Endoscopic Vision Augmentation Using Multiscale Bilateral-Weighted Retinex for Robotic Surgery

Xiongbiao Luo, *Senior Member, IEEE*, Hui-Qing Zeng, Ying Wan, Xiao-Bin Zhang, Yan-Ping Du, and Terry M. Peters, *Life Fellow, IEEE*

***Abstract*—Endoscopic vision plays a significant role in minimally invasive surgical procedures. The visibility and maintenance of such direct in-situ vision is paramount not only for safety by preventing inadvertent injury, but also to improve precision and reduce operating time. Unfortunately, endoscopic vision is unavoidably degraded due to illumination variations during surgery. This work aims to restore or augment such degraded visualization and quantitatively evaluate it during robotic surgery. A multiscale bilateral-weighted retinex method is proposed to remove non-uniform and highly directional illumination and enhance surgical vision, while an objective no-reference image visibility assessment method is defined in terms of sharpness, naturalness, and contrast, to quantitatively and objectively evaluate endoscopic visualization on surgical video sequences. The methods were validated on surgical data, with the experimental results showing that our method outperforms existent retinex approaches. In particular, the combined visibility was improved from 0.81 to 1.06, while three surgeons generally agreed that the results were restored with much better visibility.**

***Index Terms*—Endoscopy, endoscopic vision, retinex, illumination variation, image restoration, minimally invasive surgery, computer-assisted interventions.**

I. Endoscopic Vision

ENDOSCOPIC vision generally refers to interventional visualization of surgical sites that are intuitively observed or examined by endoscopes integrated with video cameras inserted through a port during minimally invasive surgery. Based on endoscopic vision, surgeons can not only examine abnormalities and recognize surface tissue details, but also track various surgical instruments with respect to anatomical targets. In this respect, the quality of endoscopic videos is essential for effective perception or navigation. Therefore, a high-visibility surgical field is important to prevent unintentional harm, reduce operating time, and improve clinical outcome. In addition, high-quality endoscopic vision is beneficial for endoscopic video processing such as three-dimensional (3-D) surface reconstruction, tissue deformation tracking, instrument detection and tracking, and activity recognition during surgery [1].

Robotically assisted endoscopic surgery that commonly employs the da Vinci surgical system (Intuitive Surgical, Inc., Sunnyvale, USA) is increasingly performed in various clinical procedures, e.g., robotic prostatectomy for minimally invasive prostatic tumor resection [2] and robotic urological procedure [3]. Robotic endoscopy uses stereoscopic endoscopes to intuitively visualize the organ surface and directly guide or manipulate various surgical instruments inside the body. Unfortunately, typical endoscopes have a narrow field of view that is additionally degraded by non-uniform illumination due to the light source located at the distal end of the endoscope (Fig. 1). These drawbacks unavoidably deteriorate the clear and high-quality visualization of both the organ being operated on and its anatomical surroundings. Moreover, they also lead to difficultly in distinguishing many characteristics of the visualized scene (e.g., neurovascular bundle), and prevent the surgeon from clearly observing certain structures (e.g., subtle bleeding areas). It is important therefore to explore image-processing techniques to restore such degradations.

This work aims to resolve the inherent drawbacks discussed above and enhance on-site endoscopic visibility of the surgical field. As one of well-known and commonly established digital image processing methods, Retinex is widely employed for various tasks, e.g., computational color constancy [4], intrinsic image decomposition [5], [6], image defogging [7], enhancement and restoration [8], [9]. This paper also modifies Retinex to address surgical field problems. The main contributions of this paper are clarified as follows.

- From the clinical perspective, this work is of practical surgical importance in dealing with illumination variations in robotic-assisted laparoscopic prostatectomy, digitally augmenting the laparoscopic video stream, to provide the surgeon with a wider useful field of view that is devoid of non-uniform illumination effects.
- Technically, we propose a multiscale bilateral-weight retinex method that demonstrates superior performance compared to extant retinex techniques. In particular, the method employs cross filtering to remove image streak and related artifacts, and is an effective and efficient means of enhancing the images.
- Deep learning based methods are increasingly successful in computer vision and medical imaging, particularly in image enhancement and restoration. However, training or ground truth data of surgical endoscopic video images are difficult to collect or generate. The output from our

This work was partly supported by the Fundamental Research Funds for the Central Universities (No. 20720180062), the Canadian Institutes for Health Research (CIHR) Team Grant on Prostate Imaging, and the Canadian Foundation for Innovation (LEF2099). *Corresponding author: Hui-Qing Zeng*.

X. Luo is with the Department of Computer Science, Fujian Key Laboratory of Computing and Sensing for Smart City, Xiamen University, Xiamen 361005, China (e-mail: xbluo@xmu.edu.cn)

H.-Q Zeng, X.-B. Zhang, and Y.-P Du, are with the Zhongshan Hospital, Xiamen University, Xiamen 361004, China (email: 13606080893@139.com)

Y. Wan is with the School of Electrical and Data Engineering, University of Technology Sydney, Australia (e-mail: joyee.wan@gmail.com)

T. M. Peters is with the Robarts Research Institute, Western University, London N6A 5K8, Canada (e-mail: tpeters@robarts.ca)

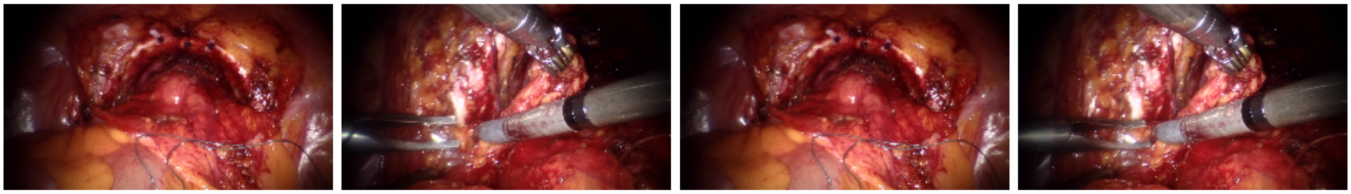


Fig. 1: Examples of degenerated surgical endoscopic images due to small viewing field and non-uniform and highly directional illumination in robotic-assisted laparoscopic prostatectomy. Note that the laparoscope providing the surgical field visualization is usually fixed at a relative distance from anatomical target regions during laparoscopic surgery. While this setting is to avoid surgical instrument fencing, it also makes the laparoscopic illumination relatively unchanged, resulting in limited access and view of surgical targets and almost identical illumination on the first and third (or second and forth) images.

proposed approach can in fact be used to train a deep-learning algorithm for surgical vision augmentation.

- Although this work develops the multiscale bilateral-weight retinex method for laparoscopic image processing, we believe that our method is also applicable to other endoscopic image enhancement such as colonoscopy and bronchoscopy, as well as to real-world or natural image nonuniform illumination removal.
- Additionally, several no-reference image visibility assessment metrics are proposed to quantitatively and objectively evaluate endoscopic video images and different surgical field augmentation algorithms.

The remainder of this paper is organized as follows. Section II reviews work related to retinex theory. Our bilateral-weighted retinex method for vision augmentation is presented in Section III, followed by the experiment settings in Section IV. Sections V and VI show and discuss the validation results before concluding this work in Section VII.

II. Related Work

Our idea of endoscopic vision augmentation is motivated by computational color constancy [10], [11]. The human visual system has a relatively constant awareness of the color of objects under illumination differences. This implies that human color perception generally depends on the reflectance of objects but not the scene illumination. Hence, the fidelity and quality of images or videos can be augmented by removing the illumination effect while retaining the reflectance component.

Retinex [12] is currently a popular and effective approach to to model human color consistency for image enhancement. Numerous papers have been published on retinex methods in the literature. Kimmel et al. [13] presented a variational retinex method for image color processing. This approach, that usually generate halos and does not preserve edge information, was further improved by Ng et al. [14], Ma et al. [15], Wang et al. [16], and Provenzi et al. [17]. Provenzi et al. [18] reported a new implementation of random spray retinex on the basis of a reset mechanism. While random spray retinex is sensitive to image noise, the number of sprays, and the pixels in the sprayed region, Banic et al. [19] modified by employing Gaussian filtering for better performance. By discussing current issues of retinex theory, Bertalmio et al. [12] proposed a kernel-based retinex, while Morel et al. [20] proved that the retinex solutions satisfy the discrete Poisson equation. Rahman et al. [21] discussed a fully automatic retinex color restoration framework that combines the multiscale retinex processing with a simple color restoration step. Jiang et al. [22] introduced the graphics processing unit technique and autolevels processing to boost retinex and further reduce computational time and improve performance. While Zosso et al. [23] presented a non-local retinex framework, the non-local retinex method with multiple predetermined parameters in optimization is a time-consuming procedure and also potentially leads to unnaturally colored images. Cai et al. [5] discussed a retinex model on the basis of joint intrinsic-extrinsic prior. A thorough review of the retinex methods was recently presented by McCann [4].

