Combining Total Variation and Nonlocal Variational Models for Low-Light Image Enhancement

Daniel Torres\textsuperscript{a}, Catalina Sbert\textsuperscript{b} and Joan Duran\textsuperscript{c}

Department of Mathematics and Computer Science & IAC3, University of the Balearic Islands, Cra. de Valldemossa, km. 7.5, E-07122 Palma, Illes Balears, Spain

Keywords: Low-Light Image Enhancement, Illumination Estimation, Reflectance Estimation, Retinex, Variational Method, Total Variation, Nonlocal Regularization.

Abstract: Images captured under low-light conditions impose significant limitations on the performance of computer vision applications. Therefore, improving their quality by discounting the effects of the illumination is crucial. In this paper, we present a low-light image enhancement method based on the Retinex theory. Our approach estimates illumination and reflectance in two steps. First, the illumination is obtained as the minimizer of an energy functional involving total variation regularization, which favours piecewise smooth solutions. Next, the reflectance component is computed as the minimizer of an energy functional involving contrast-invariant nonlocal regularization and a fidelity term preserving the largest gradients of the input image.

1 INTRODUCTION

Enhancing images captured under low-light conditions is crucial for many applications in computer vision. Various strategies have been proposed to tackle this problem (Wang et al., 2020), broadly classified into histogram equalization (Thepade et al., 2021; Paul et al., 2022), Retinex-based methods, fusion approaches (Fu et al., 2016a; Buades et al., 2020), and deep-learning techniques.

The Retinex theory (Land and McCann, 1971), which aims to explain and simulate how the human visual system perceives color independently of global illumination changes, has been an important basis for addressing low-light image enhancement. One of the most used models assumes that the observed image \( L \) is the product of the illumination \( T \), which depicts the light intensity on the objects, and the reflectance \( R \), which represents their physical characteristics:

\[
L = R \circ T,
\]

where \( \circ \) denotes pixel-wise multiplication. The illumination map is assumed to be smooth, while the reflectance component contains fine details and texture (Ng and Wang, 2011; Li et al., 2018).

Decomposing an image into illumination and reflectance is mathematically ill-posed. To address this issue, patch-based (Land and McCann, 1971), partial differential equations (Morel et al., 2010), center/surround (Jobson et al., 1996) and variational (Kimmel et al., 2001; Ng and Wang, 2011; Ma and Osher, 2012; Fu et al., 2016b; Guo et al., 2017; Gu et al., 2019) methods have been proposed.

Recently, the growing popularity of deep learning has lead to an increase in enhancement methods (Wei et al., 2018; Lv et al., 2021). However, the structure of these architectures is often non-intuitive, and their training and testing require high computational costs.

In this paper, we present a low-light image enhancement method taking into account the decomposition model given in (1). We propose to estimate illumination and reflectance separately. In the first step, the illumination component is obtained as the minimizer of an energy functional involving total variation (TV) regularization (Rudin et al., 1992), which favours piecewise smooth solutions. We assume that the channels of color images share the same illumination map. In a second step, the reflectance component is computed as the minimizer of an energy involving nonlocal regularization, which exploits image self-similarities, and a fidelity term preserving the large gradients of the input low-light image. Importantly, the nonlocal regularization depends on a weight function that is contrast-invariant.
2 RELATED WORK

In the variational framework, the enhanced image is computed as the minimizer of an energy functional that comprises data-fidelity and regularization terms. The latter quantifies the smoothness of the solution by usually prescribing priors on the gradient of the illumination or reflectance components.

Kimmel et al. (Kimmel et al., 2001) pioneered a variational model to estimate the illumination in a multiscale setting. The illumination is assumed to be spatially smooth, thus gradient oscillations are penalized through $L^2$ norm. Ma et al. (Ma and Osher, 2012) introduced TV to directly compute the reflectance. On the contrary, Ng et al. (Ng and Wang, 2011) estimate illumination and reflectance simultaneously.

One of the most celebrated works is LIME (Guo et al., 2017). The authors propose a simple TV variational model to estimate $T$. Then, the reflectance is computed as $R = \frac{L}{\gamma + |\nabla T|}$, where $\gamma > 0$ is a small constant and $|\nabla T|$ is the $\gamma$-corrected illumination. However, this strategy tends to amplify the noise, especially in dark regions. To overcome this issue, the authors proposed enhanced image $R \circ T^\gamma$, with $R$ being the denoised reflectance. Note that any discrepancy in the illumination will impact the result.

