

Generalized Random Walks for Fusion of Multi-Exposure Images

Rui Shen, *Student Member, IEEE*, Irene Cheng, *Senior Member, IEEE*, Jianbo Shi, and Anup Basu, *Senior Member, IEEE*

Abstract—A single captured image of a real-world scene is usually insufficient to reveal all the details due to under- or over-exposed regions. To solve this problem, images of the same scene can be first captured under different exposure settings and then combined into a single image using image fusion techniques. In this paper, we propose a novel probabilistic model-based fusion technique for multi-exposure images. Unlike previous multi-exposure fusion methods, our method aims to achieve an optimal balance between two quality measures, i.e., local contrast and color consistency, while combining the scene details revealed under different exposures. A generalized random walks framework is proposed to calculate a globally optimal solution subject to the two quality measures by formulating the fusion problem as probability estimation. Experiments demonstrate that our algorithm generates high-quality images at low computational cost. Comparisons with a number of other techniques show that our method generates better results in most cases.

Index Terms—Image enhancement, image fusion, multi-exposure fusion, random walks.

I. INTRODUCTION

A NATURAL scene often has a high dynamic range (HDR) that exceeds the capture range of common digital cameras. Therefore, a single digital photo is often insufficient to provide all the details in a scene due to under- or over-exposed regions. On the other hand, given an HDR image, current displays are only capable of handling a very limited dynamic range. In the last decade, researchers have explored various directions to resolve the contradiction between the HDR nature of real-world scenes and the low dynamic range (LDR) limitation of current image acquisition and display technologies. Although cameras with spatially varying pixel exposures [1], cameras that automatically adjust exposure for different parts of a scene [2], [3],

and displays that directly display HDR images [4] have been developed by previous researchers, their technologies are only at a prototyping stage and unavailable to ordinary users.

Instead of employing specialized image sensors, an HDR image can be reconstructed digitally using HDR reconstruction (HDR-R) techniques from a set of images of the same scene taken by a conventional LDR camera [5] or a panoramic camera [6] under different exposure settings. These HDR images usually have higher fidelity than LDR images, which benefits many applications, such as physically-based rendering and remote sensing [7]. As for viewing on ordinary displays, an HDR image is compressed into an LDR image using tone mapping (TM) methods [8], [9]. This two-phase workflow, HDR-R+TM (HDR-R and TM), has several advantages: no specialized hardware is required; various operations can be performed on the HDR images, such as virtual exposure; and user interactions are allowed in the TM phase to generate a tone-mapped image with desired appearance. However, this workflow is usually not as efficient as image fusion (IF, e.g., [10], [11]), which directly combines the captured multi-exposure images into a single LDR image without involving HDR-R, as shown in Fig. 1. Another advantage of IF is that IF does not need the calibration of the camera response function (CRF), which is required in HDR-R if the CRF is not linear. IF is preferred for quickly generating a well-exposed image from an input set of multi-exposure images, especially when the number of input images is small and speed is crucial.

Since its introduction in the 1980s, IF has been employed in various applications, such as multi-sensor fusion [12], [13] (combining information from multi-modality sensors), multi-focus fusion [14], [15] (extending depth-of-field from multi-focus images), and multi-exposure fusion [11], [16] (merging details of the same scene revealed under different exposures). Although some general fusion approaches [17], [18] have been proposed by previous researchers, they are not optimized for individual applications and have only been applied to gray-level images. In this paper, we only focus on multi-exposure fusion and propose a novel fusion algorithm that is both efficient and effective. This direct fusion of multi-exposure images removes the need for generating an intermediate HDR image. A fused image contains all the information present in different images and is ready for viewing on conventional displays. Unlike previous multi-exposure fusion methods [10], [11], our algorithm is based on a probabilistic model and global optimization taking neighborhood information into account. A generalized random walks (GRW) framework is proposed to calculate an optimal set of probabilities subject to two quality measures: local contrast

Manuscript received June 27, 2010; revised November 07, 2010 and January 17, 2011; accepted April 10, 2011. Date of publication May 05, 2011; date of current version November 18, 2011. This work was supported in part by Killam Trusts, iCORE, and Alberta Advanced Education and Technology. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jesus Malo.

R. Shen, I. Cheng, and A. Basu are with the Department of Computing Science, University of Alberta, Edmonton, AB T6G 2E8, Canada (e-mail: rshen@ualberta.ca; locheng@ualberta.ca; basu@ualberta.ca).

J. Shi is with the Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: jshi@cis.upenn.edu).

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/TIP.2011.2150235

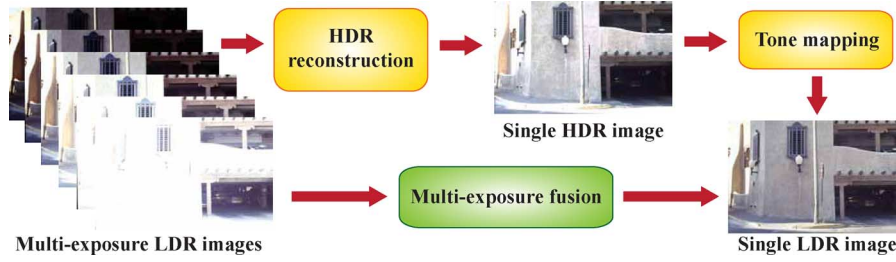


Fig. 1. Comparison between multi-exposure fusion and the HDR reconstruction and tone mapping workflow. (The Garage image sequence courtesy of Shree Nayar.)

and color consistency. The local contrast measure preserves details; the color consistency measure, which is not considered by previous methods, includes both consistency in a neighborhood and consistency with the natural scene. Although used for multi-exposure fusion in this paper, this proposed GRW provides a general framework for solving problems that can be formulated as a labeling problem [19], i.e., estimating the probability of a site (e.g., a pixel) being assigned a label based on known information. The proposed fusion algorithm has low computational complexity and produces a final fused LDR image with fine details and an optimal balance between the two quality measures. By defining problem-specific quality measures, the proposed algorithm may also be applied to other fusion problems.

The rest of the paper is organized as follows. Section II reviews previous methods. Differences between our algorithm, previous IF methods, and methods used in the HDR-R+TM workflow, are also discussed. Section III explains our algorithm in detail. Section IV discusses experimental results and performance, along with comparisons with other IF methods and the HDR-R+TM workflow. Finally, Section V gives the conclusions and future work.

II. RELATED WORK

A. Multi-Exposure Image Fusion

Image fusion methods combine information in different images into a single composite image. Here we only discuss those IF methods that are most relevant to our algorithm. Please refer to [20] for an excellent survey on IF methods in different applications. For multi-exposure images, IF methods directly work on the input LDR images and focus on enhancing dynamic range while preserving details. Goshtasby [10] partitions the input images into uniform blocks and tries to maximize the information in each block based on an entropy measure. However, the approach may generate artifacts on object boundaries if the block is not sufficiently small. Instead of dividing images into blocks, Cheng and Basu [21] combine images in a column-by-column fashion. This algorithm maximizes the contrast within a column by selecting pixels from different images. However, since no neighborhood information is considered, it cannot preserve color consistency and artifacts may occur on object boundaries. Cho and Hong [22] focus on substituting the under- or over-exposed regions in one image, which are determined by a saturation mask, with the well-exposed regions in another image. Region boundaries are blended by minimizing an energy function defined in the gradient domain. Although this algorithm works better on object boundaries, it is only applicable

to two images. Our algorithm works at the pixel level and applies a global optimization taking neighborhood information into account, which avoids the boundary artifacts.

Image fusion can also be interpreted as an analogy to alpha blending. Raman and Chaudhuri [23] generate the fused image by solving an unconstrained optimization problem. The weight function for each pixel is modeled based on local contrast/variance in a way that the fused image tends to have uniform illumination or contrast. Raman and Chaudhuri [24] generate mattes for each pixel in an image using the difference between the original pixel value and the pixel value after bilateral filtering. This measure strengthens weak edges, but may not be sufficient to enhance the overall contrast. Our algorithm defines two quality measures and finds the optimal balance between them, i.e., enhancing local contrast while imposing color consistency.

Multi-resolution analysis based fusion has demonstrated good performance in enhancing main image features. Bogoni and Hansen [16] propose a pyramid-based pattern-selective fusion technique. Laplacian and Gaussian pyramids are constructed for the luminance and chrominance components, respectively. However, the color scheme of the fused image may be very close to the average image because pixels with saturation closest to the average saturation are selected for blending. Mertens *et al.* [11] construct Laplacian pyramids for the input images and Gaussian pyramids for the weight maps. A weight for a pixel is determined by three quality measures: local contrast, saturation, and well-exposedness. The Laplacian and Gaussian pyramids are blended at each level, and then collapsed to form the final image. However, with only local calculation involved and no measure to preserve color consistency, the fusion results may be unnatural. Our algorithm also uses local contrast as one quality measure, but another quality measure that we consider is color consistency, which is not employed by [11]. Moreover, our algorithm does not use pyramid decomposition but solves a linear system defined at the pixel level.