In addition, Okuhata et al. [24] used the variational Retinex model to enhance colonoscopic images, while Khan et al. [25] employed a color reproduction and processing algorithm for colonoscopic image real-time mapping. Pedersen et al. [26] used currently available image quality metrics to evaluate wireless capsule endoscopic video sequences. Luo et al. [8] discussed a visibility-driven fusion defogging framework to enhance robotic laparoscopic video images, and more recently, Sdiri et al. [27] proposed to use joint wavelet decomposition and binocular combination for surgical image enhancement.

Although multiscale retinex processing is easily implemented with high efficiency and performance, provides dynamic range compression and preserves most of the detail, it still has two drawbacks: (1) washed out appearance: blurring the strongest edges with the remaining faint edges almost being untouched, (2) unnatural color rendition due to over-enhanced saturation. Our current work aims for edge preservation and natural color rendition and further addresses the drawbacks of the multiscale retinex color restoration method. We propose a multiscale bilateral-weighted retinex method to achieve better color fidelity and apply this method to deal with the aforementioned problems that occur in endoscopic vision during robotic surgery. Augmented visualization enables surgeons to accurately identify subtle surface details and results in improved perception of the anatomical surroundings of the organ being operated on.

III. Approaches

This section discusses multiscale bilateral-weighted retinex for endoscopic vision augmentation and defines several quality metrics to evaluate surgical vision.

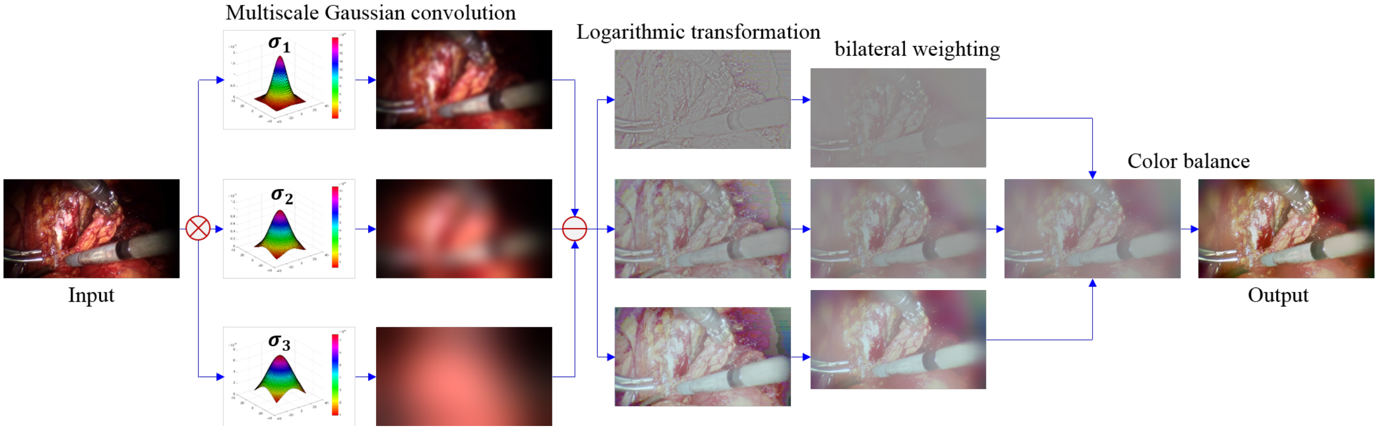


Fig. 2: Flowchart of the proposed multiscale bilateral-weighted retinex method for endoscopic vision augmentation

A. *Multiscale Bilateral-Weighted Retinex*

Our retinex approach consists of three main steps: (1) illumination decomposition, (2) bilateral weighting, and (3) color balance. Fig. 2 shows the flowchart of bilateral-weighted retinex modeling for endoscopic vision augmentation.

1) Illumination Decomposition: As discussed above, the human visual system when perceiving the color of a scene depends on the reflectance but not its illumination. Accordingly, we can decompose the illumination and reflectance elements, and remove the former but preserve the latter to enhance image visual quality. The multiscale retinex algorithm is able to separate image illumination and reflectance, and contains two operators of Gaussian convolution and logarithmic transformation.

Suppose that (x, y) is a pixel on a color image $\mathbf{I}_c$ with three channels of red, green, and blue $c \in \{r, g, b\}$. For pixel $\mathbf{I}_c(x, y)$ in each channel, the Gaussian convolution and logarithmic transformation are performed to obtain $\mathbf{J}_c(x, y)$:

$$\mathbf{J}_{m,c}(x, y) = \log \mathbf{I}_c(x, y) - \log (\mathbf{G}_m(x, y) \otimes \mathbf{I}_c(x, y)), \quad (1)$$

$$\mathbf{G}_m(x, y) = \lambda_m \exp(\frac{-(x^2 + y^2)}{2\sigma_m^2}), \quad (2)$$

where m indicates the Gaussian scale level, $\otimes$ is the convolution operator, σ_m denotes the scale, and λ_m guarantees $\int \mathbf{G}_m(x, y)dxdy = 1$. Eq. 1 aims at accurately separating the reflectance and illumination components of the image $\mathbf{I}(x, y)$. Based on the definition of retinex theory, the output $\mathbf{K}_c(x, y)$ of the multiscale retinex is the sum of weighted $\mathbf{J}_{m,c}(x, y)$ at different Gaussian levels at each channel:

$$\mathbf{K}_c(x, y) = \sum_{m=1}^{M} \omega_m \mathbf{J}_{m,c}(x, y), c \in \{r, g, b\}, \quad (3)$$

$$\mathbf{K}(x, y) = \sum_{m=1}^{M} \omega_m \mathbf{J}_m(x, y), \quad (4)$$

where $\omega_m(x, y)$ denotes a weight to balance various levels or scales, $\mathbf{K}(x, y)$ is the final enhanced color image and $\mathbf{J}_m(x, y)$ represents the output image at level m with scale σ_m.

While the approach formulated by Eq. 1 achieves the color constancy property, it assumes that

$$\mathbf{I}_c(x, y) = \mathbf{L}_c(x, y)\mathbf{R}_c(x, y), \quad (5)$$

where $\mathbf{L}_c(x, y)$ and $\mathbf{R}_c(x, y)$ represent scene illumination and reflectance, respectively. We rewrite Eq. 1:

$$\mathbf{J}_c(x, y) \simeq \log \frac{\mathbf{L}_c(x, y)\mathbf{R}_c(x, y)}{\bar{\mathbf{L}}_c(x, y)\bar{\mathbf{R}}_c(x, y)}, \quad (6)$$

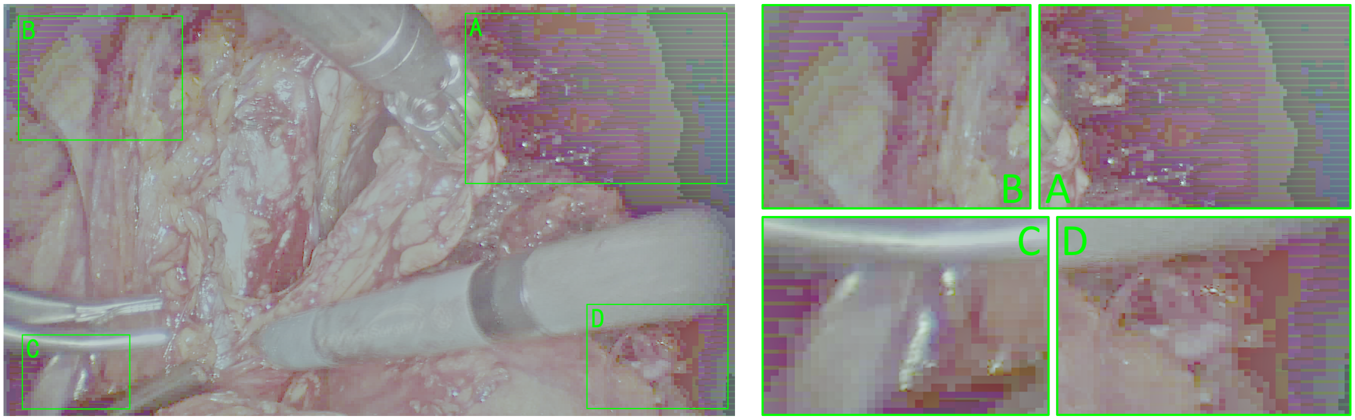
where $\bar{\mathbf{L}}_c(x, y)$ and $\bar{\mathbf{R}}_c(x, y)$ are the smoothed and average intensity after a predefined processing (which here refers to Gaussian convolution). Multiscale retinex makes a core assumption that $\bar{\mathbf{L}}_c(x, y)$ is locally smooth and constant after such processing. This implies that $\mathbf{L}_c(x, y)$ is approximately identical to $\bar{\mathbf{L}}_c(x, y)$, i.e., $\mathbf{L}_c(x, y) \simeq \bar{\mathbf{L}}_c(x, y)$. Therefore,

$$\mathbf{J}_c(x, y) \simeq \log \frac{\mathbf{R}_c(x, y)}{\bar{\mathbf{R}}_c(x, y)}, \quad (7)$$

which demonstrates that scene color is independent of the illumination $\mathbf{L}_c(x, y)$, and also explains why the retinex method should attenuate shading effects and shadows on color images.

