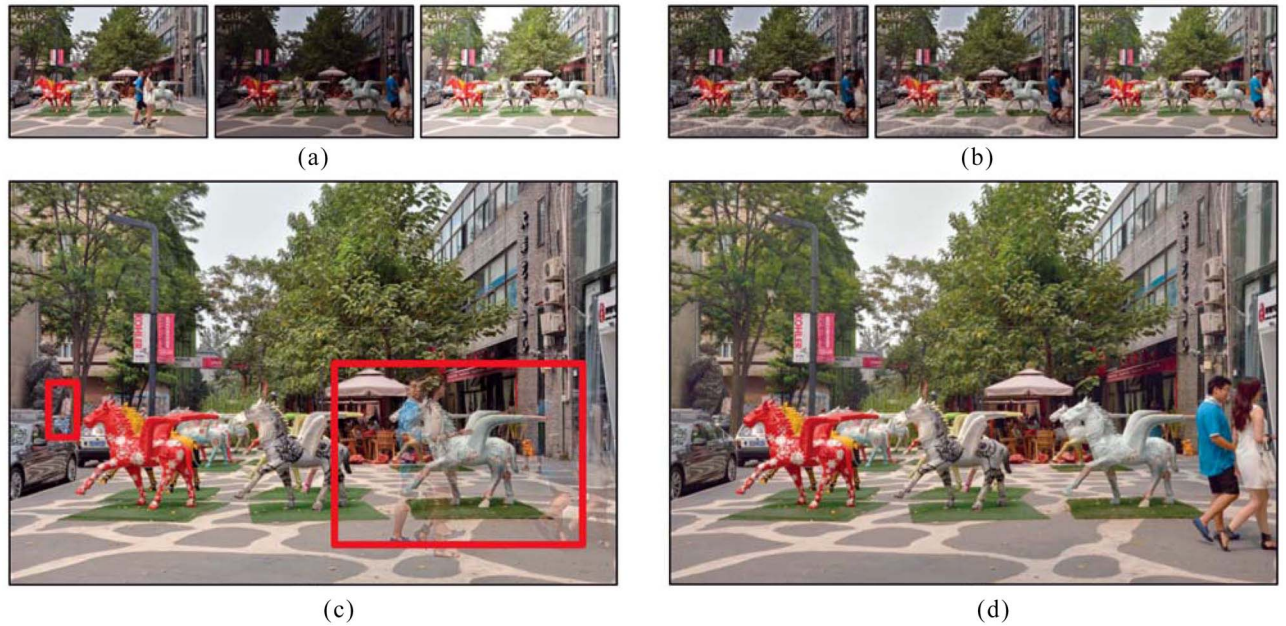


Fig. 1. Illustration of our exposure fusion approach. (a) Input exposure sequence with moving objects. (b) Fusion results with different iterative times. We just give three iteration results for simplicity. (c) Exposure fusion result by Mertens *et al.* [16]. Two rectangle regions in red indicate the ghosting artifacts. (d) Fusing result obtained by our fusion approach. The exposure sequence is captured by ourselves.

not overlap each other, the details of the exposure fusion result are not fine, or those approaches need the user interaction to indicate the moving objects tediously [35].

Based on the aforementioned analysis and motivations, we then propose a novel patch-based exposure fusion algorithm, which can fuse multiple exposure images of moving scenes using optimization. Then a patch-based match approach with a new energy cost function is developed, which considers both the color information and the geometry information such as the scale, translation and rotation information in the energy optimization procedure. A patch-based fusion approach using random walker algorithm is also developed to preserve the color and texture details from the input exposure sequence. Once a patch in the reference image has been matched with a patch in another exposure image, these two patches are then fused directly using our patch-based optimization approach. As shown in Fig. 1, the matching and fusing procedures are performed simultaneous in an iterative way [Fig. 1(b)] for the scene with the moving objects [Fig. 1(a)]. Our approach achieves the better visual quality without any ghosting artifacts [Fig. 1(d)] while the result by the exposure fusion approach in [16] introduces the distinct ghosting artifacts. Our patch-based match and fusion technique does not need the assumption that the multiple exposure images are captured by the camera mounted on a tripod or have been aligned by some registration methods. Our algorithm can handle the exposure images with the moving objects and produce a high quality fusing image without introducing any ghosting artifacts. Thus, the proposed exposure fusion algorithm can be utilized in many image and video processing tasks such as the capture of realistic and HDR images and videos for high quality video data and motion analysis for video surveillance application. Since the matching and fusion procedures are accomplished

in a uniform framework, our algorithm can detect more accurate motions and analyze them for video surveillance application.

In summary, our approach has three main contributions in the context of previous work on exposure fusion. The first contribution is to propose a patch-based match and fusion approach that directly fuses different exposure images without any image registration stage. The second contribution is to develop an iterative patch-based match algorithm by considering both the geometry and color information in a novel optimization framework. And our last contribution is a new iterative method that incorporate the match and fusion procedure so that we can fuse the image sequence using random walker algorithm with both static and dynamic scenes. Our fusion source code will be publicly available online.¹

The rest of this paper is organized as follows. Section II reviews previous work that is related to our method. Section III describes our iterative optimization algorithm in details. And the experimental results are shown and discussed in Section IV. We finally give the summary of the presented algorithm and future work in Section V.

II. RELATED WORK

Our approach is mostly related to the work of exposure fusion techniques, and HDR image reconstruction and tone mapping algorithms. The HDR image reconstruction algorithms [20] first estimate the camera response function, then reconstruct a latent radiance map to generate the final HDR image. Given a sequence of multiple exposure images, Debevec and Malik [2] estimated the camera response function and the relative radiance values using the constraint

¹<https://github.com/shenjianbing/matchfusion14>

of sensor reciprocity. Mitsunaga and Nayar [6] recovered the camera response function using a high-order polynomial, which is based on the ratios of the exposure time. When an HDR image was reconstructed, tone mapping techniques [1], [7], [10], [22] are utilized to compress its dynamic range so that it can be displayed on the dominant LDR displays. Reinhard *et al.* [7] developed a tone reproduction operator to compress the dynamic range of HDR images according to the photographic practice for a wide variety of images. Li *et al.* [10] proposed a tone mapping approach for compressing the HDR images while retaining the important visual information. The subband architecture was employed to decompose the HDR image into multiscale subbands and then applied local gain control strategy to enhance these subbands. These tone mapping techniques worked in a two-phase workflow which first constructs the HDR image and then performs the tone mapping to display the contents on the conventional LDR devices.

The exposure fusion algorithm directly fuses the multiple exposure images into a high-quality compositing image, during which, neither the calibration of camera response function nor the tone mapping algorithm is required. Exposure fusion is originally derived from the more general topic of image fusion approaches [4], [14], [19], which fuse the multisensor and multispectral images of the gray level images. Mertens *et al.* [16] fused the multiple exposure images by constructing the Laplacian and Gaussian pyramids. The fusion weights are computed by three quality measures, say contrast, saturation, and well-exposure. Zaveri and Zaveri [24] extracted the segmented regions from different registered input images, and then fused those regions by utilizing different fusion techniques, e.g., multiresolution-based approach, discrete wavelet transform, and mean max standard deviation. Recently, Song *et al.* [37] proposed an exposure fusion method in which the visible contrasts and scene gradients are calculated in a probabilistic model. This model preserved the visual details that the global tone mapping operators may fail to, and also eliminated the halo artifacts produced by the local tone mapping operators. Shen *et al.* [26] presented a generalized random walk framework to compute the probabilities of the pixels, aiming at solving the exposure fusion problem. They define the weight maps by two quality measures, say local contrast and color consistency, in the fusion process. Han *et al.* [38] employed the saliency detection to generate the saliency maps for objects or regions. The MRF model was utilized to combine the saliency map and the co-occurrence of hot spots and motion. Shen *et al.* [39] proposed an exposure fusion method using a boosting Laplacian pyramid (BLP) and a hybrid exposure weight. Other techniques have also been developed for effective exposure fusion [31], [33], [34], [40]. However, these methods have made the same assumption that the input exposure sequence has been first aligned well. The ghosting artifacts will be introduced when the moving objects or small misalignments existed in the multiple exposure images.

In order to overcome the aforementioned disadvantages and limitations, several methods have been developed to improve the fusion quality. Zhang and Cham [21] employed the gradient information to compose the multiple exposure images.