B. HDR Reconstruction and Tone Mapping

Although the HDR-R+TM workflow is usually used in different scenarios than IF methods, we still give a brief discussion on those HDR-R and TM methods sharing some similar features to our IF algorithm, because the original input and the final output of this workflow is the same as IF. Given an input LDR sequence and exposure times associated with each image in the sequence, HDR-R techniques [5] first recover the CRF, which is a mapping from the exposure at each pixel location to

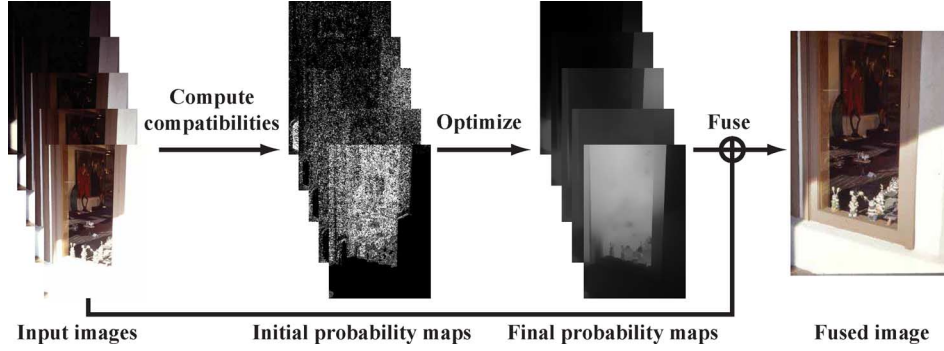


Fig. 2. Processing procedure of the proposed fusion algorithm. (The window image sequence courtesy of Shree Nayar.)

the pixel's digital LDR value and then use the CRF to reconstruct an HDR image via a weighting function. Debevec and Malik [5] recover the CRF by minimizing a quadratic objective function defined on exposures, pixels' LDR values, and exposure times. Then, a hat-shaped weighting function is used to reconstruct the HDR image. Granados *et al.* [25] assume the CRF is linear and focus on the development of an optimal weighting function using a compound-Gaussian noise model. Our IF algorithm also solves a quadratic objective function, but the function is defined on local features and the solution leads to probability maps.

Given an HDR image, tone mapping methods [26] aim to reduce its dynamic range while preserving details. TM usually works solely in the luminance channel. Global TM methods [27], [28], which apply spatially invariant mapping of luminance values, usually have high speed, while local TM methods [8], [9], [29], [30], which apply spatially varying mapping, usually produce images with better qualities especially when strong local contrast is present [9], [30]. Reinhard *et al.* [9] use a multi-scale local contrast measure to compress the luminance values. Li *et al.* [29] decompose the luminance channel of the input HDR image into multiscale subbands and apply local gain control to the subbands. Shan *et al.* [30] define a linear system for each overlapping window in the HDR image using local contrast, and the solution of each linear system are two coefficients that map luminance values from HDR to LDR. Although our IF algorithm also uses local contrast to define a linear system, the local contrast in our algorithm is calculated in a different manner and another quality measure (i.e., color consistency) is also considered. Furthermore, our IF algorithm defines a linear system on pixels from all the original LDR images, and the solution is a set of probabilities that determine the contributions from each pixel of each original LDR image to its corresponding pixel in the fused image. Krawczyk *et al.* [8] segment an HDR image into frameworks with consistent luminance and compute the belongingness of each pixel to each framework using the framework's centroid, which results in a set of probability maps. Our IF algorithm also generates probability maps, but directly from the original LDR sequence with no HDR or segmentation involved. One typical problem with some local TM methods is halo artifacts introduced due to contrast reversals [26]. Our IF algorithm balances between contrast and consistency in a neighborhood, which can prevent contrast reversals. Another problem with some TM methods is that color artifacts like oversaturation may be introduced into the results, because operations are

usually performed in the luminance channel without involving chrominance [26]. This problem does not apply to our IF algorithm, because every color channel is treated equally.

III. PROBABILISTIC FUSION

Unlike most previous multi-exposure fusion methods, we consider image fusion as a probabilistic composition process, as illustrated in Fig. 2.¹ The initial probability that a pixel in the fused image belongs to each input image is estimated based on local features. Taking neighborhood information into account, the final probabilities are obtained by global optimization using the proposed generalized random walks. These probability maps serve as weights in the linear fusion process to produce the fused image. In a probability map, the brighter a pixel is, the higher the probability. These processes are explained in detail below.

A. Problem Formulation

The fusion of a set of multi-exposure images can be formulated as a probabilistic composition process. Let $\mathcal{D} = \{\mathbf{I}^1, \dots, \mathbf{I}^K\}$ denote the set of input images and $\mathcal{L} = \{l_1, \dots, l_K\}$ the set of labels, where a label $l_k \in \mathcal{L}$ is associated with the k th input image $\mathbf{I}^k \in \mathcal{D}$. \mathbf{I}^k 's are assumed to be already registered and have the same size with N pixels each. Normally, $K \ll N$. Let us define a set \mathcal{X} of variables such that an $x_i \in \mathcal{X}$ is associated with the i th pixel p_i in the fused image \mathbf{I}^* and takes a value from \mathcal{L} . Then, each pixel in the fused image \mathbf{I}^* can be represented as

$$p_i = \sum_{k=1}^K P^k(x_i) p_i^k \quad (1)$$

where p_i^k denotes the i th pixel in \mathbf{I}^k and $P^k(x_i) \triangleq P(x_i = l_k | \mathcal{D})$ the probability of pixel p_i being assigned label l_k given \mathcal{D} with $\sum_k P^k(x_i) = 1$. This probabilistic formulation converts the fusion problem to the calculation of marginal probabilities given the input images subject to some quality measures and helps to achieve an optimal balance between different quality measures. If every pixel is given equal probability, i.e., $P^k(x_i) = 1/K, \forall i, k$, \mathbf{I}^* is simply the average of \mathbf{I}^k 's. Although it is also possible to only combine pixels with highest

¹The initial probability maps are normalized at each pixel.

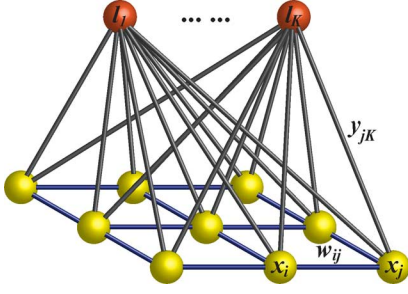


Fig. 3. Graph used in GRW. The yellow nodes are scene nodes and the orange nodes are label nodes.

probabilities instead of using (1), artifacts may appear at locations with significant brightness changes between different input images.

If $P^k(x_i)$'s are simply viewed and calculated as local weights followed by applying some relatively simple smoothing filters, such as Gaussian filter and bilateral filter, either artifacts (like halos) may appear at object boundaries or it is difficult to determine the termination criteria of the filtering; therefore, the results are usually unsatisfactory [11]. An alternative is to use multi-resolution fusion techniques [11], [16], where the weights are blended in each level to produce more satisfactory results. However, the weights in each level are still determined locally, which may not be optimal in a large neighborhood. In contrast to multi-resolution techniques, we propose an efficient single-resolution fusion technique by formulating $P^k(x_i)$'s in (1) as probabilities in GRW. A set of optimal $P^k(x_i)$'s that balances the influence of different quality measures is computed from global optimization in GRW. Experiments (Section IV) show that the results of the proposed probabilistic fusion technique are comparable to the results of multi-resolution techniques and the results of the HDR-R+TM workflow.

B. Generalized Random Walks

In this section, we propose a generalized random walks (GRW) framework based on the random walks (RW) algorithm [31], [32] and the relationship between RW and electrical networks [33].

1) *Image Representation*: As shown in Fig. 3, the variable set \mathcal{X} and the label set \mathcal{L} are represented in a weighted undirected graph similar to [31]. Each variable is associated with a pixel location, and each label is associated with an input image in our case. The graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is constructed as $\mathcal{V} = \mathcal{L} \cup \mathcal{X}$ and $\mathcal{E} = \mathcal{V} \times \mathcal{V}$ including edges both within \mathcal{X} and between \mathcal{X} and \mathcal{L} . The yellow nodes are scene nodes and the orange nodes are label nodes. For a scene node $x \in \mathcal{X}$, edges $\mathcal{E}_{\mathcal{X}}$ are drawn between it and each of its immediate neighbors in \mathcal{X} (4-connectivity is assumed in our case). In addition, edges $\mathcal{E}_{\mathcal{L}}$ are drawn between a scene node and each label node. $w_{ij} \triangleq w(x_i, x_j)$ is a function defined on $\mathcal{E}_{\mathcal{X}}$ that models the compatibility/similarity between nodes x_i and x_j , and $y_{ik} \triangleq y(x_i, l_k)$ is a function defined on $\mathcal{E}_{\mathcal{L}}$ that models the compatibility between x_i and l_k .

2) *Dirichlet Problem*: Let \mathcal{V} be arranged in a way that the first K nodes are label nodes, i.e., $\{v_1, \dots, v_K\} = \mathcal{L}$, and the rest N nodes are scene nodes, i.e., $\{v_{K+1}, \dots, v_{K+N}\} = \mathcal{X}$. With two positive coefficients γ_1 and γ_2 introduced to balance

the weights between $y(\cdot, \cdot)$ and $w(\cdot, \cdot)$, we can define a node compatibility function $c(\cdot, \cdot)$ on \mathcal{E} with the following form:

$$c_{ij} \triangleq c(v_i, v_j) = \begin{cases} \gamma_1 y_{i-K, j}, & (v_i, v_j) \in \mathcal{E}_{\mathcal{L}} \wedge v_j \in \mathcal{L} \\ \gamma_2 w_{i-K, j-K}, & (v_i, v_j) \in \mathcal{E}_{\mathcal{X}}. \end{cases} \quad (2)$$

Because the graph is undirected, we have $c_{ij} = c_{ji}$. Let $u(v_i)$ denote the potential associated with v_i . Based on the relationship between RW and electrical networks [33], the total energy of the system given in Fig. 3 is

$$E = \frac{1}{2} \sum_{(v_i, v_j) \in \mathcal{E}} c_{ij} (u(v_i) - u(v_j))^2. \quad (3)$$

Our goal is to find a function $u(\cdot)$ defined on \mathcal{X} that minimizes this quadratic energy with boundary values $u(\cdot)$ defined on \mathcal{L} . If $u(\cdot)$ satisfies $\nabla^2 u = 0$, then it is called harmonic, and the harmonic function is guaranteed to minimize such quadratic energy E [32]. The problem of finding this harmonic function is called the Dirichlet problem.