Unfortunately, the predefined processing cannot perfectly smooth and weight color images and make $\bar{\mathbf{L}}_c(x, y)$ locally constant, particularly in our case of non-uniform and highly directional illumination in robotic surgery. This implies that the core assumption $\mathbf{L}_c(x, y) \simeq \bar{\mathbf{L}}_c(x, y)$ can be violated, resulting in image artifacts. Fig. 3 illustrates this phenomenon where we see streak-like and grayish artifacts on the color image $\mathbf{K}(x, y)$. To tackle these image artifacts, we introduce a bilateral-weighting strategy in the following.

2) Bilateral Weighting: To remove streak-like and grayish artifacts while preserving structural or edge information, a weighted filtering procedure is commonly applied to such artifacts at each level, i.e., we need to re-weight all the pixels after the Gaussian convolution and logarithmic transformation. Ideally, guided filtering should be used to calculate the weights of the pixels [28]. However, this approach requires the construction of a sparse $N \times N$ matrix (where N is the number of image pixels), which is a memory- and time-consuming computation. Hence, we employ cross bilateral filtering to calculate these weights for artifact removal.



(a) Output image $\mathbf{K}(x, y)$ (b) Image artifacts

Fig. 3: The multiscale retinex generates streak-like and grayish artifacts under non-uniform and highly directional illumination

Based on cross bilateral filtering, we modify Eq. 3 by weighting each pixel $\mathbf{p} = (x, y)$ on image $\mathbf{J}_{m,c}(x, y)$:

$$\tilde{\mathbf{K}}_c(x, y) = \sum_{m=1}^{M} \omega_m \Psi_{\mathbf{p}}(\mathbf{J}_{m,c}(x, y)), c \in \{r, g, b\}, \quad (8)$$

where $\tilde{\mathbf{K}}(x, y)$ is the bilateral-weighted output. The bilateral weight $\Psi_{\mathbf{p}}(\mathbf{J}_{m,c}(x, y))$ at level m is computed by [29]

$$\Psi_{\mathbf{p}}(\mathbf{J}_{m,c}(x, y)) = \frac{1}{A_{\mathbf{p}}} \sum_{\mathbf{q} \in \Omega} \mathcal{W} \mathbf{J}_{m,c}(u, v), \mathbf{q} = (u, v), \quad (9)$$

where $\mathbf{q}$ denotes a pixel in region Ω centered at pixel $\mathbf{p} = (x, y)$ and the coefficient $\mathcal{W}$ is determined by

$$\mathcal{W} = \mathcal{P}(\|\mathbf{p} - \mathbf{q}\|) \mathcal{Q}(\|\mathbf{I}_c(x, y) - \mathbf{I}_c(u, v)\|), \quad (10)$$

where $\mathcal{P}(\cdot)$ and $\mathcal{Q}(\cdot)$ are the spatial and range filter kernels, respectively. In particular, $\mathcal{Q}(\cdot)$ is a Gaussian kernel to penalize pixels across edges that have large intensity differences. Hence, the bilateral processing can preserve edge information since $\mathcal{P}(\cdot)\mathcal{Q}(\cdot)$ takes on smaller values as the range distance and/or the spatial distance increases [29]. In addition, the normalization parameter $A_{\mathbf{p}}$ can be computed by

$$A_{\mathbf{p}} = \sum_{\mathbf{q} \in \Omega} \mathcal{P}(\|\mathbf{p} - \mathbf{q}\|) \mathcal{Q}(\|\mathbf{I}_c(x, y) - \mathbf{I}_c(u, v)\|). \quad (11)$$

Eventually, the bilateral weighted $\tilde{\mathbf{K}}(x, y)$ is written as

$$\tilde{\mathbf{K}}(x, y) = \sum_{m=1}^{M} \omega_m \Psi_{\mathbf{p}}(\mathbf{J}_m(x, y)), \mathbf{p} \in \mathbf{J}_m(x, y) \quad (12)$$

3) Color Balance: Figs. 4 (b) and (c) compare the outputs $\mathbf{K}(x, y)$ and $\tilde{\mathbf{K}}(x, y)$. While the output $\tilde{\mathbf{K}}(x, y)$ (Fig. 4 (c)) shows that the streak-like and grayish artifacts were removed, it still contains underexposed pixels and artificial light. The color balancing algorithm aims to correct such pixels and lighting effects. Various complex color balancing approaches are available in the literature [10]. This section explores a method that is simple and efficient while exhibiting higher or comparable performance to existing methods.

A color image at each channel usually has the maximal and minimal pixel intensities that correspond to some singular pixels due to underexposure or brightness artifacts. The idea of simple color balancing is to remove these singular pixels that are located at the left and right histogram tails, and stretch the dynamic range of the remaining pixels.

The output $\breve{\mathbf{K}}(x, y)$ of the color balancing step is commonly formulated by

$$\breve{\mathbf{K}}(x, y) = \begin{cases} \mu_{max}, & \mu_{max} < \mu_{min} \\ \xi_{min}, & \tilde{\mathbf{K}}(x, y) < \mu_{min} < \mu_{max} \\ \xi_{max}, & \mu_{min} < \mu_{max} < \tilde{\mathbf{K}}(x, y) \\ \varphi, & \text{otherwise} \end{cases}, \quad (13)$$

$$\varphi = \mu_{min} + \frac{(\tilde{\mathbf{K}}(x, y) - \mu_{min})(\xi_{max} - \xi_{min})}{\mu_{max} - \mu_{min}}, \quad (14)$$

$$\mu_{min} = \arg \min_{(x,y) \in \Pi_1} (\tilde{\mathbf{K}}(x, y)),\ \Pi_1 = N \frac{\zeta_1}{100}, \quad (15)$$

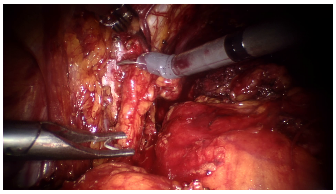
$$\mu_{max} = \arg \max_{(x,y) \in \Pi_2} (\tilde{\mathbf{K}}(x, y)),\ \Pi_2 = N(1 - \frac{\zeta_2}{100}) - 1, \quad (16)$$

where N is the number of pixels in $\tilde{\mathbf{K}}(x, y)$ and the dynamic range $[\xi_{min},\ \xi_{max}]$, Π_1 is the number of ζ_1 percent pixels located at the left side of the low intensity region in $\tilde{\mathbf{K}}(x, y)$ and ζ_2 is the percent of pixels located at the right side of the high intensity region in $\tilde{\mathbf{K}}(x, y)$. Note that the percentages ζ_1 and ζ_2 at the left and right histogram tails are experimentally determined. The processed images usually become bright or over-enhanced when the percentage is increased. We tested ζ_1 and ζ_2 from 1% to 5% and found that $\zeta_1 = \zeta_2 = 2\%$ shows relatively good performance balancing contrast and brightness. After determining ζ_1 and ζ_2, the dynamic range ξ_{min} and ξ_{max} are directly taken from the histogram at positions or quantiles $N\zeta_1/100$ and $N(1 - \zeta_1/100) - 1$, respectively.

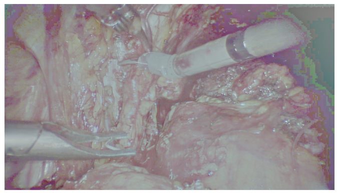
Fig. 4 (d) gives an example output of color balancing. Note that the color balance step improves contrast and brightness of the image, augmenting visualization of the surgical field without disconcerting the surgeon.

