Stochastic Capsule Endoscopy Image Enhancement

Ahmed Mohammed [1],* (iD) **, Ivar Farup** [1]**, Marius Pedersen** [1]**, Øistein Hovde** [2] **and Sule Yildirim** [3]

[1] Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway
[2] Department of Gastroenterology, Innlandet Hospital Trust, Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[3] Department of Information Security and Communication Technology Norwegian University of Science and Technology, Gjøvik, Norway
* Correspondence: mohammed.kedir@ntnu.no; Tel.: +47-486-766-16

Abstract: Capsule endoscopy, which uses a wireless camera to take images of the digestive tract, is emerging as an alternative to traditional colonoscopy. The diagnostic values of these images depend on the quality of revealed underlying tissue surfaces. In this paper, we consider the problem of enhancing the visibility of detail and shadowed tissue surfaces for capsule endoscopy images. Using concentric circles at each pixel for random walks combined with stochastic sampling, the proposed method enhances the details of vessel and tissue surfaces. The framework decomposes the image into two detail layers that contain shadowed tissue surfaces and detail features. The target pixel value is recalculated for the smooth layer using similarity of the target pixel to neighboring pixels by weighting against the total gradient variation and intensity differences. In order to evaluate the diagnostic image quality of the proposed method, we used clinical subjective evaluation with a rank order on selected KID image database and compared to state of the art enhancement methods. The result showed that the proposed method provides a better result in terms of diagnostic image quality and objective quality contrast metrics and structural similarity index.

Keywords: capsule video endoscopy; stochastic sampling; random walks; color gradient; image decomposition

1. Introduction

Capsule video endoscopy (CVE) has revolutionized the diagnostic work-up in the field of esophagus, small bowel and colon imaging. The colon traditionally has been examined via optical colonoscopy, a procedure perceived by many to be uncomfortable and embarrassing. Colon capsule endoscopy (CCE) is an alternative way to visualize the colon. Some of commercially available CCE devices include PillCam COLON I and II from Given Imaging. CCE devices are equipped with miniaturized camera, LED light source, radio transmitter and battery contained in an easy-to-swallow capsule. Unfortunately, due to power and volume limitations, capsule endoscopy does not provide image quality equivalent to traditional colonoscopy. As the capsule progresses though the colon by peristalsis of the digestive tract, orientation of the capsule is uncontrolled and images are taken under low illumination. In addition, CCE images suffer from high compression ratio, noise from CMOS (complementary metal-oxide semiconductor) image sensor and low image resolution (commonly 256 x 256). The problem of capsule image quality enhancement has been an active research topic since capsules appeared commercially in 2006. The present review here is brief due to space limitations, but is intended to highlight the broad categories of existing algorithms and to provide appropriate background for our work.

We categorize CVE image enhancement techniques based on particular image attributes they focus on for accurate diagnosis of pathologies. Regardless of the specific method, the goals of any CVE image enhancement techniques can be categorized into four main objectives: making blood vessels visible; removing or deemphasizing specular reflections and illumination variation; making tissues

visible; keeping the original color tone (because where the colors are changed, the physician requires re-training to view such images [1]).

Works on blood vessel detail enhancement mostly focus on exploiting optical properties of blood vessels for enhancement of lesions and vascular patterns. For example, the 415 nm image channel analyzes the fine surface architecture of the mucosa and the superficial capillary network, whilst the 540 nm image channel analyzes the collective vessels in the depth of the mucosa. Flexible spectral imaging color enhancement(FICE) [2] takes white-light endoscopic images from the video processor and processes them by emphasizing certain ranges of wavelengths by spectral decompositions. On the other hand, Narrow Band Imaging (NBI) [3] uses a special set of filters that are interposed after the light source to restrict the incident light to two narrow bands of wavelengths (blue centered on 415 nm and green centered on 540 nm).

The other topic discussed in the literature is illumination variation across the image. The amount of illumination from point sources incident on the scene being viewed changes slower than the reflectance. Using this concept, Ramaraj et al. [4], proposed a homomorphic filtering technique. By applying frequency domain transformation to the input image using DFT and appropriate design of Butterworth filter they claimed to obtain an enhancement result compared to contrast limited adaptive histogram equalization (CLAHE) [5]. More recently, Okuhata et al. [6] applied retinex theory to the problem of CVE image enhancement. The authors modelled the problem with a total variational model algorithm that was constructed to minimize the cost function in terms of reflectance and illuminance images. On the other hand, in order to make tissue details visible, many methods have been proposed to deal with image denoising and contrast enhancement. Palanisamy et al. [7] proposed CVE image denoising using dual tree double density complex wavelet transform. By thresholding the wavelet coefficients in all sub bands based on a maximum a posteriori probability estimator, the algorithm calculated the denoised image. Liu et al. [8] proposed CVE image de-blurring using total variation minimization. By incorporating the monotone fast iterative shrinkage or thresholding algorithm (MFISTA) combined with the fast gradient projection algorithm (proposed recently by Beck et al. [9]) total variation deblurring was extended to deal with multichannel (e.g. color) images for CVE. Considering the importance of color for the assessment of abnormalities, contrast enhancement should also preserve the color tones of CVE images. Vu et al. [10] proposes image enhancement techniques that can preserve the original color tones. This work is based on the idea that the color gamut of the human small bowel is generally restricted to a subspace of the imaging system's color space (e.g., 24 bit RGB), and which is more specific than that of typical natural images. Theoretically, an image is considered more informative if its histogram resembles a uniform distribution over a color space. Therefore, to preserve the original color tones a histogram equalization technique is applied in the proposed gastro-intestine color space. Similarly, Imtiaz et al. [11] proposed endoscopic image enhancement using an adaptive sigmoidal function and space-variant color reproduction. By using texture information, a new chrominance component was generated modifying the old chrominance component. The method was claimed to highlight some of the tissue and vascular patterns. Other works on tissue detail enhancement include Li et al. [12], in which they propose enhancement via adaptive contrast diffusion. This was done by applying diffusion in the contrast domain through eigenvalues of the Hessian at the target pixel.

In addition, several image and video processing algorithms (not tailored for a specific CVE platform) have been proposed to enhance CVE images. Attar et al. [13] combined histogram equalization with an edge preserving process to enhance WCE images. Other similar works include [5,14,15].

To summarize, most of the methods proposed for CVE image enhancement repurpose existing methods of natural image enhancement and fail to preserve detail texture for different segment of CVE images. In addition, they address the four requirements that we discussed earlier as separate and independent problems.

This paper is based on previous works by Kolås et al. [16], Black et al. [17] and Estrada et al. [18]. Kolås et al. [16] proposed a framework based on stochastic sampling for each pixel in the local neighborhood, where the local reference lightness and darkness points in each chromatic channel are calculated to find average estimate of the target pixel. On the other hand Black et al. [17] introduced connection between random walks and robust anisotropic diffusion. Following that line, Azzabou et al. [19] proposed random walks for image denoising on similar structures by recovering probabilistic densities, capturing co-occurrences of visual appearances at scale spaces. Similarly, Estrada et al. [18] proposed image denoising based on random walks across (unlike Azzabou et al. [19]) arbitrary neighborhoods surrounding a given pixel. The size and shape of each neighborhood was determined by the configuration and similarity of nearby pixels. By noting the connection between stochastic sampling and random walks, the contribution of this work can be summarized as follows. Firstly, we propose a method that can enhance shadows and detail structures for CVE images. Using concentric circles for random walks and stochastic sampling, the proposed method captures similar structures, together with local reference lightness and darkness points, simultaneously. The second contribution, which is part of the detail enhancement, is edge-aware smoothing that is based on the similarity of the target and neighboring pixels. This is estimated based on a new weighting function of total gradient variation and intensity difference.