The harmonic function $u(\cdot)$ can be computed efficiently using matrix operations. Similar to [32], a Laplacian matrix \mathbf{L} can be constructed following (4); however, unlike [32], \mathbf{L} here contains both the label nodes and the scene nodes and becomes a $(K + N) \times (K + N)$ matrix:

$$\mathbf{L}_{ij} = \begin{cases} d_i, & i = j \\ -c_{ij}, & (v_i, v_j) \in \mathcal{E} \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

where $d_i = \sum_{v_j \in \mathcal{N}_i} c_{ij}$ is the degree of the node v_i defined on its immediate neighborhood \mathcal{N}_i . Then, (3) can be rewritten in matrix form as

$$E = \begin{pmatrix} \mathbf{u}_{\mathcal{L}} \\ \mathbf{u}_{\mathcal{X}} \end{pmatrix}^T \mathbf{L} \begin{pmatrix} \mathbf{u}_{\mathcal{L}} \\ \mathbf{u}_{\mathcal{X}} \end{pmatrix} = \begin{pmatrix} \mathbf{u}_{\mathcal{L}} \\ \mathbf{u}_{\mathcal{X}} \end{pmatrix}^T \begin{pmatrix} \mathbf{L}_{\mathcal{L}} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{L}_{\mathcal{X}} \end{pmatrix} \begin{pmatrix} \mathbf{u}_{\mathcal{L}} \\ \mathbf{u}_{\mathcal{X}} \end{pmatrix} \quad (5)$$

where $\mathbf{u}_{\mathcal{L}} = (u(l_1), \dots, u(l_K))^T$ and $\mathbf{u}_{\mathcal{X}} = (u(x_1), \dots, u(x_N))^T$; $\mathbf{L}_{\mathcal{L}}$ is the upper left $K \times K$ submatrix of \mathbf{L} that encodes the interactions within \mathcal{L} ; $\mathbf{L}_{\mathcal{X}}$ is the lower right $N \times N$ submatrix that encodes the interactions within \mathcal{X} ; and \mathbf{B} is the upper right $K \times N$ submatrix that encodes the interactions between \mathcal{L} and \mathcal{X} . Hence, the minimum energy solution can be obtained by setting $\nabla E = 0$ with respect to $\mathbf{u}_{\mathcal{X}}$, i.e., solving the following equation:

$$\mathbf{L}_{\mathcal{X}} \mathbf{u}_{\mathcal{X}} = -\mathbf{B}^T \mathbf{u}_{\mathcal{L}}. \quad (6)$$

In some cases, part of \mathcal{X} may be already labeled. These pre-labeled nodes can also be represented naturally in the current framework without altering the structure of the graph. Suppose x_i is one of the pre-labeled nodes and is assigned label l_k . Then, we can simply assign a sufficiently large value to y_{ik} and solve the same (6) for the unlabeled scene nodes.

3) *Probability Calculation*: The probability $P^k(x_i)$ that a scene node x_i is assigned the k th label l_k given all the observed data \mathcal{D} can be considered as the probability that a random walker starting at a scene node $x_i \in \mathcal{X}$ first reaches the label node $l_k \in \mathcal{L}$ on the graph \mathcal{G} . Thus, $P^k(x_i)$ can be computed for each

pair of (x_i, l_k) by solving K Dirichlet problems in K iterations. Note that the probabilities here are used in the context of random walks [31]–[33], which is different from the log-probabilities used in Markov random field energy minimization [19].

$\gamma_1 y_{ik}/d_i$ can be considered as the initial probability that the scene node x_i is assigned label l_k given data \mathcal{D}_i associated with x_i and data \mathcal{D}^k associated with l_k , i.e., the probability that a random walker transits from x_i to l_k in a single move:

$$\frac{\gamma_1 y_{ik}}{d_i} = P(x_i = l_k | \mathcal{D}_i, \mathcal{D}^k). \quad (7)$$

$\gamma_2 w_{ij}/d_i$ can be considered as the probability that the scene nodes x_i and x_j are assigned the same label given \mathcal{D}_i and \mathcal{D}_j , i.e., the probability that a random walker transits from x_i to x_j in a single move:

$$\frac{\gamma_2 w_{ij}}{d_i} = P(x_i = x_j | \mathcal{D}_i, \mathcal{D}_j). \quad (8)$$

When constructing y_{ik} 's or w_{ij} 's, it is assumed that the probability that x_i takes a different label from l_k or x_j is zero. This assumption ensures the smoothness of the labeling. y_{ik} 's and w_{ij} 's are initialized at the beginning of the algorithm and remain the same in each iteration.

Let $u^k(v_i)$ be the potential associated with node v_i in the k th iteration, which we define to be proportional to $P^k(v_i)$:

$$u^k(v_i) = \gamma_3 P^k(v_i) \quad (9)$$

where γ_3 is a positive constant. Since $P^k : \mathcal{V} \rightarrow [0, 1]$, $u^k : \mathcal{V} \rightarrow [0, \gamma_3]$. $u^k(\cdot)$ is a binary function on \mathcal{L} : $u^k(l) = \gamma_3$, when $l = l_k$; $u^k(l) = 0$, otherwise. For any $x_i \in \mathcal{X}$, $\sum_{k=1}^K u^k(x_i) = \gamma_3$. Once y_{ik} 's and w_{ij} 's are defined, the probabilities $P^k(x_i)$'s can then be determined from (6) and (9).

The RW algorithm [32] requires some variables to be instantiated, i.e., some scene nodes to be pre-labeled. This requirement is relaxed in GRW. The RW algorithm with prior models (RWPM) [31] is proposed for image segmentation and derives a similar linear system to (6) from a Bayesian point of view. A small inaccuracy in [31] is that a weighting parameter γ is missing from the right-hand side in their (11). When setting $\gamma_2 = \gamma_3$ in GRW, we can get a linear system in the same format as derived in [31] with the missing parameter γ added. The nodewise priors in [31], which correspond to our compatibility function $y(\cdot, \cdot)$, are required to be defined following a probability expression. Although it is mentioned in [31] that the solution of RWPM is equivalent to that of the original RW [32] on an augmented graph considering the label nodes as the extra pre-labeled nodes, the requirement on the format of the node-wise priors limits the choice of $y(\cdot, \cdot)$. This requirement is relaxed in GRW, where we formulate the problem from the original RW point of view [33], where probabilities are considered as the transition probabilities of a random walker moving between nodes. The edge weighting function in [31], which corresponds to our compatibility function $w(\cdot, \cdot)$, serves as a regularization term. In GRW, $y(\cdot, \cdot)$ and $w(\cdot, \cdot)$ are not probability quantities; instead, they represent compatibility/similarity and are used to define the transition probabilities. In GRW, we have relaxed the extra requirements in [31], [32] and provided a more

flexible framework, where the compatibility functions (and potential function) may be defined in any form according to the need of a particular problem. For the fusion problem, the form of the compatibility functions is presented in Section III-C. Although in this paper GRW is proposed to solve the multi-exposure fusion problem, it can actually be applied to many different problems that can be formulated as estimating the probability of a site (e.g., a pixel) being assigned a label given known information.

C. Compatibility Functions

The compatibility functions $y(\cdot, \cdot)$ and $w(\cdot, \cdot)$ are defined to represent respectively the two quality measures used in the proposed fusion algorithm, i.e., local contrast and color consistency. Since image contrast is usually related to variations in image luminance [26], the local contrast measure should be biased towards pixels from the images that provide more local variations in luminance. Let g_i^k denote the second-order partial derivative computed in the luminance channel at the i th pixel in image \mathbf{I}^k , which is a indicator of local contrast. The higher the magnitude of g_i^k (denoted by $|g_i^k|$) is, the more variations occur near the pixel p_i^k , which may indicate more local contrast. In [11], a Laplacian filter is used to calculate the variations. Here both Laplacian filter and central difference in a 3×3 neighborhood were tested. With all other settings the same, central difference produces slightly better visual quality in the fused images. Therefore, central difference is used to approximate the second-order derivative. However, if the frequency (i.e., number of occurrences) of a value $|g_i^k|$ in \mathbf{I}^k is very low, the associated pixels may be noise or belong to unimportant features. Therefore, unlike previous methods [11], [23], our contrast measure is normalized by the frequencies of each $|g_i^k|$. In addition, g_i^k 's are modified using a sigmoid-shaped function to suppress the difference in high contrast regions. Because of the nonlinear human perception of contrast [34], such a mapping scheme helps us avoid overemphasis on high local variations. Hence, taking into account both the magnitude and the frequency of the contrast indicator g_i^k , the compatibility between a pixel and a label is computed as

$$y_{ik} = \theta_{ik} \left[\text{erf} \left(\frac{|g_i^k|}{\sigma_y} \right) \right]^K \quad (10)$$

where θ_{ik} represents the frequency of the value $|g_i^k|$ in \mathbf{I}^k ; $\text{erf}(\cdot)$ is the Gaussian error function, which is monotonically increasing and sigmoid shaped; the exponent K is equal to the number of input images and controls the shape of $\text{erf}(\cdot)$ by giving less emphasis on the difference in high contrast regions as the number of input images increases; and σ_y is a weighting coefficient, which we take as the variance of all g_i^k 's.