B. Surgical Vision Assessment

Endoscopic vision assessment is a challenging problem since neither gold standards nor ground truth examples are

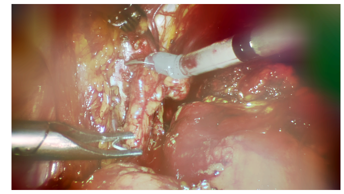


(a) Input

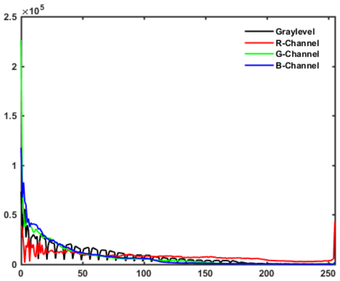


(b) $\mathbf{K}(x, y)$

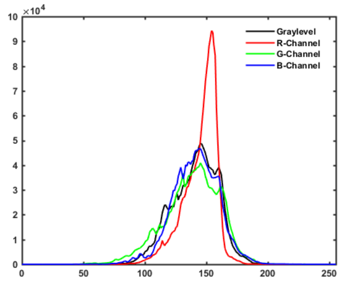
(c) $\tilde{\mathbf{K}}(x, y)$



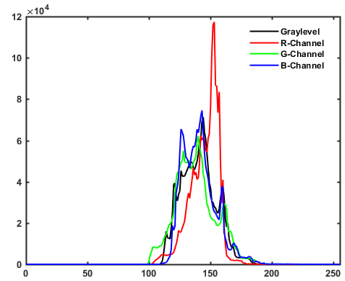
(d) $\breve{\mathbf{K}}(x, y)$



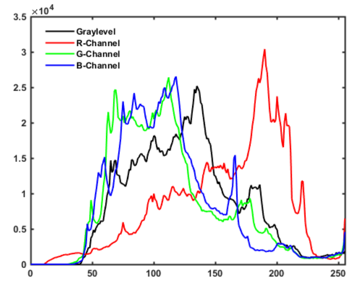
(e) Histogram of (a)



(f) Histogram of (b)



(g) Histogram of (c)



(h) Histogram of (d)

Fig. 4: Compared results of bilateral weight and color balance processing

available. Subjective assessment by surgeons is a way to manually and intuitively validate surgical field visual quality. More interestingly, quantitative objective assessment metrics without any reference are widely developed to evaluate image visual quality. In this work, we define four image metrics: (1) sharpness, (2) naturalness, (3) contrast, and (4) a hybrid metric that is a combination of all three.

The sharpness metric χ aims to quantitatively characterize how much structural information such as contours and boundaries on images can be perceived by the human visual system. Similar to previous work [30], a standard deviation of a weighted maximum local variation distribution is defined as the sharpness metric χ. On the other hand, the naturalness metric ξ depicts how natural endoscopic vision appears, and is generally a subjective judgment on images. While it is hard to characterize naturalness ξ, we define it on the basis of statistical analysis of thousands of images, for which analysis has demonstrated that the histogram curves of natural images generally yield Gaussian and Beta probability distributions [31]. In addition, contrast η describes the difference in luminance that makes regions of interest on images distinguishable. We define contrast $\mathcal{C}$ as the difference between a pixel $\mathbf{I}(x, y)$ and its average edge-weighted $\rho(x, y)$ in a patch $\mathcal{O}$ [32]:

$$\eta = G^{-1} \sum_{\mathcal{O}(x,y)} \frac{|\mathbf{I}(x, y) - \rho(x, y)|}{|\mathbf{I}(x, y) + \rho(x, y)|}, \tag{17}$$

$$\rho(x, y) = \sum_{\mathcal{O}(x,y)} \phi(x, y)\mathbf{I}(x, y), \tag{18}$$

where G is the pixel number and $\phi(x, y)$ is the detected edge.

Finally, we define a hybrid metric H_m as

$$H_m = a\chi^{\alpha} + b\xi^{\beta} + (1 - a - b)\log \eta, \tag{19}$$

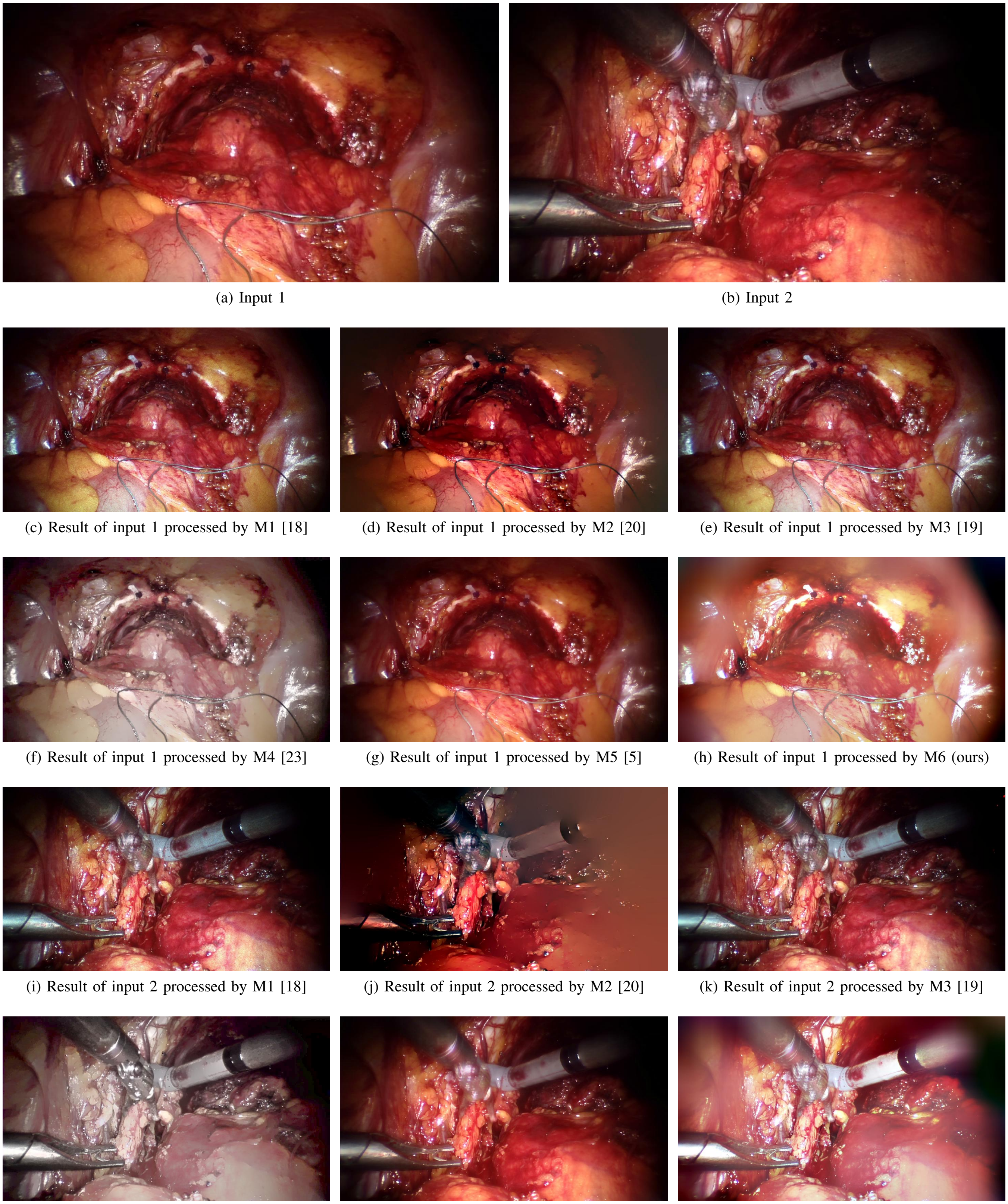
where a and b balance the three components of sharpness, naturalness, and contrast, and α and β control their sensitivities.

IV. Experimental Settings

Endoscopic video sequences were acquired during robotic-assisted laparoscopic radical prostatectomy using the da Vinci Si surgical system (Intuitive Surgical Inc., Sunnyvale, CA, USA) in the St. Josephs Hospital, London, Canada. These video data with various illumination from five different sequences were also collected under a protocol approved by the Research Ethics board of Western University, London, Canada. All the experiments were tested on a laptop installed with Windows 8.1 Professional 64-Bit System, 32.0-GB Memory, and Processor Intel Xeon CPU×8 and were implemented on the platform of Microsoft Visual Studio C++ 2008.