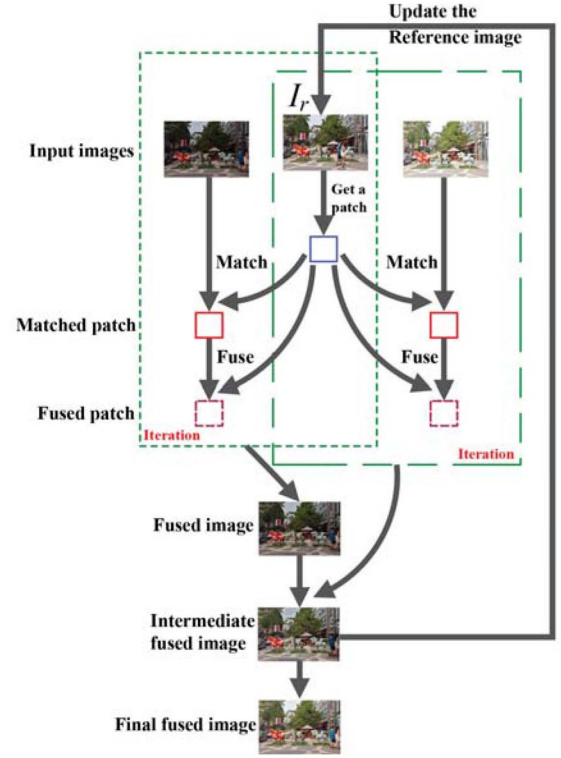


Fig. 2. Workflow of our patch-based match and fusion algorithm.

The magnitude of gradient indicates the exposure quality of pixels, and the direction of gradient was used to remove the moving objects. The weight map was computed by combining the magnitudes and the directions of the gradients. Shen *et al.* [35] presented a detail-preserving exposure fusion technique using the subband architecture and the gain control maps. The moving objects can be indicated by the user interactions so that they can resolve the exposure fusion with the moving objects. However, these methods [21], [26], [35] fail to remove the ghosting artifacts if the multiple exposure images were not aligned by some registration methods or the exposure sequence includes the moving objects. In contrast, our algorithm can automatically find the accurate and proper matches from the exposure sequence of moving scenes to make the fusion results without halo artifacts in an iterative process.

III. OUR APPROACH

In this section, a novel iterative optimization approach that simultaneously performs the matching and fusing procedures is developed to make the patch-based match and fusion algorithm more effective. The flow chart of our approach is shown in Fig. 2. The normal exposure image is first selected as the reference image (I_r) from the input sequence. The number of input exposure images is set as $n = 3$ for the illustration simplicity. Then a pair of corresponding patches are matched between the reference image and another exposure image in the sequence, and these patches are fused and optimized using the random walker algorithm. This match and fusion procedure is iterated until all the patches in the reference image are traversed, which is shown as the green

dashed rectangle in Fig. 2. The fused images are generated by fusing the patches from the reference image and other images. The intermediate fusion image is created from these fused images, and then the reference image is updated by the intermediate fusion image in each iteration. The final fused image is generated after several iterations with the termination condition. We will introduce the detailed descriptions of the proposed patch-based match and fusion algorithm of the moving scenes using optimization step-by-step in the following sections.

A. Matching

We consider both the geometry and color information during the design of the patch-based matching energy. The geometry information is used to establish the relationship between two corresponding patches while the color information is used to design a new similarity measure between these two patches. Let $\{I_1, \dots, I_N\}$ denote the set of N input images, I_r and I_s are chosen as the reference and source images, respectively, in our algorithm where $r, s \in [1, N]$. Given a patch p^r ($p^r \in I_r$), our goal is to find the reliable corresponding patch p^s ($p^s \in I_s$) in the source image. HaCohen *et al.* [29] added a constrained range of geometry information and bias values over the nearest-neighbor searching for the task of image enhancement. However, this searching approach [29] for finding correspondence patches is not reliably enough for different exposure images. In this section, we design a new similarity measure between two patches by considering both the geometry and color information. This similarity measure is optimized in our optimization framework to find more accurate and reliable corresponding patches.

The size of patch is $n \times n$, and (x_i^r, y_j^r) is a point in the patch p^r where $i, j \in [1, n]$. Then the corresponding point (x_i^s, y_j^s) in the patch p^s is computed by

$$\begin{aligned} \begin{pmatrix} x_i^s \\ y_j^s \end{pmatrix} &= \eta \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{pmatrix} x_i^r + x_T \\ y_j^r + y_T \end{pmatrix} \\ &= \eta \mathbf{R}_\theta \begin{bmatrix} x_i^r \\ y_j^r \end{bmatrix} + \mathbf{T} \end{aligned} \quad (1)$$

where \mathbf{R}_θ represents the rotation matrix and $\mathbf{T} = (x_T, y_T)^T$ denotes the translation vector. η is the scale parameter, which is set to be 1.0 in most of our experiments.

The simple way to compute the matching cost is the L2 distance and sum of the squared distance (SSD) [5]. However, these similarity measurements consider the geometric information only and ignore the color information. We utilize the log-chromaticity normalization to design a matching energy function, which is inspired by the adaptive normalized cross-correlation method for stereo matching in [28]. The patch (e.g., p^r, p^s) as 1D vector (e.g., V^r, V^s) is defined as

$$\begin{aligned} V^r &= \{v_1^r, \dots, v_t^r, \dots, v_{n^2}^r\} \\ V^s &= \{v_1^s, \dots, v_t^s, \dots, v_{n^2}^s\} \end{aligned} \quad (2)$$

where $t \in [1, n^2]$.

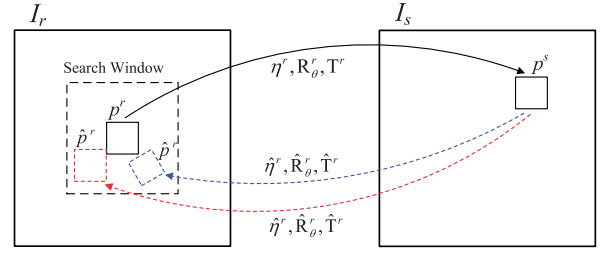


Fig. 3. Illustration of computing the reliability. The checked patch \hat{p}^r might be everywhere in the search window.

Then the similarity measurement of the red channel of the patch is computed by

$$\epsilon_{\text{red}_p} = \frac{\sum_{t \in [1, n^2]} [(v_t^r - \bar{v}^r) \times (v_t^s - \bar{v}^s)]}{\sqrt{\sum_{t \in [1, n^2]} |v_t^r - \bar{v}^r|^2} \sqrt{\sum_{t \in [1, n^2]} |v_t^s - \bar{v}^s|^2}} \quad (3)$$

$$\begin{aligned} v_t^r &= \log \frac{a_r}{\sqrt[3]{a_r b_r c_r}} \\ &+ \gamma_r \log \frac{p_{\text{red}}^r(x_i^r, y_j^r)}{\sqrt[3]{p_{\text{red}}^r(x_i^r, y_j^r) p_{\text{green}}^r(x_i^r, y_j^r) p_{\text{blue}}^r(x_i^r, y_j^r)}} \end{aligned} \quad (4)$$

$$\begin{aligned} v_t^s &= \log \frac{a_s}{\sqrt[3]{a_s b_s c_s}} \\ &+ \gamma_s \log \frac{p_{\text{red}}^s(x_i^s, y_j^s)}{\sqrt[3]{p_{\text{red}}^s(x_i^s, y_j^s) p_{\text{green}}^s(x_i^s, y_j^s) p_{\text{blue}}^s(x_i^s, y_j^s)}} \end{aligned} \quad (5)$$

where \bar{v}^r and \bar{v}^s are the mean value of V^r and V^s , respectively. $p_{\text{red}}^r(x_i^r, y_j^r)$, $p_{\text{green}}^r(x_i^r, y_j^r)$, and $p_{\text{blue}}^r(x_i^r, y_j^r)$ are, respectively, the red, green, and blue channel values of the patch in the reference image. a_r, b_r , and c_r are the global scale factors, which are defined by the illumination color. The gamma value γ depends on the camera parameters, which is set as $\gamma = 2$.

