Endoscopic video defogging using luminance blending

Xiongbiao Luo[1], Fan Yang[1] ✉, Hui-Qing Zeng[2], Yan-Ping Du[2]

[1]*School of Informatics, Xiamen University, Xiamen 361005, People's Republic of China*
[2]*Zhongshan Hospital, Xiamen University, Xiamen 361005, People's Republic of China*
✉ *E-mail: yf@xmu.edu.cn*

Endoscopic video sequences provide surgeons with direct surgical field or visualisation on anatomical targets in the patient during robotic surgery. Unfortunately, these video images are unavoidably hazy or foggy to prevent surgeons from clear surgical vision due to typical surgical operations such as ablation and cauterisation during surgery. This Letter aims at removing fog or smoke on endoscopic video sequences to enhance and maintain a direct and clear visualisation of the operating field during robotic surgery. The authors propose a new luminance blending framework that integrates contrast enhancement with visibility restoration for foggy endoscopic video processing. The proposed method was validated on clinical endoscopic videos that were collected from robotic surgery. The experimental results demonstrate that their method provides a promising means to effectively remove fog or smoke on endoscopic video images. In particular, the visual quality of defogged endoscopic images was improved from 0.5088 to 0.6475.

1. Introduction: Interventional endoscopes (e.g. bronchoscope and colonoscope) integrated with video cameras at their distal tips are widely introduced in minimally invasive surgery. The endoscope provides surgeons with real-time endoscopic video sequences that are shown on medical displays. On the basis of endoscopic vision or surgical field from these images, surgeons can directly visualise and examine abnormal tissues and treat or resect tumours in the body.

Unfortunately, the visual quality of endoscopic video images is unavoidably degraded because of surgical smoke or fog during robotic surgery. These endoscopic foggy images (Fig. 1) are generally generated from a surgical processing called cauterisation, which is usually employed to limit the bleeding vessels, while other typical operations such as laser ablation can also bring surgical smoke in surgical field. Such fog or smoke commonly distracts surgeons who may wait for a while without doing anything until surgical smoke is gone, which increases surgical time. On the other hand, surgical fog also degrades the clear visualisation of the surgical field from the endoscope, as well as covers the structural details (e.g. vessel structures) on the organ surface. This harmful issue leads to inappropriate device use and incorrectly targeted tissue, increasing surgical risks such as in tissue or tumour resection during endoscopic surgery. Therefore, endoscopic video defogging plays an essential role in enhancing and maintaining a clear field of surgical vision, not only for safety by preventing inadvertent injury, but also for improving precision and reducing operative time.

Endoscopic field defogging methods generally consists of hardware- and software-based strategies. While the former uses typical devices to remove smoke, the latter is algorithmic, i.e. computational photography techniques. This work develops a new luminance blending strategy for surgical video defogging. It combines a contrast enhancement procedure with a fast visibility recovery method to remove fog or smoke on endoscopic video sequences. We also quantitatively and objectively evaluate the experimental results of using our proposed method and others. The main contributions of this work are two-fold: (i) a new luminance blending approach with better performance than other defogging approaches and (ii) an objective image quality metric for quantitative assessment of dehazed images.

The remainder of this Letter is organised as follows. Section 2 briefly reviews work related to current dehazing methods. Our hybrid luminance blending-based dehazing method for vision augmentation is presented in Section 3, followed by the experiment settings in Section 4. Sections 5 shows and discusses the validation results before concluding this work in Section 6.

2. Related work: Real-world natural image and video dehazing or defogging techniques are widely discussed in computer vision and computational photography in the literature. Fattal [1] presented a graphical model used to calculate the atmospheric light for hazy-free image recovery. They assume that scene shading and transmission are locally independent of each other, which are not practical in applications. Tarel and Hautiere [2] introduced a fast visibility restoration strategy based on median filtering, but it usually results in colour distortion and easily fails at the image median filtering step that usually introduces null pixels. On the other hand, the fast visibility method also requires more efficient computation for real-time processing.