Several other variational methods use logarithms to linearize (1). However, this transformation amplifies errors in gradient terms. In (Fu et al., 2016b), weights are introduced to address issues arising from large gradients when either $R$ or $T$ is small.

3 PROPOSED MODELS

In this section, we introduce our low-light image enhancement method. We estimate $R$ and $T$ separately using two different variational models. Let $\Omega$ be an open and bounded domain in $\mathbb{R}^n$, with $n \geq 2$.

3.1 TV Model for Luminance

Based on the classical ROF denoising model (Rudin et al., 1992), $T$ is obtained as the minimizer of

$$\int_\Omega |DT| + \lambda \int_\Omega |T(x) - \bar{T}(x)|^2 dx,$$

where $\int_\Omega |DT|$ is the TV semi-norm, $\lambda > 0$ is a trade-off parameter, and $\bar{T}$ is an initial illumination estimate computed as $\bar{T}(x) = \max_{\epsilon \in \{R,G,B\}} L'(x)$. The assumption underlying TV is that images consist of connected smooth regions (objects) surrounded by sharp contours. Accordingly, TV is a good prior for the luminance since it is optimal to reduce noise and reconstruct the main geometrical shape.

The energy in (2) is derived from LIME (Guo et al., 2017). However, we calculate the minimizer for the exact functional instead of an approximation.

3.2 Nonlocal Model for Reflectance

To estimate the reflectance component, we use nonlocal regularization (Gilboa and Osher, 2009; Duran et al., 2014), which assumes that images are self-similar, thereby preserving fine details and texture.

3.2.1 Basic Definitions and Notations

Let $u = (u_1, \ldots, u_C) : \Omega \rightarrow \mathbb{R}^C$ be a color image, where $C$ denotes the number of channels. We also consider nonlocal functions $p = (p_1, \ldots, p_C) : \Omega \times \Omega \rightarrow \mathbb{R}$. Let $\omega : \Omega \times \Omega \rightarrow [0, 1]$ be a weight function, commonly defined in terms of differences between patches in $u$.

Definition 3.1. Given $u : \Omega \rightarrow \mathbb{R}^C$, its nonlocal gradient $\nabla_{\omega u} : \Omega \times \Omega \rightarrow \mathbb{R}^C$ is, for each $k \in \{1, \ldots, C\}$,

$$\nabla_{\omega u_k}(x, y) = \sqrt{\omega(x,y)} (u_k(y) - u_k(x)).$$

Given $p : \Omega \times \Omega \rightarrow \mathbb{R}$, its nonlocal divergence $div_{\omega p} : \Omega \rightarrow \mathbb{R}$ is, for each $k \in \{1, \ldots, C\}$,

$$(div_{\omega p_k}(x)) = \int_\Omega \left( p_k(x,y) \sqrt{\omega(x,y)} - p_k(y,x) \sqrt{\omega(y,x)} \right) dy.$$

Definition 3.2. The nonlocal vectorial total variation (NLVTV) of $u \in L^1(\Omega; \mathbb{R}^C)$ is defined as

$$\int_\Omega |Du| = \sup_{p \in F} \left\{ -\int_\Omega \langle u(x), (div_{\omega p}(x)) \rangle dx \right\},$$

with $Y = \{ p \in C^0(\Omega \times \Omega; \mathbb{R}^C) : |p|_{NL}(x) \leq 1, \forall x \in \Omega \}$ and $|p|_{NL}(x) = \sum_{k=1}^C \int_\Omega (p_k(x,y))^2 dy$.

3.2.2 Nonlocal Energy Functional

We propose to estimate the reflectance $R : \Omega \rightarrow \mathbb{R}^C$, usually $C = 3$, as the minimizer of the functional

$$\int_\Omega |DR| + \lambda \int_\Omega |R(x) - \bar{R}(x)|^2 dx,$$

where $\lambda > 0$ is a trade-off parameter, and $\bar{R}$ is an initial illumination estimate computed as $\bar{R}(x) = \max_{\epsilon \in \{R,G,B\}} R'(x)$, $R'$ being the denoised reflectance. Note that any discrepancy in the illumination will impact the result.

