QoE-Based Multi-Exposure Fusion in Hierarchical Multivariate Gaussian CRF

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

Abstract

Many state-of-the-art fusion methods, combining details in images taken under different exposures into one well-exposed image, can be found in the literature. However, insufficient study has been conducted to explore how perceptual factors can provide viewers better Quality of Experience (QoE) on fused images. We propose two perceptual quality measures: perceived local contrast and color saturation, which are embedded in our novel hierarchical multivariate Gaussian Conditional Random Field (CRF) model, to illustrate improved performance for multi-exposure fusion. We show that our method generates images with better quality than existing methods for a variety of scenes.

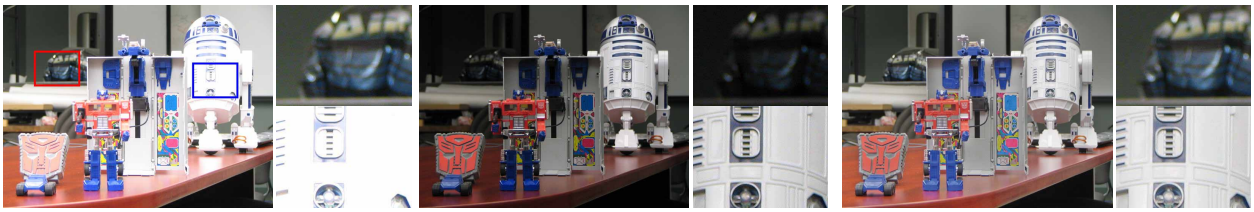
Index Terms

Multi-exposure fusion, human perception, QoE, conditional random field, MAP estimation

I. Introduction

Human perceptual factors have attracted increasing attention in research on visual communication techniques [1], [2]. The rationale behind this trend is to appeal to human observers with high visual quality images, videos, and graphics. Thus, it is important for applications to take the human visual system into consideration when designing image processing algorithms. Contrast and color are generally recognized to be important parameters [3], [4] in image quality. Motivated by these research findings, we study the visual impact of perceived local contrast and color saturation on fused images.

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



(a) (b)

Fig. 1. Details from multi-exposure images in the source sequence are combined into a single image. Because of the high dynamic range of the scene, a single image suffers from either over-exposure or under-exposure in some regions (*e.g.,* in the insets) and fails to present all the details. Using our fusion technique, details from different regions in each source image are transferred into one well-exposed image. (a) Source images taken under different exposures; (b) Fused image.

Multi-Exposure Fusion (MEF) is necessary because a conventional digital camera often produces images with insufficient details on a natural scene due to the incompatibility of its Low Dynamic Range (LDR) relative to the High Dynamic Range (HDR) of the scene. As shown in Figure 1(a), neither of the source images captured under different exposures is able to present all of the details in the scene, although the individual images combined contain complementary high-quality details of the scene like in Figure 1(b). This composition of local details can be achieved using MEF techniques [5], [6] or HDR imaging techniques [7]. In the HDR imaging approach, a radiance map of the scene is constructed, which allows a physical interpretation of the pixel values, but the process needs to adapt the reconstructed HDR image on consumer displays for viewing by applying Tone Mapping (TM) methods [8], [9]. In contrast, the MEF approach bypasses the HDR generation process and directly builds a visually appealing result based on certain perceptual criteria requiring minimal user intervention.

We propose a novel MEF method based on perceptual quality measures that exploit both contrast and color information. In order to deliver maximum image details, we model the probability for the human visual system to detect local contrast based on physiological findings. Incorporating these perceptual measures, the optimal fusion weights are then derived using Maximum *A Posteriori* (MAP) estimation in our Hierarchical Multivariate Gaussian Conditional Random Field (HMGCRF) model. The remainder of this paper is organized as follows. Section II presents the background of MEF. Section III explains our proposed Quality of Experience (QoE) based MAP-HMGCRF fusion method. Experimental results are summarized in Section IV. Conclusions and future work are given in Section V.

II. Background of Multi-Exposure Fusion

The history of image fusion research dates back to 1984 when Burt [10] proposed Laplacian pyramid-based fusion for binocular grayscale images. In 1993, Burt and Kolczynski [11] applied this method to fuse multi-exposure grayscale images. Mertens *et al.* [5] used local variation, saturation and well-exposedness based measures in a Laplacian pyramid-based fusion scheme for color images. Raman and Chaudhuri [12] estimated the fused pixel values by solving an optimization problem. Local variation and gradient based measures were used in [13] to infer the luminance components of the fused pixels in a Bayesian model. A gradient-directed MEF method [14] was proposed for dynamic scenes. In contrast, our focus is the study of perceptual factors on static scenes. In our previous work [6], a probabilistic fusion method was proposed, which applies local variation and neighborhood consistency computation. Although the technique generates high-quality results, we believe that images delivered by MEF techniques can be more visually appealing by considering human perceptual parameters. To address some common issues, which include loss of local details and poor color scheme leading to the loss of vividness, we propose using two perceptual quality measures (*i.e.,* perceived local contrast and color saturation) to give a more accurate evaluation of pixel contributions, in order to achieve higher-quality fused images. After validating the effectiveness of the perceptual measures using our previous model [6], we then propose a more flexible fusion model, where the fusion weights are computed as the MAP estimate in a hierarchical multivariate Gaussian CRF model to further illustrate the effectiveness of the perceptual parameters.

III. QoE-Based Multi-Exposure Fusion

A. Overview

Given a source image sequence, the contributions from individual pixels to the fused image are perceptually tuned by two locally-defined quality measures, *i.e.,* perceived local contrast (Section III-B) and color saturation (Section III-C). First, physical contrasts are calculated for each pixel in the luminance channel. Visual responses to these physical contrasts, *i.e.,* perceived contrasts, are then modeled using a transducer function followed by a psychometric function. These perceived contrasts, together with color saturation, are used in our MAP-HMGCRF model for pixel contribution evaluation, where the contributions are modeled as the MAP estimate of a multidimensional potential field (Section III-E).

B. Perceived Local Contrast

Earlier physiological studies of contrast sensitivity in three opponent color channels, *i.e.,* black-white (luminance), red-green, and yellow-blue, show that luminance sensitivity is normally higher than chro-

matic sensitivity [15], which inspires us to employ a luminance contrast measure to help preserve details. Local contrast represents the perception of local luminance variations with respect to the surrounding luminance, and different measures of local contrast exist in the literature. Simple definitions like Weber contrast normally assume small targets on a large uniform background [16]. In order to deal with complex images of natural scenes in MEF, we modified the local band-limited contrast proposed by Peli [16], which defines local contrast as the ratio between the band-pass filtered image and the low-pass filtered image. We perform contrast calculation in the luminance channel of the LHS color space.