The outline of the article is as follows: in Section 2, we re-visit the subjects of random walks and stochastic sampling. In Section 3, we present our approach and detail derivation by showing both theoretically and numerically. In Section 4, we present implementation and pseudocode of the proposed method. In Section 5, we evaluate the framework's results both subjectively and with objective metrics, along with comparison to other works. Finally, in Section 6 we present discussion and conclusion respectively.

2. BACKGROUND

2.1. Anisotropic diffusion and random walks

Inspired by the heat equation, Perona and Malik et al. [20], proposed a diffusion process that encourages intra-regional smoothing in preference to inter-regional smoothing. Given a heat equation, described by $\partial I(x,y,t)/\partial t = div(c(x,y,t)\nabla I)$ where $I(x,y,0)$ is the original image at time $t = 0$, ∇I is the image gradient and $c(x,y,t)$ is a constant conduction coefficient. By modifying the original formulation to include a variable conduction coefficient $c(x,y,t) = g(\|\nabla I\|)$ where g is a monotonically decreasing function, they showed that edge aware smoothing could be obtained. This work has motivated large number of researchers to study the mathematical properties and implementation of the diffusion process for denoising. Despite numerous advantages, theoretical justification [17], and numerous provisions of such a method one can claim that it can not deal with image textures [21]. Following this work many alternative formulation of diffusion process have been formulated using a total variation approach [22] and iterative wavelet shrinkage [23,24]. The most interesting alternative formulation of the diffusion process for our current work involves random walks [17–19]. Black et al. [16] points out a connection between random walks in neighborhoods and robust anisotropic diffusion. Consider image intensity difference between target pixel s and neighboring pixel p. Within one of piecewise constant image regions, these neighbor differences will be small, zero-mean, and normally distributed. Hence, an optimal estimator for the "true" value of the image intensity s is equivalent to choosing the mean of the neighboring intensity values. The neighbor differences will not be normally distributed, however, for an image region that includes a boundary (intensity discontinuity).

2.2. Retinex-inspired Envelope with Stochastic Sampling

The Human Visual System (HVS) has complex and robust mechanisms to acquire useful information from the physical environment. The intensity of a target pixel is heavily influenced
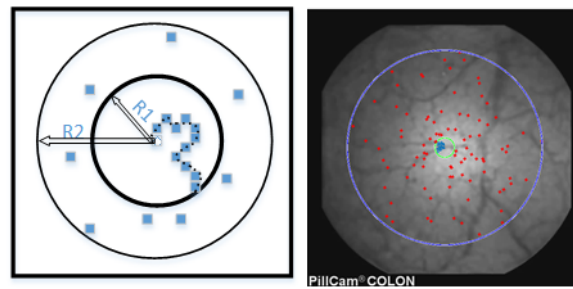
Figure 1. Concentric sampling regions for the target pixel at the center to characterize local visual context. As shown in the right-hand image, once the random walk is beyond the inner circle random sampling is done, marked by the red dots. All of the samples are aggregated to form one chain.

by neighboring pixel values. In [16] Kolås proposed a framework, where the random sprays were used to calculate two envelope functions representing the local reference black and white points. Given the two envelopes, target pixel s is adjusted by contrast stretching as $S = \frac{s - E_{\min}}{E_{\max} - E_{\min}}$, where S is the modified intensity value, $E_{\min}$ and $E_{\max}$ are the minimum and maximum values of the envelope at the target pixel. For further details see [16]. Inspired by anisotropic diffusion and stochastic sampling in that work, a new method to enhance CVE image shadow and detail structures is presented. In the next section, we discuss the proposed method.

3. STOCHASTIC CVE IMAGE ENHANCEMENT

The HVS uses saccades several times per second to move the fovea between points of interest and build an understanding of our visual environment. This has been the origin of many local contrast enhancement techniques. Different spatial sampling techniques have been applied to estimate the relative value of the target pixel [16]. In this article, we propose a different type of sampling for the neighborhood of the target pixel versus other image-wide regions.

Our approach involves exploring similar local neighborhood and lightness as well as darkness pixel. Consider a target pixel at the center of the circle in Fig. (1). In order to smooth and contrast-enhance the target pixel simultaneously, similar local neighborhood pixels are explored through random walk within the inner circle, $R1$ whilst local lightness and darkness pixels are explored in the outer circle, $R2$. Starting from each pixel, a random walk is initialized in a 3x3 neighborhood until it passes the inner circle. This enables us to determine similar local texture. Once the random walk is out of the inner circle, it samples randomly in-between the inner circle and outer circle to estimate local lightness and darkness see Fig. (2). Finally, both the random walk and random sampling are combined to estimate the final local lightness and darkness, which result in local contrast enhanced image. In the next subsections, we will go through the details of the proposed sampling method.

3.1. Smoothing

Given initial pixel position x_0, the random walk explores the local neighborhood within the inner circle. Let's assume for a given iteration the random walk passed through the pixel positions $x \in X$, where X is a set of all samples in a single chain (N number of iteration is done for each pixel and in subsequent sections, we refer the total number of samples in each iteration, as a chain). Since we are interested in exploring similar texture in the neighborhood, the similarity of the target pixel and x depends on the image gradient between the starting pixel, the neighboring pixel, and their intensity difference. By noting this fact, for each chain m the similarity of neighboring pixel x_{j+1} and initial pixel x_0 can be expressed by Eq. (1). Introducing artifacts in smoothing could lead to artificial texture on the tissue surface, which might result in inaccurate diagnosis. Hence, by using a structural tensor description of the gradient features, which allows a more precise description of the local gradient, the similarity is given as follows:

(a) Stochastic sampling

(b) Weight estimation
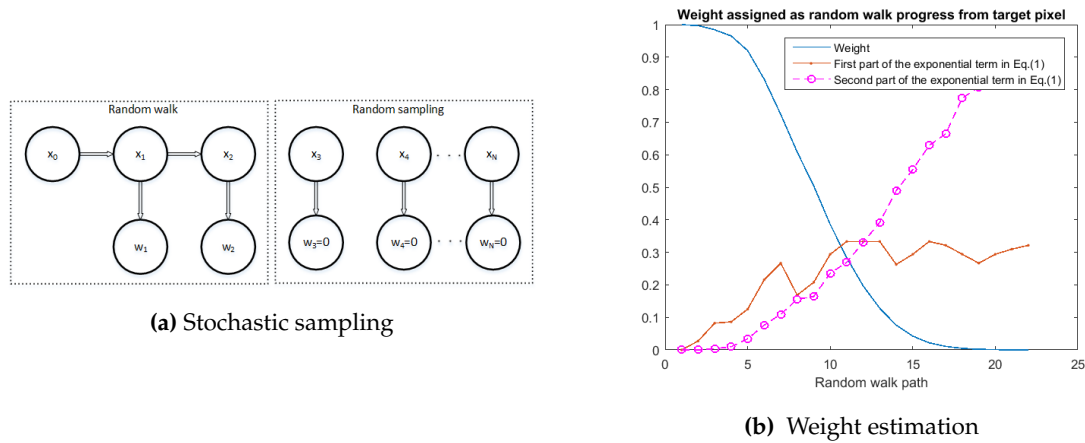
Figure 2. (a) Stochastic sampling: Random walks capture low variance neighbors that are essential for smoothing whilst random sampling captures high variance intensity variations that are essential for local contrast enhancement Therefore, random walks are used for smoothing whilst random samples weights are set to zero. (b) weight estimation of random walks along path starting from target pixel. Pixels inside the inner circle are assigned weights according to Eq. (1)

$$
{}^m w_0^{j+1}(x_0|x_{j+1}) = e^{-\left\{ \left\| \frac{x_0 - x_{j+1}}{\sigma_I} \right\|_1 + \left\| \frac{TV(\nabla I)}{\sigma_g} \right\|_1 \right\}}
\tag{1}
$$

