RESEARCH

Error reduction through post processing for wireless capsule endoscope video

Pål Anders Floor[1]*, Ivar Farup[1], Marius Pedersen[1] and Øistein Hovde[2]

Abstract

The wireless capsule endoscope (WCE) is a pill-sized device taking images, which are transmitting to an on-body receiver, while traveling through the digestive system. Since image data is transmitted through the human body, which is a harsh medium for electromagnetic wave propagation, noise may at times heavily corrupt the reconstructed image frames. A common way to combat noise is to use error-correcting codes. In addition one may also utilize inter- and intra frame correlation to reduce the impact of noise at the receiver side, placing no extra demand on the WCE. However, it is then of great importance that the chosen post processing methods do not alter the content of the image as this can lead to miss-detection by gastroenterologists. In this paper we will investigate the possibility for additional noise suppression and error concealment at the receiver side in a high intensity error regime. Due to the high correlation generally inherent in WCE video, satisfactory results are obtained, as concluded from both subjective tests with gastroenterologists as well as the structural similarity (SSIM) metric. More surprisingly, the subjective tests indicate that the inpainted frames in many cases can be used for clinical assessment. These results indicate that one can apply error reduction through post processing together with error-correcting codes to obtain a more noise-robust system without any further demand on the WCE.

Keywords: Wireless capsule endoscope, Error reduction and concealment, Image post processing, Subjective testing, Ultra wide band, SSIM

1 Introduction

Severe diseases in the digestive system like inflammatory bowel disease and cancer reduce the quality of life, or even the length of life, in a huge number of patients. In Europe, colorectal cancer is the second most common cause of cancer death [1]. One way to detect such diseases at an early stage, is to make screening of the digestive system a common procedure beyond a certain age. However, fear of pain and difficulties caused by screening methods like colonoscopy [2], is a major factor limiting the number of people who would volunteer for such screening. In order to motivate as many people as possible to perform preemptive screening, it is vital that the procedure is not unpleasant.

Wireless capsule endoscopy (WCE) is a good option for screening of the digestive system since it is less unpleasant than traditional methods like colonoscopy and gastroscopy. The drawback with current standard WCEs is the low resolution of images as well as the framerate that can be supported given the power capability of the small battery carried onboard [3]. Since the human body is a poor transmission medium for electromagnetic waves, some frames will also be heavily corrupted by noise.

One way to cope with the larger amount of data that would result from increased frame rate and image resolution is to apply compression algorithms that reduces the large correlation among pixels typical for WCE frames, without introducing visible distortions. One such algorithm was proposed by Kim et. al. in [4]. It is also possible to obtain higher transmission rates through the human body over current standard WCEs without increasing the transmission power by using *ultra wide band* (UWB)

*Correspondence: paal.anders.floor@ntnu.no
[1]Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway
Full list of author information is available at the end of the article

communication [5, 6]. This can further allow for an increase in the frame rate.

One will still face problems with heavy noise corruption of some frames: human organs have varying dielectric properties that may cause rapid increase in attenuation due to scattering [7] implying bad reception and thereby severe distortions due to noise. At a certain threshold named *channel outage*, all communication becomes impossible [8]. To cope with these problems, it is common to apply error-correcting codes. An additional possibility that places no further demand on the WCE are post processing techniques at the receiver (located outside the body) where there are no significant restrictions on power usage and computational complexity. Typical WCE video sequences have significant inter- and intra-frame redundancy which can be utilized to suppress noise, conceal and remove errors (in a similar way as redundancy in error-correcting codes are utilized).

If one can find post processing techniques of acceptable quality, which does not introduce false artifacts into the image frames, one can obtain a more noise-robust WCE system through combination with state-of-the-art error-correcting codes.

In this paper, we demonstrate that the error scenarios mentioned above can be dealt with in a satisfactory manner through post processing at the receiver side using combinations and variants of known inpainting algorithms. We analyze both non-compressed and compressed video transmitted using UWB communication over the abdominal channel model derived in [7, 9]. The algorithms are assessed through (1) subjective tests by gastroenterologists, and (2) objective tests through the structural similarity (SSIM) metric [10]. To our knowledge, few existing efforts consider error reduction through post processing for WCE applications, especially for UWB-based systems. Kim et al. addresses this issue in [11], but different from that effort, we deal with high error density.