Figure 1: Pairs of low-light and ground-truth images from the LOL dataset (Wei et al., 2018) used for the experiments.
Figure 2: Visual impact of the trade-off parameter $\lambda$ in (2). Each row contains the illumination, reflectance and enhanced image, respectively. We observe a piecewise smooth illumination in all cases. However, the differences are more noticeable in the reflectances and the enhanced images. Indeed, the smaller $\lambda$ is, the brighter the result.

$$\beta \int_{\Omega} |D_\omega R| + \frac{\alpha}{2} \int_{\Omega} |\nabla R - G|^2 + \frac{1}{2} \int_{\Omega} |R - R_0|^2, \quad (3)$$

where $\alpha, \beta > 0$ are trade-off parameters, $\int_{\Omega} |D_\omega R|$ is the NLVTV, and $R_0$ is an initial estimate of the reflectance computed as $R_0(x) = \frac{L(x)}{\Gamma(x)}$, with $T$ being the minimizer of (2) and $\epsilon > 0$ a small constant. The weight function $\omega$ will be defined in terms of differences between patches in $L$. Therefore, the underlying assumption behind NLVTV is that images are self-similar, making it a good prior for the reflectance.

The second term in (3) enforces closeness between the gradients of the reflectance and those of an adjusted version of the low-light image, strengthening the structural information. Following (Li et al., 2018), we define $G = \left(1 + \lambda_G e^{-|L|/\sigma_G}\right) \circ \nabla L$, with

$$\nabla \hat{L} = \begin{cases} 0 & \text{if } |\nabla L| < \epsilon_G, \\ \nabla L & \text{otherwise,} \end{cases}$$

where $\sigma_G$ controls the amplification rate of different gradients, $\lambda_G$ controls the degree of the amplification, and $\epsilon_G$ is the threshold that filters small gradients.

### 3.2.3 Contrast-Invariant Weights

For the function $\omega : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ involved in NLVTV, we propose to use bilateral weights that consider both the spatial closeness between points and the similarity in $L : \Omega \to \mathbb{R}^C$. This similarity is computed by considering a whole patch around each point and using the Euclidean distance across color channels:

$$d(L(x), L(y)) = \int_{\Omega} |(L(x + z) - \mu(x)) - (L(y + z) - \mu(y))|^2 dz, \quad (4)$$

where $\mu(x)$ denotes the mean value of the patch centered at $x$. This allows the distance (4) to be contrast invariant between the selected patches.

The weights are defined as

$$\omega(x, y) = \frac{1}{\Gamma(x)} \exp \left( - \frac{|x - y|^2}{h_{\text{spt}}^2} - \frac{d(L(x), L(y))}{h_{\text{sim}}^2} \right) \quad (5)$$

where $h_{\text{spt}}, h_{\text{sim}} > 0$ are filtering parameters that control how fast the weights decay with increasing spatial distance or dissimilarity between patches, respectively, and $\Gamma(x)$ is the normalization factor. Note that $0 < \omega(x, y) \leq 1$ and $\int_{\Omega} \omega(x, y) dy = 1$, but $\Gamma(x)$ breaks down the symmetry of $\omega$. The average between very similar regions preserves the integrity of the image but reduces small oscillations, which contain noise.

For computational purposes, NLVTV is limited to interact only between points at a certain distance. Let
The discrete variational model related to (2) is
$$\min_{T \in \mathbb{R}^N} \|\nabla T\|_1 + \lambda \|T - \hat{T}\|_2^2,$$
(7)

where $\nabla T$ and $\nabla R$ are computed via forward differences and Neumann boundary conditions, and denoted for each pixel $i$ and channel $k$ by $(\nabla T)_i = ((\nabla T)_i,1, (\nabla T)_i,2)$ and $(\nabla R)_i,k = ((\nabla R)_i,k,1, (\nabla R)_i,k,2)$, respectively.

Let $M$ be the size of the neighborhood around each pixel where the weights of the NLVTV regularization term are nonzero. The nonlocal gradient $\nabla_R \in \mathbb{R}^{N \times C \times M}$, denoted for each pixel $i$ and channel $k$ by $(\nabla_R)_{i,k} = ((\nabla_R)_{i,k,1}, \ldots, (\nabla_R)_{i,k,M})$, is defined as $(\nabla_R)_{i,k} = \sqrt{\omega_{i,k}} (R_{i,k} - R_{i,k})$, where $\{ \omega_{i,j} \}$ contains the discretization of the weights (5). In practice, the weight of the reference pixel is set to the maximum of the weights in the neighborhood, i.e., $\omega_{i,j} = \max_{1 \leq j \leq M} \omega_{i,j}$. This setting avoids the excessive weighting of the reference pixel.

