

Variational Bayesian Method for Retinex

Liqian Wang, Liang Xiao, *Member, IEEE*, Hongyi Liu, and Zihui Wei

Abstract—In this paper, we propose a variational Bayesian method for Retinex to simulate and interpret how the human visual system perceives color. To construct a hierarchical Bayesian model, we use the Gibbs distributions as prior distributions for the reflectance and the illumination, and the gamma distributions for the model parameters. By assuming that the reflection function is piecewise continuous and illumination function is spatially smooth, we define the energy functions in the Gibbs distributions as a total variation function and a smooth function for the reflectance and the illumination, respectively. We then apply the variational Bayes approximation to obtain the approximation of the posterior distribution of unknowns so that the unknown images and hyperparameters are estimated simultaneously. Experimental results demonstrate the efficiency of the proposed method for providing competitive performance without additional information about the unknown parameters, and when prior information is added the proposed method outperforms the non-Bayesian-based Retinex methods we compared.

Index Terms—Bayesian methods, Retinex, image enhancement, parameter estimation, variational methods.

I. INTRODUCTION

COLOR image enhancement is an important preprocessing phase of many image analysis tasks such as image segmentation, pattern recognition, and so on. A number of methods have been proposed to solve the image enhancement problem. Some methods enhance images by manipulating image histogram [1], [2], some methods transform the gray level of an image [3], and some methods develop a nonlinear transformation for dynamic range compression [4]. In current color image enhancement research, methods based on the Retinex theory have been commonly used in many applications. Retinex was first proposed by Land and McCann [5], [6] as a model of color and luminance perception of human

Manuscript received August 18, 2013; revised January 10, 2014; accepted May 6, 2014. Date of publication May 16, 2014; date of current version July 1, 2014. This work was supported in part by the Natural Science Foundation of China under Grant 61171165, Grant 60802039, and Grant 61301215, in part by the National Scientific Equipment Developing Project of China under Grant 2012YQ050250, in part by the Qing Lan Project of Jiangsu Province, in part by the Six Top Talents of Jiangsu Province Grant, in part by the Jiangsu Planned Projects for Post-Doctoral Research Funds under Grant 1301025C, and in part by the Innovative Research Project for Postgraduates in Colleges of Jiangsu Province under Grant CXZZ13_0212. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xin Li.

L. Wang is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: alisonwlq@gmail.com).

L. Xiao and Z. Wei are with the Jiangsu Province Key Laboratory of Spectral Imaging and Intelligent Sensing, Nanjing 210094, China (e-mail: xiaoliang@mail.njust.edu.cn; gswei@mail.njust.edu.cn).

H. Liu is with the School of Science, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: hyliu@mail.njust.edu.cn).

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

Digital Object Identifier 10.1109/TIP.2014.2324813

visual system (HVS). The Retinex theory was concluded in two aspects: first, the color sensation of HVS involves the interaction of at least three independent cone systems; second, the color sensation is dependent on reflectance rather than non-uniform illumination. Therefore, Retinex is effective in enhancing images with non-uniform lighting.

Based on Land and McCann's Retinex, many variants of Retinex algorithm have been proposed to solve the ill-posed problem of separating reflectance and illumination. These algorithms are usually categorized [7]–[9] as path-based algorithms [5], [6], [10], [11], recursive algorithms [12]–[14], center/surround algorithms [15]–[17], PDE-based algorithms [9], [18], [19], and variational algorithms [8], [20], [21]. The original Retinex algorithm, which performs a ratio reset computation on a random path, is a typical path-based algorithm. The path-based algorithms have high computational complexity. The mainly difference between them is the path geometry. Frankle and McCann [12], [13] proposed recursive algorithms by replacing the path computation with efficient recursive matrix calculation. However, the results of the recursive algorithms are strongly affected by the iteration number. The representative algorithms of the center/surround approach are the single-scale Retinex (SSR) [15] and the multi-scale Retinex (MSR) [16], [22]. Although the SSR and the MSR are simple and effective, they tend to produce halo artifacts near strong edges. The multiscale Retinex with color restoration (MSRCR) [22] method is an improvement of the MSR; however, it does not solve the problem of the halo artifacts. In PDE-based algorithm [9], the ill-posed problem is modeled as a Poisson equation based on the assumption that the illumination varies smoothly, while the reflectance changes at sharp edges. Based on the same assumptions in [9], variational Retinex models [8], [21] are established by using different regularization terms of illumination function and reflection function. However, the model parameters are not easy to measure.

In recent years, variational Bayesian methods have been applied in non-blind and blind image deconvolution [23]–[25] and super resolution [26] problems. Within the hierarchical Bayesian formulation, the deconvolution or the super resolution methods are able to simultaneously estimate model parameters along with the unknown images, and do not require tuning parameters. Moreover, by applying the variational method, we are able to obtain an approximation of the posterior distribution. To the best of our knowledge, no work has been reported on the variational Bayesian method for simultaneous estimation of the parameters and images for the Retinex.

In this paper, we propose a variational Bayesian method for Retinex to simulate and interpret how the human vision system perceives color. We use the Gibbs distributions as prior

distributions for the reflectance image and the illumination image, and the gamma distributions for the model parameters. By assuming that the reflection function is piecewise continuous and illumination function is spatially smooth, we define the energy functions in the Gibbs distribution as a total variation (TV) function and a smooth function for the reflectance function and the illumination function, respectively. Applying the variational Bayes approximation approach, we obtain the approximation of the posterior distribution of unknowns. Then with the Bayesian inference the proposed method simultaneously estimates model parameters along with the unknown illumination image and reflectance image, and does not require parameters tuning.

The rest of the paper is organized as follows. In Section II, we review the variational approximation method used in the Bayesian formulation based methods. In Section III we present the Bayesian model and the prior models on the observed image, the hidden variables and the hyperparameters, respectively. Section IV describes the algorithm of the proposed method. In Section V, we show experimental results demonstrating the efficiency of our method. Finally, the conclusions of this paper are given in Section VI, and some computational details are given in Appendix.

II. BACKGROUND ON VARIATIONAL BAYES APPROXIMATION

In this section, we briefly review the variational Bayes approximation approach which will be used to solve the color enhancement problem. Assume a model with a set of unobserved variables $\mathbf{x} = \{x_1, \dots, x_n\}$, a observed variable \mathbf{y} and the model parameters θ . Let $\Theta = (\mathbf{x}, \theta)$ denotes all unknown variables, the probability density function parameterized by Θ is $p(\mathbf{y}|\Theta)$. Bayesian methods treat the unknown parameter as random variable and assume a prior distribution for the unknowns. As is widely known, based on the application of the Bayes' rule, Bayesian inference is as follows:

$$p(\Theta|\mathbf{y}) = \frac{p(\Theta, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\Theta)p(\Theta)}{\int_{\Theta} p(\mathbf{y}|\Theta)p(\Theta)d\Theta}, \quad (1)$$

where $p(\Theta|\mathbf{y})$ is the posterior distribution, $p(\Theta, \mathbf{y})$ is the joint probability distribution.

Since $p(\mathbf{y}) = \int_{\Theta} p(\mathbf{y}|\Theta)p(\Theta)d\Theta$ cannot be calculated analytically in general, the computation of $p(\Theta|\mathbf{y})$ is intractable. This difficult problem can be avoided by finding an approximation $q(\Theta)$ of the true posterior distribution $p(\Theta|\mathbf{y})$. The variational Bayes approximation which is also known as mean-field variational Bayes or ensemble learning [27] uses the Kullback-Leibler (K-L) distance [28] to measure the similarity between $p(\Theta|\mathbf{y})$ and $q(\Theta)$. The main intention of the variational Bayes approximation is finding $q(\Theta)$ that minimizing the K-L distance between $q(\Theta)$ and $p(\Theta|\mathbf{y})$. The K-L distance is defined as

$$D_{KL}(q(\Theta)||p(\Theta|\mathbf{y})) = \int_{\Theta} q(\Theta) \log \frac{q(\Theta)}{p(\Theta|\mathbf{y})} d\Theta. \quad (2)$$

Notice that the K-L distance is not symmetric, and is actually a 'reverse' K-L distance from $q(\Theta)$ to $p(\Theta|\mathbf{y})$ instead of the 'original' K-L distance from $p(\Theta|\mathbf{y})$ to $q(\Theta)$. The advantage

of the K-L distance in (2) will lead to an analytical solution, and parameters can be easily obtained [29]. Since the observed data \mathbf{y} is known, (2) can be rewritten as

$$D_{KL}(q(\Theta)||p(\Theta|\mathbf{y})) = \int_{\Theta} q(\Theta) \log \frac{q(\Theta)}{p(\Theta, \mathbf{y})} d\Theta + \log p(\mathbf{y}) \geq 0. \quad (3)$$

The variational Bayes approximation is based on the assumption that $q(\Theta)$ can be factorized into independent factors, i.e. $q(\Theta) = q(\mathbf{x})q(\theta)$. With this assumption the distributions of the unknown variables will be decoupled, and consequently the overall computational complexity will be greatly decreased. Therefore the disadvantage of ignoring the interaction between the unknown variables is outweighed by the computational advantages: (i) one can obtain an analytical approximation $q(\Theta)$, and (ii) all the unknown variables can be evaluated using an alternating algorithm [29], [30].

Then by denoting

$$\tilde{\mathcal{L}}(q(\mathbf{x}), q(\theta)) = \int_{\Theta} q(\mathbf{x})q(\theta) \log \frac{q(\mathbf{x})q(\theta)}{p(\mathbf{x}, \theta, \mathbf{y})} d\Theta$$

and rearranging (3), we have the Algorithm 1 in which $q(\mathbf{x})$ and $q(\theta)$ can be estimated at iteration k .

Notice that $\mathcal{L}(q(\mathbf{x}), q(\theta)) = -\tilde{\mathcal{L}}(q(\mathbf{x}), q(\theta)) = \log p(\mathbf{y}) - D_{KL}(q(\Theta)||p(\Theta|\mathbf{y}))$ provides a lower bound for $\log p(\mathbf{y})$. When the K-L distance $D_{KL}(q(\Theta)||p(\Theta|\mathbf{y}))$ becomes zero, $p(\mathbf{y}) = \exp \mathcal{L}(q(\mathbf{x}), q(\theta))$.

III. BAYESIAN MODELING AND INFERENCE FOR RETINEX

It is known that the output of image acquisition device is an effect of both the surface reflectance and the illumination, which can be modeled as

$$\mathbf{S} = \mathbf{L} \cdot \mathbf{R}, \quad (4)$$

where the $N \times 1$ vectors \mathbf{S} , \mathbf{L} , and \mathbf{R} represent the original image intensity, the light illumination, and the surface reflectance, respectively. The symbol \cdot denotes element-wise multiplication. Equation (4) is usually converted into logarithmic domain as $\mathbf{s} = \mathbf{l} + \mathbf{r}'$, where $\mathbf{s} = \log(\mathbf{S})$, $\mathbf{l} = \log(\mathbf{L})$, $\mathbf{r}' = \log(\mathbf{R})$. According to the assumptions in [8] and [21], we have $0 < \mathbf{R} < 1$, $0 < \mathbf{L} < \infty$, and $\mathbf{L} \geq \mathbf{S}$. Since the log function is monotone, we also have $\mathbf{r}' < 0$, $\mathbf{l} \geq \mathbf{s}$. By denoting $\mathbf{r} = -\mathbf{r}'$, finally we have

$$\mathbf{l} = \mathbf{s} + \mathbf{r}. \quad (5)$$

For simplicity, we decompose the observed image into hue, saturation, and value three channels, and we then propose the hierarchical Bayesian model and discuss our method based on the gray level image in the value channel. Finally, the results in the HSV color space are converted into the images in RGB color space.

From Algorithm 1, we know that the computation of the approximation of posterior distributions of unknown variables requires the joint probability distribution of all variables. For the Retinex model, the joint probability distribution is defined as:

$$p(\mathbf{s}, \mathbf{l}, \mathbf{r}, \alpha, \beta, \sigma) = p(\alpha)p(\beta)p(\sigma)p(\mathbf{r}|\alpha)p(\mathbf{l}|\sigma)p(\mathbf{s}|\mathbf{l}, \mathbf{r}, \beta), \quad (6)$$

Algorithm 1 Algorithm for the Variational Bayes Approximation

Step (1): Compute the approximate distribution of hidden variables \mathbf{x} at iteration k :

$$q^{k+1}(\mathbf{x}) = \arg \min_{q(\mathbf{x})} D_{KL}(q(\Theta) \| p(\Theta | \mathbf{y})) = \arg \min_{q(\mathbf{x})} \tilde{\mathcal{L}}(q(\mathbf{x}), q^k(\theta)).$$

Step (2): Compute $q^{k+1}(\theta)$ using $q^{k+1}(\mathbf{x})$:

$$q^{k+1}(\theta) = \arg \min_{q(\theta)} D_{KL}(q(\Theta) \| p(\Theta | \mathbf{y})) = \arg \min_{q(\theta)} \tilde{\mathcal{L}}(q^{k+1}(\mathbf{x}), q(\theta)).$$

where \mathbf{s} is the observed data, illumination \mathbf{l} and reflectance \mathbf{r} are the hidden variables, and α , β and σ are model hyperparameters which are also unknown. In Bayesian models, all unknown variables need to be assigned probability distributions. The prior distributions $p(\mathbf{l}|\sigma)$ and $p(\mathbf{r}|\alpha)$ are assigned to the unknown variables \mathbf{l} and \mathbf{r} to model the knowledge about the nature of the light illumination and the surface reflectance, respectively. The original logarithmic image \mathbf{s} is considered as a random process with the corresponding conditional distribution $p(\mathbf{s}|\mathbf{l}, \mathbf{r}, \beta)$. These distributions are depend on the unknown parameters α , β and σ . Since these unknown parameters are treated as random variables, the prior distributions $p(\alpha)$, $p(\beta)$ and $p(\sigma)$ are assigned to α , β and σ , respectively, and α , β and σ are called hyperparameters. In the following subsections, we propose the hierarchical Bayesian model for Retinex. We first describe the observation model as well as the prior models for the illumination and the reflectance. Then we define the hyperprior distributions for the three hyperparameters.