If we directly use Peli's contrast in MEF, under-exposed regions, which are normally noisy, may produce stronger responses than well-exposed regions. This makes under-exposed regions contribute more to the fused image and reduce the overall brightness. Thus, if the local background luminance at a pixel is below a threshold θ, we weight its contrast by the background luminance to suppress noise. When θ is no less than 0.2, the fused image is brighter and preserves more details. When θ is above 0.4, the image shows less vivid colors. Therefore, we suggest using $\theta \in [0.2, 0.4]$.

When combined with the Gaussian pyramid representation of a luminance image, we can construct a contrast pyramid. Let $C^n_{i,k}$ denote the weighted contrast at the i-th pixel location at level n of the Gaussian pyramid, where $n \in [0, N_c - 1]$. Then, $C^n_{i,k}$ can be calculated as:

$$C^n_{i,k} = \begin{cases} G^n_{i,k} - [\phi * \mathbf{G}^n_k]_i, & [\phi * \mathbf{G}^n_k]_i < \theta; \\ (G^n_{i,k} - [\phi * \mathbf{G}^n_k]_i)/[\phi * \mathbf{G}^n_k]_i, & \text{otherwise.} \end{cases} \tag{1}$$

where $\mathbf{G}^n_k$ denotes the n-th level of the Gaussian pyramid and $G^n_{i,k}$ the i-th coefficient in $\mathbf{G}^n_k$; and we take ϕ as a 5×5 Gaussian filter with variance 1. Figure 2 gives a comparison between our weighted and Peli's contrasts. Peli's contrast produces noisy responses in the under-exposed regions of the low-exposure image, which reduces the brightness of the fused image. This issue is resolved using our weighted contrast.

Furthermore, in order to obtain the best representative information from lower levels, the contrast magnitude $\hat{C}^n_{i,k}$ at a higher-level coefficient is determined as the maximum contrast magnitude among those associated with that coefficient and its corresponding lower-level coefficients.

1) Transducer and Psychometric Functions: The nonlinearity of human perception of contrast has been studied by many researchers [17]–[19]. According to [17], contrast perception can be considered as a two-stage procedure. In the first stage, the stimulus contrast is mapped to the internal/physiological response of the sensory system via a transducer function μ. In the second stage, the probability of correctly discriminating a stimulus with certain contrast from the standard stimulus with a fixed contrast C_s is expressed by a psychometric function Ψ, where we take $C_s = 0$. A formal relationship between the psychometric function Ψ and the transducer function μ was developed in [18], where Ψ is determined

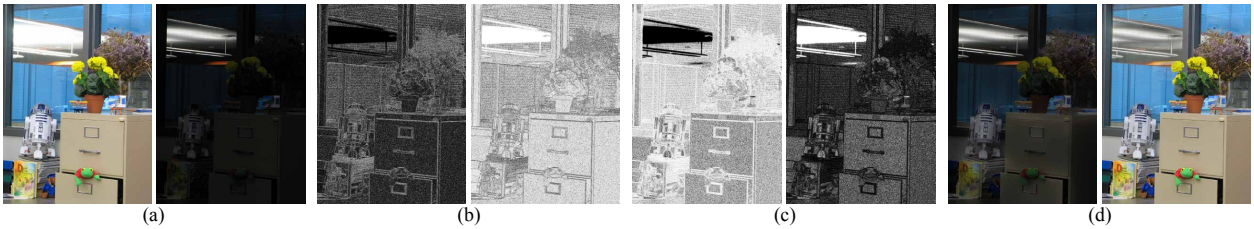


Fig. 2. Comparison of Peli's and our weighted contrasts for MEF. Peli's contrast produces noisy responses in the under-exposed regions of the low-exposure image, which reduces the brightness of the fused image. This issue is resolved using our weighted contrast. (a) High- and low-exposure source images; (b) Magnitude of Peli's contrast; (c) Magnitude of our weighted contrast; (d) Fusion results using Peli's (left) and our weighted (right) contrasts (both with transducer and psychometric functions applied).

by μ and the distribution of internal responses. We normalize $\hat{C}^n_{i,k}$'s to $[0, 1]$ before using it in μ to fulfill the assumption of the stimulus contrast range in [18].

Because the most representative information is passed upwards along the contrast pyramids, the transducer and psychometric functions are only applied to level $N_c - 1$, which also reduces the computational cost. We adopt the transducer function proposed in [17]:

$$\mu(\hat{C}^{N_c-1}_{i,k}) = \frac{(\hat{C}^{N_c-1}_{i,k} S_E)^p}{(\hat{C}^{N_c-1}_{i,k} S_I)^q + Z}, \tag{2}$$

where $S_E = 100$ is a constant; S_I, p, q, Z are four free parameters, which we set to the mean values reported from the experiments in [17], *i.e.,* $S_I = 75.70, p = 4.03, q = 3.59, Z = 24.87$. Two other forms of μ were also tested. The three-parameter function in [18] did not produce comparable results. Wilson's [19] transducer function for threshold and suprathreshold vision produced similar results. We adopted the psychometric function for contrast detection proposed in [18]:

$$\Psi(\hat{C}^{N_c-1}_{i,k}) = 1/(1 + \exp(-\mu(\hat{C}^{N_c-1}_{i,k})/b)), \tag{3}$$

where we take $b = \sqrt{6}/\pi$ as in [18]. In our implementation, $\Psi(\hat{C}^{N_c-1}_{i,k})$'s are normalized to $[0, 1]$.

C. Color Saturation

Since the perceived local contrast measure only works in the luminance channel, using it alone may not produce satisfactory results for color images in some cases where high local contrasts are achieved at the cost of low colorfulness/saturation. Objects captured at proper exposures normally exhibit more saturated colors. For instance, as shown in Figure 1, the red autobot symbol in one source image presents more saturated red than in the other. Therefore, we employ color saturation as another quality measure.

We incorporated the saturation definition in the LHS color space [4] to measure the colorfulness of a given pixel: $S = 1 - 3\min(R, G, B)/(R + G + B)$, where $S \in [0, 1]$ and R, G, B denote the red, green, and blue components in the RGB color space, respectively. Other saturation measures (*e.g.,* the saturation definition in the HSV space and Lübbe's definition in the CIELAB space [20]) produce similar results in our fusion scheme.