Both minimization problems (2) and (3) are convex but non-smooth. To find a global optimal solution, we use the first-order primal-dual algorithm introduced in (Chambolle and Pock, 2011). To do so, we rewrite each problem in a saddle-point formulation by introducing dual variables. The algorithm consists of alternating a gradient ascent in the dual variable, a gradient descent in the primal variable and an over-relaxation for convergence purposes. The gradient steps are given in terms of the proximity operator, which is defined for any proper convex function $\varphi$ as $\text{prox}_\varphi(x) = \arg\min_y \{ \varphi(y) + \frac{1}{2\tau} \|x - y\|_2^2 \}$. The efficiency of the algorithm is based on the assumption that proximity operators have closed-form representations or can be efficiently solved. For all details on convex analysis omitted in this section, we refer to (Chambolle and Pock, 2016).

### 4.1 Estimation of the Illumination

The discrete variational model related to (2) is
$$\min_{T \in \mathbb{R}^N} \|\nabla T\|_1 + \lambda \|T - \hat{T}\|_2^2,$$
(7)
Figure 4 illustrates the impact of the trade-off parameter $\lambda$ in the variational model (2). We observe that a piecewise smooth illumination is obtained in all cases. The slight variation in the illumination is amplified in the reflectance maps and the enhanced images, resulting in darker or brighter images, depending on whether more or less weight is given to the fidelity term compared to the regularization term.

The discretization of the variational model (3) is

$$\beta \| \nabla w R \|_1 + \frac{\alpha}{2} \| \nabla R - G \|_2^2 + \frac{1}{2} \| R - R_0 \|_2^2,$$

where $\| \nabla w R \|_1 = \sum_{k=1}^N \sum_{i=1}^{|I_k|} |(\nabla w R)_{i,k}|$ and $\| \cdot \|_2$ applies across channel and pixel dimensions.

The NLTV term is treated analogously to the TV case. We also need to dualize the $\alpha$-term since its proximity operator has no closed-form. By taking into account that $\frac{\alpha}{2} \| x \|_2^2 = sup_{p,q} \langle x,y \rangle - \frac{\alpha}{2} \| y \|_2^2$, the saddle-point formulation of (8) is

$$\min_{R \in \mathbb{R}^{N \times C}} \max_{\mathbb{R}^{N \times C \times 2}} \langle \nabla R, p \rangle + \langle \nabla w R, q \rangle$$

$$- \frac{1}{2\alpha} \| p - G \|_2^2 - \delta_{Q}(q) + \frac{1}{2} \| R - R_0 \|_2^2,$$

with $Q = \{ q \in \mathbb{R}^{N \times C \times M} : \sum_{k=1}^N \sum_{j=1}^{|I_k|} q_{i,j,k} \leq \beta^2, \forall i \}$. Therefore, $R$ is computed through the following primal-dual iterates:

$$p^{n+1} = \frac{\alpha (p^n + \sigma \nabla R') + \sigma G}{\alpha + \sigma},$$

$$q^{n+1} = \frac{\beta (q^n + \sigma (\nabla w R')_{i,k,j})}{\max (\beta, |q^n + \sigma (\nabla w R')_{i,k,j}|)},$$

$$R^{n+1} = R^n + \tau \div p + \tau \div w q + \tau R_0,$$

$$R_{i,j,k}^{n+1} = 2R_{i,j,k}^{n+1} - R_{i,j,k}^n.$$
Figure 5: Comparison between state-of-the-art techniques and our method on dataset in Figure 1. We observe that MSR, SRIE and RetinexNet are not robust to noise and exhibit color issues in all experiments. The method by Kimmel et al. is not able to correctly remove the effect of the illumination, while LIME produces oversaturated results. AGLLNet produces greyish images and introduces color artifacts as seen for instance in the blue bottle and the red ball of yarn. Our method provides the best compromise between discounting the illumination effect, avoiding the amplification of noise, and preserving the color and geometry structure of the scene.
In Figure 3, we discuss about the influence of the trade-off parameters \( \alpha \) and \( \beta \) in (3) on the enhanced image produced by the proposed method. We observe that larger values of \( \alpha \) or \( \beta \) result in an over-smoothed image. On the other hand, when the role of the regularization term is less relevant, as it is the case for \( \beta = 0.05 \), the noise is amplified.