Our surgical vision augmentation and assessment approaches need to determine a set of parameters as discussed in Section III. While these parameters can affect the performance of the proposed methods, unfortunately, there are no standards to optimally determine these parameters. This work follows the previous work [21], [31]. In the multiscale retinex processing step, three levels and scales are sufficient for most images and the weights can be identical [21]. Therefore, we set $M = 3$ and $\omega_m(x, y) = 1/M, m = 1, 2, 3$ in Eq. 3. On the other hand, three scales were experimentally determined: $\sigma_m \in \{15, 80, 250\}$ in this work. Additionally, based on the work of Yeganeh and Wang [31], we set the parameters in Eq. 19 as:$a = 0.5$, $b = 0.4$, $\alpha = 0.3$, and $\beta = 0.7$. While these parameters were only tested on real-world (natural) images, they also can provide acceptable results in this study.

It is quite reasonable and natural to compare our proposed approach to various Retinex algorithms since ours is also one of the Retinex family methods. We compare our approach to currently available retinex techniques: (1) M1 [18], a random spray retinex method, (2) M2 [20], retinex formulated by the discrete Poisson equation, (3) M3 [19], a modified random spray retinex method, (4) M4 [23], an unifying retinex framework, (5) M5 [5], a joint intrinsic-extrinsic prior based retinex method; (7) M6, our proposed method, as discussed in



(a) Input 1

(b) Input 2

(c) Result of input 1 processed by M1 [18]

(d) Result of input 1 processed by M2 [20]

(e) Result of input 1 processed by M3 [19]

(f) Result of input 1 processed by M4 [23]

(g) Result of input 1 processed by M5 [5]

(h) Result of input 1 processed by M6 (ours)

(i) Result of input 2 processed by M1 [18]

(j) Result of input 2 processed by M2 [20]

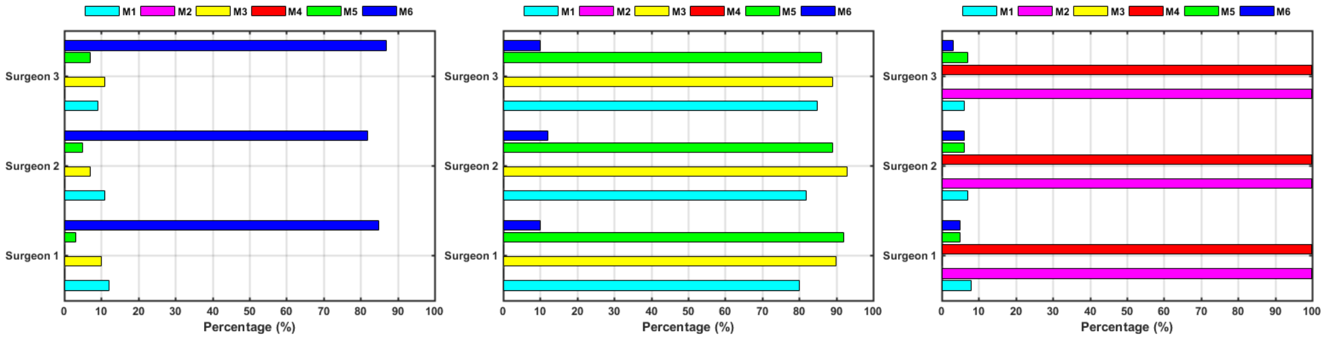
(k) Result of input 2 processed by M3 [19]

(l) Result of input 2 processed by M4 [23]

(m) Result of input 2 processed by M5 [5]

(n) Result of input 2 processed by M6 (ours)

Fig. 5: Examples of visual comparison of processed endoscopic images of using the different retinex methods.The input images (a) and (b) show different illumination on the images: (c)∼(h) and (i)∼(n) correspond to the results of using these compared methods M1 [18], M2 [20], M3 [19], M4 [23], M5 [5], and M6 (ours) to process the input images 1 and 2, respectively. Our proposed endoscopic vision augmentation approach shows the best performance.



(a) Better (b) Comparable (c) Worse

Fig. 6: Subjective assessment of the processed endoscopic video images of using methods M1 [18], M2 [20], M3 [19], M4 [23], M5 [5], and M6 (ours). The three endoscopic surgeons independently directly compared all the given processed endoscopic video images to the original images and subjectively classified those processed images into three categories of *better*, *comparable*, and *worse*.

Section III. All the methods were tested on 12000 frames.

Besides the objective quantitative assessment of the experimental results, we also introduce subjective assessment. Three surgeons manually inspected 1000 frames of surgical endoscopic images that were processed by each of the methods mentioned above. We calculate the percentage of images from 1000 frames for which the surgeon has divided them into three classes: (1) the processed image is visually better than the input image, (2) the processed image and the input image are visually comparable, and (3) the processed image is visually worse than the input image.

V. Results

Fig. 5 visually compares the experimental results of using the different retinex methods M1 [18], M2 [20], M3 [19], M4 [23], M5 [5], and ours (M6). From Fig. 5, our method outperforms other methods since it provides better visual quality. Moreover, Fig. 6 shows the results of the subjective assessment that was manually performed by the three endoscopic surgeons. The subjective assessment shows that methods M1, M3, and M5 give comparable visual quality, M2 and M4 provide worse visual quality, and our proposed approach achieves the best visual quality, compared to the input images. While M2 and M4 in all cases obtain worse visual quality than the input images (Fig. 6(c)), M1, M3, and M5 provide comparable visualization in more than 80% of the cases (Fig. 6(b)). In particular, the three endoscopic surgeons generally agree that on average, our approach provides improved visualization over 80% of the time.

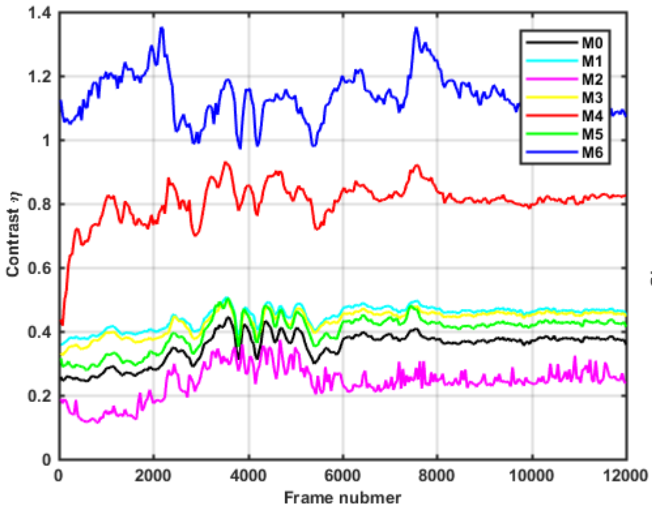
Fig. 7 plots the quantitative objective evaluation of the original or input endoscopic video images and their corresponding results of using the compared methods M1 [18], M2 [20], M3 [19], M4 [23], M5 [5], and ours (M6). It shows the obviously different assessment scores between our method and others. The surgeons intuitively perceived that the image contrast and naturalness of using our method are much better than others (Fig. 6 justifies that), although the image sharpness was comparable. Hence, these obvious differences were generally consistent with the subjective evaluation. Table I further summarizes image visual properties or metrics of the average contrast η, sharpness χ, naturalness ξ, and hybrid index H_m. According to the non-parametric statistical hypothesis Wilcoxon signed-rank test [33], we also computed p-values for the various quality metrics evaluating the processed images from the different methods. Because most of p-values were large, the null hypothesis cannot be rejected. Note that M0 in Fig. 7 and Table I denotes the visual properties of the original endoscopic images. From both Fig. 7 and Table I, we found that our method improved the visual quality of the original endoscopic images with non-uniform and highly directional illumination, while other retinex approaches either provided the comparable endoscopic vision or deteriorated the image visual quality. More interestingly, the average four metric values (1.13, 0.82, 0.29, 1.06) of our proposed method were much better than (0.35, 0.56, 0.03, 0.81) of the original images.

The computational time of the six retinex approaches M1$\sim$M6 was 6.3, 1.1, 6.5, 138.6, 21.6, and 3.1 seconds per frame, respectively. Note that our acquired endoscopic videos were high-definition with an image size of 1920×1080.

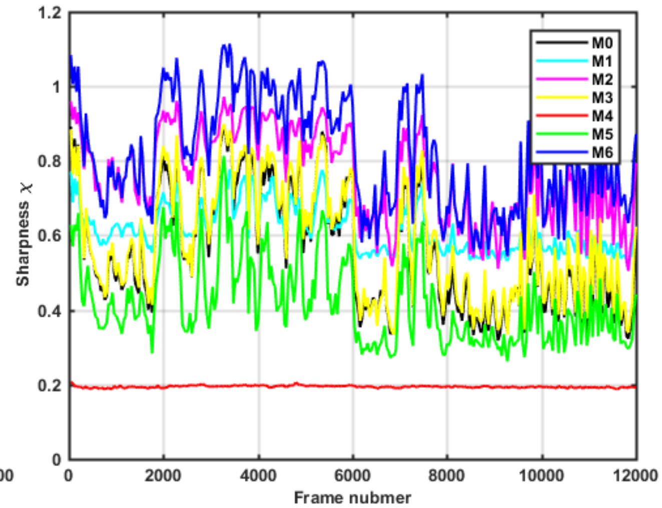
Additionally, we investigated various image enhancement methods reported recently in computer vision. Fig. 8 compares the image results of using ours and currently available approaches [34], [35], [36], [37], [38], [39], [40], [41], [8], [9]. While the two methods [37], [40] barely manage to enhance the the image, another two approaches [34], [39] slightly improved image brightness. The approaches [36], [8] generally augmented the image but introduced less naturalness. Unfortunately, the left four approaches [35], [38], [41], [9] over-enhanced the image, resulting in worse quality than the original image. Our method still works better than others.