Thus, the final matching energy of the patch can be obtained from (1) and (3) as follows:

$$E_{p-m} = \underset{\{\eta, \mathbf{R}, \mathbf{T}\}}{\text{argmin}} \frac{\epsilon_{\text{red}_p} + \epsilon_{\text{green}_p} + \epsilon_{\text{blue}_p}}{3} \quad (6)$$

where $\epsilon_{\text{green}_p}$ and ϵ_{blue_p} are the similarity measure of the green and blue channels, respectively.

In order to improve the matching accuracy, we present an effective method to compute the reliability of the matched patches. This reliability guides the iteration process of the patch-based match (see Section III-C). As shown in Fig. 3, the patches p^r and p^s are matched using the above patch-based matching procedure. Therefore, the improved correspondence between these patches is computed as follows:

$$p^s = \eta^r \mathbf{R}_\theta^r (p^r + \mathbf{T}^r) \quad (7)$$

$$\hat{p}^r = \hat{\eta}^r \hat{\mathbf{R}}_\theta^r p^s + \hat{\mathbf{T}}^r \quad (8)$$

where $\hat{\eta}^r = 1/\eta^r$, $\hat{\mathbf{R}}_\theta^r = \mathbf{R}_\theta^{r-1}$, and $\hat{\mathbf{T}}^r = -\mathbf{T}^r$ are obtained from (7). η^r , \mathbf{R}_θ^r , and \mathbf{T}^r are the initial values of the left-right consistency check.

Then we use (6) to search the optimal checked patch. As shown in Fig. 3, the accurate checked patch that corresponds with the patch p^s can be any position in the search window, such as the positions of red and blue patches or the position of patch p^r in Fig. 3. According to the above analysis, the matching reliability of patch p^r is defined by the spatial distance between the patches p^r and \hat{p}^r . The value of reliability increases when the spatial distance between p^r and \hat{p}^r decreases. If the positions of p^r and \hat{p}^r are the same, the patches p^r and p^s are the most reliable and accurate matches. Thus, the matching reliability is defined as

$$Rel^r = \exp\left(-\frac{d(\hat{p}^r, p^r)}{\rho}\right) \quad (9)$$

where $d(\hat{p}^r, p^r)$ is the Euclidean distance between \hat{p}^r and p^r , and ρ is a parameter of the maximum search radius.

The aforementioned left-right consistency check strategy is a very common and effective technique in stereo matching [32]. This strategy can detect the mismatch pixels efficiently between the left and right images, which improves the accuracy of matching in stereo vision. These advantages inspire us to employ the left-right consistency check strategy to design our matching reliability algorithm for exposure fusion.

B. Fusion

Once a patch (e.g., p^r) is matched with the patch (e.g., p^s) in the other image, we fuse these two patches with a precomputed weight. In terms of the common exposure fusion model, each pixel in the fused patch p^f is represented as

$$p_i^f = w^r(x_i^r, y_i^r) p^r(x_i^r, y_i^r) + w^s(x_i^s, y_i^s) p^s(x_i^s, y_i^s) \quad (10)$$

where $w^r(x_i^r, y_i^r)$ and $w^s(x_i^s, y_i^s)$ are the fusion weights corresponding to the reference and source patches, and $p^r(x_i^r, y_i^r)$ and $p^s(x_i^s, y_i^s)$ are the intensity values of the pixel (x_i^r, y_i^r) and (x_i^s, y_i^s) , respectively.

The performance of exposure fusion result depends on the computation of the blending weights. We need to design an effective quality assessment algorithm where the well exposure regions should be preserved more while the under-exposure and over-exposure regions should have less weights. Most of exposure fusion algorithms [21], [26], [35], [37] have made an assumption that the input images have been aligned by some registration algorithm like [8]. However, our algorithm does not need the above assumption and each patch is first matched and then fused in an iterative procedure, which is applicable for the application of exposure fusion with hand-held cameras. Thus, we develop a novel patch-based match and exposure fusion approach using random walker algorithm. Unlike the generalized random walk fusion algorithm in [26], which considers the pixel-based fusion only for the static scenes without moving objects, we design an effective approach to fuse the matched patches using the random walker algorithm with dynamic scenes.

Formally, the graph G is defined as $G = \{\mathcal{X}_p \cup \mathcal{L}_p, \varepsilon_p\}$, where $\mathcal{X}_p = \{v_1, \dots, v_N\}$ and $\mathcal{L}_p = \{v_{N+1}, \dots, v_{N+K}\}$ are, respectively, the set of scene nodes and label nodes of a patch, $\varepsilon_p = \varepsilon_{\mathcal{X}_p} \cup \varepsilon_{\mathcal{L}_p}$ is the edge set that contains the edges both within \mathcal{X}_p and between \mathcal{X}_p and \mathcal{L}_p . We first define the Laplacian of the graph by a $(N+K) \times (N+K)$ matrix as follows:

$$L(i, j) = \begin{cases} 1, & i = j \\ -\frac{\beta_1 c_{i,j}}{(d_i d_j)^{\frac{1}{2}}}, & (v_i, v_j) \in \varepsilon_{\mathcal{L}_p} \wedge v_i \in \mathcal{X}_p \\ -\frac{\beta_2 g_{i,j}}{(d_i d_j)^{\frac{1}{2}}}, & (v_i, v_j) \in \varepsilon_{\mathcal{X}_p} \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$$d_i = \sum_{v_j \in \mathcal{N}_i} (\beta_1 c_{i,j} + \beta_2 g_{i,j})^2 = \sum_j^{K+N} L(i, j) \quad (12)$$

where $i, j \in [1, N+K]$, and \mathcal{N}_i is the neighborhood nodes of v_i in the patch. $c_{(\cdot, \cdot)}$ and $g_{(\cdot, \cdot)}$ are the compatibility functions that can be modeled by any similarity measurements, e.g., local contrast and color consistency. β_1 and β_2 are two positive coefficients. Based on the spectral graph [3], the fusion energy of a patch can be given by

$$\begin{aligned} E_{p-f} &= \langle \mathbf{f}, \mathbf{L} \mathbf{f} \rangle \\ &= \sum_{(v_i, v_j) \in \varepsilon_{\mathcal{X}_p}} \frac{\beta_2 g_{i,j}}{(d_i d_j)^{\frac{1}{2}}} (\mathbf{f}(v_i) - \mathbf{f}(v_j))^2 \\ &\quad + \sum_{(v_i, v_j) \in \varepsilon_{\mathcal{L}_p}} \frac{\beta_1 c_{i,j}}{(d_i d_j)^{\frac{1}{2}}} (\mathbf{f}(v_i) - \mathbf{f}(v_j))^2 \end{aligned} \quad (13)$$

where \mathbf{f} is a harmonic function which can be regarded as a column vector in practice. In our algorithm, based on (2), the harmonic function can be written as

$$\left\{ \overbrace{\mathbf{f}(v_1), \dots, \mathbf{f}(v_N)}^{\mathbf{f}_{\mathcal{X}_p}}, \overbrace{\mathbf{f}(v_{N+1}), \dots, \mathbf{f}(v_{N+K})}^{\mathbf{f}_{\mathcal{L}_p}} \right\} = \left\{ \mathbf{f}_{\mathcal{X}_p}^T, \mathbf{f}_{\mathcal{L}_p}^T \right\}. \quad (14)$$

Actually, we must find the optimal $\mathbf{f}_{\mathcal{X}_p}$ to minimize the fusion energy (E_{p-f}) of a patch. Thus, we make the partial derivative of E_{p-f} for $\mathbf{f}_{\mathcal{X}_p}$ to be zero, i.e., $(\partial E_{p-f} / \partial \mathbf{f}_{\mathcal{X}_p}) = 0$. At last, we can obtain the following equation:

$$\mathbf{L}_{\mathcal{X}_p} \mathbf{f}_{\mathcal{X}_p} + \mathbf{Q} \mathbf{f}_{\mathcal{L}_p} = 0 \quad (15)$$

where

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{\mathcal{X}_p(N \times N)} & \mathbf{Q}_{(N \times K)} \\ \mathbf{Q}_{(K \times N)}^T & \mathbf{L}_{\mathcal{L}_p(K \times K)} \end{bmatrix}^3. \quad (16)$$

The harmonic function ($\mathbf{f}_{\mathcal{X}_p}$) can be computed by solving the above (15), and simultaneously, the fusion energy of the

² Without loss of generality, we assume that $v_i \in \mathcal{X}_p$. Thus, if $v_j \in \mathcal{L}_p$, $g_{i,j} = 0$. And if $v_j \in \mathcal{X}_p$, $c_{i,j} = 0$.

