

RESEARCH

Open Access



# Variational approach for capsule video frame interpolation

Ahmed Mohammed<sup>1\*</sup> , Ivar Farup<sup>1</sup>, Sule Yildirim<sup>2</sup>, Marius Pedersen<sup>1</sup> and Øistein Hovde<sup>3</sup>

## Abstract

Capsule video endoscopy, which uses a wireless camera to visualize the digestive tract, is emerging as an alternative to traditional colonoscopy. Colonoscopy is considered as the gold standard for visualizing the colon and takes 30 frames per second. Capsule images, on the other hand, are taken with low frame rate (average five frames per second), which makes it difficult to find pathology and results in eye fatigue for viewing. In this paper, we propose a variational algorithm to smooth the video temporally and create a visually pleasant video. The main objective of the paper is to increase the frame rate to be closer to that of the colonoscopy. We propose variational energy that takes into consideration both motion estimation and intermediate frame intensity interpolation using the surrounding frames. The proposed formulation incorporates both pixel intensity and texture feature in the optical flow objective function such that the interpolation at the intermediate frame is directly modeled. The main feature of this formulation is that error in motion estimation is incorporated in our model, so that only robust motion estimation are used in estimating the intensity of the intermediate frame. We derived Euler-Lagrange equations and showed an efficient numerical scheme that can be implemented on graphics hardware. Finally, a motion compensated frame rate doubling version of our method is implemented. We evaluate the quality of both 90 and 100% of the frames for medical diagnosis domain through objective image quality metrics. Our method improves state-of-the-art result for 90% frames while performing equivalent for the remaining cases with other existing methods. In the last section, we show application of frame interpolation to informative frame segment visualization and to reduce the power consumption.

**Keywords:** Total variation, Capsule endoscopy, Temporal super-resolution, Optical flow, The law of texture energy

## 1 Introduction

Capsule video endoscopy has proven to be a powerful tool for diagnosis of the digestive tract diseases. It has many advantages over traditional colonoscopy, as being less invasive and requires no sedation. There are different types of capsules that are currently available on the market. These include esophageal, small bowel, and colon capsules. Colon capsule video endoscopy (CCE) has been used to diagnose inflammatory bowel disease [1] (i.e., Chron's disease and ulcerative colitis), gastrointestinal bleeding, and polyps. Since it is less invasive than colonoscopy, it might also increase participation in colorectal cancer screening. It has been shown to have high sensitivity for the detection of clinically relevant lesions [2].

The second generation CCE developed by [3] is about  $11 \times 31$  mm and takes 14 frames/min until the first frame of the small bowel then captures frames at adaptive frame rate of 4–35 frames per second depending on the speed of the capsule. Although, adaptive frame rate improves the visualization, the video appears jagged; sample videos can be obtained from [4]. The images are at a lower resolution compared to traditional colonoscopy (usually full-HD). The images produced by capsule video endoscopy suffer from several problems, such as uneven and low illumination, low resolution, high compression ratio, and noise. The problem of capsule image enhancement has been an active research topic over the past decade [5, 6], but there are few publications that consider the low temporal frame rate aspect of capsule videos. Frame interpolation is a technique of creating intermediate frames based on overlapping neighboring frames in sequence. The CCE video reader softwares such as RapidReader [4] support viewing from 2–40 frames per second with an option to

\*Correspondence: [mohammed.kedir@ntnu.no](mailto:mohammed.kedir@ntnu.no)

<sup>1</sup>Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway  
Full list of author information is available at the end of the article

pause, rewind, and play the video. Watching the video at two frames per second is the most robust way to find pathologies, but the videos are not smooth for watching and hence takes more time to view. It is important to note that frame interpolation does not increase the duration of CCE videos rather makes the video more smooth and natural to view. CCE frame interpolation should in general consider the following three conditions: Firstly, the interpolated frame should not contain motion and image artifacts that could lead to wrong diagnosis. Secondly, flickering and blurring of frames should be avoided when displaying image sequences. Thirdly, the interpolated frames need to compensate apparent motion of camera and give a natural motion portrayal. Karargyris et al. [7] proposed three-dimensional reconstruction of the digestive wall in capsule endoscopy videos using elastic video interpolation. In their work, they propose a methodology that creates the intermediate (interpolated) frames between two CCE frames followed by three-dimensional reconstruction, given that these two frames carry mutual information of some degree. The interpolation is done by computing optical flow from the region-based matching technique. The segmentation of the video frames is performed by the fuzzy region growing segmentation followed by matching of the segments in consecutive frames based on color, texture, and geometry information. Similar work was presented in [8] where the authors presented frame interpolation as a post-processing technique at the receiver for saving battery power at the transmitter. This is done by transmitting frames at a lower frame rate, given it can be reconstructed by using neighboring frames at the receiver, hence saving power at the transmitter. In their work, they used unidirectional and bidirectional block-matching motion estimation and compensation method to create the intermediate frames.

In this paper, we further explore this direction of estimating intermediate frames and propose a different parametrization of variational energy for CCE frame interpolation. We show experimentally that the proposed variational energy formulation improves the quality of interpolated CCE frames. The contributions of this work are three folds. Firstly, we combine motion estimation and compensation in a single energy formulation and such formulation improves the quality of interpolated frame with less computation since we do not compute forward and backward optical flow. The proposed energy formulation includes symmetric motion estimation and compensation which can be solved by primal dual approach [9]. Secondly, by exploiting the law of texture energy, we introduce a symmetric textural and intensity constraint for computing robust interpolation of CCE frames. Unlike previous approaches that compute one directional optical flow [7], our formulation considers symmetric optical flow under the symmetric textural and intensity constraint,

which gives a better interpolation of CCE frames in terms of objective quality metrics. Thirdly, we evaluate and analyze appropriateness of CCE frame interpolation for medical diagnosis using objective image quality metrics.

The outline of the article is as follows: in Section 2, we re-visit earlier works in variational formulation for frame interpolation. In Section 3, we present our approach and detail derivation both theoretically and numerically. In Section 4, we present implementation of the proposed method. In Section 5, we evaluate the proposed method, along with comparison to other works. Finally, in Section 6, we present discussion and conclusion.

## 2 Background

Before going into the details of our method, we present a short review of variational frame interpolation approaches as used in natural image sequences. There are two general steps for frame interpolation. These are motion estimation and motion compensation, respectively. Motion estimation involves computing the temporal movement between the current frame and the previous frame, in the form of motion vectors. The motion vectors can be block-based (sparse) or pixel-based(dense). Keller et al. [10] proposed a variational method for both optical flow calculation (motion estimation) and the actual new frame interpolation (motion compensation). The flow and intensities are calculated simultaneously in a multiresolution setting. Using standard maximum a posteriori to variational formulation rationale, the authors derived a minimum energy formulation for the estimation of a reconstructed sequence as well as motion recovery. Similarly, Rakèt et al. [11] proposed motion compensated frame interpolation with a symmetric optical flow constraint. Motion vectors are computed using TV-L1 energy, and interpolated frame is computed by averaging the warped flow to current and previous frame. Once motion vectors are computed, motion compensation is used to estimate the intermediate frame as follows:

$$I\left(\mathbf{x} + \frac{1}{2}\mathbf{u}, n\right) = \frac{1}{2}(I(\mathbf{x}, n-1) + I(\mathbf{x} + \mathbf{u}, n+1)) \quad (1)$$

Motion estimation involves computation of displacement vectors between two neighboring frames. Let  $I(\mathbf{x}, n)$  be a video sequence and  $\mathbf{u} = (u, v)$  be a displacement vector of pixel position  $\mathbf{x} = (x, y)$  of frame number  $n$ . Assuming the intensity of the pixel did not change due to displacement, we can write the optical flow constraint as

$$I(\mathbf{x}, n) = I(\mathbf{x} + \mathbf{u}, n+1), \quad I : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R} \quad (2)$$

Taking the Taylor series expansion of the right-hand side of Eq. (2), we get the known optical flow constraint as

$$\nabla I \cdot \mathbf{u} + I_t = 0 \quad (3)$$

where  $\nabla I$  and  $I_t$  are spatial and temporal derivative operators, respectively. Horn and Schunck [12] proposed a variational approach to solving Eq. (3) using L2-norm under smoothness assumption of flow field as shown in Eq. (4). This quadratic cost function can easily be solved using Euler-Lagrange equations. The main disadvantage of this formulation is that it penalizes high gradients of  $\mathbf{u}$  and disallows discontinuities.

$$E_{HS}(\mathbf{u}) = \int_{\Omega} \{ \lambda (|\nabla \mathbf{u}|)^2 + (\nabla I \cdot \mathbf{u} + I_t)^2 \} dA \quad (4)$$

where  $dA = dx dy$  and  $\lambda$  is a constant.