VI. Discussion

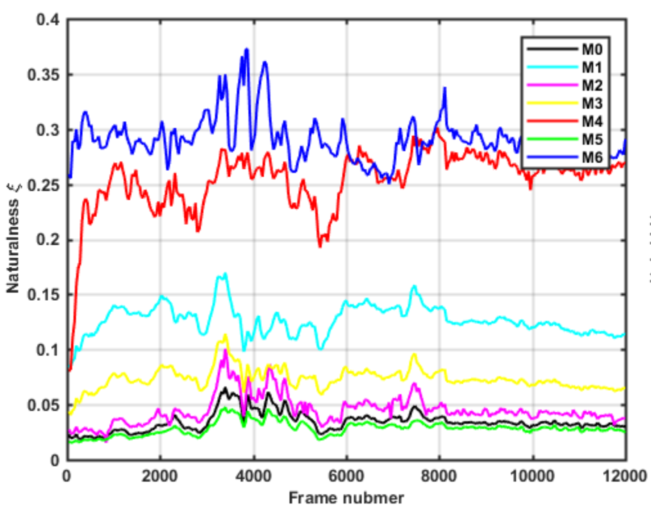
This work showed the first study on augmenting endoscopic field visualization for realizing subtle surface structures and anatomical surroundings of the organ being operated on during robotic-assisted endoscopic procedures. In endoscopic surgery, the organ sub-structures or surroundings are hidden and hardly perceived by surgeons due to limited (non-uniform and highly



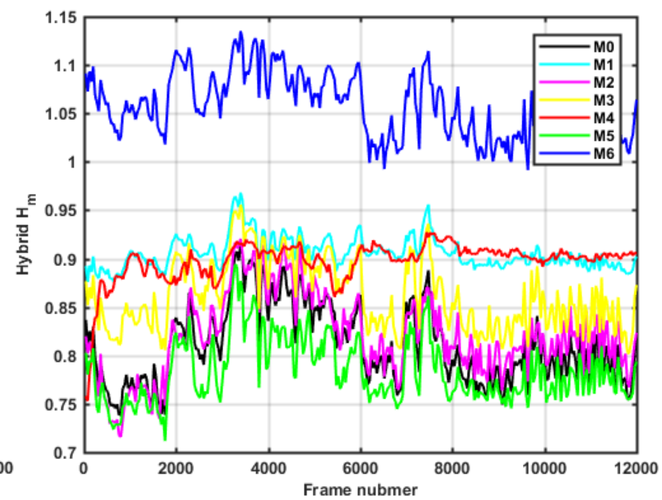
(a) Contrast η



(b) Sharpness χ



(c) Naturalness ξ

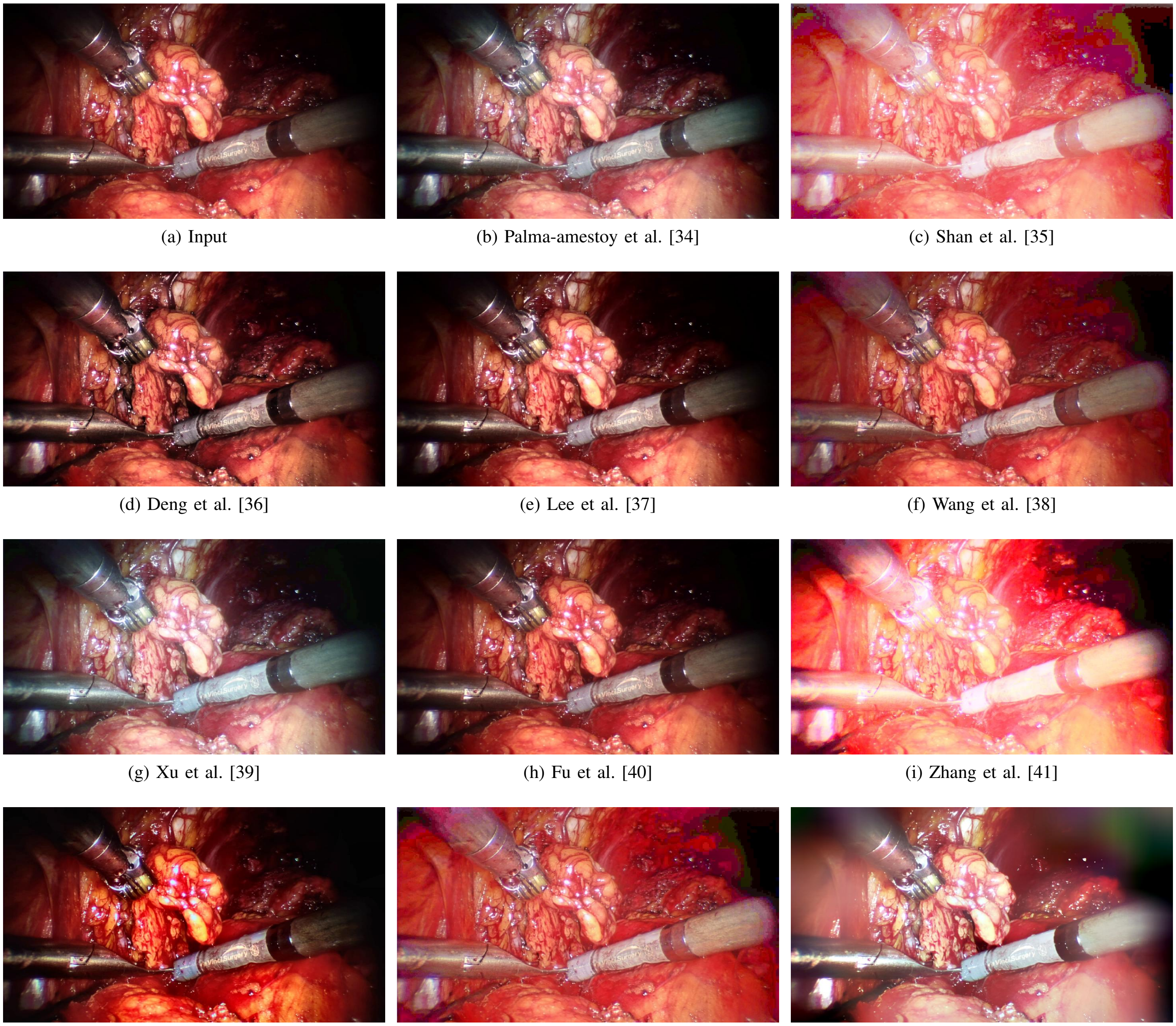


(d) Hybrid H_m

Fig. 7: Quantitative objective assessment of the processed endoscopic video images of using methods M1 [18], M2 [20], M3 [19], M4 [23], M5 [5], and M6 (ours). M0 indicates the quantitative quality of the input endoscopic video images. All the sharpness, naturalness, contrast, and hybrid metrics demonstrate that our proposed method works much better than others.

TABLE I: Quantitative evaluation of calculated average contrast η, sharpness χ, naturalness ξ, hybrid metric H_m (Eq.19), and p-values of the processed images by the different methods of M1 [18], M2 [20], M3 [19], M4 [23], M5 [5], and M6. M0 denotes the quantitative assessment of the original or input endoscopic video images from robotic prostatectomy.