The paper is organized as follows: in Section 2, the system block diagram and simulation setup are presented. In Section 3, the relevant post processing methods are presented and analyzed. In Section 4, we evaluate the proposed methods both by subjective and objective tests and discuss the results. Conclusions are given in Section 5.

2 Simulation setup

The architecture of a typical WCE device is usually based on known algorithms [12]. In this paper, we provide results for a specific choice of encoder, modulator, and channel model. Figure 1 shows the relevant communication system.

The video source is images taken from *PillCam*® *Colon* [13], that is, current standard WCE images. From a futuristic perspective, it would be convenient to analyze *high definition* (HD) images. However, HD WCE images are not yet available as this require new imaging sensors. On the other hand, another likely future improvement of WCEs would be increased framerate using same sized images.

The main problem with currently available pillcam video streams is that they are already processed (i.e., it is difficult to obtain raw format images), for example, through compression. This is disadvantageous with respect to assessment of new compression schemes and therefore also inpainting methods constructed for particular compression methods (as we will come back to later).
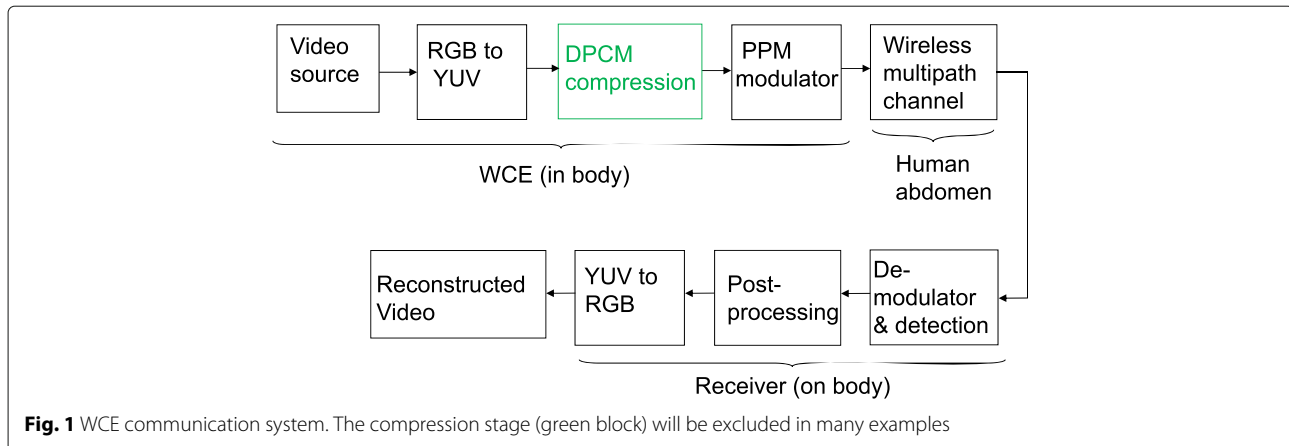
The RGB data is transformed to luminance and chrominance components (YUV) using the transform proposed in [14]. This transform is constructed for integer processing, making it less computational demanding, and thereby suitable for WCE application.

When the compression stage is included, compression is performed frame by frame using the algorithm in [4], which is based on *differential pulse-coded modulation* (DPCM). DPCM is built around simple prediction filters and reduces correlation among pixels. This simple scheme has demonstrated a compression ratio of 95% for colonoscopy frames where decent quality is still provided in the decompressed frames [4].

Further, we assume direct modulation of the video sequence, or the compressed data sequence, using *pulse-position modulation* (PPM). In PPM, amplitude levels are represented as the position of a pulse in time within some fixed (symbol) time window (see [15, 364–373] for illustrations). When a large bandwidth is available, the temporal pulses can be made sharper and several bits can be coded into each pulse without increasing the symbol power, making PPM a simple and power efficient modulation scheme, especially when higher rates are needed in a power limited scenario. The performance of PPM for UWB in-body to on-body communication is evaluated in [5] and [16].