As in the case of creating the contrast pyramid, we can construct a saturation pyramid for each source image by first building a Gaussian pyramid and then calculating the saturation components at every level. To be consistent with the contrast pyramid, this saturation pyramid also has N_c levels. In practice, we observe that saturation calculation performed only at the highest pyramid level without information passing between levels is sufficient to produce satisfactory fusion results with no noticeable difference. This is because: 1) Gaussian smoothing has little influence on the objects/regions' color information when the filter's size and variance parameters are small (we use the same Gaussian filter parameters as for local contrast, *i.e.,* size 5×5 and variance 1); 2) Gaussian smoothing mainly affects object/region boundaries, where relatively large color and luminance changes occur after filtering, and these changes are captured by the perceived local contrast measure.

D. Perceptual Impact of the Proposed Quality Measures

In order to illustrate the quality contribution of the proposed perceived local contrast and color saturation measures, we incorporate these two measures in our previously published GRW model [6]. The results on one scene are presented in Figure 3. Instead of employing local variations in a non-linear function to indicate contrasts [6], we believe modeling the probability of the human visual system to perceive a given contrast will generate better perceptual quality, because this new modeling scheme offers a more accurate estimation of the amount of visual stimuli delivered from each image region, which leads to better detail preservation, as shown in the insets. Together with the color saturation measure, the fused image can exhibit more vivid colors (*e.g.,* for the sky and the street lamp).

E. MAP-HMGCRF Model

In order to effectively integrate the proposed perceptual measures, we introduce a new Multivariate Gaussian Conditional Random Field (MGCRF) model for lattice graphs (*e.g.,* a lattice of pixels), in which pixel contributions/fusion weights are evaluated as the MAP estimation. To improve computational efficiency and memory usage, we perform the computation in a hierarchical version of the MGCRF.

With local variation measure | With perceived local contrast measure | With both perceptual measures

Fig. 3. Incorporation of the proposed quality measures in MEF using the GRW model as an example. With the perceived contrast measure, more local details are preserved, as shown in the insets. With the saturation measure, the fused image exhibits more vivid colors, *e.g.,* for the sky and the street lamp. (Source sequence courtesy of HDRsoft.com.)

TABLE I

COMPARISON BETWEEN SOLVING AN HMGCRF AND DIRECTLY SOLVING AN MGCRF

Input	Size	Time (sec)		Memory (MB)		RMSE (%)
		HMGCRF	MGCRF	HMGCRF	MGCRF	
House	$752 \times 500 \times 4$	**1.176**	2.998	**43**	245	1.297
Chateau	$1500 \times 644 \times 5$	**3.995**	8.418	**152**	955	1.108
Belgium House	$1025 \times 769 \times 9$	**5.362**	7.449	**148**	773	0.800
Lamp	$1600 \times 1200 \times 15$	**23.08**	130.7	**757**	2386	0.904

1) MGCRF: Before introducing the hierarchical version, we first introduce a single-level model, the multivariate Gaussian conditional random field, combining the multivariate Gaussian Markov random field [21] and the Gaussian conditional random field [22]. Let $\mathbf{x} = (\mathbf{x}_1^T, \ldots, \mathbf{x}_N^T)^T$ denote a K-dimensional (K-D) potential field, where each $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,K})^T$ is a random vector that denotes a K-D potential, and let $\mathcal{D}$ denote the observed data. Let $\mathbf{x}_i$'s be arranged in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the i-th node in $\mathcal{V}$ represents $\mathbf{x}_i$ and each edge $e_{ij} \in \mathcal{E}$ represents the presence of interaction between its incident nodes $\mathbf{x}_i$ and $\mathbf{x}_j$. Then, $(\mathbf{x}, \mathcal{D})$ is called an MGCRF with respect to $\mathcal{G}$, if $\mathbf{x}$ given $\mathcal{D}$ satisfies the Markov property (with positivity implicitly assumed). Let $\mathbf{b} = (\mathbf{b}_1^T, \ldots, \mathbf{b}_K^T)^T$ define a K-D boundary for the potentials. Let us denote the maximum and minimum allowable potentials in the k-th dimension as $U_{\max,k}$ and $U_{\min,k}$, respectively. Then, $\mathbf{b}_k = (U_{\min,1}, \ldots, U_{\max,k}, \ldots, U_{\min,K})^T$. We assume that $\mathbf{x}$ is piecewise smooth and that the difference between any two neighboring nodes follows a multivariate Gaussian distribution with zero mean and precision matrix $\mathbf{S}_{ij}$: $\Delta_{ij}^x \triangleq \mathbf{x}_i - \mathbf{x}_j \sim N(\mathbf{0}, {\mathbf{S}_{ij}}^{-1})$. We also assume that the difference between a variable $\mathbf{x}_i$ and a boundary potential $\mathbf{b}_k$ follows a multivariate Gaussian distribution with zero mean and precision matrix $\mathbf{T}_{ik}$: $\Delta_{ik}^b \triangleq \mathbf{x}_i - \mathbf{b}_k \sim N(\mathbf{0}, {\mathbf{T}_{ik}}^{-1})$. Then,

the posterior density $p(\mathbf{x}|\mathbf{b},\mathcal{D})$ follows a multivariate Gaussian distribution:

$$\begin{aligned} p(\mathbf{x}|\mathbf{b},\mathcal{D}) &\propto \exp(-\frac{1}{2}(\sum_{k=1}^{K}\sum_{i=1}^{N}{\Delta_{ik}^{b}}^{T}\mathbf{T}_{ik}\Delta_{ik}^{b} + \\ &\quad \frac{1}{2}\sum_{e_{ij}\in\mathcal{E}}{\Delta_{ij}^{x}}^{T}\mathbf{S}_{ij}\Delta_{ij}^{x})) \\ &\propto \exp(-\frac{1}{2}\mathbf{x}^{T}\mathbf{P}\mathbf{x} - \mathbf{x}^{T}\mathbf{Q}\mathbf{b}). \end{aligned} \tag{4}$$