To circumvent this problem, L1-norm was exploited in [9]. L1-norm is a better choice for optical flow computation as it is robust for outliers and allows discontinuities in the flow field. Zach et al. [9] proposed an algorithm that can be understood as a minimization of Eq. (5), which is the sum of the total variation of flow field  $\mathbf{u} = (u, v)$  and an L1 attachment term:

$$E_Z(\mathbf{u}) = \int_{\Omega} |\nabla \mathbf{u}| + \lambda |I(\mathbf{x}, n) - I(\mathbf{x} + \mathbf{u}, n + 1)| dA \quad (5)$$

Once motion vectors are estimated, interpolated frame is computed as in Eq. (1). Our proposed approach is different from [10] in that forward and backward optical flow is not computed on every multiresolution scale. Rather, symmetric flow is employed, hence avoiding redundant flow computation. Unlike [11], we model motion vector estimation and intermediate frame interpolation as a single energy formulation. In the next section, we discuss the proposed method.

### 3 Method

A high resolution video sequence can be considered as continuous space-time image volume. Frame interpolation problem, on the other hand, can be modeled as an inpainting problem along temporal axes in this volume. 2D inpainting may be viewed as denoising with a binary mask  $\beta(x)$  which is set to zero in the missing region of the image and non-zero otherwise. Following the same reasoning, let us define original video sequence and desired high frame rate sequence as  $I_0$  and  $I$  respectively. From mathematical point of view,  $I_0$  and  $I$  are piece-wise continuous functions of  $\mathbb{R}^3 \rightarrow \mathbb{R}$  defined in space of bounded variation. Assuming the intensity of a pixel is constant in the direction of optical flow, a general formulation to estimate symmetric optical flow (SOF) and interpolated frame can be written as

$$E(\mathbf{u}, I(\mathbf{x}, n)) = \int_{\Omega} |\nabla I(\mathbf{x}, n)| dA + \int_{\Omega} \lambda_2 |I(\mathbf{x} + \mathbf{u}, n + 1) - I(\mathbf{x} - \mathbf{u}, n - 1)| dA + \int_{\Omega} \lambda_3 |\mathbf{u}| dA \quad (6)$$

where  $I(\mathbf{x}, n)$  is the required interpolated frame,  $\nabla$  is the spatial derivative operator, and  $\mathbf{u} = (u, v)$  are the  $x$  and  $y$  components of the SOF field. In order to find the intermediate frame  $I(\mathbf{x}, n)$ , we need to minimize the energy given by Eq. (6). Taking the derivative of Eq. (6) with respect to  $\mathbf{u}$  yields SOF estimation:

$$\begin{aligned} \frac{\partial E_f(\mathbf{u})}{\partial \mathbf{u}} &= \frac{\partial E(\mathbf{u}, I(\mathbf{x}, n))}{\partial \mathbf{u}} \\ &= \frac{\partial}{\partial \mathbf{u}} \int_{\Omega} \lambda_2 |I(\mathbf{x} + \mathbf{u}, n + 1) - I(\mathbf{x} - \mathbf{u}, n - 1)| dA \\ &\quad + \frac{\partial}{\partial \mathbf{u}} \int_{\Omega} \lambda_3 |\mathbf{u}| dA \end{aligned} \quad (7)$$

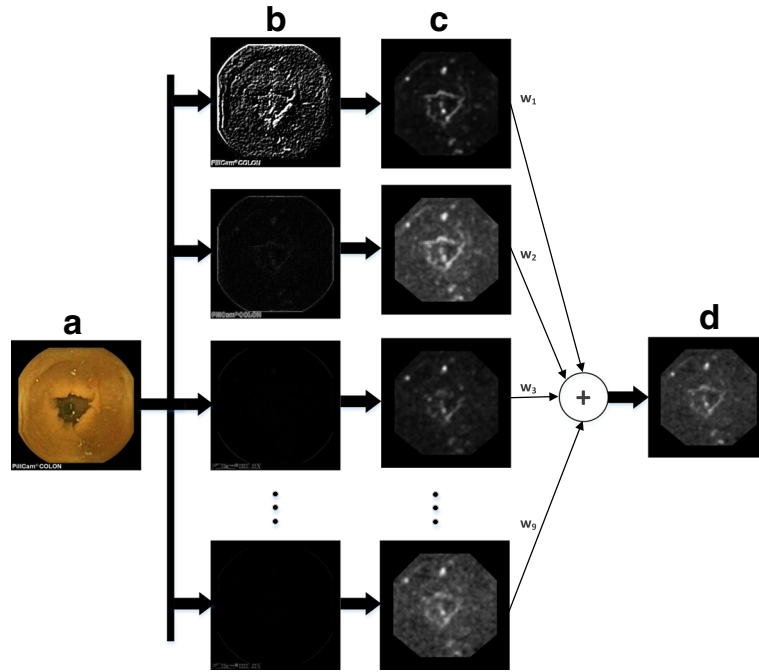
We can recognize the above expression as SOF constraint. In ideal case, where the flow field is accurate, the intermediate frame can be computed directly as  $I(\mathbf{x}, n) = I(\mathbf{x} + \mathbf{u}, n + 1) = I(\mathbf{x} - \mathbf{u}, n - 1)$ . However, for practical capsule videos, with photometric variations (i.e., shadow, shading, specular reflection, and light source changes) as well as geometrical variations (i.e., viewpoint and object orientation), these conditions do not hold. In order to improve the smoothness and accuracy of the optical flow, we propose to improve optical flow estimation by including information from textural features. In this work, we explored nine filters of size  $5 \times 5$ , which are constructed from four basic vectors  $L5 = [1, 4, 6, 4, 1]$ ;  $E5 = [-1, -2, 0, 2, 1]$ ;  $S5 = [-1, 0, 2, 0, -1]$ ; and  $R5 = [1, -4, 6, -4, 1]$ , as suggested by the law of texture energy [13]. By multiplying these four vectors mutually, textural features such as center-weighted local average ( $L5$ ), edges ( $E5$ ), spots ( $S5$ ), ripples, and waves in texture ( $R5$ ) are included in estimating robust SOF. Once the nine textural maps are computed, the final texture is estimated by weighted summation of the texture using Eq. (8). Texture map estimation scheme is shown in Fig. 1.

$$I_{th} = \sum_{n=1}^{n=9} I_{th}^i \cdot w_i \quad (8)$$

where  $w_i$  is

$$w_i = \frac{\nabla I_{th}^i}{1 + \sum_{n=1}^{n=9} \nabla I_{th}^i} \quad (9)$$

In addition, local binary kernel  $[1, 1, 1; 1, -1, 1; 1, 1, 1]$  is explored for robust optical flow computation. Incorporating the textural constraint to Eq. (7) results in our final energy for optical flow computation:



**Fig. 1** Image texture computation scheme. Given input image (a), nine filter responses (b), are computed by applying nine masks as explained in Section 3, then followed by computing energy map at each pixel by summing absolute value of filter output across neighborhood around pixel. The image responses are masked to remove margins containing text and normalized between [0, 1], (c). Finally, the final texture map (d) is estimated using Eq. (26)

$$\begin{aligned} \frac{\partial E_f(\mathbf{u})}{\partial \mathbf{u}} &= \frac{\partial}{\partial \mathbf{u}} \int_{\Omega} \lambda_2 |I(\mathbf{x} + \mathbf{u}, n + 1) - I(\mathbf{x} - \mathbf{u}, n - 1)| dA \\ &+ \frac{\partial}{\partial \mathbf{u}} \int_{\Omega} \lambda_t |I_{th}(\mathbf{x} + \mathbf{u}, n + 1) - I_{th}(\mathbf{x} - \mathbf{u}, n - 1)| dA \\ &+ \frac{\partial}{\partial \mathbf{u}} \int_{\Omega} \lambda_3 |\mathbf{u}| dA \end{aligned} \quad (10)$$

where  $I_{th}$  is a texture feature as in Eq. (8) and  $\lambda_t$  is the weighting factor. This formulation has an advantage in that it incorporates robust textural features. Moreover, in the standard formulation of variational optical flow, the estimated motion vector field depends on the reference image and is asymmetric as can be seen from Eq. (5).