³ The subscript $(N \times N)$ indicates the size of the submatrix $\mathbf{L}_{\mathcal{X}_p}$. The other submatrix with subscript can be also interpreted in the same way.

patch will reach to the minimum. After the energy optimization, we obtain the weight of each pixel in the patch, and the fused result of patches is computed from (10). The weight can be defined to be proportional to the harmonic function, and can be written as

$$\mathbf{f}_{\mathcal{X}_p} = \tau \mathbf{W}_{\mathcal{X}_p} \quad (17)$$

where τ is a positive constant. $\mathbf{W}_{\mathcal{X}_p}$ is the weight corresponding to the harmonic function $\mathbf{f}_{\mathcal{X}_p}$, and can be defined as

$$\begin{aligned} \mathbf{W}_{\mathcal{X}_p} &= \{w(v_1), \dots, w(v_N)\} \\ &= \{w(x_1, y_1), \dots, w(x_N, y_N)\}. \end{aligned} \quad (18)$$

Then the fusion weights of the reference patch $w^r(x_i^r, y_i^r)$ and the source patch $w^s(x_i^s, y_i^s)$, which have been defined in (10), can be estimated by (17) and (18). What's more, the patch-based fusion approach is a global optimal method, which also fuses the intermediate image after finishing the matching iterations.

C. Iteration

As shown in Fig. 2, two types of iterative processes are performed to improve the matching and fusing performance. One type is the patch-based matching and fusing operations that are indicated by the green dashed box in Fig. 2, and the other type is the updating of the reference image. During the procedure of the patch-based matching and fusing, the matching reliability is also computed by (9), which plays a key role to improve the efficiency of searching and the accuracy of matching. A correlation between the search range and the reliability is established to guide the random search.

Let $(x_0, y_0) \in p^s$ denote the initial searching position, the new position (x_i, y_i) is given by

$$(x_i, y_i) = (x_0, y_0) + \varphi \alpha^{st} \mathcal{R}_i \quad (19)$$

$$st = \sum_{p^m \in \mathcal{N}(p^r)} C(p^m) \quad (20)$$

$$C(p^m) = \begin{cases} 1, & Rel^m > \xi \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

where φ is the maximum search radius defined in (9), α is a ratio parameter, \mathcal{R}_i is a uniform random range in $[-1, 1] \times [-1, 1]$ and $\mathcal{N}(p^r)$ is the neighboring patches of p^r . ξ is a threshold parameter to remove the unreliable patches, and we set ξ to be 0.86 in most of our experiments.

In each iteration of the patch-based matching and fusing procedure, the reliability of the prematched patches is computed. A patch will be searched and matched quickly in a small range when there are enough reliable matched patches in the neighborhood. Fig. 4 shows the fusion results in different iterations. Our optimization converges fast and most of regions are matched after two iterations. Our iterative optimization process will terminate until the whole reference image is matched and fused. The second type of iterative process in our algorithm is to replace the reference image with the intermediate fused image in each iteration (see Fig. 2). Then this whole

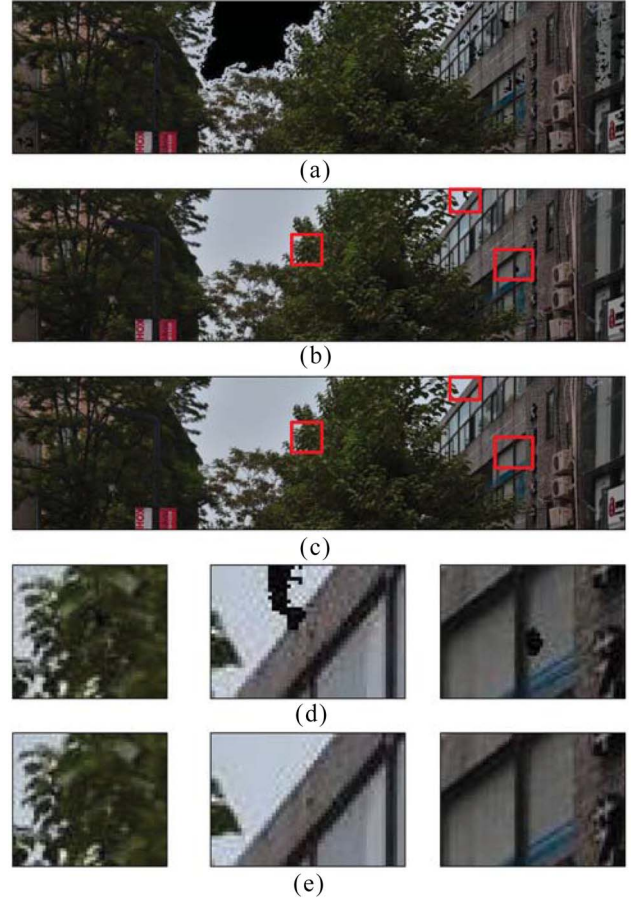


Fig. 4. Matched results in different iterations. (a) Result of the first iteration. (b) Result after two iterations. (c) Matching result of the final iteration. (d) and (e) Closeup comparison of three mismatched regions. The input exposure sequence is listed in Fig. 1.

procedure is iterated until the termination condition is satisfied. In our experiments, the termination condition is set as the number of iterations exceeds the maximum number of iterations. However, we can define the termination condition as a user-defined threshold, which computes the error between the current intermediate fused image and the intermediate fused image by last iteration. The overview of our algorithm is summarized in Algorithm 1.

The close-up comparison of fusion results by different iterations is shown in Fig. 5. Along with more matched regions, the quality of fusion result is becoming better. There are heavy dark speckles in the image after the first iteration [Fig. 5(a)], since the corresponding patches of the patches in these dark speckle regions are not found reliably.

The dark speckle regions are decreased quickly after the second iteration [Fig. 5(b)]. Finally, all the speckles are removed after the final iteration [Fig. 5(c)], which means all the patches are matched and fused well. It is worth to note that the patch-based matching and fusing are not separated in our algorithm, actually. The comparisons in both Figs. 4 and 5 demonstrate the effectiveness and efficiency of the proposed matching and fusing procedure using optimization. When we match and fuse the image sequences that contain moving objects, the occlusion problem may occur in some regions.

Algorithm 1: Patch-Based Match and Fusion Algorithm**Input:** Unregistered input exposure sequence $\{I_1, \dots, I_N\}$ **Output:** The final fused image FF

```

1 for iteration  $k = 1 \rightarrow K$  do
2   for image  $i = 1 \rightarrow N$  do
3     if  $k == 1$  and  $i == r$  then
4       continue;
5     end
6     Initialization :  $\eta^0, \mathbf{R}_\theta^0, \mathbf{T}^0$ ;
7     /*Find the reliable matches, and fuse them */
8     while  $p^r \in I_r$  do
9       Search a patch  $p^i \in I_i$  by (6) and (7);
10      Choose the reliable patch by (9) and (21);
11       $p^f \leftarrow \text{fusion}(p^f, \text{fusion}(p^r, p^i))$ ;
12       $(\eta^0, \mathbf{R}_\theta^0, \mathbf{T}^0) \leftarrow (\eta^i, \mathbf{R}_\theta^i, \mathbf{T}^i)$ ;
13    end
14    /* Update the reference image */
15    Construct the fused image  $IF_k$  using  $p^f$ ;
16    Update  $I_r \leftarrow IF_k$  (Sec. III-C);
17  end
18 end

```

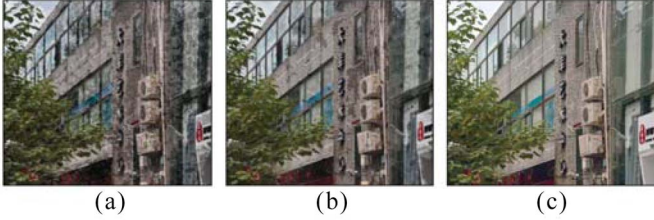


Fig. 5. Close-up comparison in different fusion iterations of the results in Fig. 1(b). Results after the (a) first, (b) second, and (c) final iterations. The input exposure sequence is listed in Fig. 1(a).