The first part of the exponential term in Eq. (1) represents the l_1 norm of the intensity difference between the initial pixel and neighboring pixel normalized to a constant. The second term represents the total variation of eigenvalues of the structural tensors at each pixel normalized to a constant. The total variation term measures whether the random walk has crossed edges or not. Similar to anisotropic diffusion, by controlling the normalization constant, edge aware smoothing is obtained. The total variation of the gradient can be rewritten as

$$
TV(\nabla I) = \left\| \sum_{n=x_0}^{x_{j+1}} \sqrt{\left(\lambda_+^{n+1} - \lambda_-^{n+1} \right)} - \sqrt{\left(\lambda_+^n - \lambda_-^n \right)} \right\| = \|DG\|_1
\tag{2}
$$

where $\|.\|_1$ is the l_1 norm, G is the gradient of the neighboring pixel along the random walk path (chain) and D is given by

$$
D = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \bullet & \bullet & \\ & & & & \bullet \end{bmatrix} \quad
G = \begin{bmatrix} \left(\lambda_+^{n+1} - \lambda_-^{n+1} \right) \\ \left(\lambda_+^n - \lambda_-^n \right) \\ \bullet \\ \bullet \\ \bullet \\ \left(\lambda_+^0 - \lambda_-^0 \right) \end{bmatrix}
\tag{3}
$$

Formulating the similarity of the target and neighboring pixels in this way gives us the flexibility to quantify if the random walk is in a different regions or texture, since the total variation of the gradient accumulates as the random walk moves to the next pixel. Moreover, regions of similar texture with gradual intensity variation due to uneven illumination variation from the point light source onboard the CVE are given equal weights when estimating the final value of the target pixel (see Fig. (2)).

As discussed in the introduction section, color distortions are undesirable in CVE image enhancement. Therefore, smoothing is done on the luma channel only in the YCbCr color space. One of

(a) Y channel of input image   (b) smoothed version

(c) Smoothed vs. original image intensity variation
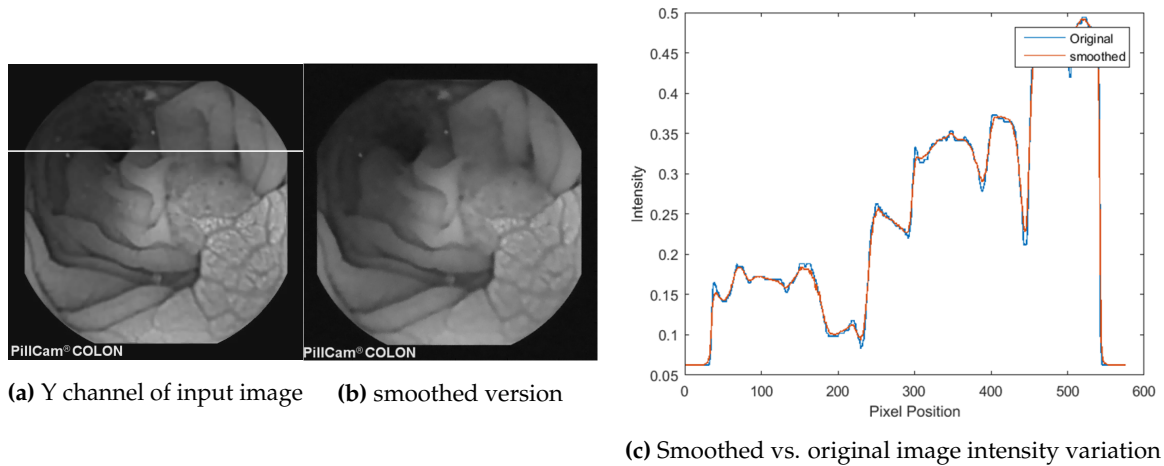
Figure 3. (a) Y channel of input image. (b) shows the smoothed image using random walk. From the intensity variation across, row 165 of the image, in (a) the smoothing preserves edges and smooths only detail variation on the surface. Hence captures the surface texture of the tissues.

the motivations for using a color gradient is the extra photometric information that is retained, which would be lost by utilizing a luminance-based gradient. In addition, for isotropic structures, where there is no preferred direction of gradient, as the directional derivative results in a zero magnitude. Hence, for a given RGB image I, the structural tensor is given by Eq. (4).

$$H = \begin{bmatrix} R_x^2 + G_x^2 + B_x^2 & R_x R_y + G_x G_y + B_x B_y \\ R_x R_y + G_x G_y + B_x B_y & R_y^2 + G_y^2 + B_y^2 \end{bmatrix} \tag{4}$$

Where $R_x, G_x, B_x, R_y, G_y, B_y$ are horizontal and vertical spatial derivatives of RGB color channels respectively. The eigenvalue analysis of the H leads to two eigenvalues $\lambda_+ \lambda_-$, which can be solved using standard techniques. Finally, given number of samples, and number of iterations, the final smoothed estimate of the target pixel, ,can computed as a weighted sum of pixels in the chain as in Eq. (5).

$$x_{den} = \frac{\sum\limits_{n=1}^{N} \sum\limits_{j=0}^{M} {}^n w_0{}^{j+1}(x_j|x_{j+1}) x_j}{\sum\limits_{n=1}^{N} \sum\limits_{j=0}^{M} {}^n w_0{}^{j+1}(x_j|x_{j+1})} \tag{5}$$

In computing the smoothed image, Fig. (3), we found that the contribution of pixels from the random sampling was very small as might be expected and so they were set to zero for numerical stability.

3.2. Local contrast enhancement

As discussed in the introduction section, local contrast enhancement is obtained by finding the local darkness and lightness. From M samples using random walk and random sampling in chain X, the minimum and maximum pixels of chain M are found as:

$$I^n_{\max} = \max_{i \in \{0,...,M\}} x^n_i$$

$$I^n_{\min} = \min_{i \in \{0,...,M\}} x^n_i \tag{6}$$

Where $I^n_{\max}$ and $I^n_{\min}$ are maximum and minimum intensity values of chain $n \in \{1,...,N\}$ respectively. It is noted that as the target pixel is also a sample point it is bounded by the maximum and minimum intensity of the samples. Similar to [16][16] the variation intensity of the samples, the range,R^n , and the relative value of the target pixel, V^n, is given as follows:

$$V^n = \begin{cases} 1/2 & \text{if} R^n = 0, \\ (x - I_{\min})/R^n & \text{else} \end{cases}$$

$$R^n = I^n_{\max} - I^n_{\min} \tag{7}$$

The final estimate of local contrast enhanced target pixel and range is obtained by averaging to the total number of iterations.

$$\bar{v} = \frac{1}{N} \sum_{n=1}^{N} V^n$$

$$\bar{r} = \frac{1}{N} \sum_{n=1}^{N} R^n \tag{8}$$

The extremal envelopes can simply be constructed from the estimated average and pixel value as follows:

$$E_{\min} = x_0 - \bar{v}\bar{r}$$

$$E_{\max} = x_0 + (1 - \bar{v})\bar{r} = E_{\min} + \bar{r} \tag{9}$$

3.3. Image decomposition

Exploiting the information provided by local contrast enhancement of the shadow details in the image and edge preserving smoothing, our framework is based on two key observations. Firstly, Shadow details are characterized by a large variation in the contrast-enhanced image and secondly, detail vessel texture is characterized by local intensity variation in the original image. Hence, the two layers containing vessel detail texture D_1 and shadow details D_2 are given as

$$D_1 = I_o - I_{rws}$$

$$D_2 = I_{CE} - I_{rws} \tag{10}$$

where I_o, I_{CE}, I_{rws} and are the original, local contrast enhanced (using Eq. (8)) and smoothed image (using random walk Eq. (5)) respectively. The final detail and shadow tissue texture enhanced image I_{enh} is obtained by convex linear combination of the two layers of detail and adding them back to the smoothed image as follows:

$$D = \gamma D_1 + (1 - \gamma)D_2$$

$$I_{enh} = KD + I_{rws} \tag{11}$$

where γ is a mixing coefficient that controls the amount of shadow details against tissue details and K is a scalar constant. Fig. (4) shows the enhanced image along its smooth , detail, and shadow layers.