While He *et al.* [3] proposed dark channel-based atmospheric light and transmission estimation with soft editing, Meng *et al.* [4] employed the boundary constraint and contextual regularisation to modify this dark channel-based method, especially, they improved the computational efficiency and skipped soft editing. Nishino *et al.* [5] estimated two statistically independent components of the scene albedo and depth by using the Bayesian defogging model. While this Bayesian-based method works well, it also results in colour distortion. Ancuti and Ancuti [6] discussed a multi-scale fusion approach that combines the white balance with linearly transformed images extracted from hazy images. This multi-scale fusion approach is generally trapped in dealing with inhomogeneous fog due to loss of transmission depth information. While Sulami *et al.* [7] proposed a reduced formation model to describe image pixels in small patches as lines that are used to recover the atmospheric light orientation, Galdran *et al.* [8] presented an improved variational framework using inter-channel contrast in optimisation. More interestingly, fusion-based defogging is generally recognised as a promising framework to address the disadvantages of various dehazing methods [9].

More recently, deep learning-driven methods are increasingly developed for single image dehazing. While Ren *et al.* [10] employed multi-scale convolutional neural networks for single image dehazing, Li *et al.* [11] proposed All-In-One Dehazing Network (AOD-NET) to directly create the clean image through a lightweight convolutional neural network instead of separately
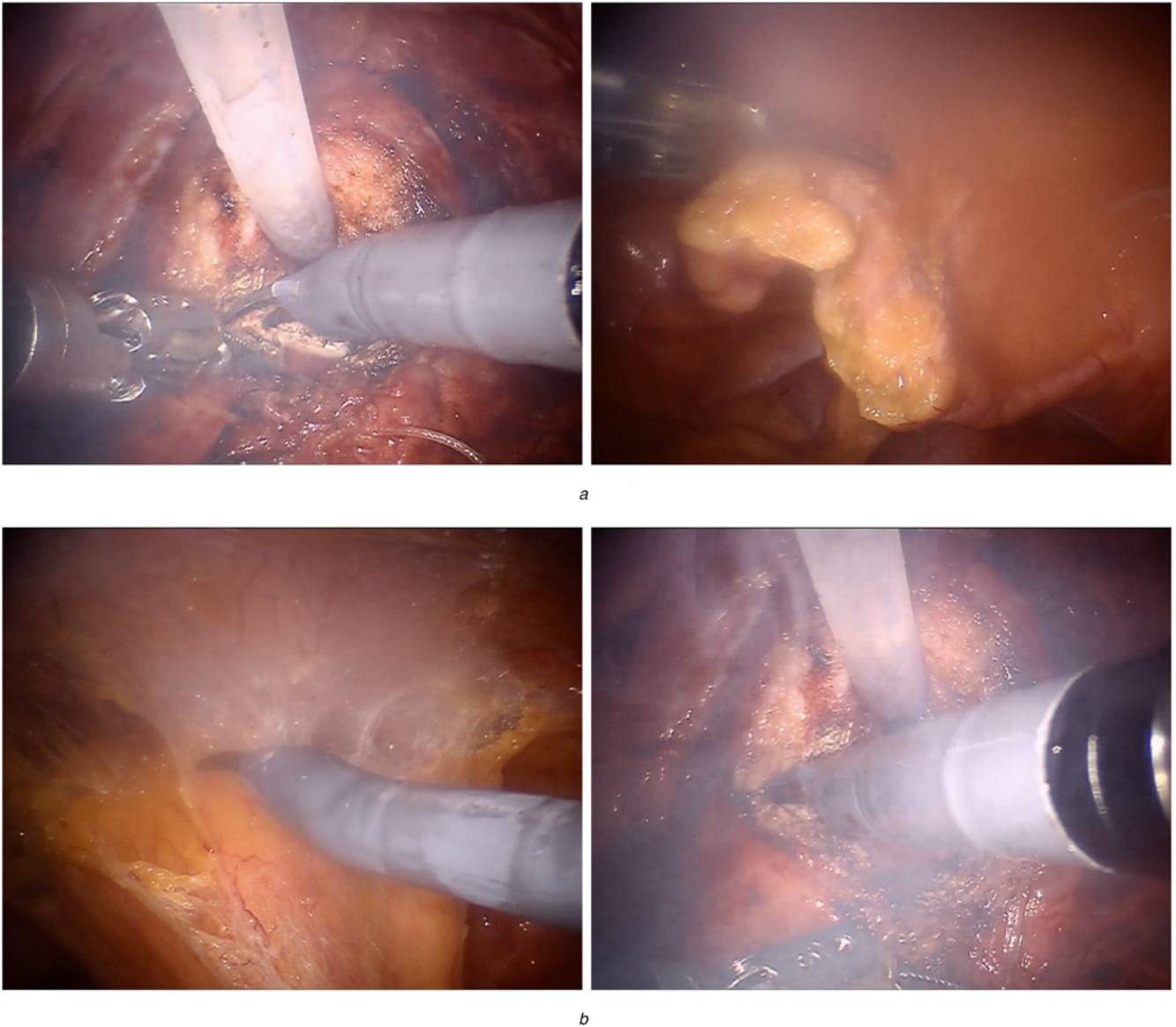
Fig. 1 *Hazy images in robotic-assisted endoscopic surgery*
a Thin smoke
b Heavy smoke