Here, $\mathbf{P}$ is an $NK \times NK$ matrix with the (i,j)-th $K \times K$ element defined as: $\mathbf{P}_{ij} = \sum_{k=1}^{K}\mathbf{T}_{ik} + \sum_{\mathbf{x}_m\in\mathcal{N}_i}\mathbf{S}_{im}$, if $i = j$, where $\mathcal{N}_i$ denotes the neighborhood around $\mathbf{x}_i$; $\mathbf{P}_{ij} = -\mathbf{S}_{ij}$, if $e_{ij} \in \mathcal{E}$; $\mathbf{P}_{ij} = 0$, otherwise. $\mathbf{Q}$ is an $NK \times KK$ matrix with the (i,k)-th $K \times K$ element defined as $\mathbf{Q}_{ik} = -\mathbf{T}_{ik}$. Then, the MAP estimate $\mathbf{f}^*$ of the posterior density $p(\mathbf{x}|\mathbf{b},\mathcal{D})$ is:

$$\begin{aligned} \mathbf{f}^* &= \arg\max_{\mathbf{f}} \exp(-\frac{1}{2}\mathbf{f}^{T}\mathbf{P}\mathbf{f} - \mathbf{f}^{T}\mathbf{Q}\mathbf{b}) \\ &= \arg\min_{\mathbf{f}} \frac{1}{2}\mathbf{f}^{T}\mathbf{P}\mathbf{f} + \mathbf{f}^{T}\mathbf{Q}\mathbf{b}. \end{aligned} \tag{5}$$

This minimization problem is equivalent to solving the linear system $\mathbf{P}\mathbf{f}^* = -\mathbf{Q}\mathbf{b}$.

2) MGCRF for MEF: We consider that the fused image is derived as a pixel-wise weighted composition of the source images:

$$\bar{p}_i = {\mathbf{u}_i}^{T}\mathbf{p}_i = (u_{i,1},\ldots,u_{i,K})(p_{i,1},\ldots,p_{i,K})^{T}, \tag{6}$$

where $\bar{p}_i$ and $p_{i,k}$ denote the i-th pixels in the fused image and the k-th source image, respectively; $u_{i,k}$ is a fusion weight that measures the contribution from pixel $p_{i,k}$; and $\mathbf{u}_i$ and $\mathbf{p}_i$ are a K-D weight vector and a K-D pixel vector, respectively. If $\bar{p}_i$'s obtained in Equation (6) exceed the dynamic range of the target device, they are truncated. If we model $\mathbf{u}_i$'s as the K-D potential field $\mathbf{x}$ on a lattice graph and $\mathbf{p}_i$'s the observed data $\mathcal{D}$, then estimating the fusion weights is equivalent to estimating the MAP configuration of an MGCRF. To fully specify the MGCRF, we also need to define the precision matrices and the potential boundary. We assume that the precision matrices $\mathbf{T}_{ik}$ and $\mathbf{S}_{ij}$ are identity matrices subject to individual scaling factors, *i.e.,* $\mathbf{T}_{ik} = \gamma_1 Y_{ik}\mathbf{I}, \mathbf{S}_{ij} = \gamma_2 W_{ij}\mathbf{I}$. Here, $\mathbf{I}$ represents a $K \times K$ identity matrix; γ_1 and γ_2 are data-independent scaling factors; and Y_{ik} and W_{ij} are data-dependent scaling factors defined as:

$$Y_{ik} = \Psi(\hat{C}_{i,k}^{N_c-1}) \cdot S_{i,k}^{N_c-1},\ W_{ij} = \prod_{k=1}^{K}\exp(-\frac{\|p_{i,k} - p_{j,k}\|}{\sigma}) \tag{7}$$

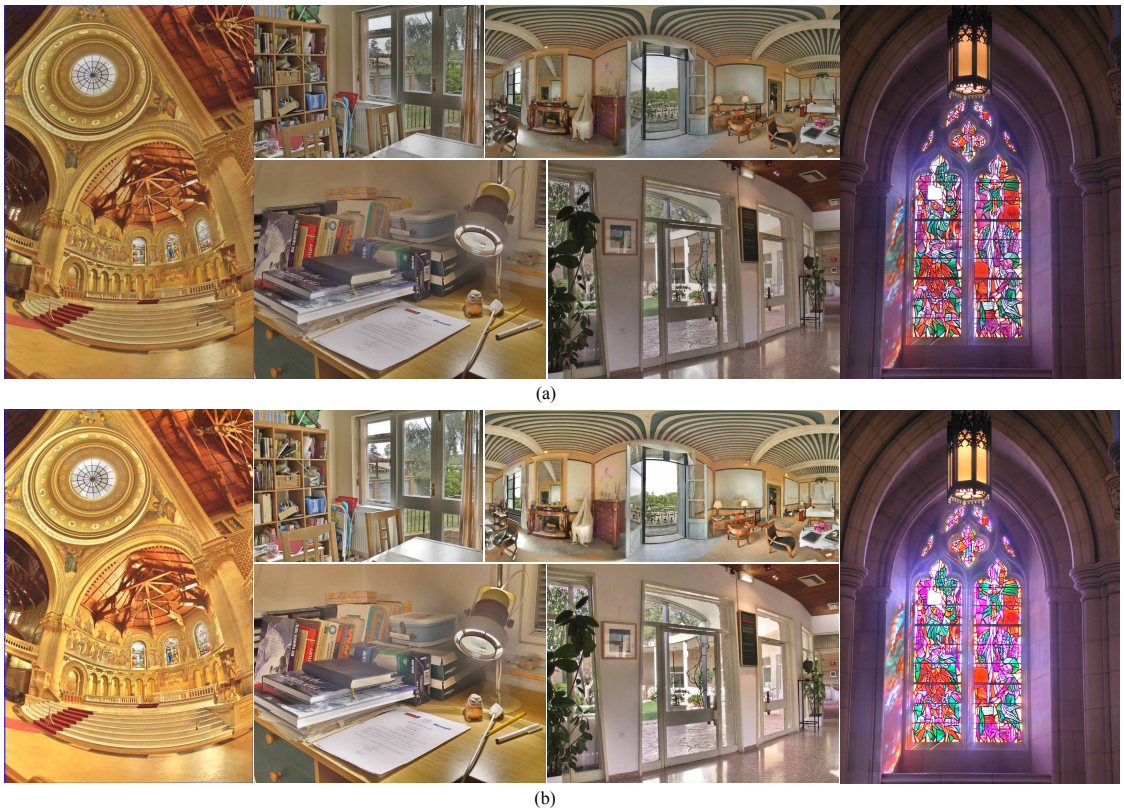


Fig. 4. Fusion results of QBF-1 and QBF-2 on the six standard test scenes. The scenes are: Memorial Church (left most), House (top left), Chateau (top right), Lamp (bottom left), Belgium House (bottom right), and National Cathedral (right most). (a) Results of QBF-1; (b) Results of QBF-2. (Source sequences courtesy of Paul Debevec, Tom Mertens, HDRsoft.com, Martin Čadík, Dani Lischinski and Max Lyons, respectively.)

where $\|\cdot\|$ denotes Euclidean distance; and σ is a free parameter. For the potential boundary, we assume that the maximum allowable potentials in each dimension are equal and so do the minimum allowable potentials, *i.e.*, $U_{\max,k} = \alpha_1, U_{\min,k} = \alpha_2, \forall k$. Setting $\alpha_2 = 0$ and with the identity precision matrix assumption, the MAP-MGCRF model degenerates to the GRW model in terms of steady-state probability calculation, but here we do not restrict the range of the boundary values.