Since our method incorporates the matching and fusing procedure in a unified optimization framework, the intuitive and effective techniques could be employed to resolve the occlusion problem. As shown in Fig. 6, the red patch on object A is occluded by object B in the source image so that the wrong patch might be found (the red patch in the source image). Since the matching reliability is computed in terms of the left-right consistency check strategy by (9), we can detect the occlusion regions. Then the reliable matched patches are propagated to the unreliable occlusion regions. Thus, our approach can match and fuse the input image sequence including moving objects with occlusions, and obtain the visually pleasing fusion results.

IV. EXPERIMENTAL RESULTS

In this section, the proposed patch-based match and fusion algorithm is performed on a variety of exposure sequences. The reference image is selected as the normal-exposure image from the input sequence, since the normal-exposure image contains more well-exposed and useful regions. We compare with several state-of-the-art techniques of both the exposure fusion algorithms [16], [21], [26], [39] and tone mapping approaches [12], [15], [18], [23] using HDR images.

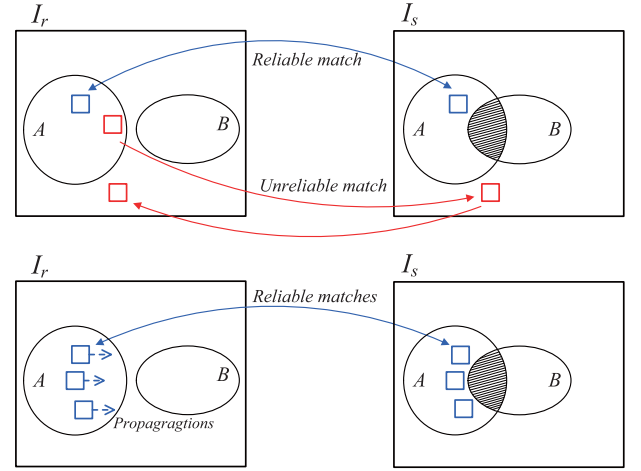


Fig. 6. Illustration of matching and fusing the occlusion regions. Object B occludes object A in the shadow region. First row: unreliable matches (small red rectangles) can be detected by the left-right consistency check. Second row: the reliable matches (small blue rectangles) are propagated to the occlusion regions.

We directly use the MATLAB implementation [16] from their webpage to produce the exposure fusion results in our comparison experiments. Both the gradient-directed fusion method [21] and the generalized random walks based exposure fusion approach [26] are implemented by ourselves. In order to obtain a relatively fair fusion quality comparison with these approaches, the optimal parameter settings are used to run the above algorithms so that the final fusion results can be as good as possible in our comparison experiments.

The main parameters of our method are setting as follows. The patch size is set as 8×8 for the resolution 640×480 in our experiments. The rotation parameter θ , which represents the rotation matrix \mathbf{R}_θ , is set as $\theta \in [-\pi/4, \pi/4]$, and the translation parameter \mathbf{T} is selected automatically according to (19). The maximum searching radius ρ is set as the half length of the largest dimension in reference image. During the iterations, the searching range is reduced automatically in terms of (19). We keep the same setting parameters in all the comparison experiments, which avoids the tedious adjustments for the parameters.

A. Qualitative Results

The comparison between our approach and the gradient-directed exposure fusion method in [21] is shown in Fig. 7. The magnitude of gradients reflects the exposure quality and is used to compute the fusing weights. The changes of the gradient directions among different exposure images are employed to detect the moving objects in the dynamic scene, which helps to preserve the consistent pixels of well exposure regions in the composite image. Comparing to the gradient-directed composition method in [21], our method provides better visual quality results without any halo artifacts [Fig. 7(c)]. As shown in Fig. 7, the input sequence contains three exposure images, there are moving objects of persons in all of these three exposure images. The moving objects are not correctly detected by the gradient direction information, which leads to the halo

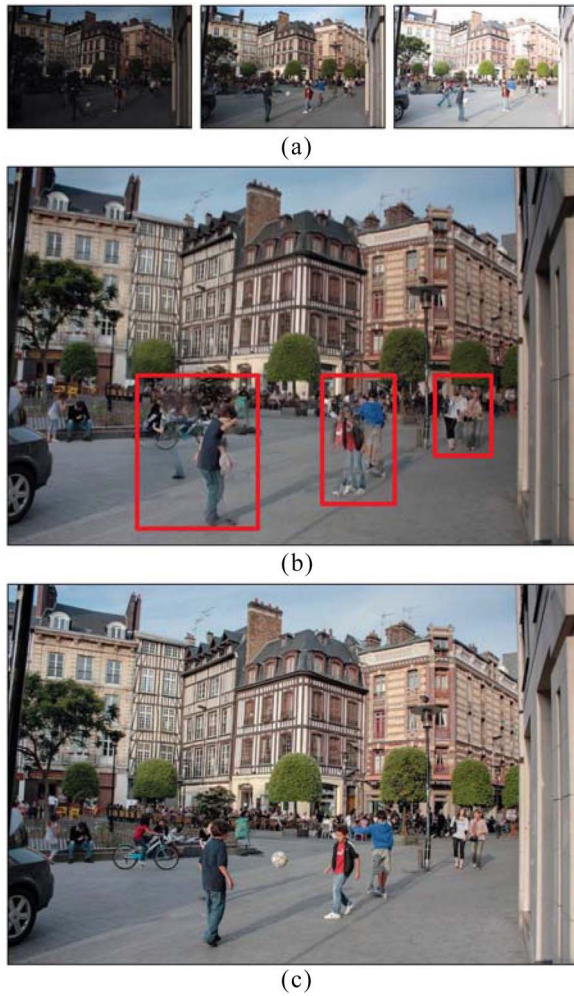


Fig. 7. Comparison results by the gradient-directed exposure fusion method [21] and our approach. (a) Input sequence. (b) Result by [21]. (c) Our result.

artifacts of the final fusion results in Fig. 7(b) (in the red rectangle regions).

Besides the gradient-directed exposure fusion approach, another important exposure fusion technique has been proposed by Shen *et al.* [26], in which the authors introduce the generalized random walks (GRW) technique to optimize the weight map and fuse the exposure images for the static scenes. Fig. 8 shows the comparison results between our approach and GRW algorithm in [26]. A GRW framework is presented to calculate the probabilities of pixels by considering two quality measures that include the local contrast and color consistency. This framework achieves an optimal balance between these two quality measures when compositing the exposure sequence. However, the result by this method still has the ghosting artifacts, produced by the weight map computation which does not consider the influence of moving scenes and objects [Fig. 8(b)]. In contrast, our approach achieves the better fusion quality for preserving the moving foreground person and the moving background without introducing any ghosting artifacts [Fig. 8(c)]. In addition, our result exhibits more color and texture details such as the regions of trees and piano.

The comparison between detail-preserving exposure fusion method [35] and our method is shown in Fig. 9. The input

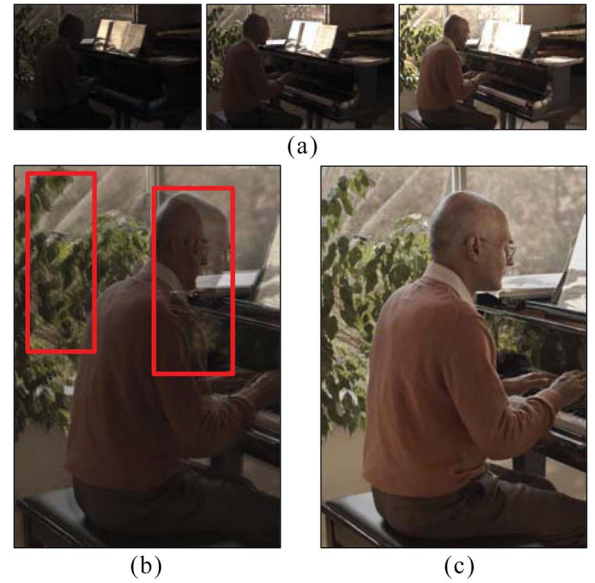


Fig. 8. Comparison of the generalized random walks based exposure fusion method [26] with our method. (a) Input exposure sequence. (b) Result by [26] with ghosting artifacts in the red rectangle region. (c) Our result.