Similarly, the derivative of Eq. (6) with respect to  $I(\mathbf{x}, n)$  using the notation  $\hat{I} = I(\mathbf{x}, n)$  for simplicity gives

$$\begin{aligned} \frac{\partial E_i(\hat{I})}{\partial \hat{I}} &= \frac{\partial E(\mathbf{u}, \hat{I} = I(\mathbf{x}, n))}{\partial \hat{I}} \\ &= \frac{\partial}{\partial \hat{I}} \int_{\Omega} |\nabla I(\mathbf{x}, n)| dA \\ &+ \frac{\partial}{\partial \hat{I}} \int_{\Omega} \lambda_2 |I(\mathbf{x} + \mathbf{u}, n + 1) - I(\mathbf{x} - \mathbf{u}, n - 1)| dA \end{aligned} \quad (11)$$

which can be rewritten to include quality measure of the optical flow. This can be expressed by dividing the optical

flow constraint equation into regions of intensity interpolation, where warping of the flow to both neighboring frames  $I(\mathbf{x} - \mathbf{u}, n - 1)$  and  $I(\mathbf{x} + \mathbf{u}, n + 1)$  are equal or not.

$$\frac{\partial E_i(\hat{I})}{\partial \hat{I}} = \frac{\partial}{\partial \hat{I}} \int_{\Omega} |\nabla \hat{I}| dA + \quad (12)$$

$$\frac{\partial}{\partial \hat{I}} \int_{\Omega} \overline{\beta(\mathbf{x})} \lambda_2 |I(\mathbf{x} + \mathbf{u}, n + 1) - I(\mathbf{x} - \mathbf{u}, n - 1)| dA + \quad (13)$$

$$\overline{\beta(\mathbf{x})} \frac{\lambda_2}{2} \|I(\mathbf{x} + \mathbf{u}, n + 1) - I(\mathbf{x} - \mathbf{u}, n - 1)\|_2 \quad (14)$$

where  $\beta(\mathbf{x}) = 1$  if  $|I(\mathbf{x} - \mathbf{u}, n - 1) - I(\mathbf{x} + \mathbf{u}, n + 1)| < \epsilon$ , and  $\epsilon$  is a small positive constant; otherwise,  $\beta(\mathbf{x})$  is set to a zero similar to 2D inpainting.  $\overline{\beta(\mathbf{x})}$  represents a negation operator. In the above formulation, Eq. (12) represents spatial diffusion on the interpolated image and Eq. (13) on the other hand is diffusion along the flow line. The extra fidelity term Eq. (14) is added to avoid blurring of the interpolated frames in regions defined by  $\beta(\mathbf{x}) = 1$ , i.e., correct intensity estimation. Here, we make an assumption that the intensity of the pixel at intermediate frame as estimated from both neighboring frame is robust, if the optical flow estimation at that pixel is accurate; otherwise, we fill the missing region using flow direction through

inpainting along the flow lines. Following the above formulation, the problem is thus to find a motion vector  $\mathbf{u}$  and intermediate frame  $I(\mathbf{x}, n)$  that minimizes Eq. (10) and Eq. (12), respectively. In order to minimize Eq. (10), we take first-order Taylor series expansion on data fidelity term of Eq. (10) and it becomes:

$$E_f(\mathbf{u}) = \int_{\Omega} \lambda_3 |\mathbf{u}| + \lambda_2 |\rho(\mathbf{u}, q)| dA \quad (15)$$

where

$$\begin{aligned} \rho(\mathbf{u}, q) = & I(\mathbf{x} + \mathbf{u}_0, n + 1) - I(\mathbf{x} - \mathbf{u}_0, n - 1) \\ & + \lambda_t (I_{th}(\mathbf{x} + \mathbf{u}_0, n + 1) - I_{th}(\mathbf{x} - \mathbf{u}_0, n - 1)) \\ & + \nabla I(\mathbf{x} + \mathbf{u}_0, n + 1)(\mathbf{u} - \mathbf{u}_0) \\ & + \nabla I(\mathbf{x} - \mathbf{u}_0, n - 1)(\mathbf{u} - \mathbf{u}_0) \\ & + \lambda_t (\nabla I_{th}(\mathbf{x} + \mathbf{u}_0, n + 1)(\mathbf{u} - \mathbf{u}_0) \\ & + \nabla I_{th}(\mathbf{x} - \mathbf{u}_0, n - 1)(\mathbf{u} - \mathbf{u}_0)) + \sigma q(\mathbf{x}) \end{aligned} \quad (16)$$

$\sigma q(\mathbf{x})$  is added to the fidelity term of our energy function to account for small intensity variation in the image. Introducing auxiliary variable  $\mathbf{w}$  to Eq. (15) and applying convex relaxation similar to [9], Eq. (15) can be decoupled into two energies as:

$$E_{1f}(\mathbf{u}) = \int_{\Omega} \lambda_3 |\mathbf{u}| + \frac{1}{2\theta} (\mathbf{u} - \mathbf{w})^2 \quad (17)$$

$$E_{2f}(\mathbf{w}) = \int_{\Omega} \frac{1}{2\theta} (\mathbf{u} - \mathbf{w})^2 + \lambda_2 |\rho(\mathbf{w}, q)| \quad (18)$$

This convex relation was first proposed by [9], coupling the two energies by quadratic link function. Setting  $\theta$  low forces the minima to occur when  $\mathbf{u} = \mathbf{w}$ . In Eq. (17), minimization problem is identical to denoising problem except that the integral is taken over a motion vector  $\mathbf{u}$  and can be solved by using Chambolles projection algorithm, and Eq. (18) can be solved simply by pointwise thresholding method.

To minimize the intensity inpainting energy given by Eq. (12), we derive the Euler-Lagrange equations. This can be written as:

$$\frac{\partial E_i}{\partial I}(I(\mathbf{x}, n)) = 0 \quad (19)$$

Minimizing the energy based on this L1 norm requires that the function to be convex and differentiable. Hence, we write absolute value function in Eq. (12) as  $\varphi(\mathbf{I}^2) = \sqrt{(\mathbf{I})^2 + \epsilon_1} \equiv |\mathbf{I}|$ , where  $\epsilon_1$  is a small positive constant regularizer.  $\varphi(\mathbf{I}^2)$  is a convex and differentiable function which meets the mentioned requirement in the process of searching minimum. Therefore, the Euler-Lagrange equation becomes

$$\begin{aligned} \frac{\partial E_i}{\partial I}(I(\mathbf{x}, n)) = & \nabla \cdot (A \nabla I(\mathbf{x}, n)) - \beta(\mathbf{x}) \lambda_2 \nabla \cdot \left( B \frac{\partial I}{\partial V} V \right) \\ & + \overline{\beta(\mathbf{x})} \lambda_2 (I(\mathbf{x} + \mathbf{u}, n + 1) - I(\mathbf{x} - \mathbf{u}, n - 1)) \end{aligned} \quad (20)$$

where

$$A = \varphi'(|\nabla I(\mathbf{x}, n)|^2), B = \varphi'(|I(\mathbf{x} + \mathbf{u}, n + 1) - I(\mathbf{x} - \mathbf{u}, n - 1)|^2), \text{ and } V = (\mathbf{u}, 1).$$

In the above equation, the first term represents a diffusion term with diffusion velocity of  $A$ . The second part of the equation represents diffusion of intensity in the direction of the computed optical flow. This term act as a transportation of intensity along the flow line. In our case, we are only interested with diffusion along the flow lines as defined by mask  $\beta(\mathbf{x})$ . It is possible to apply diffusion on the whole image along the flow lines, but we found that a good initial solution can easily be estimated by setting  $I(\mathbf{x}, n) = \frac{1}{2}(I(\mathbf{x} - u, n - 1) + I(\mathbf{x} + u, n + 1))$ . The last part of the equation avoids smoothing the images where intensities are correctly computed from optical flow.