3) HMGCRF: In order to efficiently estimate the MAP configuration on a lattice graph, we construct an N_h-level hierarchical MGCRF and perform the calculation in a coarse-to-fine fashion. A coarser-level lattice graph is obtained by downsampling the finer-level graph by a factor of 2 in each dimension. A precision matrix $\mathbf{T}_{sk}$ between variable/node $\mathbf{x}_s$ and boundary vector $\mathbf{b}_k$ at a coarser-level MGCRF is obtained as a weighted average of the precision matrices of variables in the second-order neighborhood of $\mathbf{x}_s$'s projection at the finer-level MGCRF. A precision matrix $\mathbf{S}_{st}$ between two adjacent variables $\mathbf{x}_s$ and

$\mathbf{x}_t$ at a coarser-level MGCRF is obtained as the precision matrix with the minimum determinant among those defined in the common neighborhood of the projections of $\mathbf{x}_s$ and $\mathbf{x}_t$ at the finer-level MGCRF. At the coarsest level, the MAP estimate is obtained using a direct linear system solver, and then the solution is interpolated downwards along the hierarchy to the finest level.

We evaluated the performance of this HMGCRF with $N_h = 5$ on four multi-exposure sequences of increasing size. Compared with directly solving an MGCRF, solving an HMGCRF requires less time and memory but with good solution accuracy, as shown in Table I. With all the other settings the same, the hierarchical version took 44.081% of the time and 21.085% of the memory needed by directly solving an MGCRF and produced an average root mean squared error of only 1.027%.

IV. Experimental Results

In this section, we summarize the evaluation results. Please refer to the supplementary material for more details and high-resolution images. Our method has eight free parameters, *i.e.,* $\theta, N_c, N_h, \sigma, \gamma_1, \gamma_2, \alpha_1, \alpha_2$. We take $\theta = 0.3$, $N_c = 2$, $\gamma_1 = 1$. Let a source sequence contain K $M \times N$ images. Then, we compute $\sigma = 0.1K$, $\gamma = 0.2K\sqrt{MN}/\kappa$, $N_h = \lceil \log_2(\min(M, N)/\kappa) \rceil$, where $\kappa = 32$ is the maximum number of nodes allowed along the shorter dimension of the coarsest-level lattice. Depending on the size of the source sequence, the value of N_h ranged between 4 and 6 in our experiments. We tested two sets of α_1 and α_2. In the first set, $\alpha_1 = 1, \alpha_2 = 0$, and we denote this algorithm as QBF-1. In the second set, we compute $\alpha_1 = 0.6 + \exp(-\bar{L}), \alpha_2 = 0.02\exp(-\bar{L})$, where $\bar{L}$ is the average luminance of $\mathcal{D}$, and we denote this algorithm as QBF-2. These parameter settings were used in all experiments.

Six standard test sequences were used. The fusion results of the proposed QBF-1 and QBF-2 were compared with two other MEF methods, *i.e.,* Probabilistic Fusion (PF) [6] and Exposure Fusion (EF) [5], which have previously demonstrated better performance than many other methods. In addition, two TM methods were also considered: the Photographic Tone Reproduction (PTR) local operator [8] and the iCAM06 operator [9], which have demonstrated good performance in various evaluations. The default parameter settings in PF, EF, and iCAM06 were used. The parameters in PTR were estimated by the method in [23]. The results by EF, PTR, and iCAM06, were generated by the programs provided by their respective authors. The HDR images for PTR and iCAM06 were generated using HDR reconstruction [7]. Both the objective and subjective evaluations were performed in a reference-free manner, where no ideal fused images were available to serve as references.

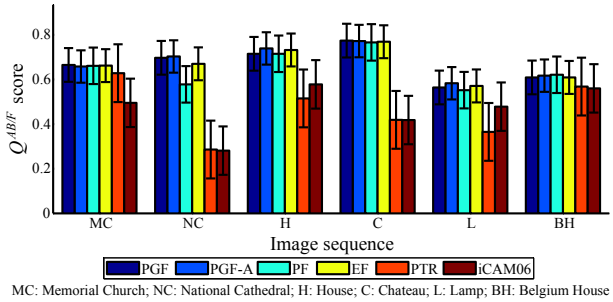


Fig. 5. Objective evaluation using $Q^{AB/F}$. All of the compared MEF methods successfully transferred most of the edge information, and they have very close performance according to $Q^{AB/F}$. On the average, QBF-2 has slightly better performance than the others.

A. *Objective Evaluation*

1) Evaluation Using $Q^{AB/F}$: Two objective evaluation metrics were employed to assess the fusion quality. The first one is the $Q^{AB/F}$ metric [24], which is widely used in the image fusion literature to measure correctly transferred edge information in the luminance channel from source images to a fused image. This metric gives a performance score between $[0, 1]$, where a higher score indicates better performance. Traditional fusion quality metrics, including $Q^{AB/F}$, were not designed for cases with more than two source images in MEF. Nevertheless, $Q^{AB/F}$ has been shown to be one of the most robust and consistent metrics [25]. Most evaluation metrics that are designed for the case of two source images, including the other two metrics recommended in [25] (*i.e.,* Cvejic's metric [26] and Yang's metric [27]), largely depend on the calculation and manipulation of covariance (or similar statistics) between the two source images and/or between the two source images and the fused image. Therefore, it is relatively difficult to extend such metrics to cases with multiple source images. The advantage of $Q^{AB/F}$ is that it does not rely on calculating statistical score between two source images, and thus it can be directly extended to processes involving multiple source images, such as MEF. This metric has also been proven to correspond best with subjective tests among several other popular metrics [28]. Therefore, we adapted $Q^{AB/F}$ in our evaluation. To the best of our knowledge, our evaluation is the first attempt of extending a traditional fusion quality metric to MEF.

The fusion results of QBF-1 and QBF-2 on the six test scenes, along with objective evaluation results using $Q^{AB/F}$, are shown in Figure 5. Although PTR and iCAM06 are not MEF methods, they are included in this evaluation for reference purposes only. All of the compared MEF methods successfully transferred most of the edge information, and they have very close performance according to $Q^{AB/F}$.