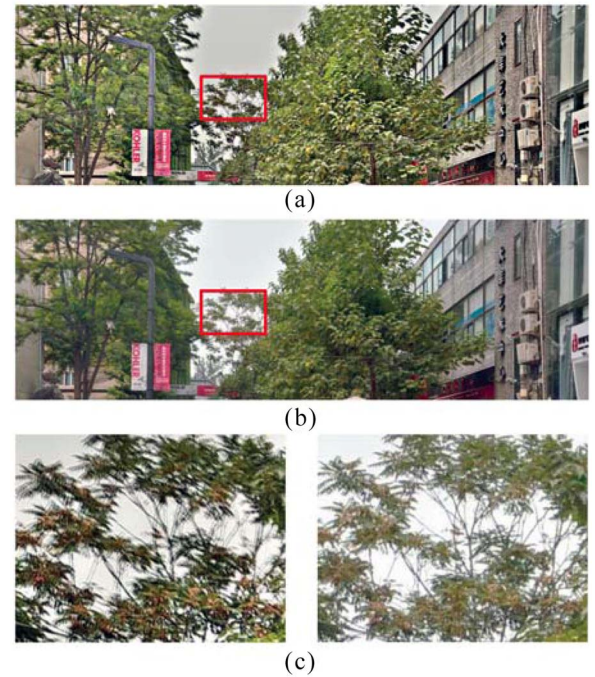


Fig. 9. Comparison between the detail-preserving exposure fusion method [35] and our approach. (a) Fusion result by [35]. (b) Our result. (c) Left and right images are the close-up of the red rectangle regions in (a) and (b), respectively. The input exposure sequence is listed in Fig. 1(a).

exposure sequence is captured by ourselves as shown in Fig. 1(a). In Shen *et al.*'s fusion method [35], the weight maps are computed by considering the image quality measurements, such as the contrast, saturation measurement and exposure measurement. Then the gain control maps are employed to modify the coefficients of subband signals to enhance the details according to the subband architecture. The moving objects need to be first indicated by user interactions and then removed. However, it is very difficult to indicate the tiny

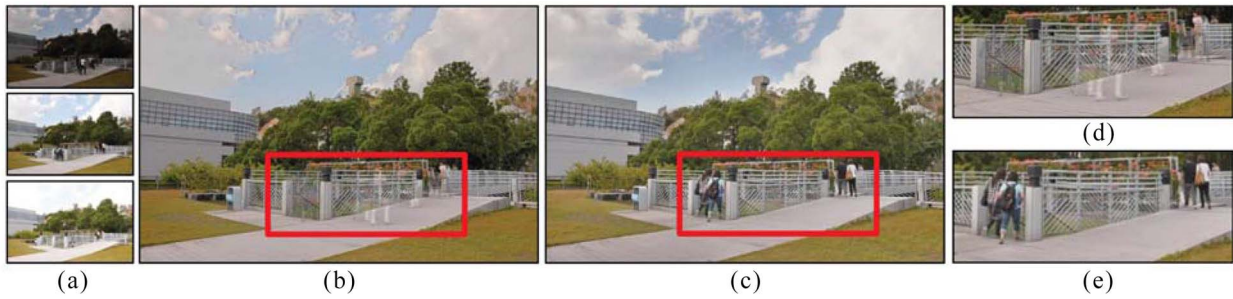


Fig. 10. Comparison of the exposure fusion method using (b) BLP [39] and (c) our method. (a) Exposure sequence. (d) and (e) Close-ups.

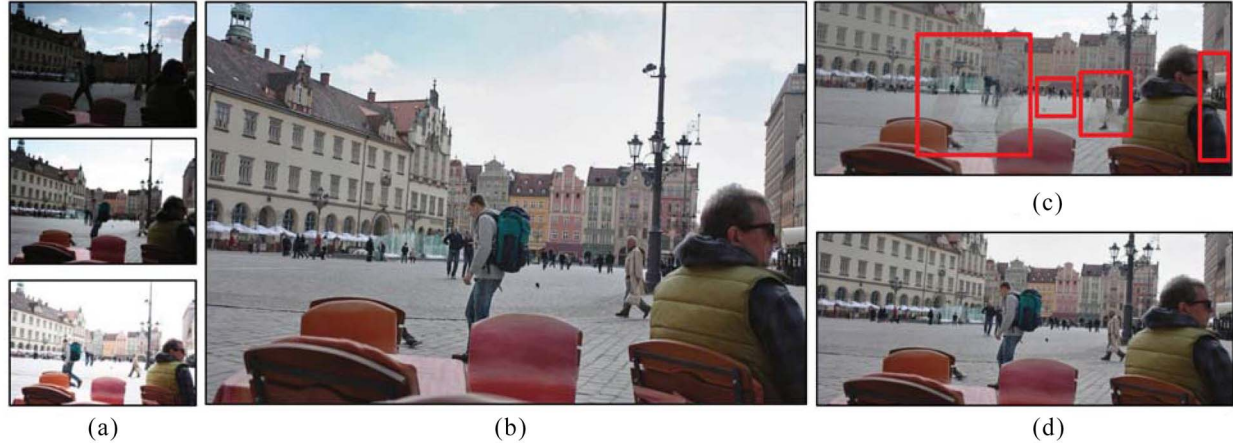


Fig. 11. Comparison of the tone mapping algorithm [11] and our method. (a) Input exposure sequence. (b) Our result. Close-ups obtained by (c) [11] and (d) our method.

moving objects such as the tree branches and the leaves by their user interaction approach [35]. As shown in Fig. 9(a), the ghosting artifacts caused by moving trees still exist, and the color of the sky is unrealistic because of the amplification of subband signals. The left image of Fig. 9(c) is the close-up of the red rectangle region in Fig. 9(a), we can see the obvious ghosting artifacts. Comparing to this detail-preserving fusion method [35], our approach achieves more pleasing and realistic results without introducing any ghosting artifacts [the right image of Fig. 9(c)].

We also compare the recent exposure fusion method using BLP [39] with our approach. Shen *et al.* [39] fused the input image sequence using the BLP that is guided by a new hybrid exposure weight measurement. The local weight, the global weight and the just noticeable distortion-based saliency weight are all considered in their exposure weight measurement. Fig. 10 shows the comparison fusion results between BLP [39] and our method. Though the fusion results by BLP can enhance the details in some regions, the ghosting artifacts still exist in the regions of moving objects since their BLP method does not consider the influence of moving objects [Fig. 10(b) and (d)] in the exposure weight computation. In contrast, our method produces visually pleasing fusion results without introducing any ghosting artifacts [Fig. 10(c) and (e)].

Then, we compare our method with some state-of-the-art tone mapping techniques using HDR image in the following experiments (Figs. 11 and 12). Fig. 11 shows the comparison

between our method and the tone mapping algorithm using the photoreceptor physiology proposed by Reinhard *et al.* [11]. It is one of the best developed techniques for compressing the dynamic range images, which is developed by simulating the photoreceptor adaptation processes in the human visual system. A computational model of photoreceptor behavior is provided to design a global tone reproduction operator. As shown in Fig. 11(c), the color appearance of the tone mapping result by Reinhard *et al.*'s method [11] is pleasingly, but the ghosting artifacts in the red rectangle regions are still serious. Fig. 11(d) gives the close-up of Fig. 11(b). We can observe that the color appearance of our result is more vivid, and moreover, there are not any ghosting artifacts in our result [Fig. 11(d)].