4. Implementation

The proposed framework was tested with the following parameter settings. Denoising and local contrast enhancement can be controlled by choosing appropriate parameters as follows. The number

(a) (b)

Input and enhanced image

(c) Intensity variation of enhanced image compared to the details
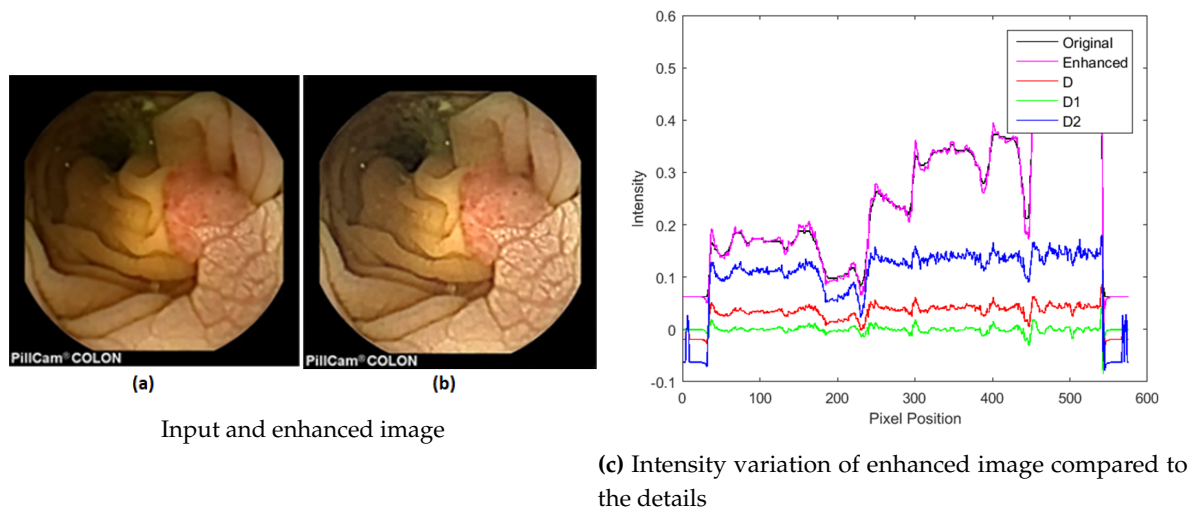
Figure 4. (a) Input image from PillCam COLON camera showing a Polyp of 9 mm size (b) Enhanced image using proposed method, in which color tone is kept whilst the details are enhanced. Compared to the original it can be seen that lumen and shadow details of the surface texture are more visible on the enhanced image. (c) Shows different layers of image decomposition D_1, D_2, and D of Eq. (10) as compared to the original and enhanced image Eq. (11). As it is shown the two detail layers captures low contrast shadows and surface details.

of samples and iterations controls the robustness of the estimated local contrast and smoothed pixel value. For our experiment, we used $N = 50$ iterations and $M = 150$ sample pixels per iteration. Edge aware smoothing was controlled though intensity and gradient normalization constants σ_I and σ_g respectively. σ_I controls intensity difference, while σ_g controls how total variation of gradient penalizes crossing edges. Moreover, it is possible to control the locality of smoothing and local contrast enhancement by choosing appropriate sampling dimensions for $R1$ and $R2$. For this experiment $R1$ is set to 10 and $R2$ is set to length of the diagonal of the image. The Pseudo-code for the proposed method is given in algorithm (1).

5. RESULT AND EVALUATION

5.1. Experimental setup and procedure

Subjective image quality assessment is the most reliable way to evaluate the visual quality of digital images perceived by trained medical doctor. To assess the performance of the proposed method for clinical application, rank order, an ordinal scaling method were used. The observers, who are trained physicians, were asked to rank the image samples in order of, from best to worst along diagnostic value of an image. The test images are reproductions of the same original image using proposed, Bilateral [25] and weighted least square (WLS) [26] image decomposition techniques. The experiments took place in a controlled room simulating a colonoscopy examination room. We used BENQ BL series display, with screen resolution of 3840 x 2160. The display is color managed for sRGB with luminance level of 80 cd/m^2 . The images were positioned side by side in a random order position as shown on Fig. (5). Moreover, to measure screen uniformity, a middle gray patch is used, and three points are sampled from left to right of the display. Our benchmark shows a 3.8 standard deviation in CIE XYZ values. To measure color uniformity different patches of red, green and blue patches were measured along black and white patches with average CIE2000 value of 1.57, which is on the order of the just noticeable difference (JND).

Data:

- $Ycbcr$: Input image
- $R1, R2$: Inner and outer sampling circles, Fig. (1)
- N, M : Number of samples and total number of iterations

Result: Enhanced image
PROCEDURE:
G = color gradient Eq. (4)
foreach *pixel* **do**
    Sampling:
    foreach *sample n=1:N* **do**
        foreach *chain < M* **do**
            while (random walk <R1)
                do random walk
            end
            random sampling
        end
        save to chain m
        Enhancing and smoothing:
        foreach *chain m=1:M* **do**

            - x = Y pixel intensity values of chain m
            - G = Pixel color gradients Eq. (4) of chain m
            - W = weight using Eq. (1). For chain pixels less than R1
            - I_{min} = Minimum of chain m Eq. (6).
            - I_{max} = Maximum of chain m Eq. (6).
            - Find range and relative value Eq. (7).

        end
    end

    - Smoothed pixel estimate, Eq. (5);
    - Contrast enhanced pixel estimate, Eq. (8).
    - Apply Eq. (10) & Eq. (11) to estimate the enhanced image

end

Algorithm 1: Stochastic CVE image enhancement

5.2. Dataset

Several guidelines have been given in the literature for the selection of images for psychophysical experiments. Holmet al. [27] recommend the use of a broad range of images as well as test charts to reveal the quality issues. 30 sample images were chosen by a medical doctor from the KID dataset [28] with pathologies and normal images from different parts of the colon. The sample images are selected based on lack of clarity and details for visual diagnosis and should be enhanced for better visualization. The images were taken by GivenImaging Pillcam COLON capsules with a resolution of 576 x 576. In addition, three sample images from Pillcam COLON II were also included having a resolution of 256 x 256.