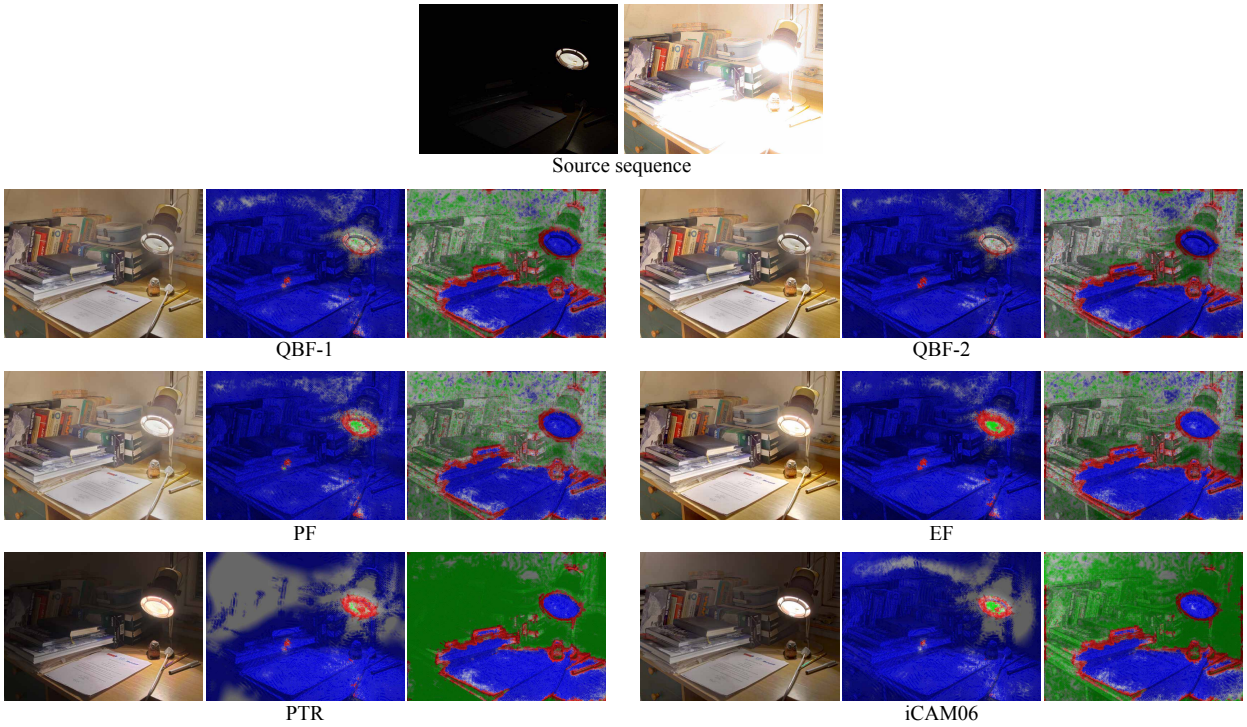


Fig. 6. Comparison of QBF-1 and QBF-2 with PF, EF, PTR, and iCAM06 on the Lamp sequence using DRIVDP. The two source images give good exposures for the bulb and the books, respectively. In a distortion map, green, blue, red, and gray pixels indicate contrast loss, amplification, reversal, and no distortion, respectively. QBF-1 and QBF-2 are more effective in preserving local details and color schemes than the others. For the bulb, QBF-2 shows the least distortion, followed by QBF-1, PTR, PF, EF, and iCAM06. For the books, QBF-2 shows the least distortion, followed by EF, PF, QBF-1, iCAM06, and PTR.

On the average, QBF-2 has slightly better performance than the others.

2) Evaluation Using DRIVDP: To strengthen the evaluation capability of $Q^{AB/F}$, we incorporate the Dynamic Range Independent Visible Difference Predictor (DRIVDP) [29] to assess per-pixel fusion quality. DRIVDP evaluates visual local contrast distortions (*i.e.,* loss of visible contrast, amplification of invisible contrast, and reversal of visible contrast) between images under a specific viewing condition and is widely used in the TM literature. Here, we use it to assess the visual distortions between a test image and each source image. We assume that the images were viewed on a typical LCD with a maximum luminance equivalent to $100cd/m^2$, a gamma value of 2.2, and a visual resolution of 30 pixels per degree at a viewing distance of 0.5 meter and that the peak contrast sensitivity of the viewer is 0.25%.

We chose two images from each of the six source sequences for this evaluation. The evaluation result on one sequence is given in Figure 6. The two source images with good exposures respectively for the