A. Formulation of the Image Prior

Considering the difference between the estimated \mathbf{l} , \mathbf{r} and the original logarithmic image \mathbf{s} , we introduce an estimated error $\boldsymbol{\epsilon}$ and assume that $\boldsymbol{\epsilon}$ is Gaussian with zero mean and precision β (the inverse of variance). The conditional distribution of \mathbf{s} is given by

$$p(\mathbf{s}|\mathbf{l}, \mathbf{r}, \beta) \propto \beta^{N/2} \exp\left[-\frac{\beta}{2} (\mathbf{l} - \mathbf{s} - \mathbf{r})^2\right]. \quad (7)$$

Based on the assumption that the reflectance changes at sharp edges, for simplicity, we may consider that when the horizontal first order difference operator Δ_i^h is applied to the reflectance image \mathbf{r} , the result $\Delta_i^h \mathbf{r}$ is sparse. Thus the super-Gaussian sparse prior can be used to model the reflectance \mathbf{r} as

$$p(\mathbf{r}) = Z_r \exp\left[-\sum_i^N \log |\Delta_i^h(\mathbf{r})|\right], \quad (8)$$

where Z_r is the normalization constant, Δ_i^h is the horizontal first order differences at pixel i , and the logarithm function is used to enforce sparsity [31].

Formula (8) enforces image sparsity in only one direction, which may not enough to preserve image features. Then we further consider a more complicated model. It is known that a Markov random field (MRF) is used to model various tasks in image processing and computer vision. For the prior

distribution $p(\mathbf{x})$ of an image \mathbf{x} , the MRF is characterized by the Gibbs distribution

$$p(\mathbf{x}|a) = \frac{1}{Z(a)} \exp\{-a F(\mathbf{x})\}, \quad (9)$$

where $Z(a) = \sum_{\mathbf{x}} \exp\{-a F(\mathbf{x})\}$ is the partition function, $F(\mathbf{x}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{x})$ is the energy function, V_c is a potential function defined on a set of connected pixels, and a is a constant.

We assume that the reflection function is piecewise continuous. Therefore, the total variation prior [32] is preferred for the reflection prior. For the total variation prior, the energy function is

$$F_{TV}(\mathbf{r}) = \sum_i^N \sqrt{(\Delta_i^h(\mathbf{r}))^2 + (\Delta_i^v(\mathbf{r}))^2}, \quad (10)$$

where Δ_i^h and Δ_i^v are linear operators corresponding to horizontal and vertical first order differences at pixel i , respectively. Then the prior model on the reflectance \mathbf{r} is given by

$$p(\mathbf{r}|\alpha) \propto \frac{1}{Z_{TV}(\alpha)} \exp[-\alpha F_{TV}(\mathbf{r})]. \quad (11)$$

Because the partition function $Z_{TV}(\alpha)$ is difficult to compute, if we assume that the difference pair $(\Delta_i^h(\cdot), \Delta_i^v(\cdot))$ at each pixel is independent of the difference pairs at all the other pixels [33], and we use the fact that

$$\iint \exp[-\alpha \sqrt{u^2 + v^2}] du dv = 2\pi/\alpha^2, \quad (12)$$

then we can approximate the partition function by

$$Z_{TV}(\alpha) \simeq c_r \alpha^{-Nt}, \quad (13)$$

where c_r is a constant independent of α . Because the dependence really exists among the difference pairs, we empirically found that using the adjustable parameter $t = 1/2$ produces favorable results and we therefore approximate the prior model on \mathbf{r} by

$$p(\mathbf{r}|\alpha) = c_r \alpha^{N/2} \exp[-\alpha F_{TV}(\mathbf{r})]. \quad (14)$$

In [24] and [25] the simultaneous autoregression (SAR) model [34], [35], which describes the spatial correlation of the image, is utilized for the blur prior in image restoration. Since we assume that the illumination function is spatially smooth, the prior of the illumination \mathbf{l} can also be modeled by the SAR as follows

$$p(\mathbf{l}|\sigma) \propto \sigma^{N/2} \exp\left[-\frac{1}{2}\sigma \|\mathbf{Cl}\|^2\right], \quad (15)$$

where \mathbf{C} denotes the discrete Laplacian operator.

Inspired by [21], we define another spatially smoothness energy function for the illumination prior as follows

$$F(\mathbf{l}) = \sum_i^N (\Delta_i^h(\mathbf{l}))^2 + (\Delta_i^v(\mathbf{l}))^2, \quad (16)$$

and the prior model on the illumination image \mathbf{l} is defined as

$$p(\mathbf{l}|\sigma) \propto \frac{1}{Z(\sigma)} \exp[-\sigma F(\mathbf{l})]. \quad (17)$$

Then by using an approximation of the partition function $Z(\sigma)$ with setting the adjustable parameter equal to 1/2, we approximate the prior model on the illumination image as

$$p(\mathbf{l}|\sigma) = c_l \sigma^{N/2} \exp \left[-\sigma \sum_i^N (\Delta_i^h(\mathbf{l}))^2 + (\Delta_i^v(\mathbf{l}))^2 \right], \quad (18)$$

where c_l is a constant. We finally use (14) and (18) for the priors of \mathbf{r} and \mathbf{l} , respectively, in our algorithm. As we will see later in the algorithm, we correct the estimated \mathbf{r} and \mathbf{l} at each iteration, and thus the adjustable parameter will not have great effect on the results.

B. Hyperpriors on the Hyperparameters

In the above conditional distribution and the *a priori* probability distribution, α , β and σ are hyperparameters assumed unknown and need to be computed. These three hyperparameters are important in determining the performance of the color enhancement algorithm. Under the Bayesian framework, we can estimate the illumination, the reflectance and the hyperparameters simultaneously. We would like to have the form of the hyperprior distributions provides an easy calculation of the posterior distribution $p(\mathbf{l}, \mathbf{r}, \alpha, \beta, \sigma | \mathbf{s})$. A reasonable practice is making the functional form of the hyperprior distribution the same as the product $p(\mathbf{r}|\alpha)p(\mathbf{l}|\sigma)p(\mathbf{s}|\mathbf{l}, \mathbf{r}, \beta)$, so that the posterior distribution will have the same functional form as the prior distribution and the prior and the posterior will be conjugate distributions. The conjugate prior provides an algebraic convenience and results in a closed-form of the posterior [36]. Since the Gamma distribution is the conjugate prior for the inverse variance of the Gaussian distribution, we use the Gamma distribution for the hyperparameters α , β and σ . The Gamma distribution for the hyperprior is defined as

$$p(h) = \Gamma(h|a_h^o, b_h^o) = \frac{h^{a_h^o-1} (b_h^o)^{a_h^o} \exp[-hb_h^o]}{\Gamma(a_h^o)}, \quad (19)$$

where $h > 0$, $h \in \{\alpha, \beta, \sigma\}$ denotes a hyperparameter, $a_h^o > 0$ and $b_h^o > 0$ denote the shape parameter and the scale parameter, respectively. a_h^o and b_h^o are assumed to be known.

Finally, the global distribution is given by (6), i.e.

$$p(\mathbf{s}, \mathbf{l}, \mathbf{r}, \alpha, \beta, \sigma) = p(\alpha)p(\beta)p(\sigma)p(\mathbf{r}|\alpha)p(\mathbf{l}|\sigma)p(\mathbf{s}|\mathbf{l}, \mathbf{r}, \beta),$$

where $p(\alpha)$, $p(\beta)$ and $p(\sigma)$ are defined in (19); $p(\mathbf{r}|\alpha)$, $p(\mathbf{l}|\sigma)$ and $p(\mathbf{s}|\mathbf{l}, \mathbf{r}, \beta)$ are defined in (14), (18) and (7), respectively. The dependencies in the joint probability model are shown in Fig. 1 which presents the relationships between variables for Retinex in a directed acyclic graph.

IV. VARIATIONAL BAYESIAN ALGORITHM FOR RETINEX

In this section, we compute the poster distribution by a variational approximation, and then we describe the algorithm of the variational Bayesian method for Retinex in Algorithm 2.

We first denote the set of all unknowns by $\Theta = \{\mathbf{l}, \mathbf{r}, \alpha, \beta, \sigma\}$ and denote the subset of Θ by Θ_θ in which the element $\theta \in \{\mathbf{l}, \mathbf{r}, \alpha, \beta, \sigma\}$ is removed from Θ . The Bayesian inference is based on the posterior distribution

$$p(\Theta|\mathbf{s}) = p(\mathbf{l}, \mathbf{r}, \alpha, \beta, \sigma | \mathbf{s}) = \frac{p(\mathbf{s}, \mathbf{l}, \mathbf{r}, \alpha, \beta, \sigma)}{p(\mathbf{s})}, \quad (20)$$

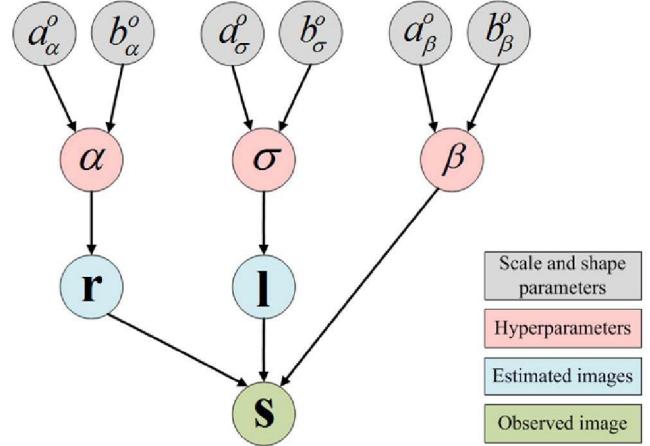


Fig. 1. Graphical model describing relationships between variables for Retinex.

where the joint probability distribution $p(\mathbf{s}, \mathbf{l}, \mathbf{r}, \alpha, \beta, \sigma)$ is given by (6). Because $p(\mathbf{s}) = \int_{\Theta_{\mathbf{l}\mathbf{r}}} \int_{\mathbf{r}} \int_{\mathbf{l}} p(\mathbf{s}, \mathbf{l}, \mathbf{r}, \alpha, \beta, \sigma) d\mathbf{l} d\mathbf{r} d\Theta_{\mathbf{l}\mathbf{r}}$ is analytically intractable, it is difficult to calculate the posterior distribution $p(\Theta|\mathbf{s})$. We then use the variational Bayes approximation approach reviewed in Section II to give an approximation $q(\Theta) = q(\mathbf{l})q(\mathbf{r})q(\alpha)q(\beta)q(\sigma)$ of a factorization form for $p(\Theta|\mathbf{s})$. Thus we need to minimize

$$\tilde{\mathcal{L}}(q(\Theta)) = \int_{\Theta} q(\Theta) \log \frac{q(\Theta)}{p(\mathbf{s}, \Theta)} d\Theta.$$

However, the TV prior model of \mathbf{r} makes the integral in $p(\mathbf{s}, \Theta)$ difficult to compute. In order to overcome this difficult, we consider the inequality in [33] that, for any $x \geq 0$ and $x_0 > 0$

$$\sqrt{x} \leq (x + x_0)/2\sqrt{x_0}. \quad (21)$$

Then by defining any N -dimensional vector \mathbf{u} whose components are $u_i > 0$, $i = 1, \dots, N$, and with $x = (\Delta_i^h(\mathbf{r}))^2 + (\Delta_i^v(\mathbf{r}))^2$ and $x_0 = u_i$, we have

$$\begin{aligned} p(\mathbf{r}|\alpha) &= c_r \alpha^{N/2} \exp \left[-\alpha \sum_i^N \sqrt{(\Delta_i^h(\mathbf{r}))^2 + (\Delta_i^v(\mathbf{r}))^2} \right] \\ &\geq c_r \alpha^{N/2} \exp \left[-\frac{\alpha}{2} \sum_i^N \frac{(\Delta_i^h(\mathbf{r}))^2 + (\Delta_i^v(\mathbf{r}))^2 + u_i}{\sqrt{u_i}} \right] \\ &\triangleq \hat{p}(\mathbf{r}, \alpha, \mathbf{u}). \end{aligned} \quad (22)$$

Thus, $p(\mathbf{s}, \Theta) \geq p(\alpha)p(\beta)p(\sigma)\hat{p}(\mathbf{r}, \alpha, \mathbf{u})p(\mathbf{l}|\sigma)p(\mathbf{s}|\mathbf{l}, \mathbf{r}, \beta) = \hat{p}(\mathbf{s}, \Theta, \mathbf{u})$ and consequently

$$\tilde{\mathcal{L}}(q(\Theta)) \leq \int_{\Theta} q(\Theta) \log \frac{q(\Theta)}{\hat{p}(\mathbf{s}, \Theta, \mathbf{u})} d\Theta \hat{\mathcal{L}}(q(\Theta), \mathbf{u}). \quad (23)$$

We then minimize the upper bound $\hat{\mathcal{L}}(q(\Theta), \mathbf{u})$ instead of minimizing $\tilde{\mathcal{L}}(q(\Theta))$. The introduced vector \mathbf{u} is an auxiliary variable that needs to be computed and, as will be shown later, has an intuitive interpretation related to the unknown variable \mathbf{r} . In resemblance to the algorithm in [23], the minimization process here implies a strategy that guarantees the computed minimum approaches the minimum of

$\tilde{\mathcal{L}}(q(\Theta))$. Notice that the Algorithm 1 in Section II provides an alternating optimization scheme. By using this scheme we are actually finding a sequence of distributions $\{q^k(\Theta)\}$ that monotonically decreases $\hat{\mathcal{L}}(q(\Theta), \mathbf{u})$ with a fixed \mathbf{u} , and consequently we obtain an ever decreasing upper bound of $\tilde{\mathcal{L}}(q^k(\Theta))$. Minimizing $\hat{\mathcal{L}}(q(\Theta), \mathbf{u})$ with respect to \mathbf{u} with fixed $q(\Theta)$ also generates a sequence of vectors $\{\mathbf{u}^k\}$ that tightens the upper bound for each $q^k(\Theta)$. Therefore, we finally find two coupled sequences $\{q^k(\Theta)\}$ and $\{\mathbf{u}^k\}$ that assure the decrease of $\hat{\mathcal{L}}(q(\Theta), \mathbf{u})$, and the process of finding these sequences is a natural extension of the majorization-minimization approach to function optimization [37]. We minimize the surrogate majorizing function $\hat{\mathcal{L}}(q(\Theta), \mathbf{u})$ rather than the actual function $\tilde{\mathcal{L}}(q(\Theta))$. Because K-L distance is a strictly convex function and $\hat{\mathcal{L}}(q(\Theta), \mathbf{u})$ is continuous in both $q(\Theta)$ and \mathbf{u} , the computed $q^k(\Theta)$ converges to the global minimum of $\tilde{\mathcal{L}}(q(\Theta))$. The proof of these properties of the majorization-minimization approach is analogous to that of the EM algorithm (a special case of the MM algorithm), which can be found in [38]. Due to space limitations here we omit the proof.