5.3. Subjective evaluation

Five medical doctors who specialized in colonoscopy imaging participated in the subjective experiment. Observers were asked the standard question "Decide which image has the best diagnostic image quality. Once you make a decision click on the letters below each image indicating their rank. 'A' being the best image and 'D' being the worst image. And no tie is allowed". The user interface for the subjective evaluation is shown on Fig. (5). The rank order data were converted to into pairwise comparison for ease of analysis. The pairwise raw scores were used compute z-score ranking
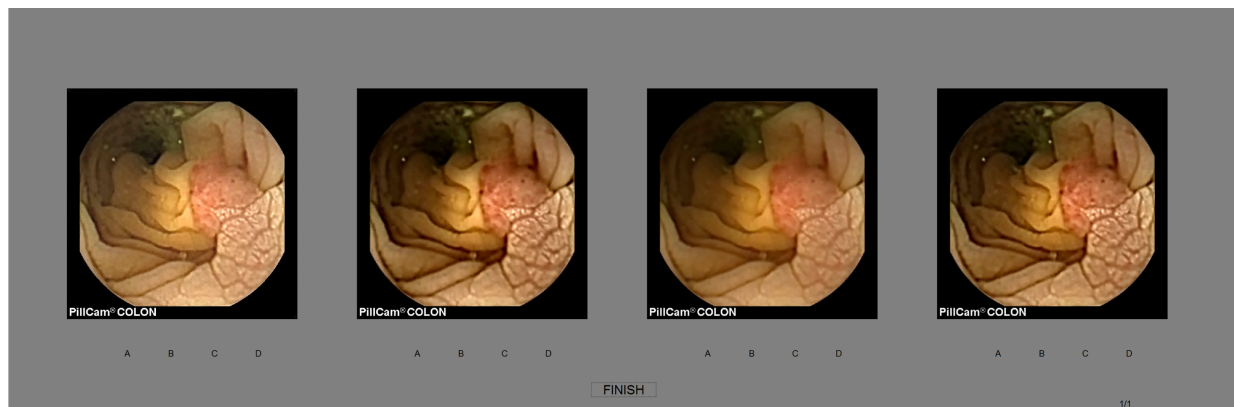
Figure 5. Psychometric experimental setup for subjective evaluation of diagnostic quality of an image. The observer gives letters from A (best) to D (worst) quality image for diagnosis.
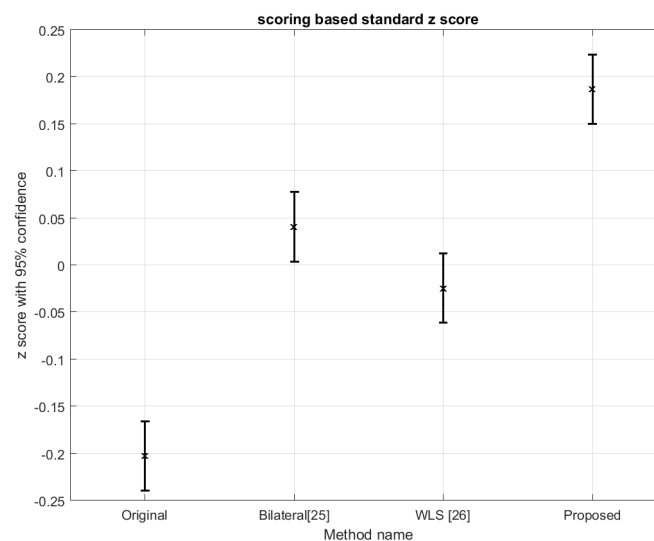
Figure 6. Rank order z-score showing observer preference for diagnostic image quality.

with Montag confidence interval [29]. The z-score ranking result from the subjective experiment is summarized on Fig. (6).

The subjective evaluation across different observers were also consistent. As it shown on Fig. (7), the proposed method performed better for the sample images in the dataset. In general, it can be deduced from Fig. (6) and Fig. (7) that image enhancement provides a better image quality as compared to original image for diagnosis.

5.4. Objective evaluation and comparison

Vision research does not offer an answer as the which objective image quality metrics corresponds to diagnostic qualities of capsule images. As there is no a single image quality metric, we rely on metrics that are correlated with subjective image quality for natural images. Among these metrics the proposed method is evaluated and compared objectively on the basis of weighted-level framework(WLF) [30] and SSIM[31]. Assuming the readers are familiar with SSIM, we limit the discussion to WLF. WLF is a no reference image contrast metric, which measure perceptual contrast by computing the global contrast. This metric is used as it has a good correlation in relation to observer perceived contrast. It is done by subsampling each channel separately into a pyramidal structure and obtaining a contrast map
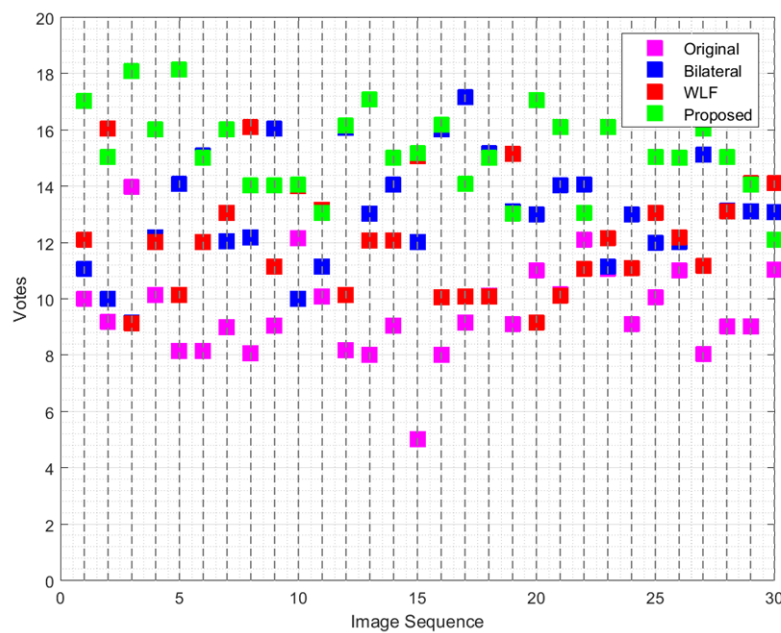
Figure 7. Average observers preference for the sample images selected for the subjective experiment.

of each level. The overall measure of each channel is a weighted recombination of the average contrast for each level. Structural similarity index (SSIM) is also used for objective evaluation of the contrast enhancement methods since SSIM has been shown to be a better predictor of subjective image quality than PSNR. In addition, SSIM is applied to see how the enhanced image differ from the reproduction images. the CIE2000 metric is applied to understand the perceived color difference between the enhancement and the original image. We compared the method against methods that fall into image decomposition techniques. Mainly image detail enhancement categories such as Bilateral[25] and weighted least square (WLS) [26]image decomposition techniques. Furthermore, we also compared against contrast limited adaptive histogram equalization (CLAHE) [32], which represent detail and contrast enhancement techniques. The average performance is summarized on Table I. The WLF value displayed on Table I, is a ratio of average WLF value of the method to the WLF value of the original image. As it can be seen from Table I, the proposed method has the highest WLF ratio compared to other image decomposition techniques and equivalent results on CIE2000 and SSIM values to other methods. However, CLAHE enhanced images lack details of tissue surfaces Fig. (9). It is interesting to note that based on a single objective quality metric it difficult to assess the diagnosis values of these images in relation to the subjective evaluation of these images by medical doctors. Many of other related works in capsule image enhancement use standard natural image quality metrics [5–8]. However, these metrics are proposed for natural images where relative smoothness is preferred. Therefore, further research is required to investigate a single image quality metrics for visual inspection of capsule endoscopy images. As the capsule images exist in smaller color space and are taken under low light conditions, quantifying the diagnostic value of these images is essential. More results can be downloaded from $http://www.colorlab.no/research_and_development/software$.

5.5. The effect of parameter selection

Sampling method controls the way the algorithm enhances the contrast and details texture features. Large number of samples and iteration gives smoother and better result. This is due to a high number of samples give a robust representative of the neighborhood pixel values. The inner radius $R1$, determines the size of the random walk. It defines the size of the neighborhood circular window for exploring similar textural features. If $R1 = 0$, the method performs only contrast enhancement

Table 1. Average performance of different methods using WLF, SSIM and CIE2000 metrics respectively.