The other quality measure used in our algorithm is color consistency, which is not considered in previous methods [10], [11]. Although Bogoni and Hansen [16] also suggested a color consistency criterion by assuming that the hue component for all the input images is constant, this assumption is not true if the images are not taken with close exposure times. In addition, they

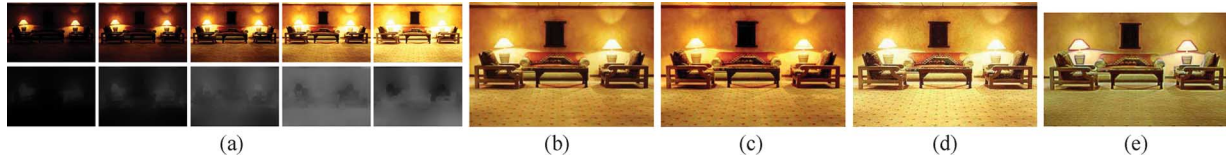


Fig. 4. Comparison of our method with EF, EntropyF, and VC using the Chairs image sequence. The result of EF is comparable to ours. The result of EntropyF suffers a little over-exposure. The result of VC shows serious halo artifacts on object boundaries. (a) Input sequence (top) and final probability maps (bottom). (b) Proposed. (c) EF [11]. (d) EntropyF [10]. (e) VC [23]. (Input sequence courtesy of Shree Nayar.)

do not consider consistency in a neighborhood. Our color consistency measure imposes not only color consistency in a neighborhood but also consistency with the natural scene. If two adjacent pixels in most images have similar colors, then it is very likely that they have similar colors in the fused image. Also, if the colors at the same pixel location in different images under appropriate exposures (not under- or over-exposed) are similar, they indicate the true color of the scene and the fused color should not deviate from these colors. Therefore, the following equation is used to evaluate the similarity/compatibility between adjacent pixels in the input image set using all three channels in the RGB color space:

$$w_{ij} = \prod_k \exp \left(-\frac{\|p_i^k - p_j^k\|}{\sigma_w} \right) \approx \exp \left(-\frac{\|\bar{p}_i - \bar{p}_j\|}{\bar{\sigma}_w} \right) \quad (11)$$

where p_i^k and p_j^k are adjacent pixels in image \mathbf{I}^k , $\exp(\cdot)$ is the exponential function; $\|\cdot\|$ denotes Euclidean distance, $\bar{p}_i = (1/K) \sum_k p_i^k$ denotes the average pixel, and σ_w and $\bar{\sigma}_w = \sigma_w/K$ are free parameters. Although the two quality measures are defined locally, a global optimization using GRW is carried out to produce a fused image that maximizes contrast and details, as well as imposing color consistency. Once $y(\cdot, \cdot)$ and $w(\cdot, \cdot)$ are defined using (10) and (11), the probabilities $P^k(x_i)$'s are calculated using (2)–(6). Here, we fix $\gamma_3/\gamma_1 = 1$ and only use the ratio $\gamma = \gamma_2/\gamma_1$ to determine the relative weight between $y(\cdot, \cdot)$ and $w(\cdot, \cdot)$. The basic steps of our algorithm are summarized in Algorithm 1.

Algorithm 1 Basic steps of the proposed fusion algorithm

- 1: Construct function $y(\cdot, \cdot)$ from (10)
 - 2: Construct function $w(\cdot, \cdot)$ from (11)
 - 3: Construct function $c(\cdot, \cdot)$ from (2)
 - 4: Construct \mathbf{L} from (4)
 - 5: **for** $k = 1$ to K **do**
 - 6: Calculate $u^k(\cdot)$ ($P^k(\cdot)$) for all $x_i \in \mathcal{X}$ by solving (6)
 - 7: **end for**
 - 8: Compute the fused image from (1)
-

D. Acceleration

To accelerate the computation as well as reduce memory usage, the final probability maps are computed at a lower resolution of the Laplacian matrix and then interpolated back to the original resolution before being used in (1). The contrast indicator g_i^k of each pixel is calculated at the original resolution. Then, the images are divided into uniform blocks of size $\eta \times \eta$. The average of g_i^k 's in a block is used to calculate

the compatibility y_{ik} between that block and the label l_k . The compatibility between two adjacent blocks is computed as w_{ij} of the average pixel in one block and the average pixel in the other block.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Comparison with Other Image Fusion Methods and Some Tone Mapping Methods

Only three free parameters are used in our algorithm, and we take $\bar{\sigma}_w = 0.1$, $\gamma = 1.0$, $\eta = 4$ in all experiments unless otherwise mentioned. 12 LDR image sequences were used in the experiments. Fig. 4 shows the comparison of our IF algorithm with three other IF methods on the Chairs image sequence. The input image sequence and final probability maps are given in Fig. 4(a). Brighter pixels in a probability map stand for higher probabilities. The result of entropy fusion (EntropyF) [10] is taken from its project webpage.² The result of variational composition (VC) [23] is taken from its paper. The results of exposure fusion (EF) [11] in all experiments are generated by the Matlab implementation provided by its authors. Its default parameter setting is used in all experiments. The result of EF is comparable to ours. The result of EntropyF suffers a little over-exposure. The result of VC shows serious halo artifacts on object boundaries.

Figs. 5 and 6 give comparison of our IF algorithm with EF and three TM methods on two image sequences. In all experiments, the intermediate HDR images for photographic tone reproduction (PTR) [9], subband compression (SC) [29] and linear windowed (LW) [30] are generated from the corresponding LDR sequences using HDRShop,³ which employs the HDR-R algorithm in [5]. The results of PTR are generated by an HDRShop plugin.⁴ The results of SC and LW are generated by the Matlab implementations provided by their respective authors. There is no constant set of parameters in their original papers, because TM methods usually depend on user-controlled parameters to generate desired tone-mapped images. However, in order to give a relatively fair image quality comparison with our and other IF methods where constant sets of parameters are used throughout the experiments, we use the default parameter settings in their programs in all experiments.

Fig. 5 shows the experimental results on the Belgium House sequence. The last row gives a closeup view of the window regions. Although the result of SC preserves as much detail as ours, it looks a little pink due to color distortion. The results of EF and PTR suffer over-exposure for all the window regions. The result of LW shows some color artifacts, e.g., color reversal

²<http://www.cs.wright.edu/~agoshtas/hdr.html>.

³<http://www.hdrshop.com/>.

⁴<http://www.gregdowning.com/HDRI/tonemap/Reinhard/>.

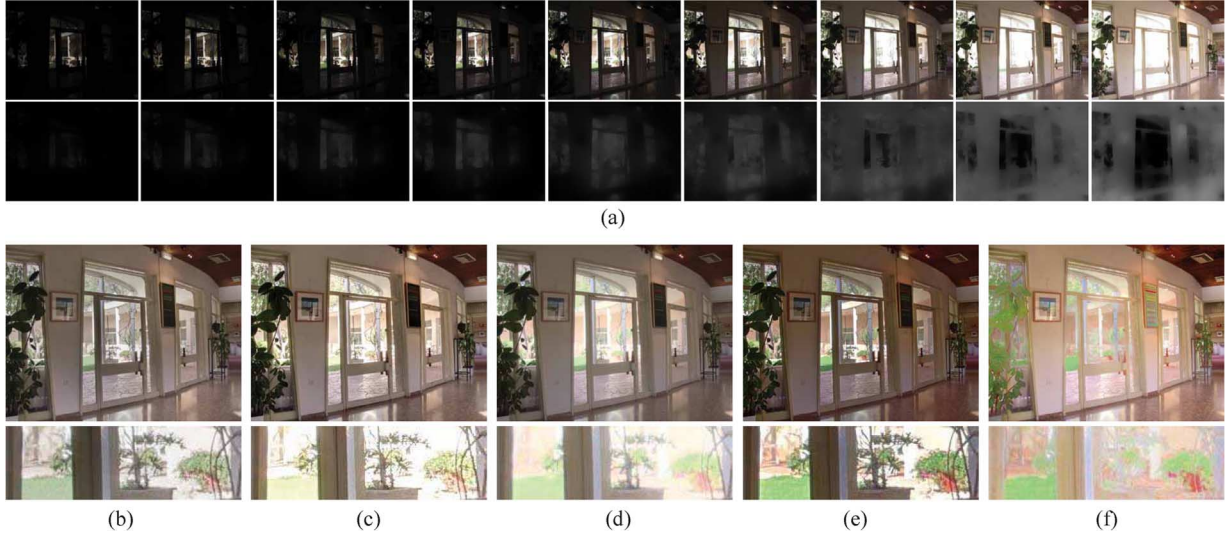


Fig. 5. Comparison of our algorithm with EF, PTR, SC, and LW using the Belgium House image sequence. The intermediate HDR image for PTR, SC, and LW is generated by HDRShop. The results of EF and PTR suffer over-exposure in the window regions. The results of SC and LW show some color distortion (looking pink). (a) Input image sequence (top) and final probability maps (bottom). (b) Proposed. (c) EF [11]. (d) PTR [9]. (e) SC [29]. (f) LW [30]. (Input sequence courtesy of Dani Lischinski.)

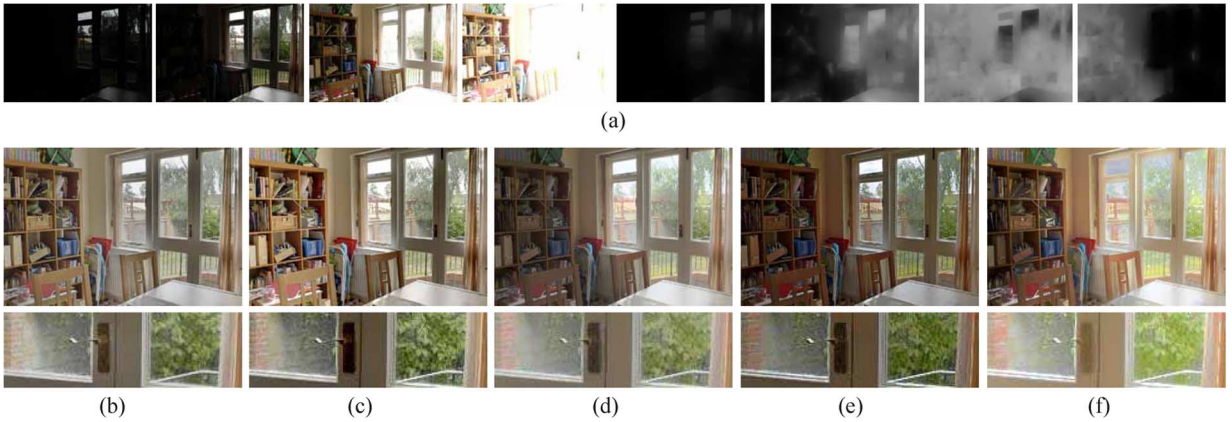


Fig. 6. Comparison of our algorithm with EF, PTR, SC and LW using the House image sequence. The intermediate HDR image for PTR, SC, and LW is generated by HDRShop. EF introduces color artifacts that assign two chairs of the same type different colors. The result of PTR is a little dark for the indoor scene. The result of SC looks a little pink and dark for the entire image. The result of LW shows some color distortion. Our result reveals more detail than EF, especially in the door lock region. (a) Input image sequence (left four) and final probability maps (right four). (b) Proposed. (c) EF [11]. (d) PTR [9]. (e) SC [29]. (f) LW [30]. (Input sequence courtesy of Tom Mertens.)

of the blackboard. Color artifacts in the results of TM methods are usually caused by operations carried out solely in the luminance channel without involving chrominance [26]. Our method treats each color channel equally and imposes color consistency, which helps to avoid color artifacts. Note that adjusting the parameters in SC and LW may reduce color distortion and generate more pleasing images, which is considered as a common user interaction in TM methods. However, we use constant sets of parameters for both TM and IF methods in our experiments in order to give a relatively fair comparison of these two types of methods.