According to the Algorithm 1 in Section II, we now describe the alternant iterative scheme involving all unknown variables in our variational Bayesian method for Retinex in Algorithm 2.

As we will see in Section V, because the reflectance image obtained directly from Retinex is usually over enhanced, in practice we only need the illumination image for displaying the results. Therefore we first estimate \mathbf{l} , and give the stopping criteria only by \mathbf{l} , although we also obtain the other unknown variables from Algorithm 2. After the estimation of \mathbf{l} , we then estimate \mathbf{r} , then \mathbf{u} and finally the hyper parameters in a conventional order in variational Bayes algorithm.

The posterior approximation $q(\Theta)$ can be computed by iteratively solving the minimization problems (24), (26), and (29). We now take (24) for instance to show the existence result for solutions of these problems [29].

Theorem 1: The minimization of (24) is reached for

$$q(\mathbf{l}) \propto \exp(E_{q(\Theta_l)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u})]). \quad (31)$$

Proof: We can rewrite $\hat{\mathcal{L}}_{q(\mathbf{l})}(q(\Theta), \mathbf{u})$ as follows:

$$\begin{aligned} \hat{\mathcal{L}}_{q(\mathbf{l})}(q(\Theta), \mathbf{u}) &= \int_{\Theta} q(\Theta) \log \frac{q(\Theta)}{\hat{p}(\mathbf{s}, \Theta, \mathbf{u})} d\Theta \\ &= \int_{\mathbf{l}} \int_{\Theta_l} q(\mathbf{l}) q(\Theta_l) \log \frac{q(\mathbf{l}) q(\Theta_l)}{\hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u})} d\Theta_l d\mathbf{l} \\ &= \int_{\mathbf{l}} \int_{\Theta_l} q(\mathbf{l}) q(\Theta_l) \log q(\mathbf{l}) d\Theta_l d\mathbf{l} \\ &\quad + \int_{\mathbf{l}} \int_{\Theta_l} q(\mathbf{l}) q(\Theta_l) \log q(\Theta_l) d\Theta_l d\mathbf{l} \\ &\quad - \int_{\mathbf{l}} \int_{\Theta_l} q(\mathbf{l}) q(\Theta_l) \log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u}) d\Theta_l d\mathbf{l} \\ &= \int_{\mathbf{l}} q(\mathbf{l}) \log q(\mathbf{l}) d\mathbf{l} + E_{q(\Theta_l)}[\log q(\Theta_l)] \\ &\quad - \int_{\mathbf{l}} q(\mathbf{l}) \int_{\Theta_l} q(\Theta_l) \log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u}) d\Theta_l d\mathbf{l}. \end{aligned} \quad (32)$$

Algorithm 2 Algorithm for the Variational Bayesian Method for Retinex

Algorithm's inputs: the original logarithmic image \mathbf{s} , the max iteration number $iter_{max}$ and the iteration stopping criterion parameter ξ_{stop} .

Algorithm's output: The logarithmic illumination image \mathbf{l} , the logarithmic reflectance image $\mathbf{r}, \mathbf{u}, \alpha, \beta$, and σ .

Initializations: Initialize $\mathbf{u}^1, q^1(\mathbf{r}), q^1(\alpha), q^1(\beta)$, and $q^1(\sigma)$. Set $k = 1$.

Iterations:

Step(1): Find $q^k(\mathbf{l})$ by

$$\begin{aligned} q^k(\mathbf{l}) &= \arg \min_{q(\mathbf{l})} \hat{\mathcal{L}}_{q(\mathbf{l})}(q(\Theta), \mathbf{u}) \\ &= \arg \min_{q(\mathbf{l})} \int_{\mathbf{l}} \int_{\Theta_l} q(\mathbf{l}) q^k(\Theta_l) \log \frac{q(\mathbf{l}) q^k(\Theta_l)}{\hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u}^k)} d\Theta_l d\mathbf{l}, \end{aligned} \quad (24)$$

and update \mathbf{l}^k by using

$$\mathbf{l}^k = \max \left\{ E_{q^k(\mathbf{l})}[\mathbf{l}], \mathbf{s} \right\}; \quad (25)$$

Step(2): Find $q^{k+1}(\mathbf{r})$ by

$$\begin{aligned} q^k(\mathbf{r}) &= \arg \min_{q(\mathbf{r})} \hat{\mathcal{L}}_{q(\mathbf{r})}(q(\Theta), \mathbf{u}) \\ &= \arg \min_{q(\mathbf{r})} \int_{\mathbf{r}} \int_{\Theta_r} q(\mathbf{r}) q^k(\Theta_r) \log \frac{q(\mathbf{r}) q^k(\Theta_r)}{\hat{p}(\mathbf{s}, \mathbf{r}, \Theta_r, \mathbf{u}^k)} d\Theta_r d\mathbf{r}, \end{aligned} \quad (26)$$

and update \mathbf{r}^{k+1} by using

$$\mathbf{r}^{k+1} = \max \left\{ E_{q^{k+1}(\mathbf{r})}[\mathbf{r}], 0 \right\}; \quad (27)$$

Step(3): Find \mathbf{u}^{k+1} by

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \int q^{k+1}(\mathbf{r}) q^k(\Theta_r) \log \frac{q^{k+1}(\mathbf{r}) q^k(\Theta_r)}{\hat{p}(\mathbf{s}, \mathbf{r}^{k+1}, \Theta_r, \mathbf{u})} d\Theta_r; \quad (28)$$

Step(4): For $h \in \{\alpha, \beta, \sigma\}$, find $q^{k+1}(h)$ by

$$\begin{aligned} q^{k+1}(h) &= \arg \min_{q(h)} \hat{\mathcal{L}}_{q(h)}(q(\Theta), \mathbf{u}) \\ &= \arg \min_{q(h)} \int_h \int_{\Theta_h} q(h) q^{k+1}(\mathbf{r}) q^k(\Theta_{h,r}) \\ &\quad \times \log \frac{q^{k+1}(\mathbf{r}) q^k(\Theta_{h,r})}{\hat{p}(\mathbf{s}, \mathbf{r}^{k+1}, h, \Theta_{h,r}, \mathbf{u}^{k+1})} d\Theta_h dh, \end{aligned} \quad (29)$$

and update h^{k+1} by using

$$h^{k+1} = E_{q^{k+1}(h)}[h]. \quad (30)$$

Termination criteria: If $\|\mathbf{l}^k - \mathbf{l}^{k+1}\| / \|\mathbf{l}^k\| \leq \xi_{stop}$ or $k > iter_{max}$, stop iteration, otherwise $k = k + 1$ and go to the iterations' step.

Let $\eta = E_{q(\Theta_l)}[\log q(\Theta_l)]$, for any non-zero scalar λ_i it holds:

$$\begin{aligned} \hat{\mathcal{L}}_{q(\mathbf{l})}(q(\Theta), \mathbf{u}) &= \int_{\mathbf{l}} q(\mathbf{l}) \log q(\mathbf{l}) d\mathbf{l} - \int_{\mathbf{l}} q(\mathbf{l}) \log \\ &\quad \left\{ \frac{\lambda_i}{\lambda_i} \times \exp(E_{q(\Theta_l)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u})]) \right\} d\mathbf{l} + \eta \\ &= \int_{\mathbf{l}} q(\mathbf{l}) \log \frac{q(\mathbf{l})}{\exp(E_{q(\Theta_l)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u})]) / \lambda_i} d\mathbf{l} \end{aligned}$$

$$+\eta - \log \lambda_i, \quad (33)$$

λ_i can be chosen as a normalization coefficient as $\lambda_i = \int_{\mathbf{l}} \exp(E_{q(\Theta_l)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u})]) d\mathbf{l}$, the last equality in (33) can be rewritten as

$$\hat{\mathcal{L}}_{q(\mathbf{l})}(q(\Theta), \mathbf{u}) = D_{KL}(q(\mathbf{l}) || \exp(E_{q(\Theta_l)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u})]) / \lambda_i) + \eta - \log \lambda_i. \quad (34)$$

Since η and $\log \lambda_i$ do not depend on $q(\mathbf{l})$, minimizing $\hat{\mathcal{L}}_{q(\mathbf{l})}(q(\Theta), \mathbf{u})$ is equivalent to minimize the first term of the right hand side of (34). It is known that the K-L distance is always non-negative. The K-L distance $D_{KL}(q(\mathbf{l}) || \exp(E_{q(\Theta_l)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u})]) / \lambda_i) = 0$ if and only if $q(\mathbf{l}) = \exp(E_{q(\Theta_l)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u})]) / \lambda_i$.

Therefore, the minimization is reached for $q(\mathbf{l}) \propto \exp(E_{q(\Theta_l)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta_l, \mathbf{u})])$. ■

A. Explicit Calculations

For the other minimization problems in Algorithm 2, we are able to get similar conclusions. Then the minimizers $q(\Theta)$ and \mathbf{u} are obtained by differentiating the integrals in (24), (26), (28), and (29) with respect to the corresponding $q(\Theta)$ and \mathbf{u} , respectively [27]. Now we derive the close form solution of each minimizer. The details of the differentiation and the computational details are given in the Appendix.

From (24), we obtain

$$q^k(\mathbf{l}) = \mathcal{N}(\mathbf{l} | E_{q^k(\mathbf{l})}[\mathbf{l}], \text{cov}_{q^k(\mathbf{l})}[\mathbf{l}]), \quad (35)$$

where the mean and covariance are

$$E_{q^k(\mathbf{l})}[\mathbf{l}] = \text{cov}_{q^k(\mathbf{l})}[\mathbf{l}] E_{q^k(\beta)}[\beta](\mathbf{r} + \mathbf{s}), \quad (36)$$

$$\begin{aligned} \text{cov}_{q^k(\mathbf{l})}[\mathbf{l}] &= \left[E_{q^k(\beta)}[\beta] + 2E_{q^k(\sigma)}[\sigma](\Delta^h)^t(\Delta^h) \right. \\ &\quad \left. + 2E_{q^k(\sigma)}[\sigma](\Delta^v)^t(\Delta^v) \right]^{-1}, \end{aligned} \quad (37)$$

where $(\cdot)^t$ denotes the transpose operation.

Similar result for $q^{k+1}(\mathbf{r})$ is obtained from (26) as

$$q^{k+1}(\mathbf{r}) = \mathcal{N}(\mathbf{r} | E_{q^{k+1}(\mathbf{r})}[\mathbf{r}], \text{cov}_{q^{k+1}(\mathbf{r})}[\mathbf{r}]), \quad (38)$$

$$E_{q^{k+1}(\mathbf{r})}[\mathbf{r}] = \text{cov}_{q^{k+1}(\mathbf{r})}[\mathbf{r}] E_{q^k(\beta)}[\beta](\mathbf{l} - \mathbf{s}), \quad (39)$$

$$\begin{aligned} \text{cov}_{q^{k+1}(\mathbf{r})}[\mathbf{r}] &= \left[E_{q^k(\beta)}[\beta] + E_{q^k(\alpha)}[\alpha](\Delta^h)^t W(\mathbf{u}^k)(\Delta^h) \right. \\ &\quad \left. + E_{q^k(\alpha)}[\alpha](\Delta^v)^t W(\mathbf{u}^k)(\Delta^v) \right]^{-1}, \end{aligned} \quad (40)$$

where

$$W(\mathbf{u}^k) = \text{diag}\left(1/\sqrt{u_i^k}\right), i = 1, \dots, N$$

is a $N \times N$ diagonal matrix with diagonal entries equal to $1/\sqrt{u^k}$. The computation of (37) and (40) requires the inversion of a $N \times N$ matrix, which is complex and inefficient. In order to avoid the matrix inversion, we find the numerical solution of (36) and (39) by using the conjugate gradient method [39] which is similar to the method used in [26].