AVERAGE PERFORMANCE			
Methods	WLF Ratio	SSIM	CIE2000
Bilateral [25]	1.06	0.915	3.03
WLS [26]	1.28	0.916	3.29
CLAHE [32]	2.65	0.822	3.6
Proposed	1.48	0.897	3.45

$N = 5$ $N = 25$ $N = 50$

$M = 10$

$M = 50$

$M = 100$
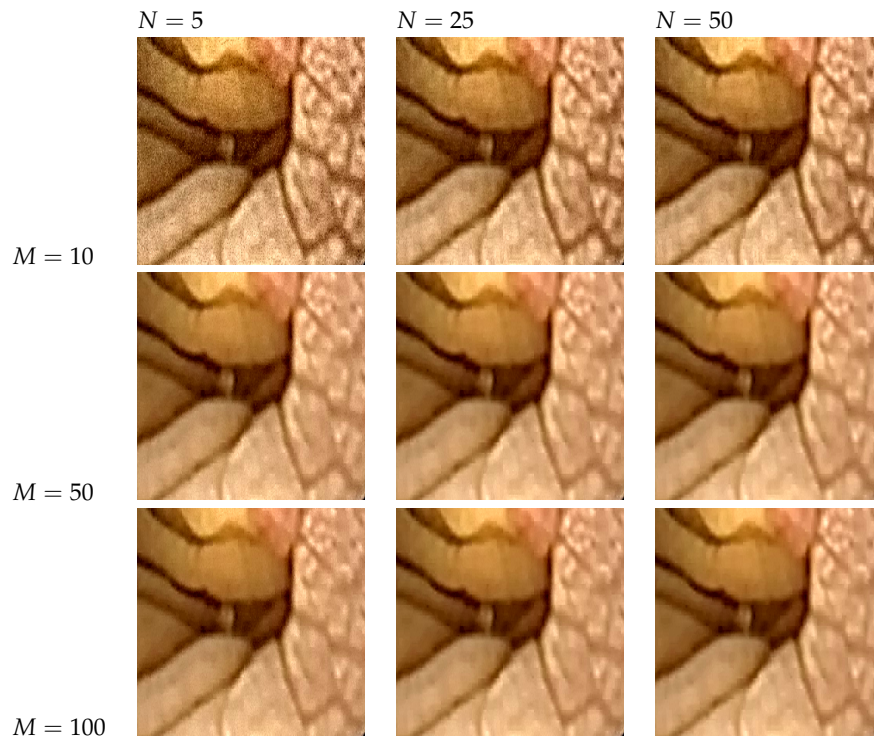
Figure 8. The effect of number of samples (N) and iteration (M). The input image is size of 174x294. R1 and R2 are set to 3 and 200 respectively with $\sigma_I = 9$ and σ_g.

similar to [16]. The other parameters that are closely related to $R1$ are the σ_I and σ_g. Taking smaller values of σ_I gives lower weights for pixels that have different intensity values from the target pixel. On the other hand, small values of σ_g weights on edges crossed during the random walk. For a pixel position that is similar to the target pixel but located on different texture or across edges are given less weights in computing the final estimate of the smoothed image. Large values of $R1$ with large values of σ_I and σ_g result in mean value of the image. These parameters do not require changing from image to image and were kept the same for all our experiment. These parameters were set to $\sigma_I = 9$ for intensity and $\sigma_I = 4$ for gradient normalization. The radius parameter is the maximum distance from the pixel where the stochastic sampling can be done. It is set in similar way to [16] and controls the locality of the spatial maxima and minima for the adjustment. It is not a critical parameter as long as it is large enough to sample reasonably across the entire image. To get a better estimate of the local lightness and blackness together with smooth estimate of the target pixel, the sampling process is iterated several times and averaged. This strongly decreases the noise level at the cost of increased time of computation. The number of iteration and samples affect the locality of the method. The effect of number of iteration and samples in presented in Fig (8).

(a) Input image   (b) Proposed   (c) WLS[26]   (d) L0 Gradient Minimization [33]

(e) Local Extrema [34]   (f) Bilateral[25]   (g) CLAHE[35]
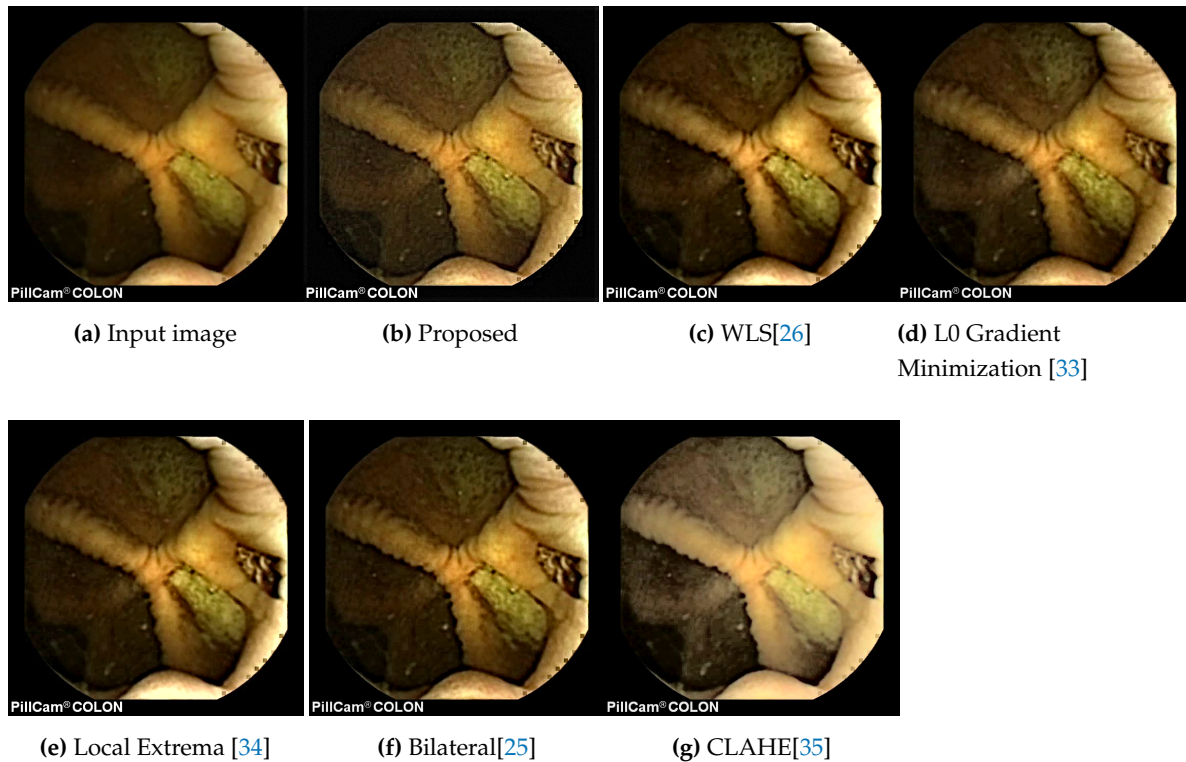
Figure 9. Comparison of different methods on PillCam COLON images. (a). Input image showing a splenic flexure (b) It is visually easy to see that our framework enhances the local contrast and the details of the tissue surface simultaneously. On both images, the proposed method gives a consistent result under different illumination variation. Bilateral [25] (f) image decomposition technique enhances the details, but creates a halo effect on edges which appear to widen the blood vessels and other tissue surfaces. Moreover, (g) enhances the local contrast but the details of the tissue surface are lost.

5.6. Smoothing

In order to elaborate more on estimation of the base layer i.e. smoothing, we tested the proposed method 3.1 against bilateral filtering for a Gaussian noise as it is a noise type for a capsule during transmission of data signal through the human body . For example, one may want to keep some level of detail in denoising capsule images. These details could be capillaries on tissue surface or vascular patterns. Detail textures are critical in defining the progress of inflammatory bowel diseases such as Ulcerative colitis in CVE. The proposed formulation for smoothing, E.q. (5), preserves texture and fine details along prominent edges that are inherent in CVE images. The level of smoothing can be controlled by choosing appropriate values of $R1$ and σ_I and σ_g which controls region for the random walk, and penalty for intensity and total gradient variation as the random walk progress from the target pixel to neighboring pixels as defined Section (3).

As it can be seen on Fig. (10), the proposed method results in smooth images yet has no halo effect on edges. Moreover, textural details are preserved. As it is noted on Fig. (9a) and in [35] other methods such as bilateral filter and weighted least square (WLS) filter methods blur edges. This is visible from Fig. (11) where the intensity variation across the sharp edge varies gradually for bilateral filtering hence blurring the edges. Bilateral filter setting was fine-tuned for a best result with parameters and radius of 5. For the proposed method, blurring is minimal. However, the proposed method produces small intensity variation in denoising which are not present in the original image. This can be explained as a trade-off between keeping the texture and smoothing the noise.