#### 4 Numerical implementation

To minimize the energies defined in Eqs. (12)–(14) and (19), we first derived the Euler-Lagrange equations. The formulations can easily be parallelized with multi-core processors [14]. The Euler-Lagrange equations for Eqs. (16)–(18) are shown in Eq. (20). The 2D divergence operator  $\nabla \cdot$  for  $N$  by  $M$  image is defined as

$$\begin{aligned} \nabla \cdot (I_x, I_y) = & \begin{cases} I_x(i, j) - I_x(i, j - 1), & \text{if } j < N. \\ 0, & \text{otherwise} \end{cases} \\ & + \begin{cases} I_y(i, j) - I_y(i - 1, j), & \text{if } i < M. \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (21)$$

Similarly, we defined derivatives using five-point stencil finite difference approximation with convolution mask  $[1, -8, 0, 8, -1]/12$ . The implementation of the second term of Eq. (20) is similar to discretization used in [15]. Deriving the Euler-Lagrange equation for Eq. (17) and setting it to zero becomes

$$\nabla \cdot \left( \frac{\nabla \mathbf{u}}{|\nabla \mathbf{u}|} \right) + \frac{1}{\theta} (\mathbf{u} - \mathbf{w}) = 0 \quad (22)$$

Let us define dual variable as

$$\mathbf{p} = \frac{\nabla \mathbf{u}}{|\nabla \mathbf{u}|} \implies \mathbf{p} |\nabla \mathbf{u}| - \nabla \mathbf{u} = 0 \quad (23)$$

Substituting Eq. (22) in Eq. (23) we get

$$\mathbf{p} \left| \nabla \left( \nabla \cdot (\mathbf{p}) - \frac{\mathbf{w}}{\theta} \right) \right| - \nabla \left( \nabla \cdot (\mathbf{p}) - \frac{\mathbf{w}}{\theta} \right) = 0 \quad (24)$$

The fixed-point iteration scheme for Eq. (24) will be

$$\mathbf{p}^{k+1} = \frac{\mathbf{p}^k + \tau \nabla \left( \nabla \cdot (\mathbf{p}^k) - \frac{\mathbf{w}^k}{\theta} \right)}{1 + \tau \left| \nabla \left( \nabla \cdot (\mathbf{p}^k) - \frac{\mathbf{w}^k}{\theta} \right) \right|} \quad (25)$$

and  $\mathbf{u}^{k+1}$  is computed from Eq. (22) as  $\mathbf{u}^{k+1} = \mathbf{w}^{k+1} + \theta \nabla \cdot (\mathbf{p}^k)$ . Finally, Eq. (18) can be solved using pointwise thresholding as in [11]. For completeness, we present the final result here. For general formulation, where  $\rho(\mathbf{u}, q) = \mathbf{g}^T \mathbf{u} + \mathbf{c}$ , the dual variable  $\mathbf{w}$  is given by  $\mathbf{w} = \mathbf{u} + TH \left( \mathbf{u} + \mathbf{g} \frac{\mathbf{c}}{|\mathbf{g}|^2} \right)$ , where TH is a thresholding operator defined as

$$TH \left( \mathbf{u} + \mathbf{g} \frac{\mathbf{c}}{|\mathbf{g}|^2} \right) = \begin{cases} -\theta \mathbf{g}, & \text{if } \mathbf{g}^T \mathbf{u} + \mathbf{c} < -\theta |\mathbf{g}|^2 \\ \theta \mathbf{g}, & \text{if } \mathbf{g}^T \mathbf{u} + \mathbf{c} > \theta |\mathbf{g}|^2 \\ \mathbf{g} \frac{\mathbf{g}^T \mathbf{u} + \mathbf{c}}{|\mathbf{g}|^2}, & \text{if } |\mathbf{g}^T \mathbf{u} + \mathbf{c}| \leq |\mathbf{g}|^2 \end{cases} \quad (26)$$

Finally, the step by step multi-scale implementation scheme is given in the resulting Algorithm 1.

```

Data:  $I(\mathbf{x}, n - 1), I(\mathbf{x}, n + 1)$ 
Result:  $I(\mathbf{x}, n)$ 
warping=2, Maxiter=20, scale=0.9, NIter=200;
foreach level= $l$  do
    /* For each pyramid level */
    Down sample both image with  $scale$ 
    foreach  $warping$  do
        Warp images to  $I(\mathbf{x}, n)$ 
        foreach  $Maxiter=k$  do
            Compute  $\mathbf{u}^{k+1}$  from Eq. (25)
            Compute  $\mathbf{w}^{k+1}$  from Eq. (26)
        end
        Upscale  $\mathbf{u}^{k+1}$  and  $\mathbf{w}^{k+1}$ .
    end
end
Initialize  $I(\mathbf{x}, n) = \frac{1}{2}(I(\mathbf{x} - u, n - 1) + I(\mathbf{x} + u, n + 1))$ 
foreach  $NIter$  do
    /* Solving Eq. (19) */
     $I(\mathbf{x}, n)_{NIter+1} = I(\mathbf{x}, n)_{NIter} + \frac{\partial E_i}{\partial I}(I(\mathbf{x}, n)_{NIter})$ 
end
    
```

**Algorithm 1:** Variational approach for capsule video frame Interpolation

## 5 Results and discussion

In CCE frames interpolation, not only the smoothness of the output video must be taken into account but also the quality of the interpolated frame for diagnosis. It must be kept in mind that the capsule moves through gastro-intestinal track with an uneven speed by muscle

peristalsis. Therefore, depending on the speed of the capsule, some of the neighboring frames might contain high overlap or very small overlap between frames. Hence, the quality of the interpolated image depends on the degree of overlap and needs to be evaluated for medical decision-making process. Therefore, the reconstruction quality in terms of objective and subjective measures is of great importance.

### 5.1 Dataset

In our experiment, we have doubled frame rates from 5 to 10 frame/second. The videos are taken with GivenImaging Pillcam Colon camera. Four sequences were extracted from GivenImaging capsule videos [4].

- Seq1: Contains 13 frames from colon with perspective passage motion of tissues. Average correlation similarity between neighboring frames is 0.8910.
- Seq2: Contains 16 frames from colon showing a 9 mm polyp on a single frame with complicated motions. Average correlation similarity between neighboring frames is 0.8078.
- Seq3: Contains 20 frames from rectum with occlusions. Average correlation similarity between neighboring frames is 0.8989.
- Seq4: Contains 18 frames from colon showing 6 mm polyp on multiple frames. Average correlation similarity between neighboring frames is 0.8621. Sample results are shown in Fig. 2.

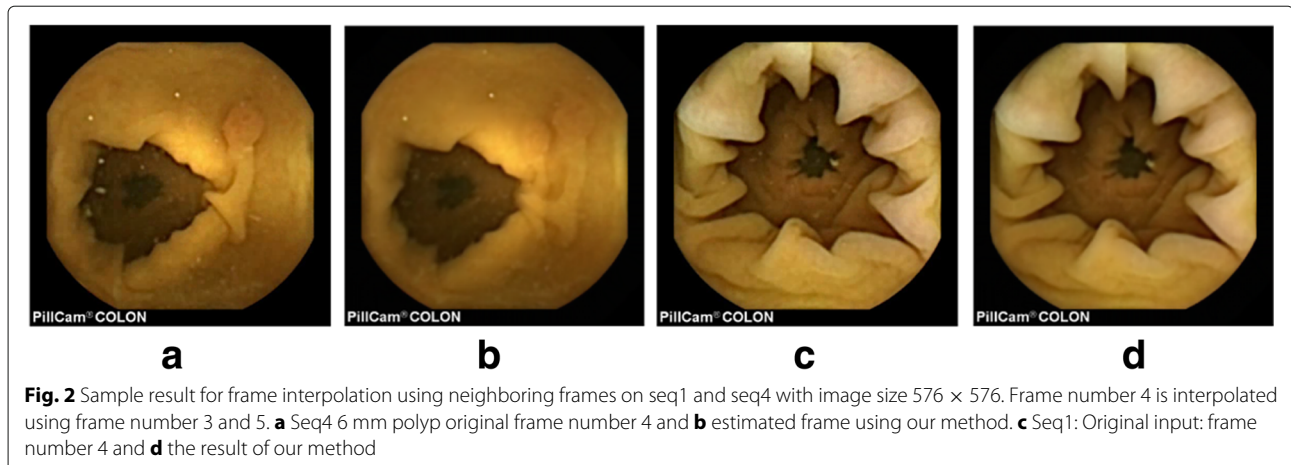
### 5.2 Objective metrics

We used the most common metrics for evaluation of interpolation error such as mean-squared error (MSE) and peak signal to noise ratio (PSNR). In addition, we have also compared using Structural SIMilarity (SSIM) [16] as a quality measure of one of the images being compared. For  $N$  by  $M$  size image, MSE and PSNR are defined as:

$$MSE = \frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M (I_{est} - I_{gr})^2 \quad (27)$$

$$PSNR = 10 \log_{10} \left( \frac{L^2}{MSE} \right) \quad (28)$$

where  $I_{est}$  is the interpolated frame,  $I_{gr}$  is the ground truth frame, and  $L$  is the peak signal strength. For SSIM, we used the implementation provided by [16]. In our experiment, as we do not have ground truth data, we interpolate odd-numbered frames using even-numbered frames and vice versa. Comparison with other methods is done for both 90 and 100% of the frames. The frames are grouped based on the similarity of the neighboring frame for interpolation. Results are summarized in Tables 1, 2 and 3 for both 90 and 100% of frame sequences. The comparison is done with state-of-the-art and traditional optical



**Fig. 2** Sample result for frame interpolation using neighboring frames on seq1 and seq4 with image size 576 × 576. Frame number 4 is interpolated using frame number 3 and 5. **a** Seq4 6 mm polyp original frame number 4 and **b** estimated frame using our method. **c** Seq1: Original input: frame number 4 and **d** the result of our method

flow variational technique TV-CLG [17]<sup>1</sup> and TV-L1[18]<sup>3</sup> respectively. Moreover, we also compared against non-variational optical flow method that is robust for large displacement optical flow computation, SFlow [19]<sup>2</sup>. In addition, frame averaging technique is included in the comparison as a baseline, as it is common in commercial products. The comparison is done using the implementation provided by the respective authors.