Fig. 6 shows the results on the House sequence. The last row gives a closeup view of the door lock region. The result of PTR is a little dark for the indoor scene. The result of SC reveals many details of the outdoor scene but looks a little pink and dark for the entire image. The result of LW also reveals many details but shows some color distortion of the outdoor scene. Although both our method and EF use local variations to indicate local contrasts, we use a nonlinear function to modify

the contrast indicators. This nonlinear mapping is to reflect the nonlinear human perception of contrast [34]. In addition, the well-exposedness measure in EF is biased towards pixels with a specific luminance. Therefore, our result reveals more detail than EF, especially in the door lock region. The two chairs in the original sequence have the same color, but for EF the color difference between them is quite obvious. Our method keeps the colors consistent while EF fails, because our method imposes consistency in large neighborhoods and consistency with the natural scene via the color consistency measure. Our algorithm works well for various scenes and combinations of exposure settings. More results on different image sequences are given in Fig. 7.

B. Computational Complexity

In the initialization step, the input image sequence is converted to gray scale for calculating y_{ik} 's, and an average image



Fig. 7. Results on different image sequences. Our algorithm works well for various scenes and combinations of exposure settings. Left most: National Cathedral; left: Chateau; right: Lizard; right most: Room. (Input sequences courtesy of Max Lyons, HDRsoft.com, Eric Reinhard, and Grzegorz Krawczyk, respectively.)

TABLE I
COMPUTATIONAL TIMES OF THE PROPOSED ALGORITHM AND EF ON THE TEST IMAGE SEQUENCES. TIMES ARE RECORDED IN SECONDS

Input sequence	Size	Initialize	Compute compatibilities	Optimize	Fuse	Total	EF [11]
Window	$226 \times 341 \times 5$	0.06	0.03	0.02	0.05	0.17	0.68
Chairs	$343 \times 231 \times 5$	0.06	0.04	0.02	0.05	0.17	0.68
Garage	$348 \times 222 \times 6$	0.07	0.05	0.02	0.06	0.20	0.80
Igloo	$236 \times 341 \times 6$	0.07	0.05	0.03	0.05	0.20	0.85
House	$752 \times 500 \times 4$	0.24	0.13	0.14	0.21	0.72	2.64
Memorial Church	$512 \times 768 \times 16$	0.75	0.46	0.20	0.78	2.18	9.82
Chateau	$1500 \times 644 \times 5$	0.78	0.40	0.57	0.76	2.51	9.73
Belgium House	$1025 \times 769 \times 9$	0.94	0.51	0.37	0.92	2.74	11.68
Room	$1024 \times 768 \times 13$	1.32	0.80	0.53	1.34	3.99	16.40
National Cathedral	$1536 \times 2048 \times 2$	1.33	0.38	1.66	1.10	4.47	12.79
Lamp	$1600 \times 1200 \times 15$	3.33	1.94	1.26	3.61	10.13	Out of Memory
Lizard	$2462 \times 1632 \times 9$	4.56	2.73	2.48	4.89	14.66	Out of Memory

is also computed for calculating w_{ij} 's. The complexity of this step is $O(NK)$, where N is the number of pixels in an input image and K is the number of images in the input sequence. Since the number of operations is proportional to the total number of input pixels, the complexity for computing the compatibilities is also $O(NK)$. We employ CHOLMOD⁵ [35], which is based on supernodal Cholesky factorization, to solve the linear system in (6). The complexity of this direct solver is proportional to the number of the nonzero entries in \mathbf{L}_X , and in our case it is $O(N/\eta^2)$. Since there are altogether K linear systems to be solved, the complexity of this step is $O(NK/\eta^2)$. With only linear operations involved, the complexity of the fusion step is $O(NK)$. Therefore, the total complexity of our algorithm is $O(NK)$.

Our algorithm is currently implemented in Matlab. The computational times for the 12 image sequences are reported in Table I, along with the comparison with EF [11]. Both our and EF's Matlab implementations were executed on the same computer with a 2.53-GHz CPU and 2-GB memory available for Matlab. Times for reading the input sequences and writing the output images are excluded. The times of EF on the National Cathedral and the Lizard sequences are not given, because its Matlab implementation requires more than 2-GB memory for the computation of those sequences. Our algorithm takes only 25% of the total time of EF on the average.

C. Analysis of Free Parameters

The effectiveness of acceleration is illustrated in Fig. 8(a). All the 12 image sequences in Table I were used in this and following analyses. For illustration, we plot three representative image sequences at different scales (i.e., Igloo, Memorial Church, and Belgium House) in the graph. We fix $\bar{\sigma}_w = 0.1$,

⁵CHOLMOD has been included in Matlab since version 7.2.

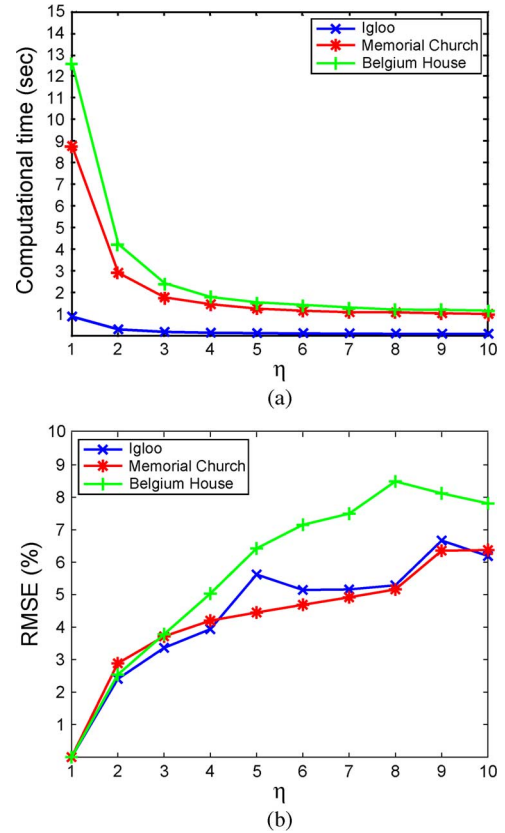


Fig. 8. Analysis of acceleration with different block width η . Error is defined as the relative difference from the results generated with $\eta = 1$. Significant efficiency improvement is only observed when $\eta \leq 5$ and significant error increase only occurs when $\eta \leq 5$. (a) Effectiveness of acceleration. (b) Error introduced.

$\gamma = 1.0$ in this analysis. The horizontal axis represents the block width η and the vertical axis represents the computational time in seconds. Time used in initialization is excluded because the

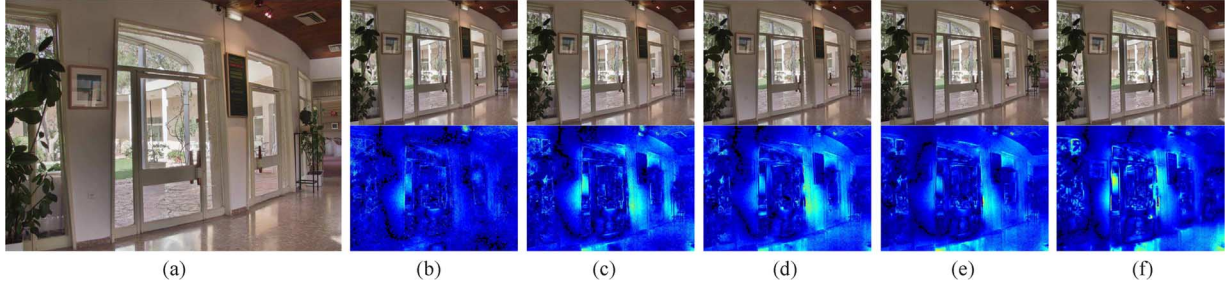


Fig. 9. Errors introduced by acceleration on the Belgium House image sequence. The result generated with $\eta = 1$ is used as the reference image. The error increases when η increases, as shown in the color-coded error maps (warmer colors indicate larger errors). However, the total error is still below 9% even when $\eta = 10$. (a) $\eta = 1$. (b) $\eta = 2$. (c) $\eta = 4$. (d) $\eta = 5$. (e) $\eta = 6$. (f) $\eta = 10$.

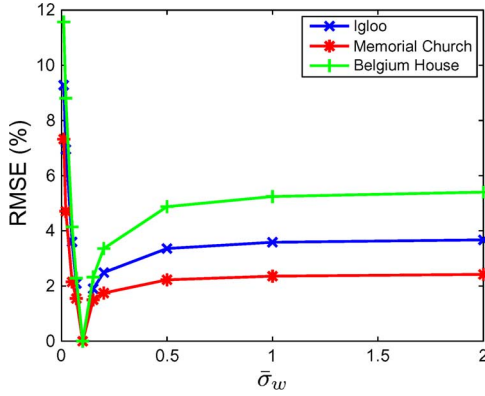


Fig. 10. Sensitivity analysis of the free parameter $\bar{\sigma}_w$. The error increases dramatically as $\bar{\sigma}_w$ becomes too small. It increases slowly when $\bar{\sigma}_w \geq 0.4$ and converges after $\bar{\sigma}_w = 1.0$.