In order to calculate \mathbf{u} , we simplify (28) by omitting the terms which are independent of \mathbf{u} . Then we obtain

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \sum_i^N \frac{E_{q^{k+1}(\mathbf{r})}[(\Delta_i^h(\mathbf{r}))^2 + (\Delta_i^v(\mathbf{r}))^2] + u_i}{\sqrt{u_i}}. \quad (41)$$

By differentiating (41) with respect to \mathbf{u} , and setting it equal to zero we finally obtain

$$u_i^{k+1} = E_{q^{k+1}(\mathbf{r})} \left[(\Delta_i^h(\mathbf{r}))^2 + (\Delta_i^v(\mathbf{r}))^2 \right], i = 1, \dots, N \quad (42)$$

where

$$\begin{aligned} E_{q^{k+1}(\mathbf{r})} \left[(\Delta_i^h(\mathbf{r}))^2 + (\Delta_i^v(\mathbf{r}))^2 \right] &= (\Delta_i^h(E_{q^{k+1}(\mathbf{r})}[\mathbf{r}]))^2 + (\Delta_i^v(E_{q^{k+1}(\mathbf{r})}[\mathbf{r}]))^2 \\ &\quad + E_{q^{k+1}(\mathbf{r})}[(\Delta_i^h(\mathbf{r} - E_{q^{k+1}(\mathbf{r})}[\mathbf{r}]))^2] \\ &\quad + E_{q^{k+1}(\mathbf{r})}[(\Delta_i^v(\mathbf{r} - E_{q^{k+1}(\mathbf{r})}[\mathbf{r}]))^2] \end{aligned} \quad (43)$$

and

$$\begin{aligned} E_{q^{k+1}(\mathbf{r})}[(\Delta_i^h(\mathbf{r} - E_{q^{k+1}(\mathbf{r})}[\mathbf{r}]))^2] &+ E_{q^{k+1}(\mathbf{r})}[(\Delta_i^v(\mathbf{r} - E_{q^{k+1}(\mathbf{r})}[\mathbf{r}]))^2] \\ &= \frac{1}{N} \text{trace} \left[\text{cov}_{q^{k+1}(\mathbf{r})}[\mathbf{r}] \times ((\Delta^h)^t(\Delta^h) + (\Delta^v)^t(\Delta^v)) \right]. \end{aligned} \quad (44)$$

From (42) we notice that estimated u_i^{k+1} is the mean of $(\Delta_i^h(\mathbf{r}))^2 + (\Delta_i^v(\mathbf{r}))^2$, which can be considered as the condition for equality in the inequality (22). Since the calculation of u_i^{k+1} involves the horizontal and vertical gradient operations applied to \mathbf{r} , the values in matrix $W(\mathbf{u}^{k+1})$ represent spatial information in image \mathbf{r} . The matrix $W(\mathbf{u}^{k+1})$ is also called the spatial adaptivity matrix [23] or the visibility matrix [40] and has been used in image deconvolution [41] and super resolution [26].

For the estimation of the hyperpriors, from (29) we obtain

$$q^{k+1}(h) \propto \exp \left(E_{q^{k+1}(\mathbf{r})} q^k(\Theta_{h,\mathbf{r}}) [\log \hat{p}(\mathbf{s}, \mathbf{r}^{k+1}, h, \Theta_{h,\mathbf{r}}, \mathbf{u}^{k+1})] \right), \quad h \in \{\alpha, \beta, \sigma\}. \quad (45)$$

Then the distributions of $q(\alpha)$, $q(\beta)$, and $q(\sigma)$ are found as gamma distributions, given by

$$q^{k+1}(\alpha) \propto \alpha^{N/2 + a_\alpha^0 - 1} \exp \left[-\alpha \left(b_\alpha^0 + \sum_i^N \sqrt{u_i^{k+1}} \right) \right], \quad (46)$$

$$\begin{aligned} q^{k+1}(\beta) &\propto \beta^{N/2 + a_\beta^0 - 1} \\ &\quad \times \exp \left[-\beta \left(b_\beta^0 + E_{q^{k+1}(\mathbf{r})}(\mathbf{l} - \mathbf{r} - \mathbf{s})^2/2 \right) \right], \end{aligned} \quad (47)$$

$$\begin{aligned} q^{k+1}(\sigma) &\propto \sigma^{N/2 + a_\sigma^0 - 1} \\ &\quad \times \exp \left[-\sigma \left(b_\sigma^0 + \sum_i^N (\Delta_i^h(\mathbf{l}))^2 + (\Delta_i^v(\mathbf{l}))^2 \right) \right]. \end{aligned} \quad (48)$$

The means of these gamma distributions are

$$E_{q^{k+1}(\alpha)}[\alpha] = \frac{N/2 + a_\alpha^0}{\sum_i^N \sqrt{u_i^{k+1}} + b_\alpha^0}, \quad (49)$$

$$E_{q^{k+1}(\beta)}[\beta] = \frac{N/2 + a_\beta^0}{E_{q^{k+1}(\mathbf{r})}(\mathbf{l} - \mathbf{r} - \mathbf{s})^2/2 + b_\beta^0}, \quad (50)$$

$$E_{q^{k+1}(\sigma)}[\sigma] = \frac{N/2 + a_\sigma^0}{\sum_i^N (\Delta_i^h(\mathbf{l}))^2 + (\Delta_i^v(\mathbf{l}))^2 + b_\sigma^0}, \quad (51)$$

where $E_{q^{k+1}(\mathbf{r})}(\mathbf{l} - \mathbf{r} - \mathbf{s})^2 = (\mathbf{l} - \mathbf{r} - \mathbf{s})^2 + \text{trace}(\text{cov}_{q^{k+1}(\mathbf{r})}[\mathbf{r}])$.

For the computation of (44) and (50), the explicit form of the inverse matrix $\text{cov}_{q^{k+1}(\mathbf{r})}[\mathbf{r}]$ is needed. We can also solve

the problem similar to the method in [26] that approximating $\text{cov}_{q^{k+1}}(\mathbf{r})$ by a diagonal matrix whose elements are the inversion of the diagonal elements of $(\text{cov}_{q^{k+1}}(\mathbf{r}))^{-1}$. A large number of experiments are also made to verify the validity of this approximation. The images whose sizes are small enough to compute the explicit form of the inverse matrix are used in these experiments, and similar to the results in [26], this approximation approach also has a minor effect in the estimation process.

Notice that by denoting $\bar{\alpha} = a_\alpha^0/b_\alpha^0$, $\bar{\beta} = a_\beta^0/b_\beta^0$, $\bar{\sigma} = a_\sigma^0/b_\sigma^0$, and $\gamma_\alpha = a_\alpha^0/(a_\alpha^0 + N/2)$, $\gamma_\beta = a_\beta^0/(a_\beta^0 + N/2)$, $\gamma_\sigma = a_\sigma^0/(a_\sigma^0 + N/2)$, we can rewrite (49)-(51) as

$$(E_{q^{k+1}}(\alpha)[\alpha])^{-1} = \gamma_\alpha \frac{1}{\bar{\alpha}} + (1 - \gamma_\alpha) \frac{\sum_i^N \sqrt{u_i^{k+1}}}{N/2}, \quad (52)$$

$$(E_{q^{k+1}}(\beta)[\beta])^{-1} = \gamma_\beta \frac{1}{\bar{\beta}} + (1 - \gamma_\beta) \frac{E_{q^{k+1}}(\mathbf{r})(\mathbf{l} - \mathbf{r} - \mathbf{s})^2}{N}, \quad (53)$$

$$(E_{q^{k+1}}(\sigma)[\sigma])^{-1} = \gamma_\sigma \frac{1}{\bar{\sigma}} + (1 - \gamma_\sigma) \frac{\sum_i^N (\Delta_i^h(\mathbf{l}))^2 + (\Delta_i^v(\mathbf{l}))^2}{N/2}, \quad (54)$$

where γ_h , $h \in \{\alpha, \beta, \sigma\}$ can be viewed as a normalized confidence parameter and its value lies in $[0, 1]$. From (52)–(54), we can see that $(E_{q^{k+1}}(h)[h])^{-1}$, $h \in \{\alpha, \beta, \sigma\}$ is a convex combination of the initial estimate of h and the maximum likelihood estimate of h . When $\gamma_h = 0$, the estimation of h is completely dependent on its maximum likelihood estimate. When $\gamma_h \rightarrow 1$ the estimation of h is dependent on its initial value and no updates will be made to h . Usually, the original estimates of parameters are random, and the estimates may have oscillations in image processing. Therefore, the introduction of the confidence parameter may balance the effects of the initial estimate and the maximum likelihood estimate and lead to a robust estimation.

Once the logarithmic illumination \mathbf{l} and the logarithmic reflectance \mathbf{r} are obtained, we need to reconstruct the enhanced image. We note that the reflectance image obtained by the Retinex is sometimes an over-enhanced image [8], [21], we first take an exponential transform to obtain the illumination image $\mathbf{L} = \exp(\mathbf{l})$ and apply a Gamma correction to \mathbf{L} as $\mathbf{L}' = \mathbf{V} \cdot [\mathbf{L}/\mathbf{V}]^{1/\gamma}$, where \mathbf{V} is the white value which is equal to 255 in 8-bit images, and γ is the adjusting parameter. Then we use the gamma-corrected illumination \mathbf{L}' to reconstruct the final result image \mathbf{S}' as $\mathbf{S}' = \mathbf{L}' \cdot \mathbf{R} = \mathbf{L}' \cdot \frac{\mathbf{S}}{\mathbf{L}} = \frac{\mathbf{S}}{(\mathbf{L}/\mathbf{V})^{1-1/\gamma}}$.

B. Relation Between the Proposed Method and the TV Variation Model for Retinex

The energy functional in the TV variation model for Retinex in [21] is given by

$$\begin{aligned} E(\mathbf{r}, \mathbf{l}) = & \int_{\Omega} |\nabla \mathbf{r}| + \frac{\alpha}{2} \int_{\Omega} |\nabla \mathbf{l}|^2 dx + \frac{\beta}{2} \int_{\Omega} (\mathbf{l} - \mathbf{r} - \mathbf{s})^2 dx \\ & + \frac{\mu}{2} \int_{\Omega} \mathbf{l}^2 dx, \end{aligned} \quad (55)$$

where the first term is the TV regularization, the second term represents the spatial smoothness prior on the illumination \mathbf{l} ,

the third term is used for the fidelity, and the last term is used only for the theoretical setting. Since the parameter μ can be set to a very small value or even zero, we do not take the last term into consideration. Then, we can easily find the corresponding relations between (7), (14), (18), and the first three terms in (55), respectively. The corresponding terms have similar functional form and play similar roles in their own models. Therefore, these two methods are likely to produce similar results, and we will confirm this conjecture by the experimental results in the next Section. The difference between the proposed method and the TV variation model for Retinex is that the proposed method treats the parameters as variables and assumes the hyperpriors on these hyperparameters, while the TV variation model for Retinex requires that these parameters are given by initial setting. Obviously, the hyperparameters in the proposed method are adapted to the input images and do not need to be set manually.

Furthermore, the proposed method provides a new perspective on Retinex. The proposed method starts with the image priors which are formed based on the intrinsic characteristics of the illumination and the reflectance, and then it separates the illumination from the reflectance by the variational Bayesian method. Based on reasonable assumptions on illumination and reflectance, various image priors can be used in the variational Bayesian method for Retinex, such as, the super-Gaussian distribution for sparse image priors, the simultaneous autoregressive model for smooth image priors, and so on. Thus, we can further construct a flexible variational Bayesian framework for Retinex, and the proposed method which uses the TV image prior for the reflectance image and a smooth image prior for the illumination image can be viewed as a special case of this framework.

V. EXPERIMENTAL RESULTS

In this section, we present the experimental results of our method, in comparison to other relevant methods, to demonstrate the performance and the effectiveness of the proposed method in non-uniform lighting image enhancement.