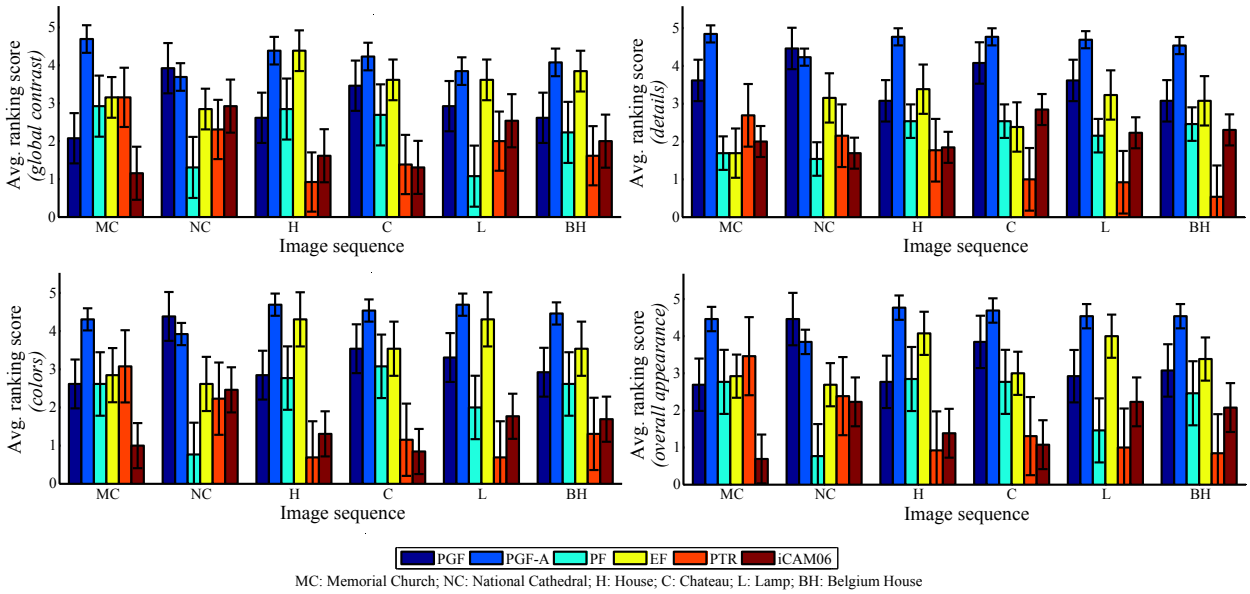


Fig. 7. Average ranking scores of different algorithms in the subjective evaluation. The *global contrast* criterion measures the global luminance variations. The *details* and the *colors* criteria measure the local details and colors reproduced and/or enhanced from the source images, respectively. The *overall appearance* criterion measures the overall impression of a fused or tone-mapped image. Our QBF-2 has the best performance under all four criteria for 5 out of 6 scenes, and shows similar performance to our QBF-1 on the other scene. QBF-1 and EF have similar performance on the average though QBF-1 gives better detail reproduction, followed by PF, iCAM06, and PTR.

bulb and the books are given in Figure 6(a). The distortion maps for each method are given in Figure 6(b)-(g), along with the fused images. In a distortion map, green, blue, red, and gray pixels indicate contrast loss, amplification, reversal, and no distortion, respectively. QBF-1 and QBF-2 are more effective in preserving local details and color schemes than the other methods. QBF-2 performs best in preventing contrast distortions for this sequence. Please note that contrast amplification is normally considered as one of the objectives in image fusion.

B. Subjective Evaluation

We also conducted a subjective evaluation, where thirteen subjects (8 males and 5 females) aged between 25 and 35 participated. All of the subjects had normal or corrected-to-normal vision and were non-experts in the field of MEF or TM. The test was performed under normal lighting conditions. For each scene, the results of different methods were anonymized and placed side by side in different orders, along with the source sequence. No other reference image, either manually or automatically

(a) (b)

Fig. 8. The results for the Lamp scene could be made more photo-realistic by increasing the weights for those source images with higher average luminance values. (a) QBF-1 with increased weights; (b) QBF-2 with increased weights.

fused or tone-mapped, was provided to guide/influence a subject's judgement. Subjects were asked to rank the results on a scale of 5 (best) to 0 (worst) in four categories: global contrast, details, colors, and overall appearance. These four criteria were also considered in [30]. The *global contrast* criterion measures the global luminance variations. The *details* and the *colors* criteria measure the local details and colors reproduced and/or enhanced from the source images, respectively. The *overall appearance* criterion measures the overall impression of a fused or tone-mapped image.

The average ranking scores of different algorithms under each criterion are reported in Figure 7. QBF-2 performs consistently well compared to the other methods. It has the best performance under all criteria on 5 out of 6 scenes, and shows similar performance to QBF-1 on the National Cathedral sequence. QBF-1 and EF have similar performance on the average though QBF-1 offers clearly better detail reproduction, followed by PF, iCAM06, and PTR. The evaluation results demonstrate that our proposed fusion method is capable of producing high-quality fused images with quality comparable, or even better quality in many cases, to tone-mapped HDR images and images by other fusion methods.

Even though photo-realistic appearance may not be always achieved due to detail maximization (*e.g.,* the appearance of shadows in the Lamp scene but with the bulb not sufficiently illuminated), for the criteria considered under the subjective evaluation, it does not really bring much negative experience to the audience. Since the perceptual quality standard is set to satisfy the average (majority) audience, the best strategy is to let an application to choose the desired parameter value. If desired by an application, a simple remedy for the Lamp scene could be to increase the weights for those source images with higher average luminance values, as shown in Figure 8.

TABLE II

COMPUTATIONAL SPEED (UNIT: SECOND)

Input	Size	QBF	PF	EF
House	$752 \times 500 \times 4$	1.504	0.539	1.706
Chateau	$1500 \times 644 \times 5$	4.837	1.665	5.378
Memorial Church	$512 \times 768 \times 16$	5.836	2.047	6.795
Belgium House	$1025 \times 769 \times 9$	6.838	2.256	7.639
National Cathedral	$1536 \times 2048 \times 2$	7.589	3.522	7.887
Lamp	$1600 \times 1200 \times 15$	28.373	7.621	30.325

C. Computational Speed

The computational speed of our QBF is proportional to the size of the source sequence. Its execution times on the six test sequences are compared with EF and PF in Table II, all of which are Matlab implementations. Times were recorded on a 2.53-GHz dual-core laptop with 4-GB memory. In the table, times for reading and writing images are excluded. QBF is a little faster than EF but slower than PF. Nevertheless, QBF produces much better fusion quality than PF as shown in the objective and subjective evaluations.

D. Discussion

Aside from $Q^{AB/F}$, we also investigated the applicability of two other traditional fusion quality metrics (Cvejic's metric and Yang's metric) in MEF. These two metrics require two source images, and produce a single global quality score for a fused image. Therefore, we first applied them to the National Cathedral sequence, which contains exactly two source images. Since such metrics estimate local structural similarity in the luminance channel between source images and the fused image, the details criterion in the subjective study provides some useful information on evaluating their performance. Their quality scores for the six compared algorithms are plotted against the average ranking scores under the details criterion of the subjective evaluation (normalized to $[0, 1]$) and the $Q^{AB/F}$ scores in Figure 9. From this plot, we can see that $Q^{AB/F}$ provides relatively better correspondence with the subjective study. We then applied Cvejic's metric, Yang's metric, and $Q^{AB/F}$ metric to the other five subsequences used in the DRIVDP-based evaluation. $Q^{AB/F}$ metric produced better correspondence with the details criterion in the subjective evaluation than the other two metrics.

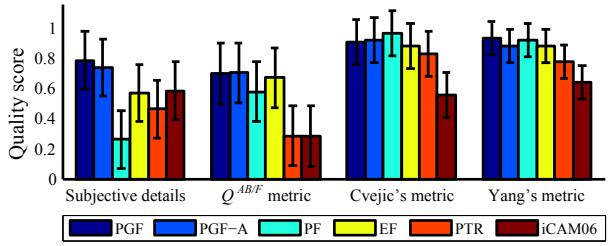


Fig. 9. Comparison of the performance of three fusion quality metrics ($Q^{AB/F}$ metric, Cvejic's metric, and Yang's metric) on the National Cathedral sequence. From this plot, we can see that $Q^{AB/F}$ provides relatively better correspondence with the subjective study.