In Figs. 12 and 13, we give a more challenging result, which compares both the tone mapping approaches and the exposure fusion algorithms [15], [16], [18], [21], [26] with our algorithm. The result by Gallo *et al.*'s method [18] is directly taken from their website in our comparison experiments. Reinhard *et al.* [12, pp. 147–152] removed the moving objects by segmenting the images and choosing the well-exposed regions to generate the HDR image. Jacobs *et al.* [15] utilized the statistical quantities to identify the moving objects, and reconstructed them using the input sequence during the HDR image generation. Gallo *et al.* [18] presented a deghosting method by choosing a reference image from the sequence and combining regions that do not cause artifacts with the reference image. Pece and Kautz [23] employed the binary

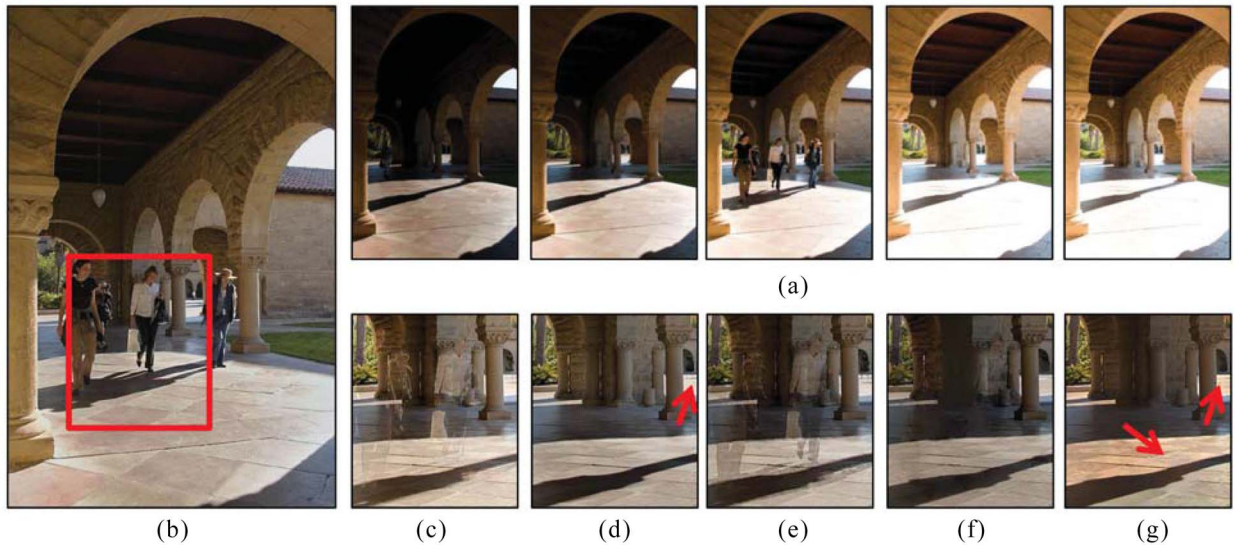


Fig. 12. Comparison of results by different methods. (c) Mertens *et al.* [16]. (d) Zhang and Cham [21]. (e) Reinhard *et al.* [11]. (f) Jacobs *et al.* [15]. (g) Gallo *et al.* [18]. (a) Input exposure sequence. (b) Our result.

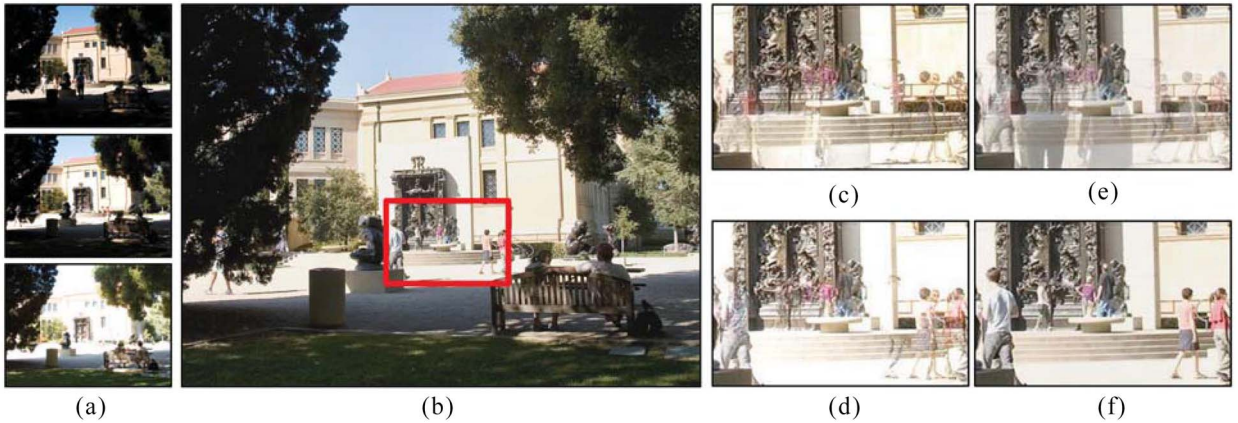


Fig. 13. Comparison results. (a) Input sequence. (b) Our result. (c) Result by Mertens *et al.* [16]. (d) Result by Pece and Kautz [23]. (e) Result by Shen *et al.* [26]. (f) Close-up of our result.

operations to detect clusters of moving pixels, which is called bitmap movement detection approach. The detected moving pixels are utilized to guide the HDR image generation. However, these approaches belong to the traditional two-phase workflow that first constructs the HDR image and then performs the tone mapping algorithms. As shown in Fig. 12(e)–(g), the results by these approaches introduce the apparent ghosting artifacts. It is apparent that the result by Gallo *et al.*'s method [18] suffers from some noticeable ghosting and noisy artifacts [see the red arrows in Fig. 12(g)]. In contrast, our approach obtains more visually pleasing images without introducing any noticeable ghosting artifacts, which preserves the moving objects of persons very well [Fig. 12(b)].

Finally, we give some more challenging comparison results with more moving objects to compare with the state-of-the-art algorithms in Fig. 13. Mertens *et al.* [16] computed the weight map to guide the blending of multiple exposure images by considering the contrast, saturation and well-exposedness measurements. As shown in Fig. 13(c), the ghosting artifacts are obvious since they directly fused the pixels of the moving objects in the same position from the input multiple

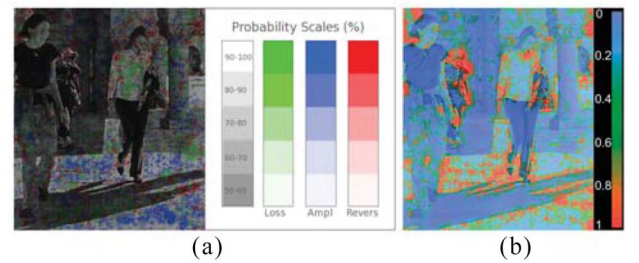


Fig. 14. Probability maps of the visible metric. (a) Three distortion maps shown partially of DRIM [17]. Green: loss of visible contrast. Blue: amplification of invisible contrast. Red: reversal of visible contrast. (b) Probability map of HDR-VDP-2 [25].

exposure images. We also provide the results by the method of Mertens *et al.* [16] in Figs. 1(c) and 12(c). For the same reason, the result by the method of Shen *et al.* [26] contains heavy ghosting artifacts [Fig. 13(e)]. Pece and Kautz [23] detected the moving objects and then removed them when the HDR image was generated, thus the result can partly remove some moving objects [Fig. 13(d)]. However, we can observe that the phenomenon of heavy ghosting artifacts still exists in













	T. Mertens	Ours	Wei Zhang	Ours	R. Shen	Ours
DRIM						
PSNR	177.675900	196.901800	178.054700	196.256900	179.692600	206.986200
HDR-VDP-2						
	(a)		(b)		(c)	

Fig. 15. Quantitative comparison results. Three quality assessment metrics are utilized to evaluate the quality of the fused images by Mertens *et al.* [16], Zhang and Cham [21], Shen *et al.* [26], and our method. (a) Exposure sequence is given in Fig. 1. (b) Exposure sequence is given in Fig. 7. (c) Exposure sequence is given in Fig. 8.

Fig. 13(d) when the moving objects are not correctly detected during the stage of HDR image generation. In contrast, our algorithm preserves the moving object correctly, and keeps many useful information without introducing the ghosting artifacts, such as the moving persons in Fig. 13(b) and (f).

B. Quantitative Evaluation

Many visual metrics [17], [25], [31] have been developed to objectively evaluate the fusion quality in recent years. We use three common visual metrics to quantitatively evaluate the performances of different fusion algorithms, which includes DRIM [17], peak signal-to-noise ratio (PSNR), and HDRVDP2 [25]. DRIM is an image quality metric which can compare the image pairs with different dynamic ranges. Mantiuk *et al.* [25] proposed a visual metric for predicting visibility and quality in different luminance conditions, which termed as HDR-VDP-2. As shown in Fig. 15, we compare the visible quality of fusion images that are produced by the state-of-the-art methods [11], [15], [16], [18], [21], [26]. It is obvious that our approach obtains the higher visible quality. The metrics of DRIM and PSNR are evaluated between the reference image and the fused image. Since the HDR-VDP-2 is limited by the dynamic ranges, an HDR image is generated by [36], and is tonemapped by Photomatix [30]. We regard this tonemapped HDR image as the reference image when evaluating the metric HDR-VDP-2. When we estimate the visual metrics DRIM and PSNR, the normal exposure image is selected as the reference image. The results by different methods are treated as the test images in all the experiments. Fig. 14(a) shows the three distortion maps of DRIM [17]. The same as in [17], we describe the loss of visible contrast with green, the amplification of invisible contrast with blue, and the reversal of visible contrast with red. Thus, the more the green, blue, and red are, the lower quality the test image is. The probability map of HDR-VDP-2 [25] is shown in Fig. 14(b). The pixels in the blue regions are much more similar with the corresponding pixels in the reference image than the pixels in the red regions. In most cases, a higher PSNR indicates that the fused image is of higher quality. As shown

in Fig. 15, comparing with other techniques, our approach can obtain more pleasing results and higher visual quality.