From the above result, we can observe that the proposed method improves the image quality by 0.3 dB compared to other state-of-the-art methods, although the difference in performance in terms of MSE between top method and our method is comparatively small ( $1.1 \times 10^{-4}$ ). Sample results of the proposed method are shown in Fig. 2. More results are shown in Fig. 3.

**5.3 Parameters**

In general, frame interpolation depends on accurate optical flow computation. Coarse-to-fine and warping techniques are frequently used tool for improving the performance of optic flow methods [9–11]. *Warping* in Algorithm 1 controls the number of times Eqs. (17) and (18) are solved iteratively, which is set by *Maxiter* on

each scale. Increasing this parameter increases the quality of computed optical flow as tradeoff with speed. *NIter* is the number of times diffusion along the optical flow lines is propagated. Interpolated frames get smoother with increase of this parameter. Finally, in order to get full advantage of the texture feature,  $\lambda_t$  in Eq. (10) represents how much textural energy map and pixel intensity contribute for estimating accurate interpolated frame. We did parameter optimization on  $\lambda_t$  against PSNR value of the interpolated and ground truth image. The result of optimization is shown in Fig. 7. A visual comparison showing results with and without texture features is shown in Fig. 6. From Figs. 7 and 6, we can see that the textural features improve the quality of interpolated frame with less motion artifacts and tissue surface blur.

**5.4 Applicability of interpolation for CCE video frames**

Colonoscopy, which is a gold standard for visualizing the colon takes 30 frames per second. The videos are in general smooth and natural to view. Currently, CCE is not recommended as a first-line colorectal cancer screening option in hospitals. Frame interpolation can be used to enhance CCE for better visualization by increasing the

**Table 1** Average PSNR for 90 and 100% of test sequences

Method	Seq1		Seq2		Seq3		Seq4		Avg.	
Frames ratio	90%	100%	90%	100%	90%	100%	90%	100%	90%	100%
TV-CLG <sup>1</sup>	26.78	24.12	<i>18.56</i>	18.33	25.26	<i>24.21</i>	25.20	23.86	23.85	22.70
SFlow <sup>2</sup>	26.12	23.71	<i>18.56</i>	<i>18.37</i>	24.71	23.76	25.10	23.94	23.54	22.52
Frame averaging	13.62	13.31	11.78	11.88	12.68	12.71	12.24	12.26	12.50	12.50
TV-L1 <sup>3</sup>	25.21	22.65	18.14	17.95	24.24	23.16	23.92	22.72	22.82	21.71
Proposed method without texture	26.71	24.49	18.26	18.03	25.03	23.86	25.69	24.27	23.82	22.70
Proposed method with texture	<i>27.26</i>	<i>24.94</i>	<i>18.45</i>	<i>18.21</i>	<i>25.34</i>	<i>24.16</i>	<i>25.29</i>	<i>23.99</i>	<i>23.95</i>	<i>22.85</i>

The PSNR value in italics indicate the best performing method

**Table 2** Average MSE ( $10^{-3}$ ) for 90 and 100% of test sequences

Method	Seq1		Seq2		Seq3		Seq4		Avg.	
Frames ratio	90%	100%	90%	100%	90%	100%	90%	100%	90%	100%
TV-CLG <sup>1</sup>	3.34	8.26	18.59	19.23	3.71	6.53	6.79	8.65	9.28	10.44
SFlow <sup>2</sup>	3.41	7.84	<i>18</i>	<i>18.63</i>	3.86	<i>6.40</i>	6.02	<i>7.48</i>	7.82	<i>9.87</i>
Frame averaging	43.63	47.28	68.12	66.47	55.43	54.91	59.85	59.67	57.78	57.52
TV-L1 <sup>3</sup>	4.05	11.38	20.20	20.49	4.50	8.21	7.73	9.74	9.13	12.13
Proposed method without texture	3.97	6.82	19.82	20.55	4.06	8.03	6.47	8.46	8.55	10.89
Proposed method with texture	<i>2.54</i>	<i>5.69</i>	18.92	19.56	<i>3.51</i>	6.99	5.99	7.91	7.78	9.98

The MSE value in italics indicate the best performing method

frame rate and improving capsule battery life. In literature, there are recent works that aim to detect informative segments automatically [20, 21]. Increasing the frame rate of these segments will assist the gastroenterologist to go through the video quickly. Moreover, frame interpolation can be used as post-processing for saving battery life. In order to reduce the power consumption of an endoscope capsule transferring still images over a wireless channel from inside human intestines to on-body receivers, the transmitted frame rate can be reduced in favor of generating the frames at the receiver side [8]. However, CCE estimation of intermediate CCE frames, with rapid changing scene and large displacement between frames, can cause problems even to the human observer. In such scenarios, it is difficult (sometimes impossible) to estimate the intermediate frame as there is no information available for reconstruction. Hence, it is important to have a frame reconstruction without undermining the diagnostic value of the video, for example Figs. 4 and 5. When a gastroenterologist examines CCE videos, he/she can play videos and pause on a given frame for examining. This begs a question, how to predict if the interpolated frame is reliable for diagnosis.

In order to examine the appropriateness of an interpolation, we analyzed different parameters of neighboring frames that could impact the quality of interpolated frame. Figure 8 shows PSNR value plot against maximum and minimum magnitude of the optical flow. From the figure, we can see that for optical flow maximum magnitude less than 25, the PSNR is stable. Similar observation can be

made from Fig. 8, as the correlation between the neighboring frame increases the PSNR value shows improvement. This is an interesting observation, in that we can use it to switch-off frame interpolation when flows are above some threshold or when the correlation between two frames are below a given value. The result shown in Fig. 8 is expected in that as the correlation between neighboring frames is an indicator to the quality of the interpolated frame. In order to compute robust threshold of correlation between neighboring frames, we collected CCE videos from nine people and extracted seven segments (50 frames each) from each video, which are marked by gastroenterologist as suspected region for different types of pathology. Table 4 shows correlation between neighboring frames as ratio of number of frames for a given correlation value to the total number of frames. By using the data from Fig. 8 and Table 4, one can estimate robust threshold for interpolation that works for significant portion of the CCE videos.