We first compare our variational Bayesian Retinex (VBR) method with some Retinex methods including the SSR [15], the MSR [16], [22], the MSRCR [22], the McCann's Retinex (MR) method [13], and the TV-Retinex (TVR) [21]. In the SSR, the scale parameter of the Gaussian is set to 200, and in the MSR the scale parameters are 20, 180 and 360. The results of the MSRCR are obtained by using an image processing software GIMP; the parameters scale, scale division, and dynamic are set to 250, 3, and 1.6, respectively. In the MR, the number of iterations is set to 3. In the TVR, the parameters are set according to [21], i.e. $\alpha = 1$, $\beta = 0.1$, $\mu = 10^{-5}$, $\lambda = 1$. In the VBR the max iteration number iter_{\max} and the iteration stopping criterion parameter ξ_{stop} are set to 50 and 10^{-4} , respectively. The adjusting parameter γ is set to 2.2 in both the TVR and the VBR. For the initial values in the VBR, we use the original image convoluted with a Gaussian lowpass filter with size 5 and standard deviation 3 as the initial value of \mathbf{L} to avoid a zero denominator at the beginning of the algorithm. In practice we do not need to initialize

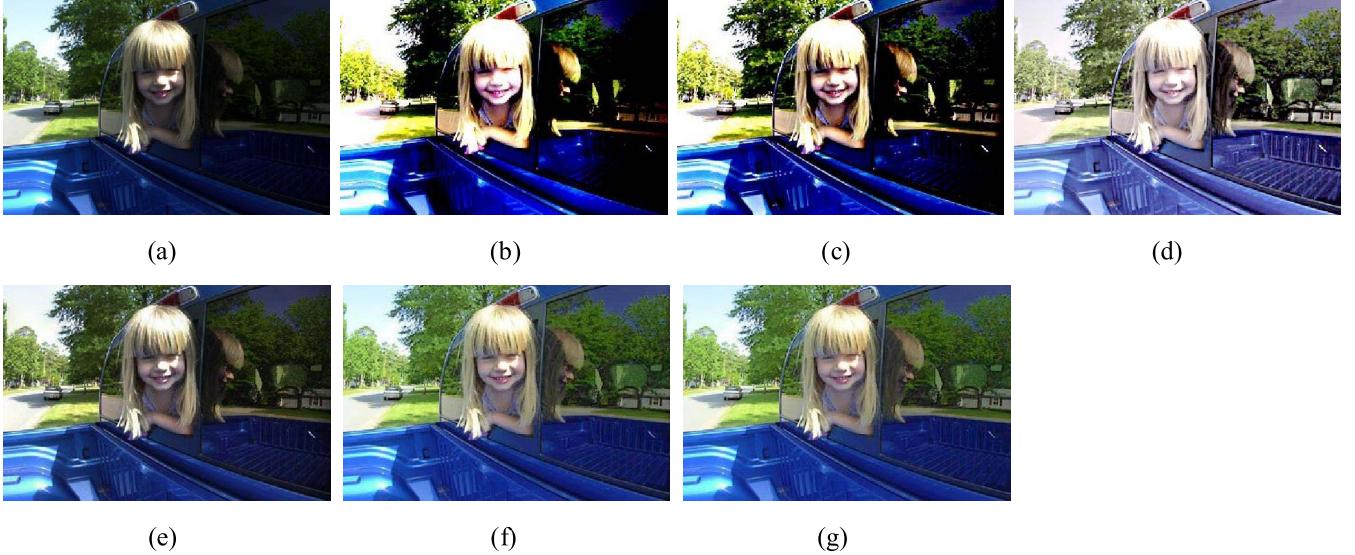


Fig. 2. Girl image. (a) Original image; (b) result by the SSR; (c) result by the MSR; (d) result by the MSRCR; (e) result by the MR; (f) result by the TVR; (g) result by the VBR.

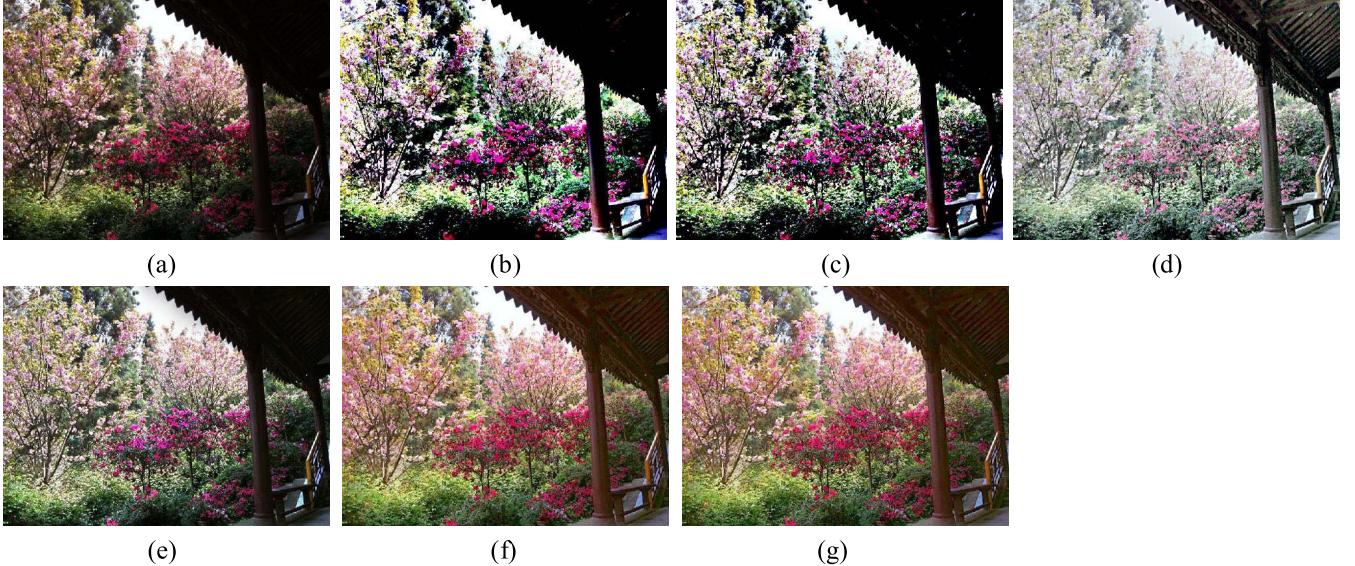


Fig. 3. Flowers image. (a) Original image; (b) result by the SSR; (c) result by the MSR; (d) result by the MSRCR; (e) result by the MR; (f) result by the TVR; (g) result by the VBR.

the probabilities $q^1(\Theta_1)$, we only initialize their means. The initial value of $E_{q^1(\mathbf{r})}[\mathbf{r}]$ is calculated by $\mathbf{r} = \log(\mathbf{L}) - \log(\mathbf{S})$. The initial values of $E_{q^1(\alpha)}[\alpha]$, $E_{q^1(\beta)}[\beta]$, and $E_{q^1(\sigma)}[\sigma]$ are calculated from (49), (50), and (51), respectively, and we set a_h^0 and b_h^0 , $h \in \{\alpha, \beta, \sigma\}$ to very small positive numbers, i.e. $a_h^0 = b_h^0 = \varepsilon_\alpha$, $a_\beta^0 = b_\beta^0 = \varepsilon_\beta$, and $a_\sigma^0 = b_\sigma^0 = \varepsilon_\sigma$, where ε_α , ε_β , and ε_σ are very small positive numbers. The initial value of \mathbf{u} is calculated from (42).

In our experiments, a large number of images are tested. Due to space limitations here we show some of the test images. Figs. 2(a) to 5(a) are the original images. The Girl image contains a human face with non-uniform illumination. The Flowers image contains a lot of details and textures, such as branches, leaves, and pavilion. The Path image is taken at night with point-source of light in dark areas. The Window image is

a backlit image with dark view indoor and bright view outdoor. Figs. 2(b)-(g) to 5(b)-(g) are the results obtained by the SSR, the MSR, the MSRCR, the MR, the TVR, and the VBR, respectively. We see from Figs. 2 to 5 that the results obtained by the SSR and the MSR look the same and have much higher contrast than other results. The results obtained by the MSRCR are not as colorful as other results, for example, the color of the sky at the top right corner of the Girl image are changed from blue to white in Fig. 2(d). The visual effect of the results obtained by the MR is not good in dark areas. The TVR and VBR produce similar results, and in terms of visual effects, the results obtained by these two methods provide better quality than the results obtained by other methods, not only in the case of outdoor images test but also in the case of indoor images test and night images test.

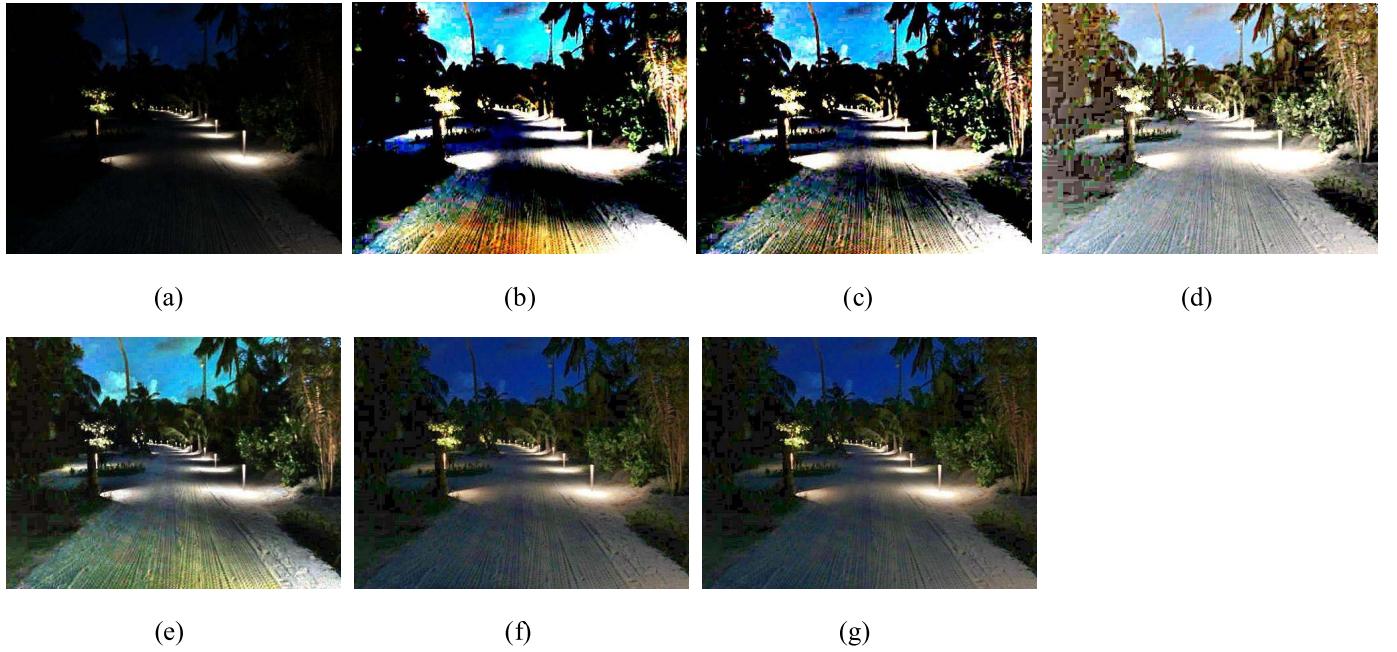


Fig. 4. Path image. (a) Original image; (b) result by the SSR; (c) result by the MSR; (d) result by the MSRCR; (e) result by the MR; (f) result by the TVR; (g) result by the VBR.

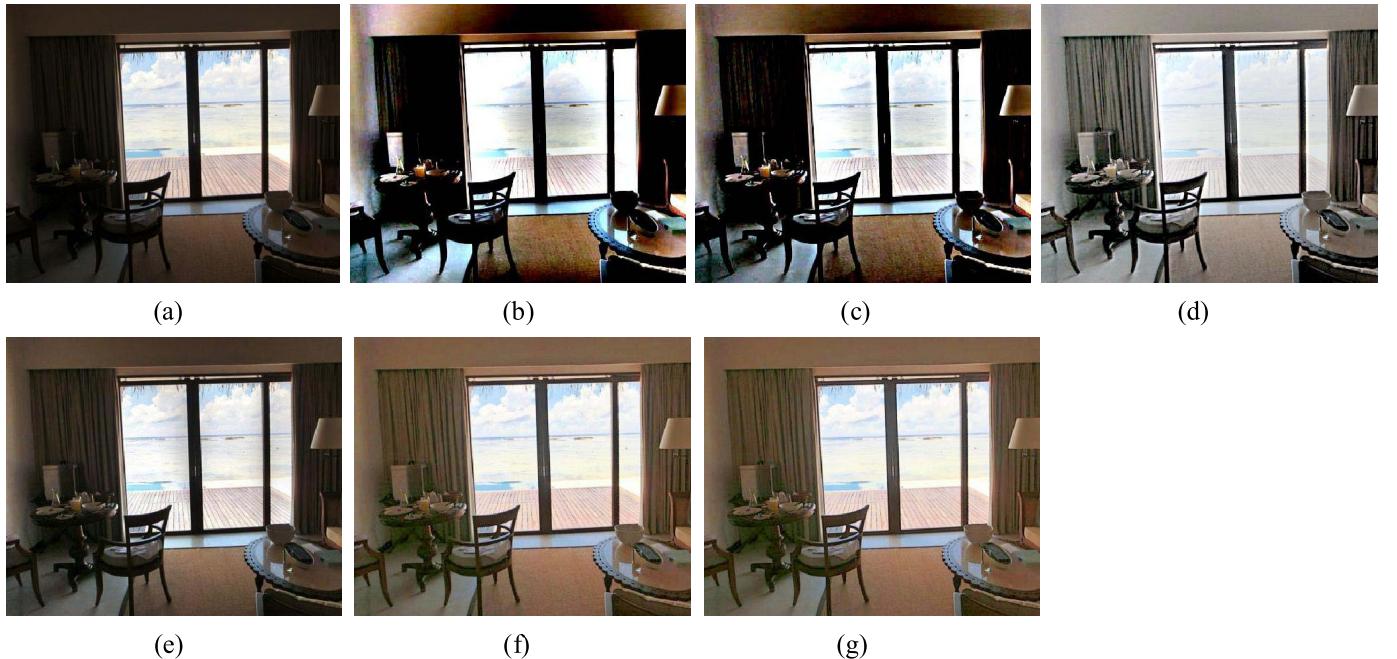


Fig. 5. Window image. (a) Original image; (b) result by the SSR; (c) result by the MSR; (d) result by the MSRCR; (e) result by the MR; (f) result by the TVR; (g) result by the VBR.

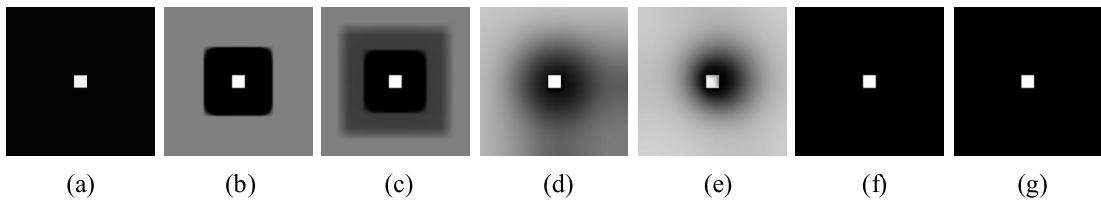


Fig. 6. Synthetic image. (a) Original image; (b) result by the SSR; (c) result by the MSR; (d) result by the MSRCR; (e) result by the MR; (f) result by the TVR; (g) result by the VBR.