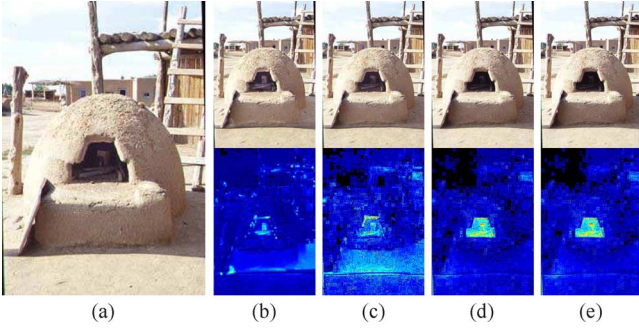


Fig. 11. Sensitivity analysis of the free parameter $\bar{\sigma}_w$ on the Igloo image sequence. The result generated with $\bar{\sigma}_w = 0.1$ is used as the reference image. When $\bar{\sigma}_w$ decreases, the image gets brighter, and therefore more information is revealed in the under-exposed regions. However, if $\bar{\sigma}_w$ becomes too small, adjacent pixels with large difference may be ignored, which leads to artifacts at object boundaries. (a) $\bar{\sigma}_w = 0.1$. (b) $\bar{\sigma}_w = 0.01$. (c) $\bar{\sigma}_w = 0.05$. (d) $\bar{\sigma}_w = 0.5$. (e) $\bar{\sigma}_w = 2.0$. (Input sequence courtesy of Shree Nayar.)

acceleration does not affect this step. Computational time decreases as η increases. However, significant efficiency improvement is only observed when $\eta \leq 5$.

Errors introduced by acceleration are shown against block width η in Fig. 8(b). The error in a pixel p_i^* is calculated using Euclidean distance between a pixel in a resulting image with $\eta > 1$ and the corresponding pixel in the reference image with $\eta = 1$, i.e., $E(p_i^*) = \|p_i^* - p_i^{ref}\|$. The pixel values are normalized between $[0, 1]$. The total error in a resulting image is measured using the root mean squared error (RMSE), i.e., $E(\mathbf{I}^*) = \sqrt{(1/N) \sum_i E(p_i^*)^2}$. The error increases as the block size η increases. However, significant error increase only occurs when

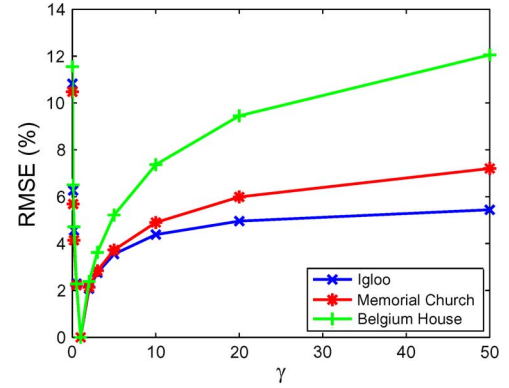


Fig. 12. Sensitivity analysis of the free parameter γ . The error increases dramatically as γ becomes too small, but increases slowly when $\gamma \geq 5.0$.

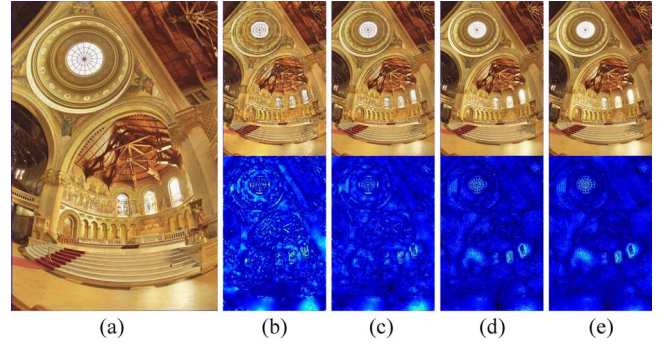


Fig. 13. Sensitivity analysis of the free parameter γ on the Memorial Church image sequence. The result generated with $\gamma = 1.0$ is used as the reference image. When γ decreases, more detail is revealed in the over-exposed regions. However, if γ is too small, artifacts may occur at object boundaries. (a) $\gamma = 1.0$. (b) $\gamma = 0.01$. (c) $\gamma = 0.2$. (d) $\gamma = 5.0$. (e) $\gamma = 10.0$. (Input sequence courtesy of Paul Debevec.)

$\eta \leq 5$. Even when $\eta = 10$, the total error is still below 9%. Some results on the Belgium House image sequence with different η 's and their corresponding color-coded error maps are shown in Fig. 9. In an error map, warmer colors indicate larger errors. The error increases as η increases. In order to balance the speed and error, we tested different values and found that using a block width $\eta \in [2, 5]$ generates reasonably good results.

In the analysis of $\bar{\sigma}_w$, we fix $\gamma = 1.0$, $\eta = 4$. The error is defined similarly as in the analysis of η and the results with $\bar{\sigma}_w = 0.1$ are used as reference images. The analysis on three representative image sequences is plotted in Fig. 10. Some results on the Igloo sequence with different $\bar{\sigma}_w$'s and their corresponding color-coded error maps are shown in Fig. 11. The error increases dramatically as $\bar{\sigma}_w$ becomes too small. It increases

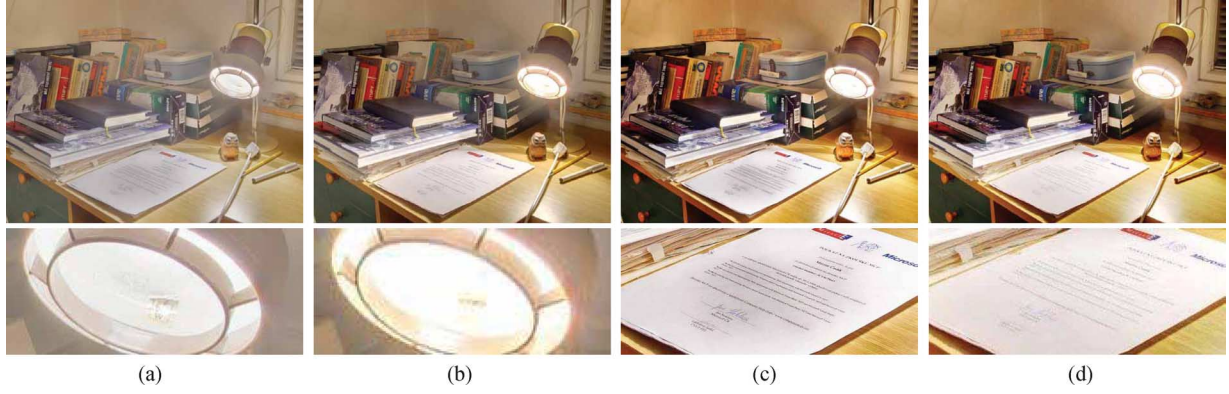


Fig. 14. Comparison of our algorithm with EF [11] using the Lamp image sequence. The results of EF are generated on a smaller scale (1/4) of the input sequence. Although the result of EF looks more colorful in some regions than our result, it preserves less detail. Especially for the lamp, the bulb and the tag beside it are clearly visible in our result but washed out in the result of EF. Although histogram equalization causes our result to lose some information in the lamp region, our result still preserves more detail than the result of EF, e.g., in the paper region. (a) Proposed before histogram equalization. (b) EF before histogram equalization. (c) Proposed after histogram equalization. (d) EF after histogram equalization. (Input sequence courtesy of Martin Cadík.)

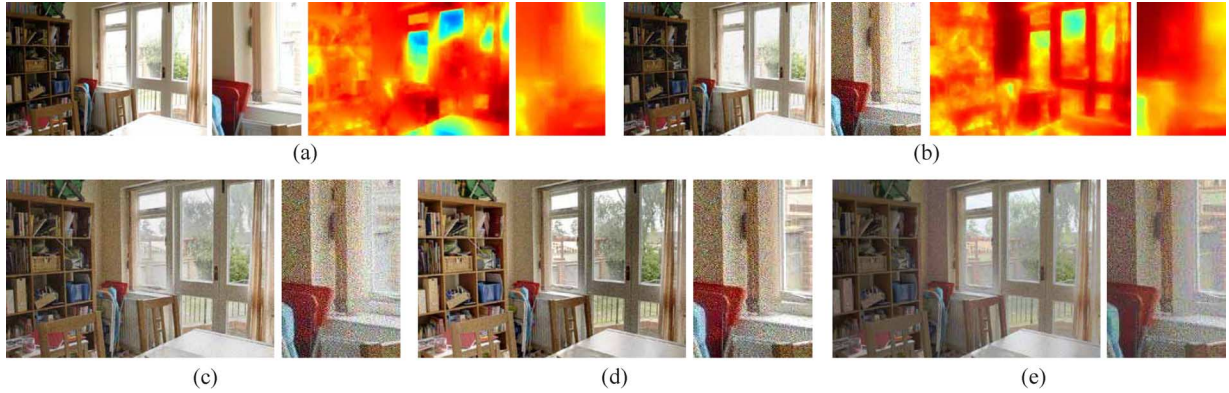


Fig. 15. Analysis of our algorithm's sensitivity to Gaussian white noise and comparison with other algorithms using the House image sequence. A close-up view of regions near the window is placed beside each image to give a clearer view of the effect of noise. As the Gaussian noise variance increases, errors introduced by the noise become more obvious. Our algorithm, EF, and PTR are all affected by the noise in the input image. (a) Original image (left) and its corresponding color-coded probability map (right) calculated by the proposed method (warmer colors represent higher probabilities). (b) Gaussian noise-corrupted image with 0 mean and 0.01 variance (left) and its corresponding color-coded probability map (right) calculated by the proposed method. (c) Fused image generated by the proposed method. (d) Fused image generated by EF [11]. (e) Tone-mapped image generated by PTR [9].

slowly when $\bar{\sigma}_w \geq 0.4$ and converges after $\bar{\sigma}_w = 1.0$. When $\bar{\sigma}_w$ decreases, the image gets brighter, and therefore more information is revealed in the under-exposed regions. However, if $\bar{\sigma}_w$ becomes too small, adjacent pixels with large difference may be ignored, which leads to artifacts at object boundaries. Therefore, we suggest using $\sigma_w \in [0.05, 0.4]$.