We will apply the $3.4 − 4.8$ GHz UWB channel model for the human abdominal region from [7, 9] throughout. UWB communication can potentially facilitate larger data rates than narrowband systems applied in current WCE standards at low transmission power [5, 6]. The human body is a harsh medium for electromagnetic waves with high attenuation and scattering that may lead to rapid reductions in data rate as well as outages. However, the channel is also *multipath*, meaning that several replicas of the signal can be received at different on-body locations. The effect of severe channel conditions can then be reduced using multiple receiver antennas, effectively increasing the average datarate [17].

Each receiver antenna corresponds to a communication path, or *receiver branch*, that picks up a replica

Fig. 1 WCE communication system. The compression stage (green block) will be excluded in many examples

of the signal. Each receiver branch uses a *matched filter* [15, p.413–417] to maximize the signal-to-noise ratio (SNR) of the received UWB signal. With multiple receiver branches, *diversity* [18, pp. 307–308] can be exploited to further improve the SNR. Here, *maximum ratio combining* [18, pp. 312-313] is applied. In maximum ratio, combining the matched filter output is multiplied by the corresponding channel gain, and by doing so, the signal from each receiver branch is weighted by a factor that is proportional to its strength. That is, contributions from a good branch is strengthened while the poor ones are weakened. After combining, the PPM symbols are detected and converted to bit streams.

Ideally, the algorithms should be tested in real time. However, as no pillcam prototype for UWB communication yet exists, we have to rely on a simulator of the system. We will use the simulator described in [16] applying PPM modulation, the above UWB channel, as well as a multiple antenna receiver as described above. The simulator simulates a moving device through the abdominal model in [7, 9]. The reader may consult [16] for a detailed description of the channel simulator. Note that for proof of concept, we have opted out channel coding in the simulations. The advantage of using a simulator is that one can enforce many non-ideal scenarios on the datastream in a controlled manner, and thereby more easily identify possible error scenarios for the relevant application.

Post processing is the focus of this paper, where variants of known methods are combined in order to reduce and conceal errors introduced through the above simulation framework. In addition to comparing original and corrected images through subjective tests by gastroenterologist in Section 4.1, we will apply the SSIM metric [10] to assess the quality of reconstructed frames. The SSIM measures the similarity between two images, which for our investigation is original versus noisy- or inpainted image. The SSIM score is within the range [0, 1] where '1' implies identical images and '0' implies no correspondence.

Throughout the paper, we will present SSIM values based on the "Y-channel," that is, the luminance channel (or frame) in the luminance/chrominance decomposition performed by the RGB to YUV transform in Fig. 1. SSIM values will be displayed together with each example image. The mean SSIM score for each method based on a given set of images is provided in Section 4.2. The reason why we choose to assess SSIM for the Y-channel is that the majority of the images energy lies in the Y-channel. That is, it is the most crucial channel when it comes to reconstruction of structures and objects in the image. The chrominance (U and V) channels are mainly about color reconstruction. As gastroenterologists are very critical to color changes, the best way of assessing the color reconstruction for WCE application is through subjective testing with experienced gastroenterologists (which we perform in Section 4.1).

3 Post processing methods

We consider two scenarios: (I) single pixel errors appearing with high density due to bad reception. (II) *Error blocks* due to channel outages where significant parts of a frame is missing.

3.1 Scenario i: spatial inpainting

The drawback of PPM modulation is that it may introduce large decoding errors, named *anomalous errors*, if the channel deteriorates from the optimal operation point (any two symbols may be exchanged with equal likelihood) [19, p. 627]. However, in images, anomalous errors will be close to *salt and pepper* noise, which can be efficiently reduced by known spatial inpainting methods utilizing intra frame correlation.

3.1.1 Uncompressed frames

We apply *median filtering* (MF) [20], which can cope with a salt and pepper noise density up to 50% [21, p.200]. A median filter runs over the entire image replacing each pixel with the median of pixel values in a certain

neighborhood of the relevant pixel [20]. We consider a quadratic ($n \times n$) neighborhood of pixels for the median computation here.

Figure 2 shows the original image, the same image corrupted by errors from the relevant simulation model, and the median filtered versions using 4×4 and 5×5 pixel blocks. The reconstruction is quite good, which is also confirmed by the SSIM values: 0.93 for 4×4 blocks and 0.96 for 5×5 blocks. These values are inline with the average SSIM provided in Section 4.2.1.