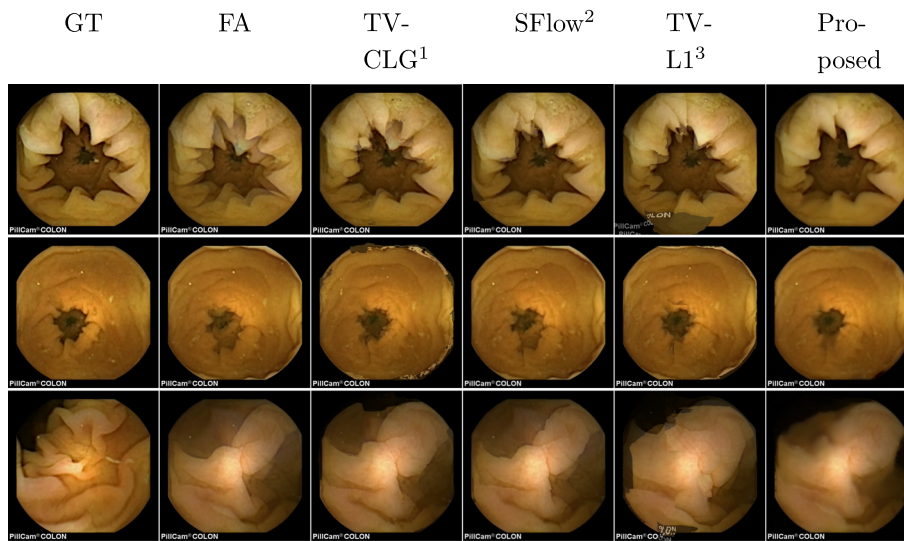
It is also important to note the performance of each method with respect to correlation of neighboring frames. Figure 5 shows the plot of correlation between neighboring frames and PSNR value between ground truth and interpolated frame. The data is curve fitted using exponential family of the form  $a - b \times e^{-cx}$ . It is easy to see that frame averaging performance above other methods where correlation between neighboring frames is less than 0.4, although the interpolated frame is blurred and has motion artifacts, as the apparent motion is not compensated. This is expected in that in case of large displacement between

**Table 3** Average SSIM for 90 and 100% of test sequences

Method	Seq1		Seq2		Seq3		Seq4		Avg.	
Frames ratio	90%	100%	90%	100%	90%	100%	90%	100%	90%	100%
TV-CLG <sup>1</sup>	0.997	0.994	0.987	0.986	<i>0.997</i>	<i>0.995</i>	0.995	<i>0.994</i>	<i>0.994</i>	0.993
SFlow <sup>2</sup>	0.997	0.995	<i>0.988</i>	<i>0.987</i>	0.997	0.995	0.996	0.994	0.994	0.993
Frame averaging	0.965	0.964	0.949	0.950	0.957	0.957	0.956	0.956	0.956	0.956
TV-L1 <sup>3</sup>	0.997	0.993	0.986	0.986	<i>0.997</i>	0.994	0.995	0.993	<i>0.994</i>	0.992
Proposed method without texture	0.997	0.995	0.986	0.985	<i>0.997</i>	0.994	0.995	<i>0.994</i>	<i>0.994</i>	0.992
Proposed method with texture	<i>0.998</i>	<i>0.996</i>	0.987	0.986	<i>0.997</i>	<i>0.995</i>	<i>0.996</i>	<i>0.994</i>	<i>0.994</i>	<i>0.993</i>

The SSIM value in italics indicate the best performing method



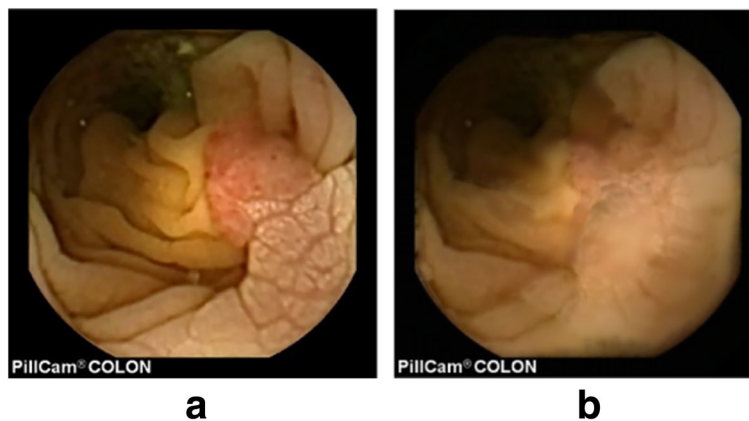


**Fig. 3** Visual comparison of interpolated images. The first and second rows show the performance of different methods. On the last row, we can observe that the interpolated frame is significantly different from the ground truth. This is due to large displacement between the neighboring frame, with correlation value between them being 0.691. In our approach, interpolation is done when we are confident on the interpolated frame. More details are given under Section 5.4

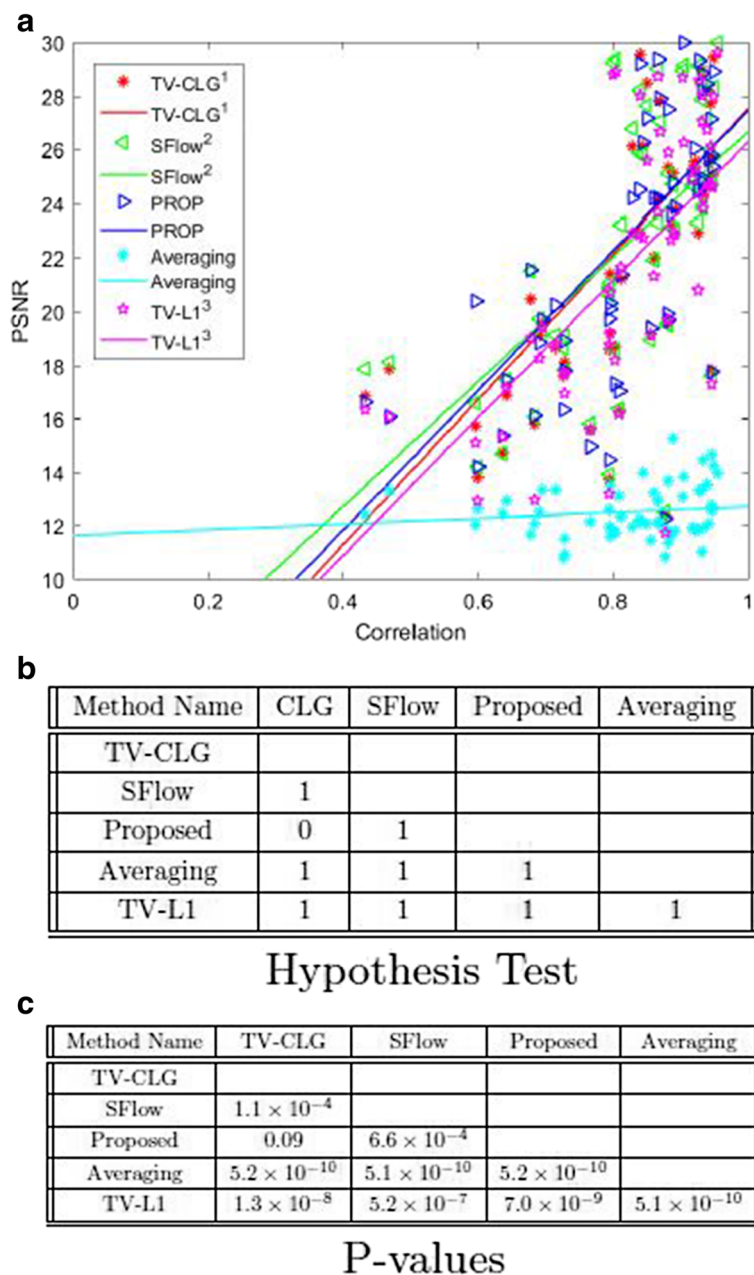
neighboring frames, it is difficult to estimate accurately the optical flow. On the other hand, using methods that are robust for large displacement optical flow as [19] gives a better result compared to variational techniques for large optical flow displacement. However, variational methods, specifically our proposed method performs better than other methods including [19] for approximately 90% of CCE video frames which has correlation greater than 0.75 as shown in Table 4.

Moreover, we performed a non-parametric paired Wilcoxon signed-rank test [22] comparing the PSNR value for each method. The null hypothesis (i.e. data in two paired methods are samples from continuous

distributions with equal medians  $H = 0$ , against the alternative that they are not  $H = 1$ ) is tested with Bonferroni correction of confidence interval [23]. Figure 5, b and c shows Wilcoxon signed rank test for all methods for correlation value between neighboring frames greater than 0.75. As it can be seen, the proposed method performs statistically better except for TV-CLG [17]. Although, the proposed method performs better against TV-CLG [17] in terms of mean PSNR, it is not statistically significant. For our experiment, we set a correlation value of 0.75 between neighboring frames to decide if the computed intermediate frame is suitable for diagnosis. As it is shown in Table 4, this threshold includes 90% of the frames in



**Fig. 4** Large displacement between neighboring frames with maximum magnitude of flow 92 pixels: frame interpolation for frame 7 of Seq2. **a** The ground truth frame. **b** Estimated based on frame 6 and 8. The polyp shown in **a** appears only in a single frame (frame number 8)