In the analysis of γ , we fix $\bar{\sigma}_w = 0.1$, $\eta = 4$. The error is defined similarly as in the analysis of η and the results with $\gamma = 1.0$ are used as reference images. The analysis on three representative image sequences is plotted in Fig. 12. Some results on the Memorial Church sequence with different γ 's and their corresponding color-coded error maps are shown in Fig. 13. The error increases dramatically as γ becomes too small, but increases slowly when $\gamma \geq 5.0$. When γ decreases, more detail is revealed in the over-exposed regions. However, if γ is too small, like in the analysis of $\bar{\sigma}_w$, artifacts may occur at object boundaries. Therefore, we suggest using $\gamma \in [0.2, 5]$.

D. Post-Processing for Further Enhancement

Although our algorithm preserves more detail than EF, the results of EF for some image sequences may have higher contrast (sharper) in certain regions. An example is given in

Fig. 14(a) and (b). The tag near the bulb is clearly visible in our result but washed out in the result of EF. The region near the red book looks more colorful in the result of EF. Higher contrast can also be obtained from our method by applying histogram equalization as a post-processing step. In order to provide a fair comparison, we added histogram equalization to both methods. The images in Fig. 14(c) and (d) illustrate that the contrast in our result is improved while EF suffers a loss of detail during the process. Although histogram equalization causes our result to lose some information in the lamp region, our result still preserves more detail than the result of EF, e.g., in the paper region.

E. Effect of Noise

One limitation of our algorithm is that it is sensitive to noise in the input image sequence. One example is given in Fig. 15. White Gaussian noise with 0 mean and variance from 0.001 to 0.01 with increments of 0.001 is added to one of the four input images (pixel values are scaled to the range [0,1]). For brevity, only the corrupted image with variance 0.01 is shown in Fig. 15, along with the original image. In our algorithm, the initial probabilities are determined by local contrast, and this local mea-

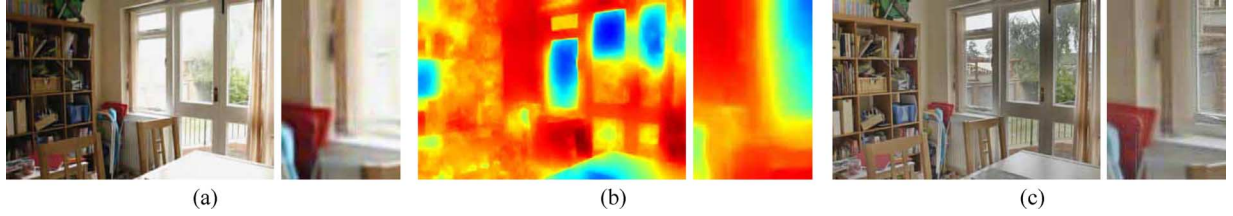


Fig. 16. Fusion result using our algorithm after adding a denoising step. The probability map calculated for this denoised image and the fused image are close to the ones computed from the clean input images. However, the object boundaries are a little blurred during the denoising process, which also affects the fusion result. (a) Denoised image. (b) Color-coded probability map. (c) Fused image.

sure is sensitive to white Gaussian noise because it is calculated by local variance. Since a global optimization scheme is used afterwards, the error caused by the noise tends to propagate to a larger neighborhood. The color-coded probability maps are given in Fig. 15(a) and (b), where warmer colors represent higher probabilities. Compared to the probability map generated for the original image, higher probabilities are assigned to pixels in the corrupted image, especially in the textureless regions like the walls. Even if we use the correctly computed probability maps, pixels from the noisy images still contribute to the fused image through the use of (1). Therefore, the noisy images have significant influence on the fused images, as shown in Fig. 15(c). A close-up view of regions near the window is placed beside each image to give a clearer view of the effect of noise. As the Gaussian noise variance increases, errors introduced by the noise become more obvious. However, the results by EF are also affected by the noise as shown in Fig. 15(d). In addition, the noise in an input image also affects the HDR reconstruction process, which results in a noisy HDR image. The tone mapping process fails to correct the noise in this HDR image and the noise remains in the tone-mapped image. PTR [9], SC [29], and LW [30] generate similarly affected results. Therefore, only the result of PTR is given in Fig. 15(e).

One solution to this problem is to add a pre-processing step to reduce the noise in the input images. One example is given in Fig. 16. The input sequence is the House image sequence, where one image is corrupted with white Gaussian noise with 0 mean and 0.01 variance as shown in Fig. 15(b). Fig. 16(a) gives the denoised image using the method proposed by Portilla *et al.* [36]. The noisy image is first decomposed into subbands using the steerable pyramid transform [37], and then Bayesian least square estimation and Gaussian scale mixture model are employed to denoise each subband. The probability map calculated for this denoised image and the fused image are given in Fig. 16(b) and (c), respectively. They are close to the ones computed from the clean input images [see Figs. 15(a) and 6(b)]. However, the object boundaries are a little blurred during the denoising process, which also affects the fusion result. In the future, we will explore the possibility of incorporating a noise term into our fusion model to make our algorithm more robust to noise.

F. Analysis of Different Forms of Compatibility Function for Contrast Measure

The compatibility function used for contrast measure can take different forms. One alternative is to use the contrast transducer

function proposed by Wilson [34], which is a sigmoid-shaped function amplifying difference at low contrasts and suppressing difference at high contrasts. Wilson's transducer function takes the following form for contrasts below or near the contrast threshold where $|c|S = 1$:

$$\mu(c) = \frac{1}{k} \left(1 + (|c|S)^Q \right)^{1/3} - \frac{1}{k} \quad (12)$$

where c represents local contrast; $k = 0.2599$ is a parameter obtained by setting the contrast detection threshold to 0.75; $Q = 3.0$ is an empirically determined parameter; S is called the contrast sensitivity and in our experiment we set $S = 1/0.75C_{max}$, where C_{max} represents the maximum magnitude of the local contrast detected from the input sequence. Because of this setting of S , all contrasts are below or near the threshold. Therefore, (12) is used in our experiment instead of the unified transducer function in [34] that combines (12) and a function for high suprathreshold contrasts. In our experiment, the local contrast c_i^k at the i th pixel in image \mathbf{I}^k is calculated based on Weber contrast

$$c_i^k = \frac{\Delta L_i^k}{L_b} \quad (13)$$

where we take ΔL_i^k as the luminance difference between the pixel p_i^k and the average pixel of its 3×3 neighborhood and L_b as the average luminance of the input sequence. L_b can also be taken as the local average luminance, but this may make c_i^k biased towards under-exposed pixels.

Fig. 17 gives a comparison between the results obtained by our current compatibility function [see (10)] and by Wilson's transducer function coupled with Weber contrast [see (12) and (13)] using the Memorial Church image sequence. With all other parameter settings the same, i.e., $\bar{\sigma}_w = 0.1$, $\gamma = 1.0$, and $\eta = 4$, Wilson's transducer function with Weber contrast preserves more detail in the window regions, but produces contrast reversals at some places near the windows. These contrast reversals are caused by combining pixels from input images with large exposure gaps. When γ is increased to give more emphasis on color consistency, the contrast reversals disappear although there is some loss of detail, as shown in Fig. 17(c). However, the brightness of the entire fused image also decreases. The current form of the compatibility function gives a better balance between local contrast and color consistency. We will analyze compatibility functions of other forms (e.g., [38]) and other quality measures (e.g., [39]) that may enhance our algorithm's performance in the future.

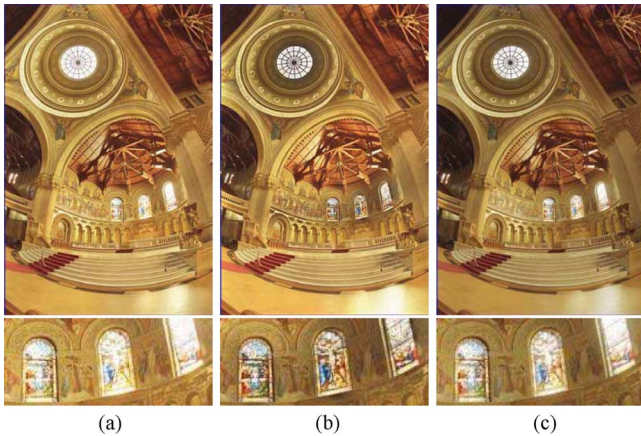


Fig. 17. Comparison between the results obtained from our current compatibility function [Equation (10)] and from Wilson's transducer function with Weber contrast [Equations (12) and (13)] using the Memorial Church image sequence. With all other parameter settings the same, Wilson's transducer function with Weber contrast preserves more detail in the window regions, but produces contrast reversals at some places near the windows. When γ is increased to 100 to give more emphasis on color consistency, the contrast reversals disappear although with some loss of detail. However, the brightness of the entire fused image also decreases. (a) Equation (10) with $\bar{\sigma}_w = 0.1$, $\gamma = 1.0$, and $\eta = 4$. (b) Equations (12) and (13) with $\bar{\sigma}_w = 0.1$, $\gamma = 1.0$, and $\eta = 4$. (c) Equations (12) and (13) with $\bar{\sigma}_w = 0.1$, $\gamma = 100$, and $\eta = 4$.