computing the transmission map and the atmospheric light for single image dehazing. Moreover, Ren *et al.* [12] developed a deep video dehazing method based on semantic segmentation, which can effectively use the abundant information that exists across neighbouring frames for precise dehazing. Liu *et al.* [13] introduced a simple generic model-agnostic convolutional neural network trained end-to-end to recover clear images from hazy inputs.

Although these methods work well on natural images, they remain challenging to deal with surgical endoscopic video image fog or smoke, particularly in the case of inhomogeneous or thick haze. This work aims to address the problem of hazy images or videos with inhomogeneous or thick haze, particularly foggy endoscopic videos.

3. Approaches: This section details our luminance blending framework for surgical endoscopic video defogging. Our method contains several steps: (i) contrast enhancement, (ii) visibility recovery, and filtering, and (iii) luminance blending. Fig. 2 shows the flowchart of our processing, as discussed in the following section.

3.1. Contrast enhancement: Surgical foggy images are of low-contrast and limited illumination, especially in hazy regions. The goal of contrast enhancement is to improve the contrast of hazy-less regions on the endoscopic image and calculate the luminance $L(u, v)$ and to enhance the luminance of the final defogged surgical image.

The contrast enhancement step assumes (i) most regions on the foggy image are hazy pixels that critically affect the mean of the foggy image and (ii) the level of haze in these regions depends on the distance between the atmospheric light and the scene, as discussed in [6]. On the basis of the assumption, we compute the enhanced luminance $L(u, v)$ by the magnifying difference between the surgical hazy image $I(u, v)$ and its average luminance value λ in the three channels $c \in \{r, g, b\}$

$$L_c(u, v) = \beta(I_c(u, v) - \lambda), \ \lambda = \frac{\sum_U \sum_V H(u, v)}{UV}, \quad (1)$$