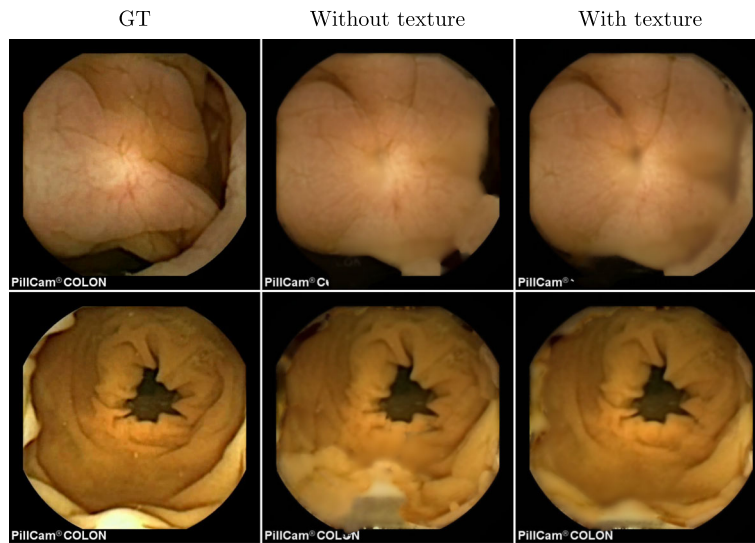
**Fig. 5** Comparison of five methods. **a** The performance of all methods plotted on correlation vs. PSNR axis. As it can be seen from the graph SFlow and frame averaging performs well for large displacement optical flow compared to other methods. The performance of the proposed method tops other methods for correlation > 0.7 which includes majority of the cases. **b** and **c** shows Wilcoxon signed rank test for 90% of the sequences in-terms of hypothesis test and *p*-value

typical CCE video. For frames below the threshold, frame interpolation is off, and frame doubling is done to make the frame rate consistent.

**5.5 Future direction: CCE video frame interpolation**

Compared with wired colonoscopy, the limited working time, the low frame rate, and the low image resolution limit the wider application of CCE. An increase in the

frame rate, angle of view, depth of field, and duration of the procedure and improvements in illumination seem likely in the future. The progress of battery technology and robust computational frame interpolation techniques can mitigate problems with the current CCE capsules. CCE needs to be small enough to be swallowable, and the battery needs to last more than 8 h [24]. The transmission of the image data occupies about 90% of the total



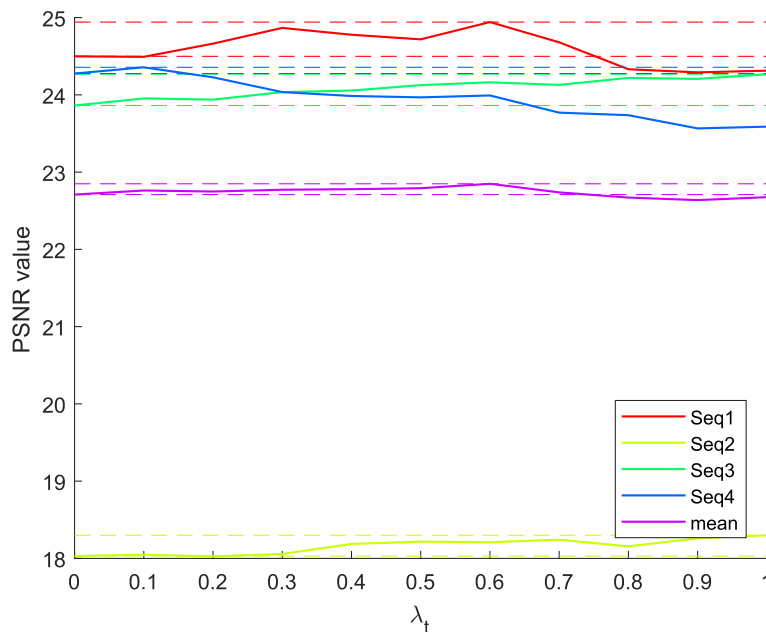
**Fig. 6** Visual comparison of proposed method with and without textural feature constraint for sample frames from seq3 and seq4. First column shows the ground truth frame 8 and 12 from seq3 and seq4. The second column and third column show interpolation result using frames 7 and 9 and frames 11 and 13 with texture and without textural feature constraints

power in CCE [25]. Hence, computational techniques can have a significant impact in improving the frame rate of future capsules. As shown on Tables 1, 2, 3 and 4, the performance the proposed method improves with correlation between the neighboring frames. With high frame rate videos, the proposed method gives more robust interpolated frames. This could significantly increases the

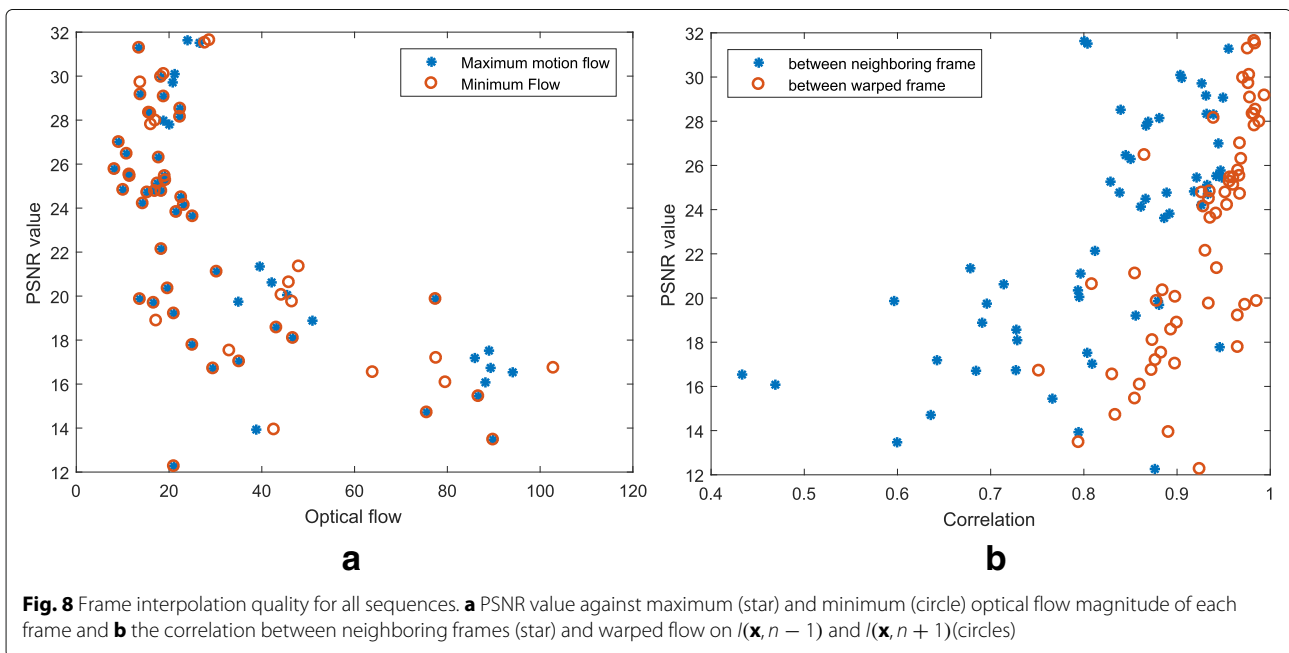
chance of finding more disease pathologies as the CCE passes through the gastrointestinal tract.

### 6 Conclusion

In this paper, we discussed the limitation of the current CCE videos regarding low frame rates. It is desirable to have a smooth video which is pleasant to view as well as



**Fig. 7** PSNR vs  $\lambda_t$ . The above plot shows PSNR value as a function of  $\lambda_t$ . The lower line shows the performance of the proposed method without texture, and the upper line shows the maximum improvement with texture features for a given sequence. We can see that the texture component improves the interpolated frame quality depending on the complexity of the scene



give a better diagnostic value by reducing eye fatigue. We proposed a variational approach to CCE frame motion estimation and intermediate frame intensity computation, simultaneously. In addition, textural features are included to make robust motion estimation. We also evaluated the quality of both 90 and 100% of the frames for medical diagnosis domain through objective image quality metrics. We found that the proposed method gives a state-of-the-art result for CCE frame interpolation. Moreover, the proposed method can be parallelized, and computationally efficient methods exist for GPU implementation. As a future work, we will explore extending variational methods to make them more robust for large displacement between neighboring frames (i.e., low correlation). In addition, objective metrics used here need to be supplemented with subjective evaluation by medical professional. Further video materials can be downloaded from <https://www.ntnu.edu/web/colourlab/software>.