We further test the effect of local details preservation of the Retinex methods we compared. Fig. 6 shows the test on a synthetic image, we see that the TVR and the proposed VBR

methods retain the small white block at the center of the big black square, while the SSR, the MSR, the MSRCR, and the MR produce a disordered boundary. The similar conclusion

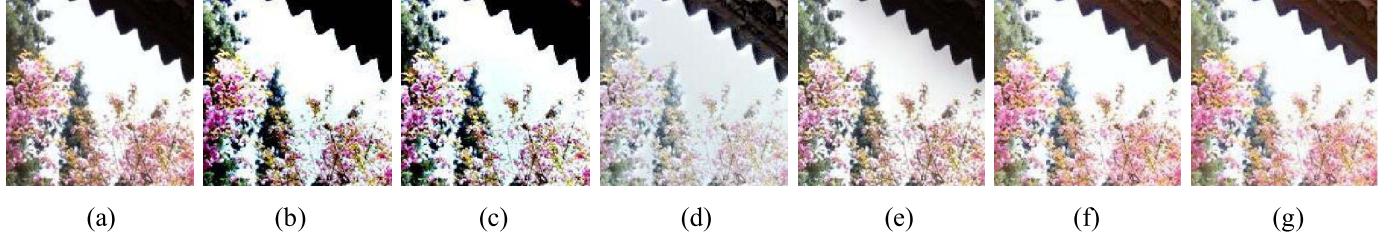


Fig. 7. Partial enlargements of the images in Fig. 3. (a) Original image; (b) result by the SSR; (c) result by the MSR; (d) result by the MSRCR; (e) result by the MR; (f) result by the TVR; (g) result by the VBR.

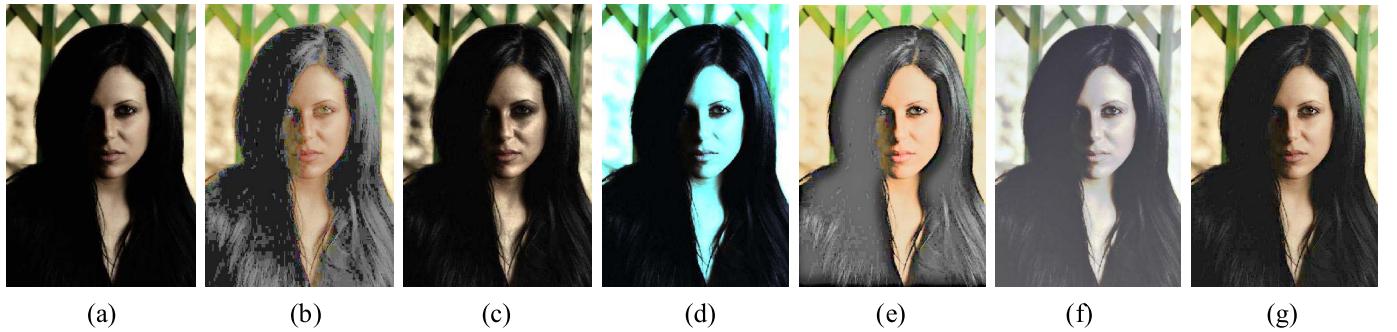


Fig. 8. Woman image. (a) Original image; (b) result by the HE; (c) result by the CLAHE; (d) result by the GLG; (e) result by the AINDANE; (f) result by the ACE; (g) result by the VBR.

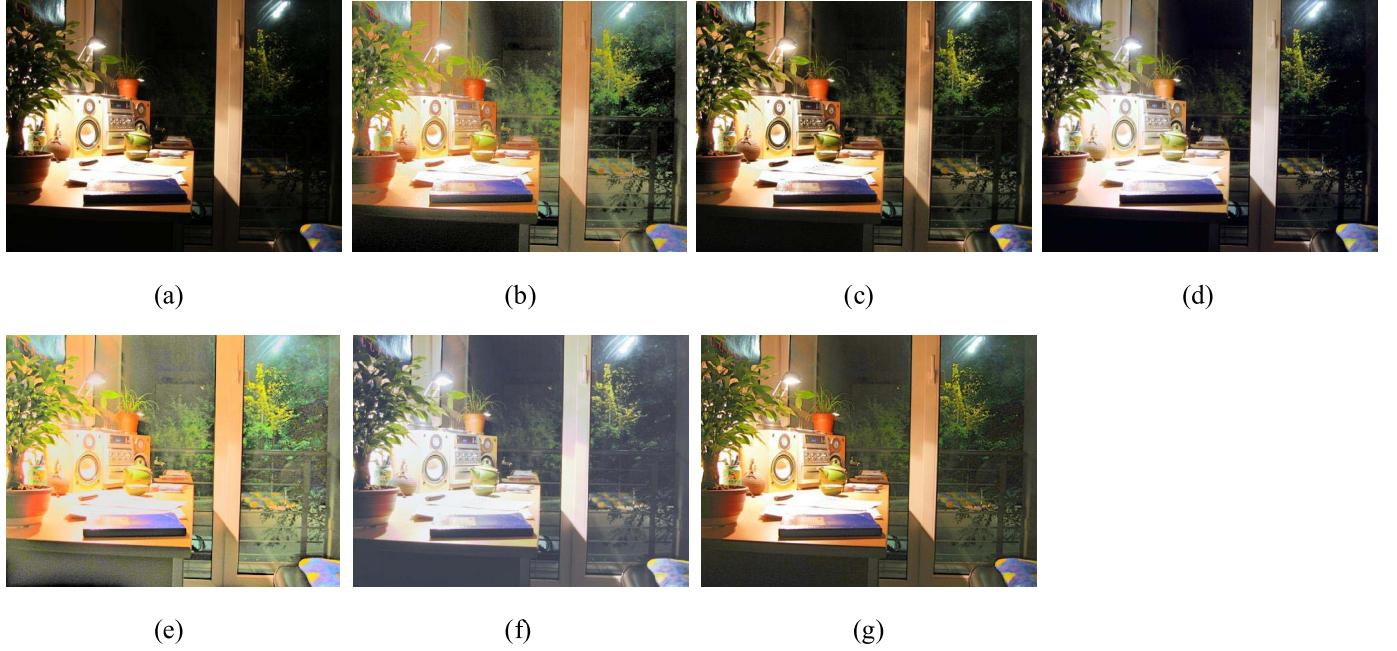


Fig. 9. Desk image. (a) Original image; (b) result by the HE; (c) result by the CLAHE; (d) result by the GLG; (e) result by the AINDANE; (f) result by the ACE; (g) result by the VBR.

can also be reached in the test on natural images. Fig. 7 shows the partial enlargements of the original Flowers image and the results in Fig. 3. We can see that the SSR and the MSR do not provide illumination compensation to the dark areas; the MSRCR and the MR also has a disorder effect around the contrasty edges; the TVR and the VBR preserve local details and make objectives such as flowers, leaves and edge of eaves distinguishable.

In the next set of experiments, we compare the proposed the VBR method with some non-Retinex methods including the conventional histogram equalization (HE) [1], the contrast-limited adaptive histogram equalization (CLAHE) [2], the gray-level grouping (GLG) [3], the adaptive integrated neighborhood dependent approach for nonlinear enhancement (AINDANE) [4], and the automatic color enhancement (ACE) [42], [43]. In the ACE, the slope parameter alpha is set to 5.

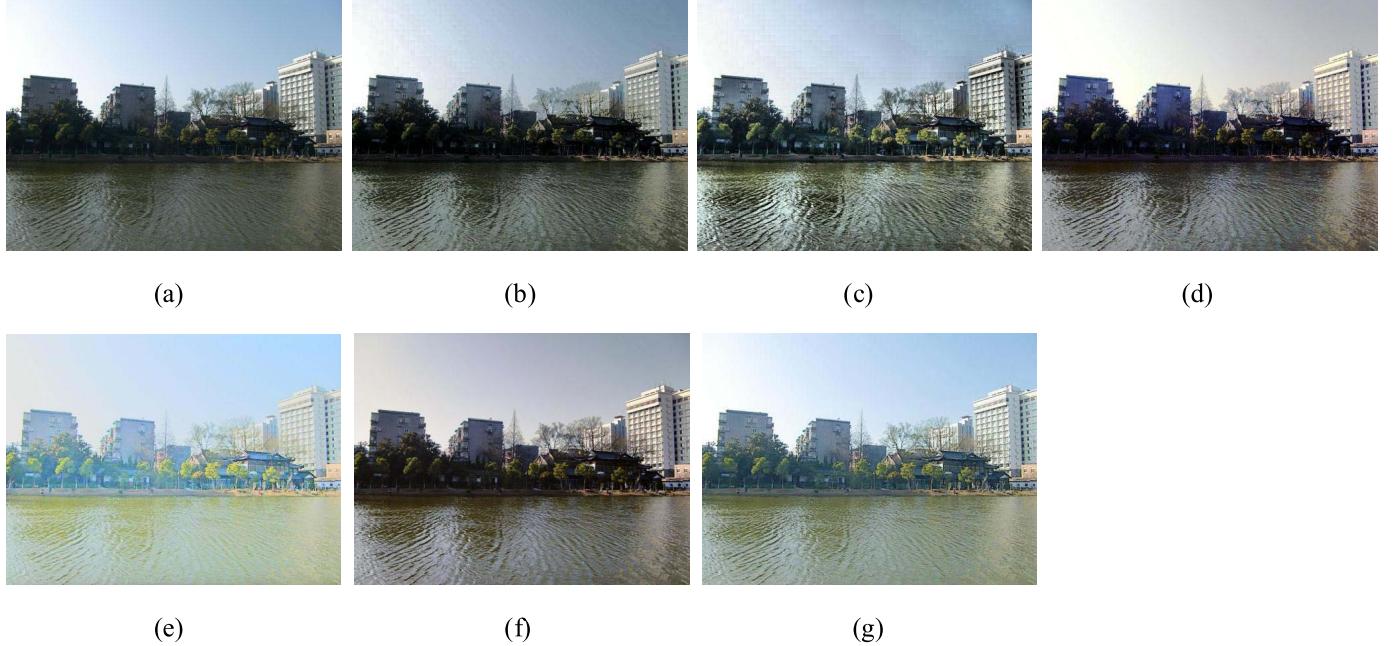


Fig. 10. Landscape image. (a) Original image; (b) result by the HE; (c) result by the CLAHE; (d) result by the GLG; (e) result by the AINDANE; (f) result by the ACE; (g) result by the VBR.

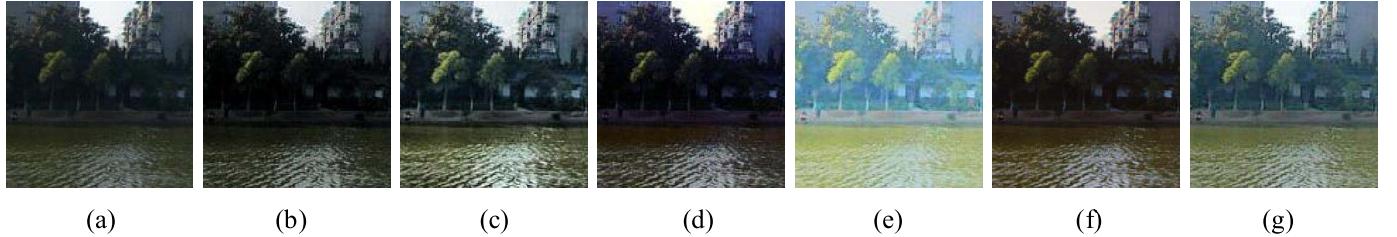


Fig. 11. Partial enlargements of the images in Fig. 10. (a) Original image; (b) result by the HE; (c) result by the CLAHE; (d) result by the GLG; (e) result by the AINDANE; (f) result by the ACE; (g) result by the VBR.

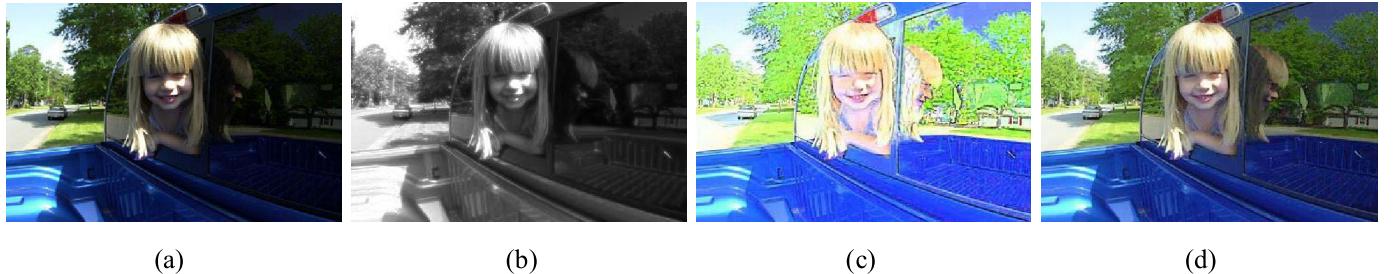


Fig. 12. (a) Original image; (b) illumination result by the VBR; (c) result with the reflection function derived from Algorithm 2; (d) result with the reflection function computed as $\mathbf{R} = \mathbf{S}/\mathbf{L}$ by the VBR.