where β is the magnification factor to control the luminance of the augmented foggy regions and $U \times V$ are the width and height of the hazy endoscopic image. The original luminance $H(u, v)$ at
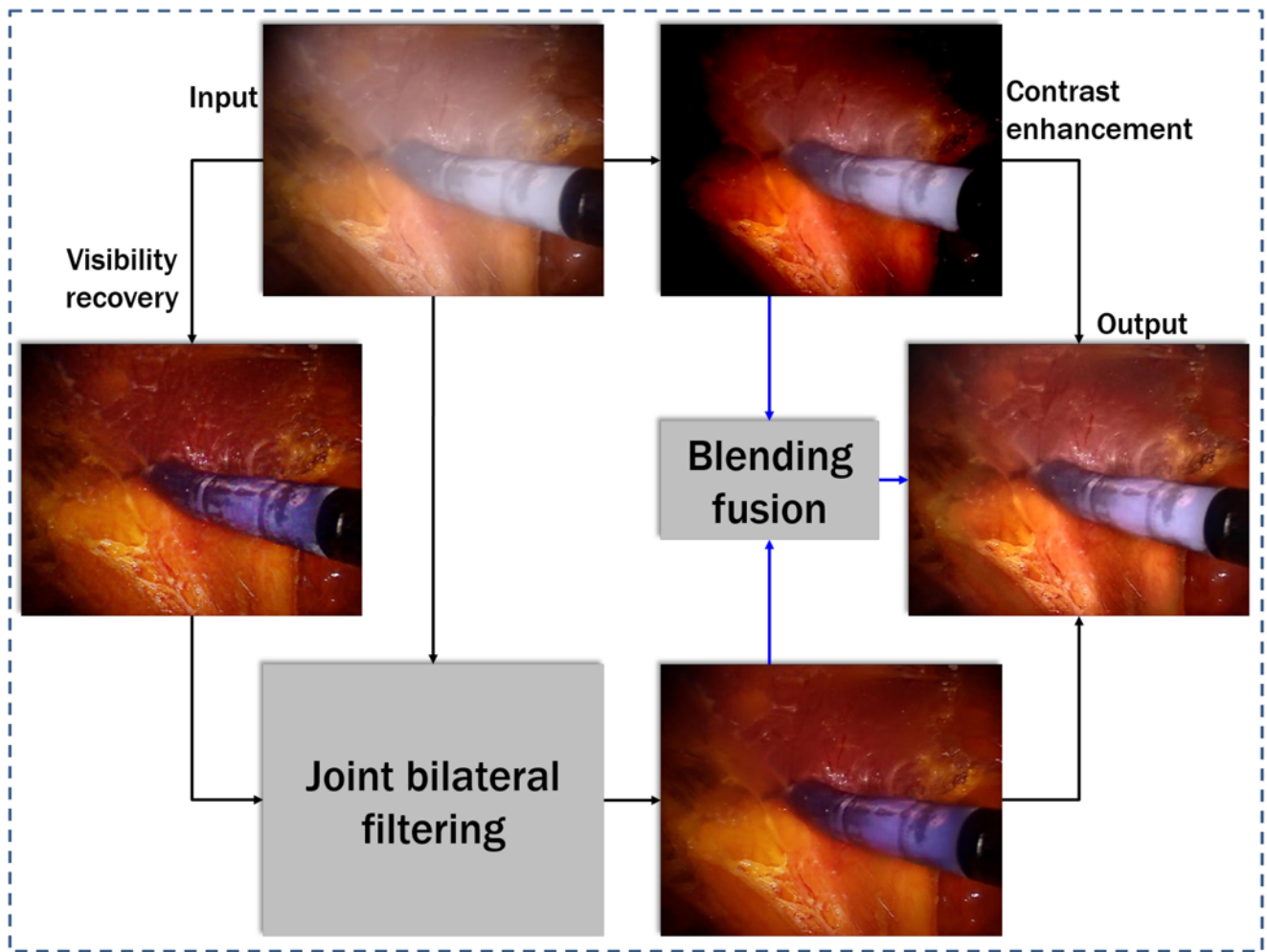
Fig. 2 *Flowchart of our proposed defogging method for night-time images*

each pixel is calculated by [14]

$$H(u, v) = a \times I_r(u, v) + b \times I_g(u, v) + c \times I_b(u, v), \quad (2)$$

where coefficients $a = 0.299$, $b = 0.587$, and $c = 0.114$.

3.2. Visibility recovery and filtering: A widely used physical imaging model is established for hazy images by Koschmieder's law [3]

$$I(u, v) = F(u, v)T(u, v) + A_\infty(1 - T(u, v)), \quad (3)$$

where $I(u, v)$ denotes an observed (foggy) image, $F(u, v)$ refers to as a haze-free image (also called scene radiance), and A_∞ indicates the atmospheric light or the sky luminance. The transmission map $T(u, v)$ describes the amount of the unscattered light entering a camera and can be computed by

$$T(u, v) = \exp(-kd(u, v)) \quad (4)$$

where k and $d(u, v)$ are the atmosphere's scattering factor and the distance between the camera and any objects in a scene.

On the basis of (3), we aim to solve hazy-free image $F(u, v)$ under the unknown variables A_∞ and $T(u, v)$. However, according to a fast visibility recovery method [2], we did not directly estimate $T(u, v)$ since it is difficult to precisely predict the transmission map related to depth information. To skip $T(u, v)$, the atmospheric veil $X(u, v)$ was employed [6]

$$X(u, v) = A_\infty(1 - T(u, v)), \quad T(u, v) = 1 - \frac{X(u, v)}{A_\infty}. \quad (5)$$

Then, (3) can be rewritten to calculate $F(u, v)$

$$F(u, v) = \frac{A_\infty(I(u, v) - X(u, v))}{A_\infty - X(u, v)}. \quad (6)$$

This requires the atmospheric light A_∞ and veil $X(u, v)$ for which robust estimates can be obtained much more easily than the depth and transmission maps in the original formulation (3). The methods that are used to determine A_∞ and veil $X(u, v)$ have been discussed in [2]. Here, we skip the technical details of how to estimate light A_∞ and veil $X(u, v)$.