Gamma correction is an essential tool for adjusting the lighting, as shown in Figure 4. Indeed, we observe how varying \( \gamma \) significantly affects the final image: an excessive gamma correction \( (\gamma = 0.25) \) leads to a saturated image, while a less significant transformation \( (\gamma = 0.55) \) results in an overly dark image.

### 5.2 Comparison with the State of the Art

We compare our method with Multiscale Retinex (MSR) (Jobson et al., 1997), Kimmel et al. (Kimmel et al., 2001), SRIE (Fu et al., 2016b), LIME (Guo et al., 2017) and the deep learning techniques RetinexNet (Wei et al., 2018) and AGLLNet (Lv et al., 2021). Kimmel et al. and LIME have been implemented by ourselves, while SRIE is sourced from (Ying et al., 2017) and MSR from (Petro et al., 2014). The trained networks of RetinexNet and AGLLNet are provided by the authors on their webpages. The most suitable parameters for state-of-the-art and our method have been chosen based on visual evaluation.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR ↑</th>
<th>SSIM ↑</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSR</td>
<td>12.41</td>
<td>0.7229</td>
</tr>
<tr>
<td>Kimmel et al.</td>
<td>14.27</td>
<td>0.8287</td>
</tr>
<tr>
<td>SRIE</td>
<td>17.27</td>
<td>0.8483</td>
</tr>
<tr>
<td>LIME</td>
<td>18.09</td>
<td>0.8592</td>
</tr>
<tr>
<td>RetinexNet</td>
<td>17.56</td>
<td>0.7819</td>
</tr>
<tr>
<td>AGLLNet</td>
<td>16.82</td>
<td>0.8661</td>
</tr>
<tr>
<td>Ours</td>
<td><strong>20.08</strong></td>
<td><strong>0.8868</strong></td>
</tr>
</tbody>
</table>

We compare our method with Multiscale Retinex (MSR) (Jobson et al., 1997), Kimmel et al. (Kimmel et al., 2001), SRIE (Fu et al., 2016b), LIME (Guo et al., 2017) and the deep learning techniques RetinexNet (Wei et al., 2018) and AGLLNet (Lv et al., 2021). Kimmel et al. and LIME have been implemented by ourselves, while SRIE is sourced from (Ying et al., 2017) and MSR from (Petro et al., 2014). The trained networks of RetinexNet and AGLLNet are provided by the authors on their webpages. The most suitable parameters for state-of-the-art and our method have been chosen based on visual evaluation.

Figure 5 displays a comparison among all methods on Bookcase, Clothes rack and Plush toy. We observe that MSR, SRIE and RetinexNet are not robust to noise and exhibit color issues in all cases. The method by Kimmel et al. is not able to correctly remove the effect of the illumination, while LIME tends to oversaturate the scene. AGLLNet produces greyish images, introduces color artifacts, as seen in the blue bottle and the red ball of yarn in Plush toy, and modifies the texture of the reference image, as seen in the wood of Bookcase. Our method provides the best compromise between discounting the illumination effect, avoiding the amplification of noise, and preserving the color and geometry structure of the scene.

We also evaluate the performance of all methods in terms of PSNR and SSIM. Table 1 displays the average values on the results in Figure 5. Our method outperforms the others in terms of both metrics.

An additional analysis has been performed with the natural low-light image Lamp in Figure 6. Only
LIME, RetinexNet and our method are able to remove the effect of the illumination. However, LIME and RetinexNet are affected by noise and also produce artifacts, like the halo surrounding the lamp.

6 CONCLUSION

In this paper, we have proposed a low-light image enhancement method that estimates illumination and reflectance separately using variational models. In particular, we have introduced a contrast-invariant non-local regularization term for recovering fine details in the reflectance component. The experiments have shown that our method obtains state-of-the-art results and performs well in terms of noise reduction, color recovery, geometry and texture preservation.

ACKNOWLEDGEMENTS

This work is part of the MaLiSat project TED2021-1326444-B-100, funded by MCIN/AEI/10.13039/501100011033/ and by the European Union NextGenerationEU/PRTR, and also of the MoLaLIP project PID2021-125711OB-I00, financed by MCIN/AEI/10.13039/501100011033/ and by the European Union 132644B-I00, funded by MCIN/AEI/10.13039/501100011033/ and by the European Union.

REFERENCES


515