(a) Input image   **(b)** Noisy image

(c) Bilateral Filtering   **(d)** Proposed denoising   **(e)** Zoomed comparison on edges
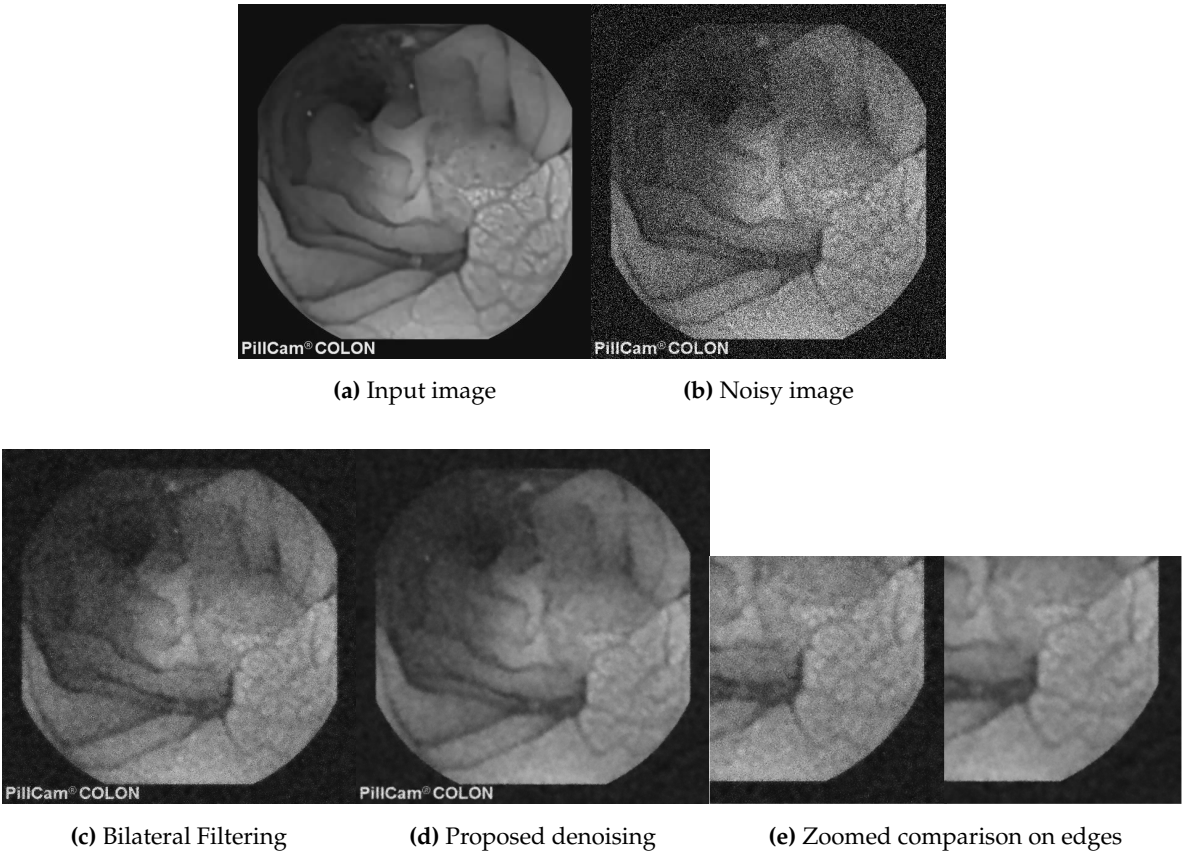
Figure 10. Comparison of smoothing methods: The proposed method results in a smooth image with smooth edges. However, it is easier to see that bilateral filtering creates a halo effect around edges.

DENOISING			
Noise STD.DEV	BILATERAL FILTERING	ANISOTROPIC DIFFUSION	PROPOSED
$\sigma = 0.03$	36.08	27.59	35.17
$\sigma = 0.09$	33.40	27.29	32.45
$\sigma = 0.27$	19.41	32.45	23.85

(a) Intensity variation across row 165 of noisy image and smoothed image $\sigma = 0.09$
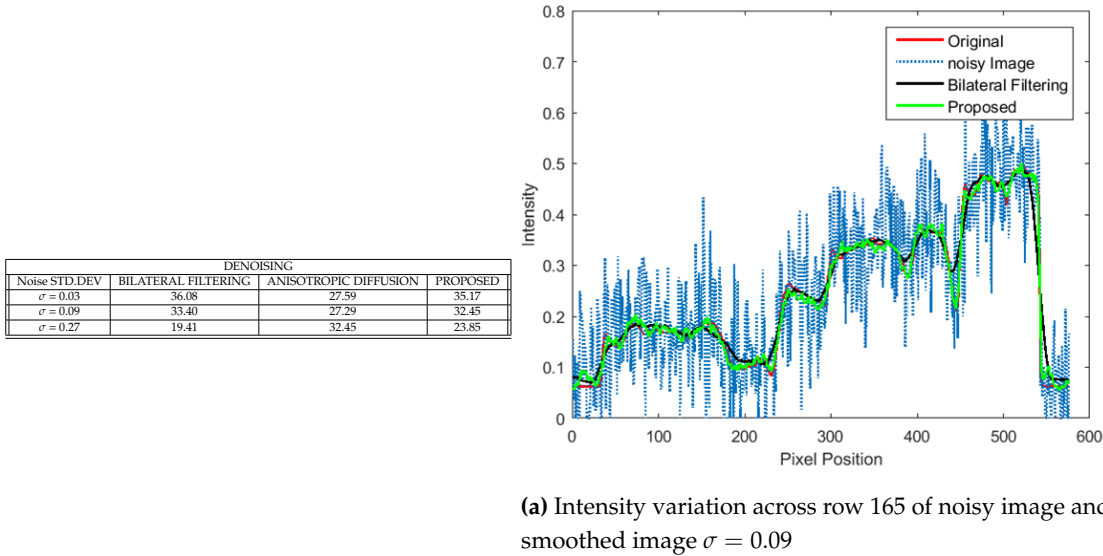
Figure 11. From the intensity variation curve, we can see that the proposed method preserves edges and smooth's the smaller variations in amplitude

In addition, we tested the proposed method for smoothing on 30 sample images from KID dataset against bilateral filtering [34] and anisotropic diffusion [20]. A Gaussian noise with standard

deviation where added to Y channel of the input image. We used PSNR to measure the perceived quality improvement. The parameters of all methods were set initially and were averaged for all test images. For bilateral filtering we used for spatial and range standard deviation and window size of 5. Anisotropic diffusion was done for 15 iterations. For the proposed method $N = 50$ and $M = 150$ with $\sigma_I = 9, \sigma_g = 4$. The results are summarized in Fig. (11)

6. CONCLUSION AND FUTURE WORK

In this work, we have proposed a framework that could enhance detail and shadow texture of tissues for CVE images. Based on stochastic sampling and edge-aware smoothing the proposed method delivers a state-of-the-art result for clinical applications. Computational complexity is a limitation of the method. Smaller random walk circle substantially reduces the time required as a trade-off between the number of samples and the level of the details that will be enhanced. To this end, random walk problem has been posited as solving a Dirichlet problem for image segmentation [36]. For our future work, similar approach could be used to reformulate the proposed method for faster implementation. Moreover, we are planning to investigate the effect of enhancement for automated detection of polyps and ulcerative colitis.

Acknowledgments: This research has been supported by the Research Council of Norway through project no. 247689 "IQ-MED: Image Quality enhancement in MEDical diagnosis, monitoring and treatment".