Since the result $F(u, v)$ of the fast visibility recovery usually contains image noise and artefacts, we employ joint bilateral filtering to process $F(u, v)$ and obtain $J(u, v)$.

The bilateral filter is an edge-aware image processing method to denoise and simultaneously preserve edge information [15, 16]. The concept of joint bilateral filtering is to perform spatial filtering (particularly a Gaussian kernel) on a low-resolution image and simultaneously apply a range filter to process a high-resolution image (here the low- and high-resolution images refer to the recovery image $F(u, v)$ and the original image $I(u, v)$, respectively) [17]

$$J(p) = \frac{1}{K_p} \sum_{q \in N_p} F(q)\Omega(p, q), \quad (7)$$

$$\Omega(p, q) = \exp\left(\frac{\|p - q\|^2}{2\sigma_s^2} + \frac{\|I_u(p) - I_u(q)\|^2}{2\sigma_c^2}\right) \quad (8)$$

where $\boldsymbol{p} = (u, v)$, $\boldsymbol{q} = (\hat{u}, \hat{v})$, variances σ_s, σ_c in the region N_p centred at the pixel $\boldsymbol{p}$, and K_p is computed by

$$K_p = \sum_{\boldsymbol{q} \in N_p} \exp\left(\frac{\|\boldsymbol{p} - \boldsymbol{q}\|^2}{2\sigma_\text{s}^2} + \frac{\|\boldsymbol{I}_\text{u}(\boldsymbol{p}) - \boldsymbol{I}_\text{u}(\boldsymbol{q})\|^2}{2\sigma_\text{c}^2} \right), \qquad (9)$$

which is the normalisation term to guarantee the sum of the weights for all the pixels to be one.

3.3. Luminance blending: This step is to estimate illumination on image $\boldsymbol{J}(u, v)$ and $\boldsymbol{L}(u, v)$ and blend their illumination to improve the illumination of the defogged endoscopic surgical image.