**Table 4** Percentage of frames with neighboring frame correlation above given value

Correlation	Percentage (%) of frames
$\geq 0.90$	55.69
$\geq 0.85$	72.21
$\geq 0.80$	84.26
$\geq 0.75$	90.16
$\geq 0.70$	93.92
$\geq 0.60$	97.51
$\geq 0.50$	99.13

**Abbreviations**

CCE: Colon capsule video endoscopy; GPU: Graphics processing unit; GT: Ground truth; HD: High definition; MSE: Mean-squared error; PSNR: Peak signal to noise ratio; SOF: Symmetric optical flow; SSIM: Structural Similarity; TV-L1: Total variation L1 norm

**Funding**

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."

**Availability of data and materials**

The dataset supporting the conclusions of this article is available in the [4] repository.

**Authors' contributions**

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

**Authors' information**

Mohammed Ahmed is PhD student at NTNU in Gjøvik, in the area of Medical Imaging for fast automatic and accurate anomaly detection and diagnosis using capsule video endoscopy. He received Master's degree in Electronics and Information Engineering from Chonbuk National University, South Korea in 2014.

Ivar Farup is a professor of computer science and study program leader for bachelor in engineering – computer science at NTNU Gjøvik. He received his MSc (siv.ing.) in technical physics from NTH, Norway, 1994 and PhD (dr. scient.) from the Department of Mathematics, University of Oslo, 2000. He is Professor of Computer Science since 2012. His work is centered on Colour science and Image processing

Sule Yıldırım is associate professor at the NISLAB, Department of Information Security and Communication Technology, NTNU Gjøvik. She was appointed as the head of computer science department, at HIHM and also worked there as associate professor before her current position. She has background in artificial intelligence and machine learning. Her work is centered on Secure Technologies and Semantic, agent based and learning systems for ontology modeling in Semantic Web and for the development of smart characters in video games.

Dr. Sule Yildirim Yayilgan is an associate professor at the Norwegian University of Science and Technology (NTNU) at the Department of Information Security and Communication Technology. Her main fields of research interests are artificial intelligence, application of machine learning in various fields, signal and image processing, and biometrics. She has participated in projects funded by EU Horizon 2020, Eurostars and Erasmus+ programs, the Research Council of Norway. She also actively takes part as PC in conferences and acts as reviewer in several journals.

Marius Pedersen received his BSc. in Computer Engineering in 2006 and MSc in Media Technology in 2007, both from Gjøvik University College, Norway. He completed a PhD program in color imaging in 2011 from the University of Oslo, Norway, sponsored by Océ. He is currently employed as professor at NTNU Gjøvik, Norway. He is also the director of the Norwegian Colour and Visual Computing Laboratory (Colourlab). His work is centered on subjective and objective image quality.

Østein Hovde, MD/PhD, is an associate professor at the Institute of Clinical medicine, University of Oslo. He is also a Senior consultant at Innlandet Hospital, Gjøvik. His main scientific work is in the fields of inflammatory bowel diseases and therapeutic endoscopy.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway. <sup>2</sup>Department of Information Security and Communication Technology, Norwegian University of Science and Technology, 2815 Gjøvik, Norway. <sup>3</sup>Department of Gastroenterology, Innlandet Hospital Trust, Gjøvik, Norway and Institute of Clinical Medicine, University of Oslo, Norway, Oslo, Norway.

Received: 5 December 2017 Accepted: 12 April 2018

Published online: 01 May 2018

#### References

1. C Parker, CE Spada, M McAlindon, C Davison, S Panter, Capsule endoscopy—not just for the small bowel: a review. *Expert Review of Gastroenterology & Hepatology*. **9**(1), 79–89 (2015). <https://doi.org/10.1586/17474124.2014.934357>
2. C Spada, GC Hassan, *J Endosc.* **44**(5), 527–536 (2012). <https://doi.org/10.1055/s-0031-1291717>
3. Medtronic, Pillcam Colon II (2009). <http://www.medtronic.com>. Accessed 15 July 2016
4. GivenImaging. Capsule Video Endoscopy: Atlas, (2016). <http://www.capsuleendoscopy.org>. Accessed 2016
5. MS Imtiaz, KA Wahid, Color enhancement in endoscopic images using adaptive sigmoid function and space variant color reproduction. *Comput. Math. Methods Med.* **2015**(2), 3905–3908 (2015)
6. J Pohl, I Aschmoneit, S Schuhmann, C Ell, Computed image modification for enhancement of small-bowel surface structures at video capsule endoscopy. *Endoscopy*. **42**(06), 490–492 (2010). <https://doi.org/10.1055/s-0029-1243994>
7. A Karargyris, N Bourbakis, Three-dimensional reconstruction of the digestive wall in capsule endoscopy videos using elastic video interpolation. *IEEE Trans. Med. Imaging*. **30**(4), 957–971 (2011). <https://doi.org/10.1109/TMI.2010.2098882>
8. EJ Daling, Reduction of Power Consumption in Video Communication based on Low Frame Rate Transmission and Decoder Frame Interpolation (2011). <https://brage.bibsys.no/xmlui/handle/11250/2370308>. Accessed 10 Feb 2017
9. C Zach, T Pock, H Bischof, A duality based approach for realtime TV-L 1 optical flow. *Pattern Recognition*. **1**(1), 214–223 (2007). <https://doi.org/10.1007/978-3-540-74936-3-22>
10. F Keller, SH Lauze, M Nielsen, Video super-resolution using simultaneous motion and intensity calculations. *IEEE Trans. Image Process.* **20**(7), 1870–1884 (2011). <https://doi.org/10.1109/TIP.2011.2106793>
11. Rakêt LL, L Roholm, A Bruhn, J Weickert, Motion compensated frame interpolation with a symmetric optical flow constraint. *Lect. Notes Comput. Sci (Incl. Subseries Lect. Notes Artif Intell. Lect Notes Bioinform)*. **7431 LNCS**(PART 1), 447–457 (2012). <https://doi.org/10.1007/978.3.642.33179.443>
12. BG Horn, BKP Schunck, Determining optical flow. *Artificial Intell.* **17**, 185–203 (1981)
13. KI Laws, in *Image processing for missile guidance*. Rapid texture identification, vol. 238 (International Society for Optics and Photonics, 1980), pp. 376–382
14. C Yan, Y Zhang, J Xu, F Dai, J Zhang, Q Dai, F Wu, Efficient parallel framework for HEVC motion estimation on many-core processors. *IEEE Trans. Syst. Video Technol.* **24**(12), 2077–2089 (2014)
15. M Nielsen. A variational algorithm for motion compensated inpainting, vol. 02 (Kingston University, London, 2004), pp. 777–787. <https://doi.org/10.5244/C.18.80>
16. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, *IEEE Trans. Image Process.* **13**(4), 600–612 (2004). <https://doi.org/10.1109/TIP.2003.819861>
17. M Drulea, S Nedevschi, in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. Total variation regularization of local-global optical flow, (2011), pp. 318–323. <https://doi.org/10.1109/ITSC.2011.6082986>
18. Sá,nchez, E Meinhardt-Llopis, G Facciolo, *Image Process. On Line*. **1**(1), 137–150 (2013). <https://doi.org/10.5201/ijpol.2013.26>
19. C Liu, J Yuen, A Torralba, Sift flow: Dense correspondence across scenes and its applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(5), 978–994 (2011). <https://doi.org/10.1109/TPAMI.2010.147>
20. Y Chen, Y Lan, H Ren, Trimming the wireless capsule endoscopic video by removing redundant frames, 1–4 (2012). <https://doi.org/10.1109/WiCOM.2012.6478729>
21. A Mohammed, S Yildirim, M Pedersen, . Hovde, F Cheikh, in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*. Sparse coded handcrafted and deep features for colon capsule video summarization, (2017), pp. 728–733. <https://doi.org/10.1109/CBMS.2017.13>
22. JD Gibbons, S Chakraborti, *Nonparametric Statistical Inference*, vol. 1. (CRC Press, 2010)
23. JH McDonald, *Handbook of biological statistics*, vol. 2. (Sparky House Publishing, Baltimore, 2009), pp. 256–259
24. G Ou, N Shahidi, C Galorport, O Takach, T Lee, R Enns, Effect of longer battery life on small bowel capsule endoscopy. *World J. Gastroenterology: WJG*. **21**(9), 2677 (2015)
25. A Moglia, A Menciassi, P Dario, Recent patents on wireless capsule endoscopy. *Recent Patents Biomed Eng.* **1**(1), 24–33 (2008)

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)