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a new fusion algorithm for multi-exposure images considering fusion as a probabilistic composition process. A generalized random walks framework was proposed to compute the probabilities. Two quality measures were considered: local contrast and color consistency. Unlike previous fusion methods, our algorithm achieves an optimal balance between the two measures via a global optimization. Experimental results demonstrated that our probabilistic fusion produces good results, in which contrast is enhanced and details are preserved with high computational efficiency. Compared to other fusion methods and tone mapping methods, our algorithm produces images with comparable or even better qualities. In future work, we will explore more effective quality measures and the possibility of incorporating multi-resolution technique in the fusion process to further enhance our technique for different fusion problems. We will also explore the possibility of applying the generalized random walks framework to other image processing problems.

ACKNOWLEDGMENT

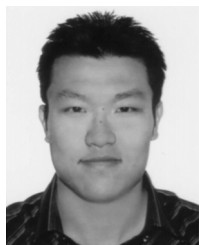
The authors would like to thank L. Grady for providing the source code of the original random walks algorithm. The authors would also like to thank the reviewers for their valuable comments.

REFERENCES

- [1] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2000, vol. 1, pp. 472–479.
- [2] H. Mannami, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, "High dynamic range camera using reflective liquid crystal," in *Proc. Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

- [3] J. Tumblin, A. Agrawal, and R. Raskar, "Why i want a gradient camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 103–110.
- [4] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," in *Proc. ACM SIGGRAPH*, 2004, pp. 760–768.
- [5] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH*, 1997, pp. 369–378.
- [6] M. Aggarwal and N. Ahuja, "High dynamic range panoramic imaging," in *Proc. Int. Conf. Comput. Vis.*, 2001, pp. 2–9.
- [7] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*, 2nd ed. Waltham, MA: Morgan Kaufmann, 2010.
- [8] G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Lightness perception in tone reproduction for high dynamic range images," *Comput. Graph. Forum*, vol. 24, no. 3, pp. 635–645, 2005.
- [9] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *Proc. ACM SIGGRAPH*, 2002, pp. 267–276.
- [10] A. Goshtasby, "Fusion of multi-exposure images," *Image Vision Comput.*, vol. 23, no. 6, pp. 611–618, 2005.
- [11] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion," in *Proc. Pacific Graphics*, 2007, pp. 382–390.
- [12] M. Kumar and S. Dass, "A total variation-based algorithm for pixel-level image fusion," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2137–2143, Sep. 2009.
- [13] S. Zheng, W.-Z. Shi, J. Liu, G.-X. Zhu, and J.-W. Tian, "Multisource image fusion method using support value transform," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1831–1839, Jul. 2007.
- [14] S. Li, J. T.-Y. Kwok, I. W.-H. Tsang, and Y. Wang, "Fusing images with different focuses using support vector machines," *IEEE Trans. Neural Netw.*, vol. 15, no. 6, pp. 1555–1561, Nov. 2004.
- [15] H. Zhao, Q. Li, and H. Feng, "Multi-focus color image fusion in the HSI space using the sum-modified-laplacian and a coarse edge map," *Image Vis. Comput.*, vol. 26, no. 9, pp. 1285–1295, 2008.
- [16] L. Bogoni and M. Hansen, "Pattern-selective color image fusion," *Pattern Recognit.*, vol. 34, no. 8, pp. 1515–1526, 2001.
- [17] G. Piella, "Image fusion for enhanced visualization: A variational approach," *Int. J. Comput. Vis.*, vol. 83, no. 1, pp. 1–11, 2009.
- [18] V. S. Petrovic and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 228–237, Feb. 2004.
- [19] S. Z. Li, "Markov random field models in computer vision," in *Proc. Eur. Conf. Comput. Vis.*, 1994, pp. 361–370.
- [20] M. I. Smith and J. P. Heather, "A review of image fusion technology in 2005," in *Proc. SPIE*, 2005, vol. 5782, pp. 29–45.
- [21] I. Cheng and A. Basu, "Contrast enhancement from multiple panoramic images," in *Proc. ICCV Workshop OMNIVIS*, 2007, pp. 1–7.
- [22] W.-H. Cho and K.-S. Hong, "Extending dynamic range of two color images under different exposures," in *Proc. Int. Conf. Pattern Recognit.*, 2004, vol. 4, pp. 853–856.
- [23] S. Raman and S. Chaudhuri, "A matte-less, variational approach to automatic scene compositing," in *Proc. Int. Conf. Comput. Vis.*, 2007, pp. 1–6.
- [24] S. Raman and S. Chaudhuri, "Bilateral filter based compositing for variable exposure photography," in *Proc. Eurographics Short Papers*, 2009.
- [25] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. P. A. Lensch, "Optimal HDR reconstruction with linear digital cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 215–222.
- [26] M. Cadik, M. Wimmer, L. Neumann, and A. Artusi, "Evaluation of HDR tone mapping methods using essential perceptual attributes," *Comput. Graph.*, vol. 32, no. 3, pp. 330–349, 2008.
- [27] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Vis. Comput. Graph.*, vol. 3, no. 4, pp. 291–306, 1997.
- [28] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Comput. Graph. Forum*, vol. 22, no. 3, pp. 419–426, 2003.
- [29] Y. Li, L. Sharan, and E. H. Adelson, "Compressing and companding high dynamic range images with subband architectures," in *Proc. ACM SIGGRAPH*, 2005, pp. 836–844.

- [30] Q. Shan, J. Jia, and M. S. Brown, "Globally optimized linear windowed tone mapping," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 4, pp. 663–675, Jul.–Aug. 2010.
- [31] L. Grady, "Multilabel random walker image segmentation using prior models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 763–770.
- [32] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
- [33] P. G. Doyle and J. L. Snell, *Random Walks and Electric Networks*. Washington DC: MAA, 1984.
- [34] H. R. Wilson, "A transducer function for threshold and suprathreshold human vision," *Biolog. Cybern.*, vol. 38, no. 3, pp. 171–178, 1980.
- [35] Y. Chen, T. A. Davis, W. W. Hager, and S. Rajamanickam, "Algorithm 887: CHOLMOD, supernodal sparse Cholesky factorization and update/downdate," *ACM Trans. Math. Softw.*, vol. 35, no. 3, pp. 1–14, 2008.
- [36] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [37] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. Int. Conf. Image Process.*, 1995, vol. 3, pp. 444–447.
- [38] M. A. García-Pérez and R. Alcalá-Quintana, "The transducer model for contrast detection and discrimination: Formal relations, implications, and an empirical test," *Spatial Vis.*, vol. 20, no. 1–2, pp. 5–43, 2007.
- [39] S. Winkler, "Vision models and quality metrics for image processing applications," Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2000.



Rui Shen (S'07) received the B.Eng. degree in computer science and technology from Beihang University, Beijing, China, and the M.S. degree in computing science from the University of Alberta, Edmonton, AB, Canada. He is currently working toward the Ph.D. degree in computing science at the University of Alberta.

He is the author or coauthor of eight papers published in international conferences and journals. His current research interests include probability models, sensor fusion, stereo vision, and image processing.

Mr. Shen is the recipient of the iCORE Graduate Student Scholarship in the Information and Communications Technology and the Izaak Walton Killam Memorial Scholarship.



Irene Cheng (M'02–SM'09) received the Ph.D. degree in computing science from the University of Alberta, Edmonton, AB, Canada.

She is now the Scientific Director of the Multimedia Research Group and an Adjunct faculty in the Department of Computing Science, University of Alberta. She also holds an Adjunct position in the Faculty of Medicine and Dentistry. She has two books and more than 100 papers published in international journals and peer-reviewed conferences. Her research interests include incorporating

human perception—Just-Noticeable-Difference—following psychophysical methodology, to improve multimedia transmission techniques. She is also engaged in research on 3-D TV and perceptually motivated technologies in multimedia, including online education.

Dr. Cheng is the Chair of the IEEE NCS EMBS Chapter, the Chair of the Communications Society MMTC 3DRPC Interest Group, and a board member of the SMC TC on Human Perception in Vision, Graphics and Multimedia. She is a General Chair in IEEE ICME 2011 and is awarded visiting professorship at INSA Lyon in 2011.



Jianbo Shi received the B.A. degree in computer science and mathematics from Cornell University, Ithaca, NY, in 1994 and the Ph.D. degree in computer science from the University of California at Berkeley in 1998 with a thesis on "Normalize Cuts image segmentation algorithm."

He joined the Robotics Institute at Carnegie Mellon University, Pittsburgh, PA, in 1999 as a research faculty, where he led the Human Identification at Distance (HumanID) project, developing vision techniques for human identification and activity inference. In January 2003, he joined the Department of Computer and Information Science at the University of Pennsylvania, where he is currently an Associate Professor. His current research focus is on object recognition-segmentation, mid-level shape representation, and human behavior analysis in video.

Prof. Shi received a U.S. National Science Foundation CAREER award on learning to see—a unified segmentation and recognition approach, in 2004.



Anup Basu (M'90–SM'02) received the Ph.D. degree in computer science from the University of Maryland, College Park.

He was a Visiting Professor at the University of California, Riverside, a Guest Professor at the Technical University of Austria, Graz, and the Director at the Hewlett-Packard Imaging Systems Instructional Laboratory, University of Alberta, Edmonton, Canada, where, since 1999, he has been a Professor at the Department of Computing Science, and is currently an iCORE-NSERC Industry Research Chair. He originated the use of foveation for image, video, stereo, and graphics communication in the early 1990s, an approach that is now widely used in industrial standards. He also developed the first robust (correspondence free) 3-D motion estimation algorithm, using multiple cameras, a robust (and the first correspondence free) active camera calibration method, a single camera panoramic stereo, and several new approaches merging foveation and stereo with application to 3-D TV visualization and better depth estimation. His current research interests include 3-D/4-D image processing and visualization especially for medical applications, multimedia in education and games, and wireless 3-D multimedia transmission.