We transfer the images $\boldsymbol{J}(u, v)$ and $\boldsymbol{L}(u, v)$ from the red, green and blue (RGB) to YCbCr colour space. For the Y-component or luminance component of them, we used recursive filtering [18] to estimate the illumination of $\boldsymbol{J}(u, v)$ and $\boldsymbol{L}(u, v)$ and obtain $G_J(u, v)$ and $G_L(u, v)$. By using image illumination $G_J(u, v)$ and $G_L(u, v)$, we seek to recognise pixels in hazy regions. So, a
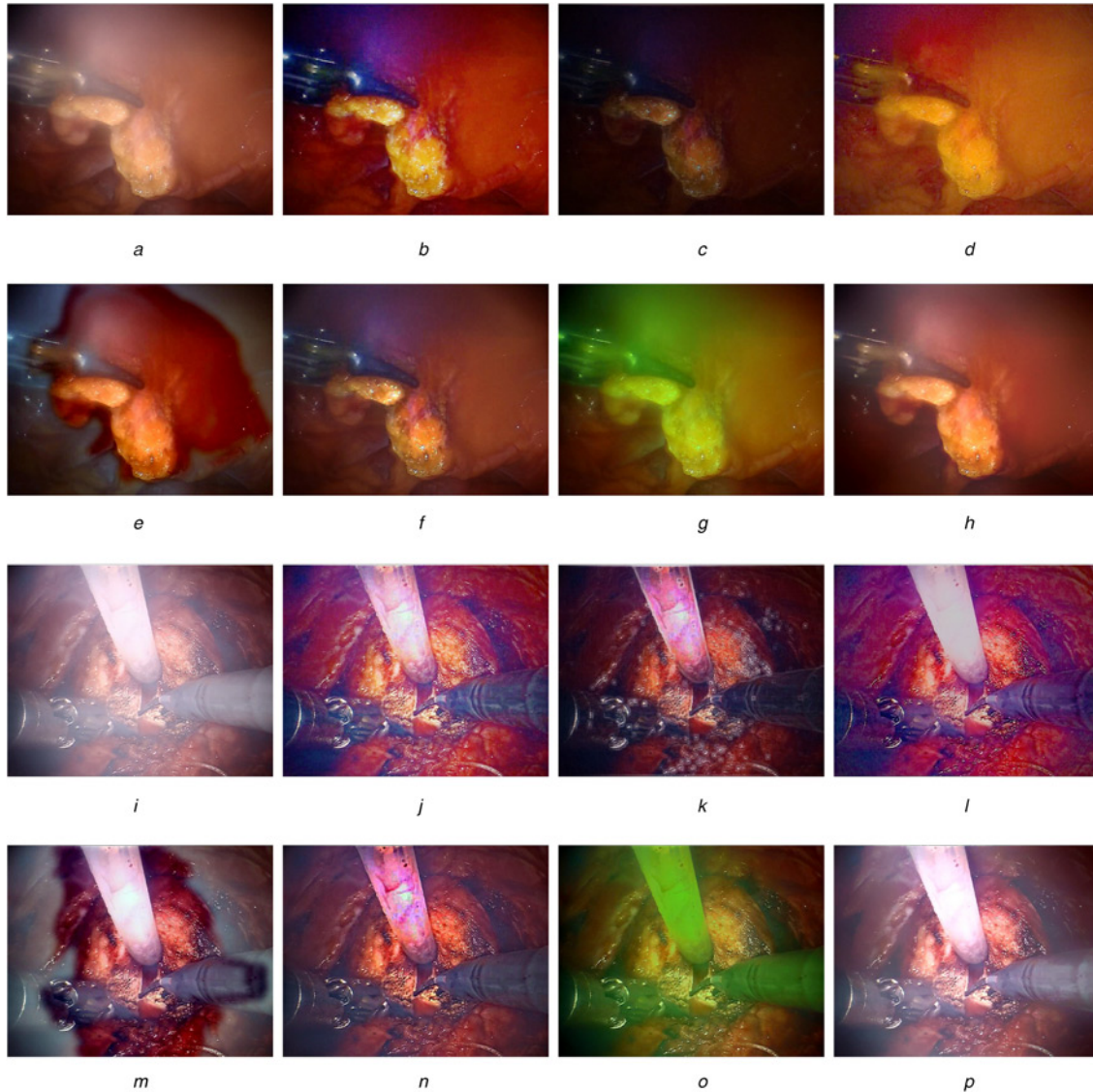

Fig. 3 *Comparison of using various defogging methods: (a)–(h) thin-fog image and (i)–(p) thick-fog image*
a Thin-fog image
b M1 [2]
c M2 [3]
d M3 [5]
e M4 [6]
f M5 [4]
g M6 [7]
h M7 (ours)
i Thick-fog image
j M1 [2]
k M2 [3]
l M3 [5]
m M4 [6]
n M5 [4]
o M6 [7]
p M7 (ours)

Table 1 Quantitatively objective assessment of the results obtained from the seven defogging approaches

Approaches	M0	M1 [2]	M2 [3]	M3 [5]	M4 [6]	M5 [4]	M6 [7]	M7 (ours)
SSIM $\mathcal{S}$	—	0.6587	0.4781	0.6676	0.6488	0.7978	0.3944	0.9275
naturalness $\mathcal{N}$	0.1411	0.2319	0.0218	0.1097	0.1439	0.0752	0.0602	0.2274
hybrid ψ	—	0.4890	0.2956	0.4445	0.4468	0.5088	0.2608	0.6475

M0 indicates the quantitative results of the original foggy images and does not have the SSIM index that is a reference-based metric

weight function $W_K(G_K(u, v))$, $K \in \{J, L\}$ is empirically introduced, and the output $O_e(u, v)$, $e \in \{Y, Cb, Cr\}$ of the blending fusion can be formulated

$$O_e(u, v) = \frac{\sum_{K \in \{J, L\}} W_K(G_K(u, v))O_e(u, v)}{\sum_{K \in \{J, L\}} W_K(G_K(u, v))}. \qquad (10)$$