Approaches	Contrast η			Sharpness χ			Naturalness ξ			Hybrid H_m		
	Mean	STD	p-value	Mean	STD	p-value	Mean	STD	p-value	Mean	STD	p-value
M0	0.35	0.05	–	0.56	0.16	–	0.0344	0.01	–	0.81	0.04	–
M1	0.45	0.03	0.97	0.62	0.06	0.88	0.13	0.01	0.89	0.91	0.02	0.38
M2	0.24	0.05	1.98	0.76	0.12	0.53	0.05	0.01	0.02	0.82	0.04	0.48
M3	0.43	0.04	0.99	0.58	0.15	0.79	0.07	0.01	0.06	0.86	0.04	0.42
M4	0.81	0.06	0.47	0.19	0.00	0.04	0.25	0.03	0.67	0.90	0.02	0.39
M5	0.40	0.05	1.13	0.42	0.11	1.08	0.03	0.01	0.01	0.78	0.03	0.64
M6	1.13	0.07	0.33	0.82	0.15	0.43	0.29	0.02	0.45	1.06	0.03	0.36



(a) Input (b) Palma-amestoy et al. [34] (c) Shan et al. [35]

(d) Deng et al. [36] (e) Lee et al. [37] (f) Wang et al. [38]

(g) Xu et al. [39] (h) Fu et al. [40] (i) Zhang et al. [41]

(j) Luo et al. [8] (k) Wang et al. [9] (l) Ours

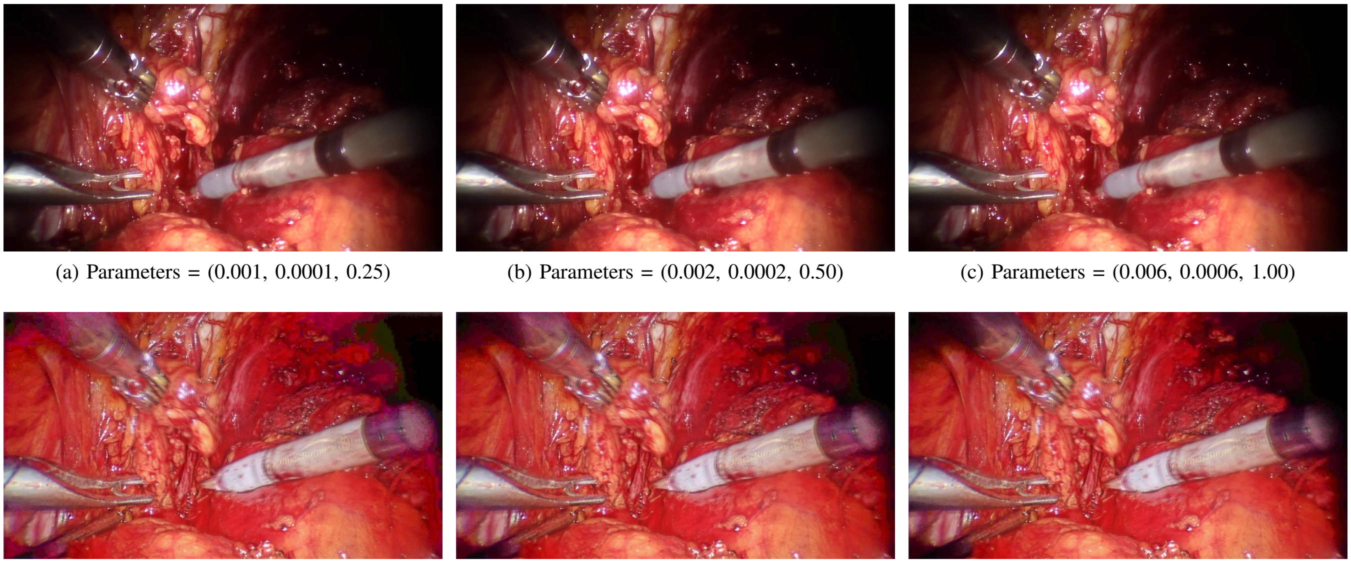
Fig. 8: A comparison of various image enhancement methods reported recently for surgical laparoscopic video augmentation. Our method generally outperforms these image enhancement approaches.

directional) illumination and narrow endoscopic view of the surgical field. Based on color constancy, this work first introduced the retinex theory to address these limitations.

A. Effectiveness

Retinex theory, a successful strategy for computational color constancy, was developed to various models or methods in the literature. Current retinex models are commonly used for natural image enhancement and illumination processing. From the experiments results we evaluated and compared these retinex models [18], [20], [19], [23], [5], we found that most current retinex models do not work well for surgical video images. One reason is that these models are constructed on the basis of natural real-world images under atmospheric lighting, while the endoscope uses synthesized light sources transmitted by optical fibers into the body for imaging the surgical field, i.e., the mechanism of illumination imaging was different. On the other hand, these different retinex models are usually derived in accordance with various assumptions, e.g., the path-wise structure on the image is substituted with the random spray structure [18], the retinex solutions satisfy a discrete Poisson equation by assuming the paths to be symmetric random walks [20], and illumination is supposed to vary smoothly [23], [5]. Unfortunately, these assumptions are commonly violated in endoscopic imaging. This also implies that these retinex models are inapplicable to enhance endoscopic vision during minimally invasive surgery.

The multiscale retinex model employs Gaussian convolution and logarithmic transformation to directly process the reflectance and illumination information. This work employed



(a) Parameters = (0.001, 0.0001, 0.25) (b) Parameters = (0.002, 0.0002, 0.50) (c) Parameters = (0.006, 0.0006, 1.00)

(d) Illumination = 10 (e) Illumination = 20 (f) Illumination = 30

Fig. 9: Parameter sensitivity analysis: The firs row shows the results of adjusting the join optimization parameters (*shape*, *texture*, *illumination*) in M5 [5] and the second row the results of changing the illumination parameter in the method [9]

multiscale retinex because it is simple and easily implemented. By removing the illumination, the image visual fidelity was augmented. However, it still violates the assumption that illumination is presumed to vary smoothly, resulting in streak-like and grayish artifacts on the image. We introduced bilateral weight to tackle the artifacts in the three levels after the Gaussian convolution and logarithmic transformation. The experimental results demonstrate that the bilateral-weight processing is effective for removing the artifacts. In particular, our improved method for robotic-assisted laparoscopic video processing is more effective and robust than other currently available retinex models [18], [20], [19], [23], [5], which is the main technical contribution of this work. More promisingly, an interesting *RetiNet* enhancement framework for surgical laparoscopic video augmentation can be constructed by combining our modified retinex model with convolutional neural networks. In addition, no-reference image visual quality assessment metrics are important to evaluate various endoscopic vision. This work defined several measures to quantitatively evaluate the original and processed surgical endoscopic video images.

B. Limitations

Although our proposed multiscale bilateral-weighted retinex algorithm outperforms other retinex approaches discussed above, it still has several potential limitations or open issues.

First, we simply select $\omega_m = 1/M$ (i.e., $\omega_1 = \omega_2 = \omega_3 = 0.33$ in Eq. 12) to combine the outputs of the three levels. It's difficult to prove that such a strategy is optimal for integrating these three-level image in a meaningful manner. We plan to introduce a more effective and accurate fusion method to combine the three-level outputs in the future. Additionally, the values of the three scales potentially affect the performance of the proposed algorithm and will be further investigated for surgical vision augmentation in our future work.

Next, it remains challenging for objectively no-reference quantitative assessment of surgical laparoscopic video images in robotic prostatectomy. Our defined quantitative objective metrics do not guarantee that they can precisely characterize image properties related to image visual quality. In particular, our metrics involve several parameters that possibly affect the performance of surgical vision assessment. We need to further investigate these parameters for accurate quality metrics of surgical video images. Also note that while Fig. 6 shows that the results of M4 is always worse than the original images, M4 achieves high scores in most of the objective metrics compared to the original images (Fig. 7 and Table 1). This inconsistency between subjective and objective evaluation also questions the robustness of our surgical vision assessment metrics. Our future study aims to develop more accurate assessment metrics to address such a validation issue and remove such an inconsistency between the subjective and objective evaluation.

On the other hand, we tested currently available image quality metrics such as cumulative probability of blur detection [42], dubbed blind/referenceless image spatial quality evaluator (BRISQUE) [43], local edge gradient analysis [44], and local phase coherence [45] on robotic-assisted prostatectomy video images. Unfortunately, they were inconsistent with the subjectively assessed results by the three surgeons. These metrics commonly work on typical real-world images that are physically and inherently different from laparoscopic video sequences. Generally speaking, it is a complicated but interesting and important issue to develop no-reference, quantitative, objective, precise quality metrics for robotic-assisted laparoscopic video images in our future work.

Third, all the compared enhancement algorithms involve various parameters that potentially affect their performance of surgical laparoscopic video processing. Even though the experimental results were comparable when using various parameters in the algorithms (Fig. 9), it is still interesting to further investigate how these algorithmic parameters affect their performance in the case of laparoscopic video enhancement.

Eventually, our proposed augmentation approach requires about 3.1 seconds to process a high-definition image with a size of 1920×1080, and it is difficult to achieve real-time operation of standard hardware. However, in the future, we are confident that multithreaded programing, code optimization, and GPU implementation will enable real-time performance. Additionally, we also strive for a user study to evaluate the practical effectiveness of our method in accordance with the real-time implementation and settings of surgical systems.