V. CONCLUSION

This paper presents a novel patch-based match and fusion algorithm that can fuse the multiple exposure images of moving scenes using optimization. A uniform iterative optimization framework is developed to integrate the match and fusion process where the reliable corresponding patches are matched and then fused in each iteration. Thus, our approach does not need the registration process for the input multiple exposure images taken by hand-hold cameras. In order to eliminate the influence of various exposure time to compute the similarity measurement, a new patch-based energy cost function is proposed to improve the accuracy and reliability of the matching patches. Then, we design an efficient patch-based exposure fusion approach using the random walker algorithm to preserve the contents of moving objects with more details. The match and fusion processes are performed simultaneously in our algorithm, which makes our approach handle both the static and dynamic scenes. Comparing to the state-of-the-art fusion methods, our approach achieves the better visual quality fusion results without introducing any ghosting artifacts. The proposed framework can be extended for the other applications such as fusing the multiple exposure videos in the future work, where the spatial-temporal measurement should be investigated to measure the quality of exposure videos.

REFERENCES

- [1] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Vis. Comput. Graphics*, vol. 3, no. 4, pp. 291–306, Oct. 1997.
- [2] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH*, 1997, pp. 369–378.
- [3] F. Chung, *Spectral Graph Theory*. Providence, RI, USA: American Mathematical Society, 1997.
- [4] L. Snidaro, R. Niu, G. L. Foresti, and P. K. Varshney, "Quality-based fusion of multiple video sensors for video surveillance," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 4, pp. 1044–1051, Aug. 2007.
- [5] A. Fusiello, V. Roberto, and E. Trucco, "Efficient stereo with multiple windowing," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, San Juan, Puerto Rico, 1997, pp. 858–863.

- [6] T. Mitsunaga and S. K. Nayar, "Radiometric self calibration," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Fort Collins, CO, USA, 1999, pp. 374–380.
- [7] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, 2002.
- [8] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from handheld exposures," *J. Graph. Tools*, vol. 8, no. 2, pp. 17–30, 2003.
- [9] H. Seetzen *et al.*, "High dynamic range display systems," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 760–768, 2004.
- [10] Y. Li, L. Sharan, and E. H. Adelson, "Compressing and companding high dynamic range images with subband architectures," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 836–844, 2005.
- [11] E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Trans. Vis. Comput. Graphics*, vol. 11, no. 1, pp. 13–24, Jan./Feb. 2005.
- [12] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. San Francisco, CA, USA: Morgan Kaufmann, 2005.
- [13] J. Tumblin, A. Agrawal, and R. Raskar, "Why I want a gradient camera," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, 2005, pp. 103–110.
- [14] M. I. Smith and J. P. Heather, "A review of image fusion technology in 2005," *Proc. SPIE*, vol. 5782, pp. 29–45, Apr. 2005.
- [15] K. Jacobs, C. Loscos, and G. Ward, "Automatic high-dynamic range image generation for dynamic scenes," *IEEE Comput. Graph. Appl.*, vol. 28, no. 2, pp. 84–93, Mar./Apr. 2008.
- [16] T. Mertens, J. Kautz, and F. V. Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, 2008.
- [17] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," *ACM Trans. Graph.*, vol. 27, Aug. 2008, Art. ID 69.
- [18] O. Gallo, N. Gelfand, W. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, 2009, pp. 1–7.
- [19] Y. Zhang and Q. Ji, "Efficient sensor selection for active information fusion," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 3, pp. 719–728, Jun. 2010.
- [20] M. Granados *et al.*, "Optimal HDR reconstruction with linear digital cameras," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, 2010, pp. 215–222.
- [21] W. Zhang and W. K. Cham, "Gradient-directed composition of multi-exposure images," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, San Francisco, CA, USA, 2010, pp. 530–536.
- [22] M. Song, D. Tao, C. Chen, X. Li, and C. W. Chen, "Color to gray: Visual cue preservation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1537–1552, Sep. 2010.
- [23] F. Pece and J. Kautz, "Bitmap movement detection: HDR for dynamic scenes," in *Proc. Conf. Vis. Media Product. (CVMP)*, London, U.K., 2010, pp. 1–8.
- [24] T. Zaveri and M. Zaveri, "A novel region based multimodality image fusion method," *J. Pattern Recognit. Res.*, vol. 6, no. 2, pp. 140–153, 2011.
- [25] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, Jul. 2011, Art. ID 40.
- [26] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3634–3646, Dec. 2011.
- [27] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video production system," *ACM Trans. Graph.*, vol. 30, Jul. 2011, Art. ID 41.
- [28] Y. S. Heo, K. M. Lee, and S. U. Lee, "Robust stereo matching using adaptive normalized cross-correlation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 807–822, Apr. 2011.
- [29] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, "Non-rigid dense correspondence with applications for image enhancement," *ACM Trans. Graph.*, vol. 30, Jul. 2011, Art. ID 70.
- [30] Photometric. (2012). *Commercially-Available HDR Processing Software* [Online]. Available: <http://www.hdrsoft.com/>
- [31] Z. Liu *et al.*, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 94–109, Jan. 2012.
- [32] Q. Yang, "A non-local cost aggregation method for stereo matching," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2012, pp. 1402–1409.
- [33] Z. Li, J. Zheng, and S. Rahardja, "Detail-enhanced exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4672–4676, Nov. 2012.
- [34] T. Jinno and M. Okuda, "Multiple exposure fusion for high dynamic range image acquisition," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 358–365, Jan. 2012.
- [35] J. Shen, Y. Zhao, and Y. He, "Detail-preserving exposure fusion using subband architecture," *Vis. Comput.*, vol. 28, no. 5, pp. 463–473, 2012.
- [36] P. Sen *et al.*, "Robust patch-based HDR reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, 2012, Art. ID 203.
- [37] M. Song *et al.*, "Probabilistic exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 341–357, Jan. 2012.
- [38] J. Han, E. Pauwels, and P. Zeeuw, "Fast saliency-aware multi-modality image fusion," *Neurocomputing*, vol. 111, pp. 70–80, Jul. 2013.
- [39] J. Shen, Y. Zhao, S. Yan, and X. Li, "Exposure fusion using boosting Laplacian pyramid," *IEEE Trans. Cybern.*, vol. 44, no. 9, pp. 1579–1590, Sep. 2014.
- [40] H. Singh, V. Kumar, and S. Bhooshan, "A novel approach for detail-enhanced exposure fusion using guided filter," *Sci. World J.*, vol. 2014, Feb. 2014, Art. ID 659217.



Xiameng Qin is currently pursuing the Ph.D. degree from the School of Computer Science, the Beijing Institute of Technology, Beijing, China.

His current research interests include high dynamic range imaging and exposure fusion.



Jianbing Shen (M'11–SM'12) is a Professor with the School of Computer Science, the Beijing Institute of Technology, Beijing, China. His current research interests include intelligent image and video processing. He has published over 40 refereed papers in journals and conference proceedings.



Xiaoyang Mao received the M.S. and the Ph.D. degrees in computer science from Tokyo University, Tokyo, Japan.

She is a Professor with the University of Yamanashi, Kofu, Japan. Her current research interests include flow visualization, texture synthesis, nonphoto-realistic rendering, and human–computer interactions.

Xuelong Li (M'02–SM'07–F'12) is a Full Professor with the Center for Optical Imagery Analysis and Learning, the State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.



Yunde Jia received the Ph.D. degree in mechatronics from the Beijing Institute of Technology, Beijing, China, in 2000.

He is currently a Professor of Computer Science and also a Director of the Laboratory of Media Computing and Intelligent Systems, School of Computer Science, the Beijing Institute of Technology. His current research interests include computer vision, media computing, human–computer interaction, and intelligent systems.