Note that the weight function $W_K(\cdot)$ (also called the weight matrix) depends on the level of smoke. In the heavy-smoke case, if the foggy pixel intensity belongs to the range of 16–128, these pixels will be assigned with weight 1. In the thin-smoke case, the pixels on the interval [128, 235] will be assigned with weight 1. The luminance output $O_Y(u, v)$ may not be distributed into the full range of pixel intensity, resulting in a low-contrast image. We implement the following linear transformation to stretch its histogram to a specific intensity range $[P, Q]$

$$\hat{O}_Y(u, v) = P + \frac{O_Y(u, v) - O_{\mathrm{Min}}(u, v)}{O_{\mathrm{Max}}(u, v) - O_{\mathrm{Min}}(u, v)}(Q - P), \qquad (11)$$

where $\hat{O}_Y(u, v)$ denotes the final luminance, $O_{\mathrm{Min}}(u, v)$ and $O_{\mathrm{Max}}(u, v)$ are the minimum and maximum intensities of the blending output $O_Y(u, v)$, respectively. We empirically set $P = 15$ and $Q = 236$ in our work. Eventually, we combine the Y-component $\hat{O}_Y(u, v)$ and the chromatic components $O_{Cb}(u, v)$ and $O_{Cr}(u, v)$ and transform them into the RGB colour space, obtaining the final defogged image.

4. Validation: Foggy endoscopic videos were acquired from robotic surgery. All the experiments were executed on a laptop installed with Windows 8.1 Professional 64 bit system, 32.0 GB memory, and processor Intel(R) Xeon(R) CPU $\times$ 8 2.8 GHz and MATLAB R2017a. We tested about 1200 frames in this Letter.

We compare the proposed method with the following approaches: (i) M1, Tarel *et al.* [2], (ii) M2, He *et al.* [3], (iii) M3, Nishino *et al.* [5], (iv) M4, Ancuti and Ancuti [6], (v) M5, Meng *et al.* [4], (vi) M6, Sulami *et al.* [7], and (vii) M7, our method.

We introduce a naturalness metric to depict how natural surgical images appear based on statistically analysing thousands of images [19]. On the other hand, we also employ structural similarity index (SSIM) [20] to evaluate structural information on images. Eventually, we define a hybrid quality metric ψ to evaluate defogged endoscopic images

$$\psi = \gamma\mathcal{S} + (1 - \gamma)\mathcal{N}, \qquad (12)$$

where $\mathcal{S}$ denotes the SSIM and $\mathcal{N}$ indicates the naturalness. The coefficient γ is set to 0.6, which was experimentally determined to balance the structural information and naturalness.

5. Results and discussion: Fig. 3 visually compares the defogged results of endoscopic images with thin and thick fogs. The visual quality of the results demonstrates that our method works better than others since it removes fog without introducing colour distortion. On the other hand, two surgeons manually inspected all the dehazed results and generally believe that our defogged

method outperforms other approaches since the subjective visual quality of using our method is more natural and colourful, which is better than the original foggy image.

Table 1 quantitatively compares the objective assessment of the dehazed results obtained from the seven approaches. The quantitative assessment results show that our proposed approach outperforms other methods. While the average naturalness of M1 and our proposed method M7 were comparable, the average SSIM was improved from 0.7978 to 0.9275. More interestingly, the average hybrid quality of our methods was 0.6475, which was much better than other approaches (M5 provides 0.5088).

Additionally, the computational times of the methods M1, M3, M4, M5, M6, and M7 were 31.3, 62.3, 1.3, 5.2, 75.6, and 1.1 s/frame, respectively. Method M2 deals with an image in more than 2700 s since soft editing was extremely slow [3]. Our method works faster than others.

This work aims to enhance the surgical field visualisation of endoscopic surgery. We developed a new luminance blending defogging algorithm. The experimental results demonstrate that our algorithm outperforms others from subjective and objective evaluations. The effectiveness of our algorithm lies in fusing the advantages of the enhancement and restoration dehazing methods. Our method has several potential limitations including unclear parameter sensitivity, effective enhancement, quality assessment, and heavy processing time. These limitations will be further investigated in the future. In addition, though our method works better than other approaches, it still introduces colour distortion, which will be further investigated.