Figs. 8(a) to 10(a) are the original images. Figs. 8(b)-(g) to 10(b)-(g) are the results obtained by the HE, the CLAHE, the GLG, the AINDANE, the ACE, and the VBR, respectively. We see from Figs. 8 to 10 that the HE produces some unexpected artifacts in images. For example, many black blocks appear on the woman's hair in Fig. 8(b). The results obtained by the AINDANE are too brighter than the other results, and then the details in the dark area look unnatural. The CLAHE and the GLG play a minor role in the illumination correction. There

is even strong color cast in the result obtained by the GLG as shown in Fig. 8(d). The ACE makes image colors change into gray. The VBR produces the most natural results, and makes image details clear and bright, such as the hair textures, the branches in Fig. 8(g) and Fig. 9(g), respectively. The partial enlargements of the original Landscape image and the results in Fig. 10 are also shown in Fig. 11. We can see that the proposed method is effective in illumination correction and detail preservation.

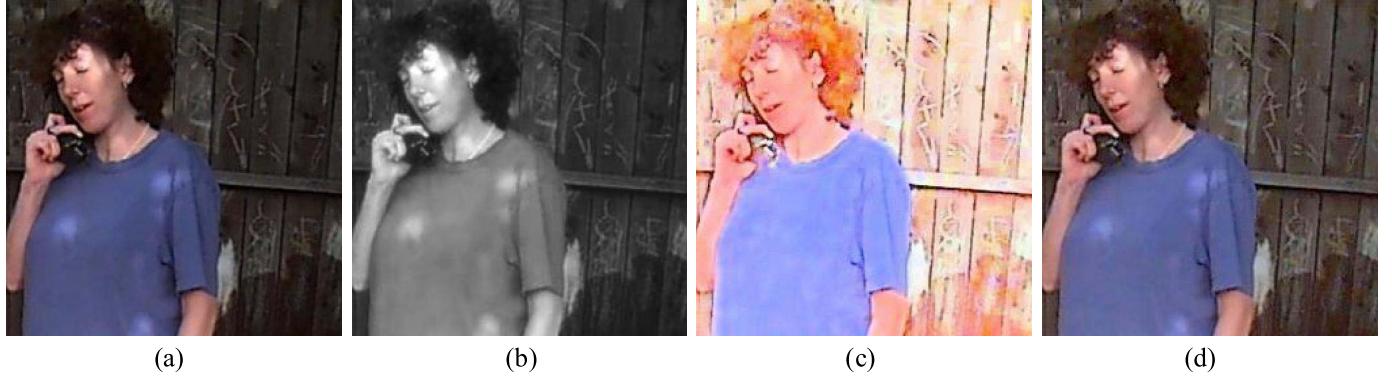


Fig. 13. (a) Original image; (b) illumination result by the VBR; (c) result with the reflection function derived from Algorithm 2; (d) result with the reflection function computed as $\mathbf{R} = \mathbf{S}/\mathbf{L}$ by the VBR.



Fig. 14. (a) Original image; (b) illumination result by the VBR; (c) result with the reflection function derived from Algorithm 2; (d) result with the reflection function computed as $\mathbf{R} = \mathbf{S}/\mathbf{L}$ by the VBR.

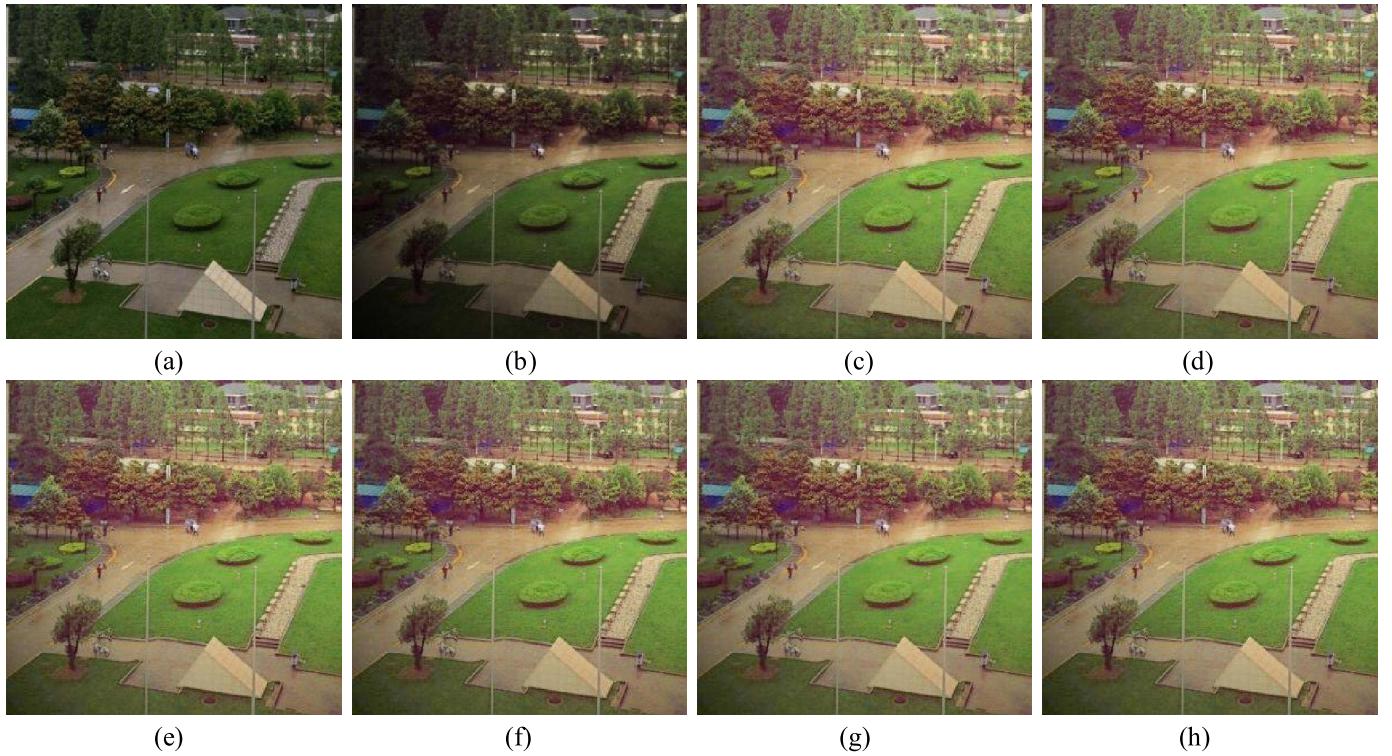


Fig. 15. (a) Original image; (b) synthetic shadowed image; (c) result by the TVR; (d) result by the VBR with $\gamma_\alpha = 0$, $\gamma_\beta = 0$, and $\gamma_\sigma = 0$; (e) result by the VBR with $\gamma_\alpha = 0$, $\gamma_\beta = 0$, and $\gamma_\sigma = 1$; (f) result by the VBR with $\gamma_\alpha = 0$, $\gamma_\beta = 1$, and $\gamma_\sigma = 1$; (g) result by the VBR with $\gamma_\alpha = 1$, $\gamma_\beta = 0$, and $\gamma_\sigma = 1$; (h) result by the VBR with $\gamma_\alpha = 1$, $\gamma_\beta = 1$, and $\gamma_\sigma = 1$.

Because the reflectance image obtained directly from Retinex is usually over enhanced, in practice we use the reflection function computed as $\mathbf{R} = \mathbf{S}/\mathbf{L}$. In this set of our

experiment, we test the differences between the results with these two types of reflection function and show the illumination images obtained by the VBR. We see from Figs. 12 to 14



Fig. 16. (a) Original image; (b) result by the VBR using (14) and (18); (c) result by the VBR using (8); (d) result by the VBR using (15).



Fig. 17. (a) Original image; (b) result by the VBR using (14) and (18); (c) result by the VBR using (8); (d) result by the VBR using (15).



Fig. 18. Partial enlargements of the images in Fig. 16 (left) and Fig. 17 (right). (a) Original image; (b) result by the VBR using (14) and (18); (c) result by the VBR using (8); (d) result by the VBR using (15).

TABLE I

NATURALNESS VALUES OBTAINED BY THE TVR AND THE VBR AND
MEANS OF POSTERIOR HYPERPARAMETER OBTAINED BY THE VBR
USING THE PRIOR HYPERPARAMETER AND DIFFERENT
CONFIDENCE PARAMETERS

Method	γ_α	γ_β	γ_σ	$E[\alpha]$	$E[\beta]$	$E[\sigma]$	Naturalness
TVR	-	-	-	-	-	-	0.7549
	0	0	0	13.85	2.40×10^6	29.91	0.7549
	0	0	1	13.94	2.54×10^6	26.49	0.7552
VBR	0	1	1	11.18	1.19×10^6	26.49	0.7553
	1	0	1	9.21	2.55×10^6	26.49	0.7554
	1	1	1	9.21	1.19×10^6	26.49	0.7554

that the results with the reflection functions derived from Algorithm 2 are obviously over enhanced and some details are smoothed out. The reflection function computed as $\mathbf{R} = \mathbf{S}/\mathbf{L}$ results in good visual effects and preserves many details. In addition the illumination images obtained by the VBR present the illumination of the objects as accurately as possible without obvious halo effects near the edges.

In this set of experiments, we test the effect of prior information and the role of the confidence parameters in the proposed method. As we have already explained in Section IV, when $\gamma_h = 0$, $h \in \{\alpha, \beta, \sigma\}$, the estimation of the hyperparameters and the image only depends on the observed image. However, some information about the original image and the imaging parameters can usually be obtained. This information can be

used in the estimation. In this test, because it is difficult to give an exact definition of a perfect lighting image, for simulation purposes, we use an image with synthetic unnatural shadow. A naturalness metric [44] which describes the degree of natural of an image is also used to quantitatively evaluate the results. The naturalness metric varies from 0 (completely unnatural image) to 1 (completely natural image). Thus the original image can be used to compute the prior hyperparameter values $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\sigma}$ which can be used as prior information in the estimation of the hyperparameters. Fig. 15(a) is the original image. We add a synthetic non-uniform shadow on Fig. 15(a) to obtain Fig. 15(b), and then we enhance Fig. 15(b) by the TVR and the VBR, respectively. Fig. 15(c) is the result obtained by the TVR. Figs. 15(d)-(h) are the results obtained by the VBR with the prior hyperparameter values $\bar{\alpha} = 9.21$, $\bar{\beta} = 1.19 \times 10^6$, $\bar{\sigma} = 26.492$ and different confidence parameters. When the confidence parameters $\gamma_h = 1$, $h \in \{\alpha, \beta, \sigma\}$, the estimation of the hyperparameters and the image is dependent on the prior hyperparameter values. The different confidence parameters and corresponding means of the posterior distributions of the hyperparameters and the naturalness values are shown in Table I. From this table we see that when no prior information is added (i.e. $\gamma_\alpha = \gamma_\beta = \gamma_\sigma = 0$) the results obtained by the VBR and the TVR have the same naturalness value, and when some prior hyperparameter value is available, the VBR is able to produce the results having higher naturalness value than that of the result obtained by the TVR.

TABLE II
AVERAGE COMPUTATIONAL TIME OF THE METHODS IMPLEMENTED BY
THE MATLAB FOR AN 236×360 IMAGE

Method	SSR	MSR	MR	TVR	VBR
Time (sec)	0.3652	0.7749	54.96	1.378	9.044
Method	HE	CLAHE	GLG	AINDANE	
Time (sec)	0.04306	0.1513	0.3691	0.1389	

In the last set of our experiments, we test two different priors (8) and (15) which we mentioned in the first part of Section III. We use (7), (18), and (19) for testing (8), and use (7), (14) and (19) for testing (15). The computational details to deduce the expressions of the approximations of posterior distributions are analogous to those in the previous subsection. We sacrifice these details for the sake of brevity; the interested reader may consult the literatures [24], [25], [31], [34] for further details. The initial value of $q^1(\mathbf{r})$ is calculated as in the previous set of experiments. The initial values of the other approximate posteriors are calculated from corresponding priors. Figs. 16(a) to 18(a) are the original images and corresponding partial enlargements. Figs. 16(b)-(d) to 18(b)-(d) are the results obtained by the VBR using the priors (14) and (18), the prior (8), and the prior (15), respectively. We see from the results that the image priors (8) and (15) works well in the variational Bayesian framework for Retinex. The obtained results by using (15) and (18) are almost the same. We can also see from Fig. 18 that because we only use the horizontal first order difference operator in (8), the prior (8) is not able to preserve as much details as the TV prior which contains both the horizontal and the vertical first order difference operators.

We finally conclude this section by showing the computational time of the methods. Besides the MSRCR and the ACE, the other nine methods are implemented by the MATLAB on a quad-core 2.8GHz Xeon PC, and for an 236×360 image we only show the average computational time of these nine methods in Table II. The VBR has longer computational time than most of the other methods, due to solving linear equations. However, by utilizing preconditioning techniques [45], [46] and C language, the computational time of the proposed method can be improved.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed a variational Bayesian method for Retinex. Based on the assumptions on the reflectance and the illumination, we construct the prior models on images. The hyperpriors on the hyperparameters are also constructed, and then the unknown illumination image, the reflectance image, and hyperparameters for the priors of the hidden variables are simultaneously estimated by using a hierarchical Bayesian model. Computational difficulty of the posterior distributions of the unknown variables is overcome by using a variational Bayesian approximation approach. The relationship between the proposed method and the TV variation model for Retinex is also analyzed. Experimental results demonstrate that the proposed method achieves better

or comparable performance when no additional information is included, and when some prior information is available the performance of the proposed method is improved. In our future work we will focus on how to select appropriate priors which are adapted to image content.