VII. Conclusions

This paper presented a multiscale bilateral-weighted retinex modeling framework to augment surgical endoscopic vision for robotic surgery. We modified the multiscale retinex method by introducing bilateral weights to improve its performance. We also define several no-reference image visual quality assessment metrics to quantitatively and objectively evaluate endoscopic vision. The experimental results demonstrate that our method provides an effective and efficient strategy for endoscopic vision augmentation. Moreover, both subjective and quantitative objective assessments demonstrate that our method outperforms other retinex algorithms. In particular, the visual quality contrast, sharpness, naturalness, and hybrid visibility of endoscopic vision was improved from (0.35, 0.56, 0.03, 0.81) to (1.13, 0.82,0.29, 1.06), respectively.

Acknowledgments

The authors thank the assistance of Dr. Stephen Pautler for facilitating the data acquisition, Dr. A. Jonathan McLeod and Dr.Uditha Jayarathne for helpful discussions.

References

[1] D. J. Mirota, G. D. Hager, and M. Ishii, "Vision-based navigation in image-guided interventions," *Annual Review of Biomedical Engineering*, vol. 13, no. 5, pp. 297–319, 2011.

[2] J. W. Yaxley, G. D. Coughlin, S. K. Chambers, S. Occhipinti, H. Samaratunga, L. Zajdlewicz, N. Dunglison, R. Carter, S. Williams, D. J. Payton, J. Perry-Keene, M. F. Lavin, and R. A. Gardiner, "Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled phase 3 study," *The Lancet*, vol. 388, no. 10049, pp. 1057–1066, 2016.

[3] J. J. Rassweiler, R. Autorino, J. Klein, A. Mottrie, A. S. Goezen, J.-U. Stolzenburg, K. H. Rha, M. Schurr, J. Kaouk, V. Patel, P. Dasgupta, and E. Liatsikos, "Future of robotic surgery in urology," *BJU International*, vol. 120, no. 6, pp. 822–841, 2017.

[4] J. J. McCann, "Retinex at 50: color theory and spatial algorithms, a review," *Journal of Electronic Imaging*, vol. 26, no. 3, pp. 031 204–1, 2017.

[5] B. Cai, X. Xu, K. Guo, K. Jia, B. Hu, and D. Tao, "A joint intrinsic-extrinsic prior model for retinex," in *Computer Vision (ICCV), IEEE International Conference on*, 2017, pp. 4020–4029.

[6] A. S. Baslamisli, H.-A. Le, and T. Gevers, "Cnn based learning using reflection and retinex models for intrinsic image," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2018, pp. 6674–6683.

[7] A. Galdran and A. Alvarez-Gila, "On the duality between retinex and image dehazing," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2018, pp. 8212–8221.

[8] X. Luo, A. J. McLeod, S. E. Pautler, C. M. Schlachta, and T. M. Peters, "Vision-based surgical field defogging," *IEEE Transactions on Medical Imaging*, vol. 36, no. 10, pp. 2021–2030, 2017.

[9] S. Wang and G. Luo, "Naturalness preserved image enhancement using a priori multi-layer lightness statistics," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 938–948, 2018.

[10] M. Ebner, *Color constancy*. John Wiley & Sons, 2007.

[11] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: survey and experiments," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2475–2489, 2011.

[12] M. Bertalmio, V. Caselles, and E. Provenzi, "Issues about retinex theory and contrast enhancement," *International Journal of Computer Vision*, vol. 83, no. 1, pp. 101–119, 2009.

[13] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel., "A variational framework for retinex," *International Journal of Computer Vision*, vol. 52, no. 1, pp. 7–23, 2003.

[14] M. K. Ng and W. Wang, "A total variation model for retinex," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 345–365, 2011.

[15] W. Ma, J.-M. Morel, S. Osher, and A. Chien, "An L_1-based variational model for retinex theory and its application to medical images," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2011, pp. 153–160.

[16] L. Wang, L. Xiao, H. Liu, and Z. Wei, "Variational bayesian method for retinex," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3381–3396, 2014.

[17] E. Provenzi and V. Caselles, "A wavelet perspective on variational perceptually-inspired color enhancement," *International Journal of Computer Vision*, vol. 106, no. 2, pp. 153–171, 2014.

[18] E. Provenzi, M. Fierro, A. Rizzi, L. Carli, D. Gadia, and D. Marini, "Random spray retinex: a new retinex implementation to investigate the local properties of the model," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 162–171, 2007.

[19] N. Banic and S. Loncaric, "Light random sprays retinex: Exploiting the noisy illumination estimation," *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 1240–1243, 2013.

[20] J. Morel, A. Petro, and C. Sbert, "A PDE formalization of retinex theory," *IEEE Transactions Image Processing*, vol. 19, no. 11, pp. 2825–2837, 2010.

[21] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 100–110, 2004.

[22] B. Jiang, G. Woodell, and D. Jobson., "Novel multi-scale retinex with color restoration on graphics processing unit," *Journal of Real-Time Image Processing*, vol. 10, no. 2, pp. 239–253, 2015.

[23] D. Zosso, G. Tran, and S. J. Osher, "Non-local retinex – unifying framework and beyond," *SIAM Journal on Imaging Sciences*, vol. 8, no. 2, pp. 787–826, 2015.

[24] H. Okuhata, H. Nakamura, S. Hara, H. Tsutsui, and T. Onoye, "Application of the real-time retinex image enhancement for endoscopic images," in *the IEEE Engineering in Medicine and Biology Society (EMBS), International Conference of*, 2013, pp. 3407–3410.

[25] T. H. Khan, S. K. Mohammed, M. S. Imtiaz, and K. A. Wahid, "Color reproduction and processing algorithm based on real-time mapping for endoscopic images," *SpringerPlus*, vol. 5, p. 17, 2016.

[26] M. Pedersen, O. Cherepkova, and A. Mohammed, "Image quality metrics for the evaluation and optimization of capsule video endoscopy enhancement techniques," *Journal of Imaging Science and Technology. 2017, 61 (4),*, vol. 61, no. 4, pp. 040 402–1–8, 2017.

[27] B. Sdiri, M. Kaaniche, F. A. Cheikh, A. Beghdadi, and O. J. Elle, "Efficient enhancement of stereo endoscopic images based on joint wavelet decomposition and binocular combination," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 33–45, 2019.

[28] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[29] J. Kopf, M. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 96–1–5, 2007.

[30] K. Bahrami and A. C. Kot, "A fast approach for no-reference image sharpness assessment based on maximum local variation," *IEEE Signal Processing Letters*, vol. 21, no. 5, pp. 751–755, 2014.

[31] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 657–667, 2013.

[32] T. Celik and T. Tjahjadi, “Automatic image equalization and contrast enhancement using gaussian mixture modeling,” *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 145–156, 2012.
[33] J. H. McDonald, *Handbook of Biological Statistics (3rd edition)*. Sparky House Publishing, 2014.
[34] R. Palma-Amestoy, E. Provenzi, M. Bertalmio, and V. Caselles, “A perceptually inspired variational framework for color enhancement,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 458–474, 2009.
[35] Q. Shan, J. Jia, and M. S. Brown, “Globally optimized linear windowed tone mapping,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 4, pp. 663–675, 2010.
[36] G. Deng, “A generalized unsharp masking algorithm,” *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1249–1261, 2011.
[37] C. Lee, C. Lee, and C.-S. Kim, “Contrast enhancement based on layered difference representation of 2D histograms,” *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5372–5384, 2013.
[38] S. Wang, J. Zheng, H.-M. Hu, and B. Li, “Naturalness preserved enhancement algorithm for non-uniform illumination images,” *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.
[39] H. Xu, G. Zhai, X. Wu, and X. Yang, “Generalized equalization model for image enhancement,” *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 68–82, 2014.
[40] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, “A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation,” *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4965–4977, 2015.
[41] Q. Zhang, Y. Nie, L. Zhang, and C. Xiao, “Underexposed video enhancement via perception-driven progressive fusion,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 6, pp. 1773–1785, 2016.
[42] N. D. Narvekar and L. J. Karam, “A no-reference image blur metric based on the cumulative probability of blur detection (cpbd),” *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2678–2683, 2011.
[43] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.
[44] C. Feichtenhofer, H. Fassold, and P. Schallauer, “A perceptual image sharpness metric based on local edge gradient analysis,” *IEEE Signal Processing Letters*, vol. 20, no. 4, pp. 379–382, 2013.
[45] R. Hassen, Z. Wang, and M. M. A. Salama, “Image sharpness assessment based on local phase coherence,” *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2798–2810, 2013.