Author Contributions: The work presented in this paper was carried out in collaboration between all authors. AM conceived and carried out the main part of this manuscript. IF contributed in analyzing and implementation, SY and MP are a supervisor of this research. ØH evaluated the result for clinical application. All authors read and approved the final manuscript.

Conflicts of Interest: "The authors declare no conflict of interest".

Abbreviations

The following abbreviations are used in this manuscript:

CVE	Capsule video endoscopy
CCE	Colon capsule endoscopy
CLAHE	Contrast limited adaptive histogram equalization
CMOS	Complementary metal-oxide semiconductor)
DFT	Discrete Fourier transform
FICE	Flexible spectral imaging color enhancement
HVS	Human Visual System
JND	Just noticeable difference
NBI	Narrow Band Imaging
SSIM	Structural similarity index
WLF	Weighted-level framework
WLS	weighted least square

References

1. Rimbaş, M.; Negreanu, L.; Ciobanu, L.; Benguş, A.; Spada, C.; Băicuş, C.R.; Costamagna, G. Is virtual chromoendoscopy useful in the evaluation of subtle ulcerative small-bowel lesions detected by video capsule endoscopy? *Endoscopy international open* **2015**, *3*, E615.
2. Sakai, E.; Endo, H.; Kato, S.; Matsuura, T.; Tomeno, W.; Taniguchi, L.; Uchiyama, T.; Hata, Y.; Yamada, E.; Ohkubo, H.; others. Capsule endoscopy with flexible spectral imaging color enhancement reduces the bile pigment effect and improves the detectability of small bowel lesions. *BMC gastroenterology* **2012**, *12*, 83.
3. Lambert, R.; Kuznetsov, K.; Rey, J. Narrow-band imaging in digestive endoscopy. *The Scientific World Journal* **2007**, *7*, 449–465.

4. Ramaraj, M.; Raghavan, S.; Khan, W.A. Homomorphic filtering techniques for WCE image enhancement. Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on. IEEE, 2013, pp. 1–5.

5. Ibrahim, H.; Kong, N.S.P. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics* **2007**, *53*.

6. Okuhata, H.; Nakamura, H.; Hara, S.; Tsutsui, H.; Onoye, T. Application of the real-time Retinex image enhancement for endoscopic images. Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE. IEEE, 2013, pp. 3407–3410.

7. Gopi, V.P.; Palanisamy, P.; Niwas, S.I. Capsule endoscopic colour image denoising using complex wavelet transform. In *Wireless Networks and Computational Intelligence*; Springer, 2012; pp. 220–229.

8. Liu, H.; Lu, W.S.; Meng, M.Q.H. De-blurring wireless capsule endoscopy images by total variation minimization. Communications, Computers and Signal Processing (PacRim), 2011 IEEE Pacific Rim Conference on. IEEE, 2011, pp. 102–106.

9. Beck, A.; Teboulle, M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing* **2009**, *18*, 2419–2434.

10. Vu, H.; Echigo, T.; Yagi, K.; Okazaki, H.; Fujiwara, Y.; Yagi, Y.; Arakawa, T. Image-Enhanced capsule endoscopy preserving the original color tones. International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging. Springer, 2011, pp. 35–43.

11. Imtiaz, M.S.; Wahid, K.A. Color enhancement in endoscopic images using adaptive sigmoid function and space variant color reproduction. *Computational and mathematical methods in medicine* **2015**, *2015*.

12. Li, B.; Meng, M.Q.H. Wireless capsule endoscopy images enhancement via adaptive contrast diffusion. *Journal of Visual Communication and Image Representation* **2012**, *23*, 222–228.

13. Attar, A.; Xie, X.; Zhang, C.; Wang, Z.; Yue, S. Wireless Micro-Ball endoscopic image enhancement using histogram information. Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. IEEE, 2014, pp. 3337–3340.

14. Srivastava, R.; Gupta, J.; Parthasarthy, H.; Srivastava, S. PDE based unsharp masking, crispening and high boost filtering of digital images. International Conference on Contemporary Computing. Springer, 2009, pp. 8–13.

15. Sheet, D.; Garud, H.; Suveer, A.; Mahadevappa, M.; Chatterjee, J. Brightness preserving dynamic fuzzy histogram equalization. *IEEE Transactions on Consumer Electronics* **2010**, *56*.

16. Kolås, Ø.; Farup, I.; Rizzi, A. Spatio-temporal Retinex-inspired envelope with stochastic sampling: a framework for spatial color algorithms. *Journal of Imaging Science and Technology* **2011**, *55*, 40503–1.

17. Black, M.J.; Sapiro, G.; Marimont, D.H.; Heeger, D. Robust anisotropic diffusion. *IEEE Transactions on image processing* **1998**, *7*, 421–432.

18. Estrada, F.J.; Fleet, D.J.; Jepson, A.D. Stochastic Image Denoising. BMVC, 2009, pp. 1–11.

19. Azzabou, N.; Paragios, N.; Guichard, F. Random walks, constrained multiple hypothesis testing and image enhancement. European Conference on Computer Vision. Springer, 2006, pp. 379–390.

20. Perona, P.; Malik, J. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence* **1990**, *12*, 629–639.

21. Azzabou, N.; Paragios, N.; Guichard, F. Random walks, constrained multiple hypothesis testing and image enhancement. European Conference on Computer Vision. Springer, 2006, pp. 379–390.

22. Easley, G.R.; Labate, D.; Colonna, F. Shearlet-based total variation diffusion for denoising. *IEEE Transactions on Image processing* **2009**, *18*, 260–268.

23. Shih, A.C.; Liao, H.Y.; Lu, C.S. A new iterated two-band diffusion equation: theory and its application. *IEEE Transactions on Image Processing* **2003**, *12*, 466–476.

24. Rajpoot, K.; Rajpoot, N.; Noble, J.A. Discrete wavelet diffusion for image denoising. International Conference on Image and Signal Processing. Springer, 2008, pp. 20–28.

25. Fattal, R.; Agrawala, M.; Rusinkiewicz, S. Multiscale shape and detail enhancement from multi-light image collections. *ACM Trans. Graph.* **2007**, *26*, 51.

26. Farbman, Z.; Fattal, R.; Lischinski, D.; Szeliski, R. Edge-preserving decompositions for multi-scale tone and detail manipulation. ACM Transactions on Graphics (TOG). ACM, 2008, Vol. 27, p. 67.

27. Holm, J.; Tastl, I.; Johnson, T. Definition & use of the ISO 12640-3 reference color gamut. Color and Imaging Conference. Society for Imaging Science and Technology, 2006, Vol. 2006, pp. 62–68.

28. Koulaouzidis, A.; Iakovidis, D. KID: Koulaouzidis-Iakovidis Database for Capsule Endoscopy, 2015.

29. Montag, E.D. Empirical formula for creating error bars for the method of paired comparison. *Journal of Electronic Imaging* **2006**, *15*, 010502.

30. Simone, G.; Pedersen, M.; Hardeberg, J.Y. Measuring perceptual contrast in digital images. *Journal of Visual Communication and Image Representation* **2012**, *23*, 491–506.

31. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **2004**, *13*, 600–612.

32. Zuiderveld, K. Contrast limited adaptive histogram equalization. *Graphics gems* **1994**, pp. 474–485.

33. Xu, L.; Lu, C.; Xu, Y.; Jia, J. Image smoothing via L 0 gradient minimization. ACM Transactions on Graphics (TOG). ACM, 2011, Vol. 30, p. 174.

34. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. Computer Vision, 1998. Sixth International Conference on. IEEE, 1998, pp. 839–846.

35. Subr, K.; Soler, C.; Durand, F. Edge-preserving multiscale image decomposition based on local extrema. ACM Transactions on Graphics (TOG). ACM, 2009, Vol. 28, p. 147.

36. Grady, L. Random walks for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* **2006**, *28*, 1768–1783.