APPENDIX

CALCULATION OF THE DIFFERENTIATION IN ALGORITHM 2

In order to obtain the minimizer $q(\Theta)$ (for simplicity here we omit the superscript k), we need to differentiate the integrals in (24), (26), and (29) with respect to the corresponding $q(\Theta)$, and setting them equal to zero. Due to space limitations here we only give the calculation details of $q(\mathbf{l})$. The details to deduce the expressions of the other approximate posteriors are similar to that of $q(\mathbf{l})$.

From (33) we know that

$$\hat{\mathcal{L}}_{q(\mathbf{l})}(q(\Theta), \mathbf{u}) = \int_{\mathbf{l}} q(\mathbf{l}) \log \frac{q(\mathbf{l})}{\exp(E_{q(\Theta)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta, \mathbf{u})]) / \lambda_i} d\mathbf{l} + \eta - \log \lambda_i,$$

where

$$\eta = E_{q(\Theta)}[\log q(\Theta)]$$

is independent of $q(\mathbf{l})$, λ_i is a normalization constant. Differentiating $\hat{\mathcal{L}}_{q(\mathbf{l})}(q(\Theta), \mathbf{u})$ with respect to $q(\mathbf{l})$ and setting it to zero, we obtain

$$\begin{aligned} \frac{\partial \hat{\mathcal{L}}_{q(\mathbf{l})}(q(\Theta), \mathbf{u})}{\partial q(\mathbf{l})} &= \log q(\mathbf{l}) - E_{q(\Theta)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta, \mathbf{u})] \\ &\quad + \log \lambda_i = 0, \end{aligned}$$

which implies $\log q(\mathbf{l}) = E_{q(\Theta)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta, \mathbf{u})] + C$, where C denotes a constant.

Since $\hat{p}(\mathbf{s}, \Theta, \mathbf{u}) = p(\alpha)p(\beta)p(\sigma)\hat{p}(\mathbf{r}, \alpha, \mathbf{u})p(\mathbf{l}|\sigma)p(\mathbf{s}|\mathbf{l}, \mathbf{r}, \beta)$, we then have

$$\begin{aligned} &\log q(\mathbf{l}) \\ &= E_{q(\Theta)}[\log \hat{p}(\mathbf{s}, \mathbf{l}, \Theta, \mathbf{u})] + C \\ &= E_{q(\Theta)}[\log p(\alpha) + \log p(\beta) + \log p(\sigma) \\ &\quad + \log \hat{p}(\mathbf{r}, \alpha, \mathbf{u}) + \log p(\mathbf{l}|\sigma) + \log p(\mathbf{s}|\mathbf{l}, \mathbf{r}, \beta)] + C \\ &= E_{q(\Theta)}[\log p(\mathbf{l}|\sigma) + \log p(\mathbf{s}|\mathbf{l}, \mathbf{r}, \beta)] + C_1 \\ &= E_{q(\Theta)}\left[\log\left(c_1\sigma^{N/2}\exp\left[-\sigma\sum_i^N (\Delta_i^h(\mathbf{l}))^2 + (\Delta_i^v(\mathbf{l}))^2\right]\right)\right] \\ &\quad + E_{q(\Theta)}\left[\log\left(\beta^{N/2}\exp\left[-\frac{\beta}{2}(\mathbf{l} - \mathbf{s} - \mathbf{r})^2\right]\right)\right] + C_1 \\ &= E_{q(\Theta)}\left[-\sigma\sum_i^N (\Delta_i^h(\mathbf{l}))^2 + (\Delta_i^v(\mathbf{l}))^2\right] \\ &\quad + E_{q(\Theta)}\left[-\frac{\beta}{2}(\mathbf{l} - \mathbf{s} - \mathbf{r})^2\right] + C_2 \\ &= -E_{q(\sigma)}[\sigma]\left[(\Delta_i^h)^t(\Delta_i^h) + (\Delta_i^v)^t(\Delta_i^v)\right]\mathbf{l}^2 \\ &\quad - \frac{1}{2}E_{q(\beta)}[\beta]\mathbf{l}^2 + E_{q(\beta)}[\beta]\mathbf{l}(\mathbf{r} + \mathbf{s}) \\ &\quad - \frac{1}{2}E_{q(\beta)}[\beta](\mathbf{r} + \mathbf{s}) + C_2 \\ &= -\frac{1}{2}[\text{cov}_{q(\mathbf{l})}[\mathbf{l}]]^{-1}[\mathbf{l} - E_{q(\mathbf{l})}[\mathbf{l}]]^2 + C_3, \end{aligned}$$

where $E_{q(\mathbf{l})}[\mathbf{l}]$ and $\text{cov}_{q(\mathbf{l})}[\mathbf{l}]$ are defined in (36) and (37), respectively, C_1 , C_2 , and C_3 are constants with respect to \mathbf{l} . The third, fifth, and seventh equations hold by putting the terms which are not the functions of \mathbf{l} into the constants. It is clear that $q(\mathbf{l}) = \mathcal{N}(\mathbf{l}|E_{q(\mathbf{l})}[\mathbf{l}], \text{cov}_{q(\mathbf{l})}[\mathbf{l}])$ and $E_{q(\mathbf{l})}[\mathbf{l}]$ and $\text{cov}_{q(\mathbf{l})}[\mathbf{l}]$ are the mean and covariance, respectively.

REFERENCES

- [1] R. A. Hummel, "Image enhancement by histogram transformation," *Comput. Graph. Imag. Process.*, vol. 6, no. 2, pp. 184–195, Apr. 1977.
- [2] Z. Karel, "Contrast limited adaptive histogram equalization," in *Graphic Gems IV*. San Diego, CA, USA: Academic, 1994, pp. 474–485.
- [3] Z. Y. Chen, B. R. Abidi, D. L. Page, and M. A. Abidi, "Gray-level grouping (GLG): An automatic method for optimized image contrast enhancement-part I: The basic method," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2290–2302, Aug. 2006.
- [4] L. Tao and V. Asari, "An integrated neighborhood dependent approach for nonlinear enhancement of color images," in *Proc. Int. Conf. ITCC*, Apr. 2004, vol. 2, pp. 138–139.
- [5] E. H. Land and J. McCann, "Lightness and the retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, Jan. 1971.
- [6] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–128, Dec. 1977.
- [7] W. Ma and S. Osher, "A TV Bregman iterative model of retinex theory," *Comput. Appl. Math.*, Univ. California, Los Angeles, CA, USA, Rep. 10-13, 2010.
- [8] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for retinex," *Int. J. Comput. Vis.*, vol. 52, no. 1, pp. 7–23, Apr. 2003.
- [9] J. M. Morel, A. B. Petro, and C. Sbert, "A PDE formalization of retinex theory," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2825–2837, Nov. 2010.
- [10] E. Provenzi, L. D. Carli, A. Rizzi, and D. Marini, "Mathematical definition and analysis of the retinex algorithm," *J. Opt. Soc. Amer. A*, vol. 22, no. 12, pp. 2613–2621, Dec. 2005.
- [11] E. Provenzi, M. Fierro, A. Rizzi, L. De Carli, D. Gadia, and D. Marini, "Random spray retinex: A new retinex implementation to investigate the local properties of the model," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 162–171, Jan. 2007.
- [12] J. Franke and J. McCann, "Method and apparatus for lightness imaging," U.S. Patent 4 384 336, May 17, 1983.
- [13] J. J. McCann, "Lessons learned from Mondrians applied to real images and color gamuts," in *Proc. 7th Color Imag. Conf.*, Jan. 1999, pp. 1–8.
- [14] B. Funt, F. Ciurea, and J. McCann, "Retinex in MATLAB," *J. Electron. Imag.*, vol. 13, no. 1, pp. 48–57, Jan. 2004.
- [15] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of the center/surround retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.
- [16] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *Proc. ICIP*, Sep. 1996, pp. 1003–1006.
- [17] Z. U. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," *J. Electron. Imag.*, vol. 13, no. 1, pp. 100–110, Jan. 2004.
- [18] A. Blake, "Boundary conditions for lightness computation in mondrian world," *Comput. Graph. Imag. Process.*, vol. 32, pp. 314–327, Dec. 1985.
- [19] B. K. P. Horn, "Determining lightness from an image," *Comput. Graph. Imag. Process.*, vol. 3, no. 4, pp. 277–299, Dec. 1974.
- [20] M. Elad, R. Kimmel, D. Shaked, and R. Keshet, "Reduced complexity retinex algorithm via the variational approach," *J. Vis. Commun. Imag. Represent.*, vol. 14, no. 4, pp. 369–388, Dec. 2003.
- [21] M. K. Ng and W. Wang, "A total variation model for retinex," *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 345–365, Feb. 2011.
- [22] D. J. Jobson, Z. U. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Imag. Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.
- [23] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. Imag. Process.*, vol. 17, no. 3, pp. 326–339, Mar. 2008.
- [24] R. Molina, J. Mateos, and A. K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3715–3727, Dec. 2006.
- [25] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian blind deconvolution using a total variation prior," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 12–26, Jan. 2009.
- [26] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian super resolution," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 984–999, Apr. 2011.
- [27] J. W. Miskin, "Ensemble learning for independent component analysis," Ph.D. dissertation, Astroph. Group, Cavendish Lab., Univ. Cambridge, Cambridge, U.K., 2000.
- [28] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, Mar. 1951.
- [29] V. Šmíd and A. Quinn, *The Variational Bayes Method in Signal Processing*. New York, NY, USA: Springer-Verlag, 2005.
- [30] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, Gatsby Comput. Neurosci. Unit, Univ. London, London, U.K., 2003.
- [31] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," in *Proc. ECCV*, 2012.
- [32] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.
- [33] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Adaptive total variation image deblurring: A majorization-minimization approach," *Signal Process.*, vol. 89, no. 9, pp. 1683–1693, Sep. 2009.
- [34] C. L. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," *IEEE Trans. Imag. Process.*, vol. 52, no. 8, pp. 2222–2233, Aug. 2004.
- [35] B. D. Ripley, *Spatial Statistics*. New York, NY, USA: Wiley, 1981, pp. 88–90.
- [36] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York, NY, USA: Springer-Verlag, 1985.
- [37] K. Lange, "Optimization," in *Springer Texts in Statistics*. New York, NY, USA: Springer-Verlag, 2004.
- [38] C. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, Mar. 1983.
- [39] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. MSR-TR-98-14, 1994.
- [40] G. L. Anderson and A. N. Netravali, "Image restoration based on a subjective criterion," *IEEE Trans. Syst., Man, Cybern.*, vol. 6, no. 6, pp. 845–853, Dec. 1976.
- [41] A. K. Katsaggelos and M. G. Kang, "A spatially adaptive iterative algorithm for the restoration of astronomical images," *Int. J. Imag. Syst. Technol.*, vol. 6, no. 4, pp. 305–313, 1995.
- [42] A. Rizzi, C. Gatta, and D. Marini, "A new algorithm for unsupervised global and local color correction," *Pattern Recognit. Lett.*, vol. 24, no. 11, pp. 1663–1677, Jul. 2003.
- [43] P. Getreuer, "Automatic color enhancement (ACE) and its fast implementation," *Imag. Process. Line*, vol. 2, pp. 266–277, Dec. 2012.
- [44] S. N. Yendrikhovski, F. J. J. Blommaert, and H. de Ridder, "Perceptually optimal color reproduction," in *Proc. SPIE*, vol. 3299, pp. 274–281, Jul. 1998.
- [45] R. H. Chan, T. F. Chan, and C.-K. Wong, "Cosine transform based preconditioners for total variation deblurring," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1472–1478, Oct. 1999.
- [46] M. K. Ng, H. Shen, E. Y. Lam, and L. Zhang, "A total variation regularization based super-resolution reconstruction algorithm for digital video," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 74585, Jun. 2007.



Liqian Wang received the B.S. and M.S. degrees from the Nanjing University of Science and Technology, Nanjing, China, in 2006 and 2008, respectively, where she is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering. Her research interest includes variational partial differential equation application in image processing, image restoration, image enhancement, and pattern recognition.

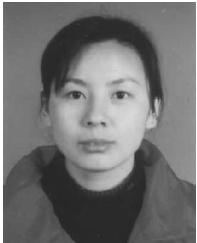


Liang Xiao (M'11) received the B.S. degree from the Department of Applied Mathematics, Nanjing University of Science and Technology (NUST), Nanjing, China, and the Ph.D. degree from the School of Computer Science and Technology, NUST, in 1999 and 2004, respectively.

From 2006 to 2008, he was a Post-Doctoral Researcher with the Pattern Recognition Laboratory at NUST. From 2009 to 2010, he was also a Post-Doctoral Researcher with the Rensselaer Polytechnic Institute, Troy, NY, USA. He is currently a Professor with the School of Computer Science and Engineering at NUST. He has authored more than 80 scientific papers in image processing, pattern recognition, and computer vision. His research interest includes variational partial differential equation application in image processing, image modeling, pattern recognition, motion estimation and tracking, virtual reality, and system simulation.



Zhihui Wei received the B.S. and M.S. degrees in mathematics from Southeast University, Nanjing, China, in 1983 and 1986, respectively, and the Ph.D. degree from the Department of Radio Electronics, Southeast University, in 2003. He is currently a Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing. His research interest covers image processing, image modeling, wavelet analysis, multiscale analysis, digital watermark, and image coding and compressing.



Hongyi Liu received the B.S. and M.S. degrees in computational mathematics from Northwest University, Xi'an, China, in 1997 and 2003, respectively, and the Ph.D. degree in pattern recognition and artificial intelligence from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2011. She is currently a Lecturer with the School of Science at NUST. Her research interest includes probabilistic graphics model, optimization algorithms in image processing, image analysis and image modeling, variational partial differential equation application in image processing, and compressive sensing.