6. Conclusion: We proposed a new luminance blending defogging framework that integrates contrast enhancement, joint bilateral filtering, and visibility recovery to remove smoke in endoscopic videos from robotic surgery. We evaluated our method on endoscopic video sequences acquired from robotic prostatectomy. The experimental results demonstrate the effectiveness of our proposed method, which outperforms other approaches. In particular, our method improved the hybrid quality of the dehazed results from 0.5088 to 0.6475.

8 References

[1] Fattal R.: 'Single image dehazing', *ACM Trans. Graph.*, 2008, **27**, (3), pp. 1–10

[2] Tarel J.P., Hautiere N.: 'Fast visibility restoration from a single color or gray level image'. IEEE Int. Conf. Computer Vision (ICCV), 2009, pp. 2201–2208

[3] He K., Sun J., Tang X.: 'Single image haze removal using dark channel prior', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **33**, (12), pp. 2341–2353

[4] Meng G., Wang Y., Duan J., ET AL.: 'Efficient image dehazing with boundary constraint and contextual regularization'. IEEE Int. Conf. Computer Vision (ICCV), 2013, pp. 617–624

[5]   Nishino K., Kratz L., Lombardi S.: 'Bayesian defogging', *Int. J. Comput. Vis.*, 2012, **98**, (3), pp. 263–278

[6]   Ancuti C.O., Ancuti C.: 'Single image dehazing by multi-scale fusion', *IEEE Trans. Image Process.*, 2013, **22**, (8), pp. 3271–3282

[7]   Sulami M., Glatzer I., Fattal R., ET AL.: 'Automatic recovery of the atmospheric light in hazy images'. IEEE Int. Conf. Computational Photography (ICCP), 2014, pp. 1–11

[8]   Galdran A., Vazquez Corral J., Pardo D., ET AL.: 'Enhanced variational image dehazing', *SIAM J. Imaging Sci.*, 2015, **8**, (3), pp. 1519–1546

[9]   Xu Y., Wen J., Fei L., ET AL.: 'Review of video and image defogging algorithms and related studies on image restoration and enhancement', *IEEE Assess*, 2016, **4**, pp. 165–188

[10] Ren W., Liu S., Zhang H., ET AL.: 'Single image dehazing via multi-scale convolutional neural networks'. European Conf. Computer Vision (ECCV), 2016, pp. 154–169

[11] Li B., Peng X., Wang Z., ET AL.: 'An all-in-one network for dehazing and beyond', arXiv preprint arXiv:170706543, 2017

[12] Ren W., Zhang J., Xu X., ET AL.: 'Deep video dehazing with semantic segmentation', *IEEE Trans. Image Process.*, 2018, **28**, (4), pp. 1895–1908

[13] Liu Z., Xiao B., Alrabeiah M., ET AL.: 'Single image dehazing with a generic model-agnostic convolutional neural network', *IEEE Signal Process. Lett.*, 2019, **26**, (6), pp. 833–837

[14] Szeliski R.: 'Computer vision: algorithms and applications' (Springer, 2011)

[15] Durand F., Dorsey J.: 'Fast bilateral filtering for the display of high-dynamic-range images', *ACM Trans. Graph.*, 2002, **21**, (3), pp. 257–266

[16] Chen J., Paris S., Durand F.: 'Real-time edge-aware image processing with the bilateral grid', *ACM Trans. Graph.*, 2007, **26**, (3), pp. 1–9

[17] Kopf J., Cohen M., Lischinski D., ET AL.: 'Joint bilateral upsampling', *ACM Trans. Graph.*, 2007, **26**, (3), pp. 961–5

[18] Ramirez J.M., Paredes J.L.: 'Recursive weighted myriad based filters and their optimizations', *IEEE Trans. Signal Process.*, 2016, **64**, (15), pp. 4027–4039

[19] Yeganeh H., Wang Z.: 'Objective quality assessment of tone-mapped images', *IEEE Trans. Image Process.*, 2013, **22**, (2), pp. 657–667

[20] Wang Z., Bovik A.C., Sheikh H.R., ET AL.: 'Image quality assessment: from error visibility to structural similarity', *IEEE Trans. Image Process.*, 2004, **13**, (4), pp. 600–612