V. Conclusion

In this paper, we proposed a novel fusion algorithm based on perceptual quality measures, *i.e.,* perceived local contrast and color saturation. To the best of our knowledge, this is the first time that the modeling of the probability for human eyes to detect local contrast is introduced to multi-exposure fusion, which helps us achieve maximum local detail preservation. A hierarchical multivariate Gaussian conditional random field model was proposed to effectively integrate the perceptual quality measures. Experiments demonstrated better performance of our algorithm compared to other methods. In future work, we will analyze the applicability of other perceptual quality measures in multi-exposure fusion.

Acknowledgment

The authors would like to thank Dr. Z. Wang, University of Waterloo, for providing the code for universal image quality index and SSIM image quality index, which was used in implementing Cvejic's and Yang's fusion quality metrics. The authors would also like to thank the reviewers for their constructive comments and suggestions.

References

[1] I. Lissner and P. Urban, "Toward a unified color space for perception-based image processing," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1153–1168, 2012.

[2] A. Firouzmanesh, I. Cheng, and A. Basu, "Perceptually guided fast compression of 3-D motion capture data," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 829–834, 2011.

[3] B. Bringier, N. Richard, M.-C. Larabi, and C. Fernandez-Maloigne, "No-reference perceptual quality assessment of color image," in *Proc. Eur. Signal Process. Conf.*, 2006.

[4] R. N. Strickland, C.-S. Kim, and W. F. McDonnell, "Digital color image enhancement based on the saturation component," *Opt. Eng.*, vol. 26, pp. 609–616, 1987.

[5] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, 2009.

[6] 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, 2011.

[7] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH*, 1997, pp. 369–378.

[8] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *Proc. ACM SIGGRAPH*, 2002, pp. 267–276.

[9] J. Kuang, G. M. Johnson, and M. D. Fairchild, "iCAM06: A refined image appearance model for hdr image rendering," *J. Vis. Comun. Image Represent.*, vol. 18, no. 5, pp. 406–414, 2007.

[10] P. Burt, *Multiresolution Image Processing and Analysis*. Springer-Verlag, 1984, ch. The pyramid as a structure for efficient computation, pp. 6–35.

[11] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *Proc. Int. Conf. Comput. Vis.*, 1993, pp. 173–182.

[12] S. Raman and S. Chaudhuri, "A matte-less, variational approach to automatic scene compositing," in *Proc. Int. Conf. Comput. Vis.*, 2007, pp. 1–6.

[13] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, "Probabilistic exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 341–357, 2012.

[14] W. Zhang and W.-K. Cham, "Gradient-directed multi-exposure composition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2318–2323, 2012.

[15] K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings," *J. Physiol.*, vol. 359, pp. 381–400, 1985.

[16] E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A*, vol. 7, no. 10, pp. 2032–2040, 1990.

[17] J. M. Foley and W. Schwarz, "Spatial attention: Effect of position uncertainty and number of distractor patterns on the threshold-versus-contrast function for contrast discrimination," *J. Opt. Soc. Am. A*, vol. 15, no. 5, pp. 1036–1047, 1998.

[18] 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.

[19] H. R. Wilson, "A transducer function for threshold and suprathreshold human vision," *Biol. Cybern.*, vol. 38, no. 3, pp. 171–178, 1980.

[20] E. Lübbe, *Colours in the mind - color systems in reality*. Books on Demand GmbH, 2010.

[21] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC, 2005.

[22] M. F. Tappen, C. Liu, E. H. Adelson, and W. T. Freeman, "Learning Gaussian conditional random fields for low-level vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.

[23] E. Reinhard, "Parameter estimation for photographic tone reproduction," *J. Graph. Tools*, vol. 7, no. 1, pp. 45–52, 2002.

[24] C. S. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, 2000.

[25] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective assessment of multiresolution fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 94–109, 2011.

[26] N. Cvejic, A. Loza, D. Bul, and N. Canagarajah, "A similarity metric for assessment of image fusion algorithms," *Int. J. Signal Process.*, vol. 2, no. 3, pp. 178–182, 2005.

[27] C. Yang, J.-Q. Zhang, X.-R. Wang, and X. Liu, "A novel similarity based quality metric for image fusion," *Inf. Fusion*, vol. 9, no. 2, pp. 156–160, 2008.

[28] V. Petrović, "Subjective tests for image fusion evaluation and objective metric validation," *Inf. Fusion*, vol. 8, no. 2, pp. 208–216, 2007.

[29] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," in *Proc. ACM SIGGRAPH*, no. 69, 2008, pp. 1–10.

[30] M. Čadík, 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.

Rui Shen (S'07) received the B.Eng. degree in computer science and technology in 2005 from Beihang University, Beijing, China, and the M.S. and Ph.D. degrees in 2007 and 2012, both in computing science, from the University of Alberta, Edmonton, Canada.

He is currently with the Multimedia Research Centre, University of Alberta. His research interests and experience span the areas of image processing, computer vision, machine learning, and computer graphics. He is on the organizing committees of IEEE International Conference on Multimedia and Expo (ICME) 2011 and IEEE ICME 2013. He was the recipient of the iCORE ICT Graduate Student Scholarship and the Izaak Walton Killam Memorial Scholarship.

Irene Cheng (M'02–SM'09) is the Scientific Director of the iCORE Multimedia Research Centre and an Adjunct Faculty in both Faculty of Medicine & Dentistry and Faculty of Science, University of Alberta, Canada. Her research interests, among others, include incorporating human perception, incorporating the concept of Just-Noticeable-Difference (JND) following psychophysical methodology, to improve multimedia, graphics and computer vision techniques. She completed her Ph.D. at the University of Alberta and conducted postdoctoral research at the University of Pennsylvania. Before joining academia, she was a regional Information Technology executive in Lloyds Bank International, Asia. She received an Alumni Recognition Award in 2008 from the University of Alberta for her R&D contributions. She has received, or been offered, many scholarships and fellowships from NSERC, iCORE and others. Dr. Cheng is the Chair of the IEEE Northern Canada Section, EMBS Chapter (2009-2011), Board Member of the IEEE System, Man and Cybernetics (SMC) Society, Human Perception in Vision, Graphics and Multimedia TC, and the Chair of the IEEE Communication Society, MMTC Interest Group on 3D rendering, processing and communications (2010-2012). She was the lead General Chair in IEEE ICME (July) 2011 and is a visiting professor funded at Institut National des Sciences Appliquees (INSA) de Lyon, France 2011. She has over 100 publications including two books.

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.