3.1.2 Compressed frames

For compressed frames, salt and pepper-like noise will be added in the compression domain. As the image is decompressed, distortions will be created.

Figure 4a shows the reconstruction of a DPCM compressed version of Fig. 2a. The algorithm [4] introduces certain artifacts into the frames. Since about 96% of the original data has been removed, the algorithm is quite good (note also that this algorithm was designed for HD colonoscopy images). However, a great deal of these artifacts likely appear due to the fact that the available videostreams for WCE is already compressed (blocking artifacts are observed when the image is magnified, and these correspond to location of the DPCM related artifacts). This is reflected in the SSIM between original and compressed frame which is around $0.8 - 0.85$. In comparison, the same algorithm provides a SSIM of about $0.9 - 0.95$ for HD colonoscopy images. This implies that the relevant compression algorithm does not have a good representative for original compressed WCE image. A fair subjective test is therefore hard to obtain in this case.

Since the DPCM decoder is a recursive filter [4], errors will have "tails" in each image dimension, resulting in the corner-like artifacts shown in Fig. 4b. As shown in [11], when the density of errors is low, they can be fully concealed by first using a corner detector in the decompressed image, like the *Harris detector* [22], then go back to the compression domain and insert one of the neighboring pixels in the corresponding (corrupted) pixel location. With numerous errors, as in Fig. 4b, this method mostly fails as seen in Fig. 4c. A median filter will also fail as it smoothens the compressed image, leading to severe decompression errors.

A way to cope with a high density of errors is through *total variation* (TV) inpainting [23] in the compression domain, as the noise there is close to salt and pepper noise. Figure 3 depicts our approach to spatial inpainting of compressed frames. As suggested in [21, pp. 201–202], TV inpainting can reduce such errors without smoothening other parts of the image as follows: with Ω_c, the compressed image domain, and D_c the inpainting domain (the
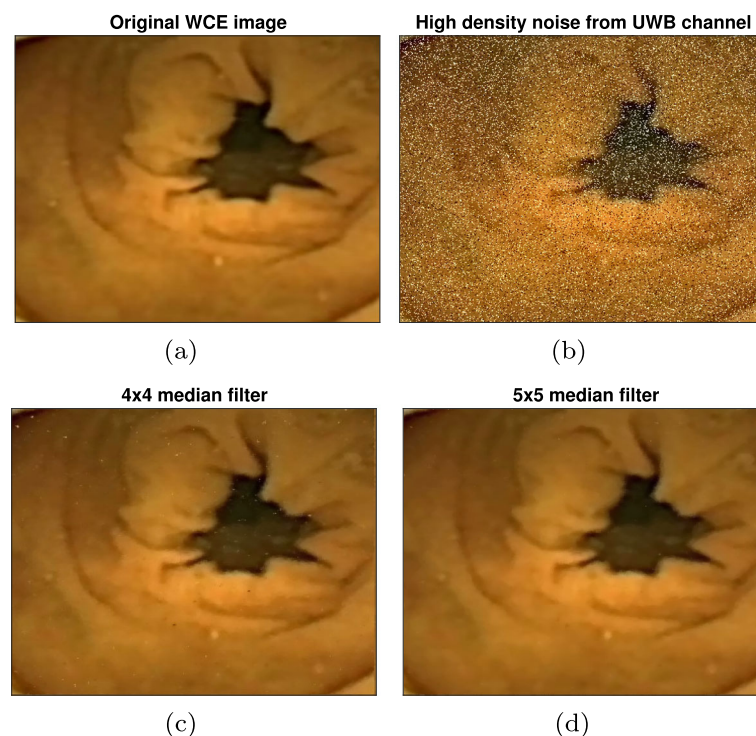
Fig. 2 a Original WCE image. **b** WCE image with massive errors from UWB PPM simulation. SSIM = 0.05 **c** Corrected image: 4×4 median filter. SSIM = 0.93. **d** Corrected image: 5×5 median filter. SSIM = 0.96

Fig. 3 Block diagram for spatial inpainting of single pixel errors in compressed frames
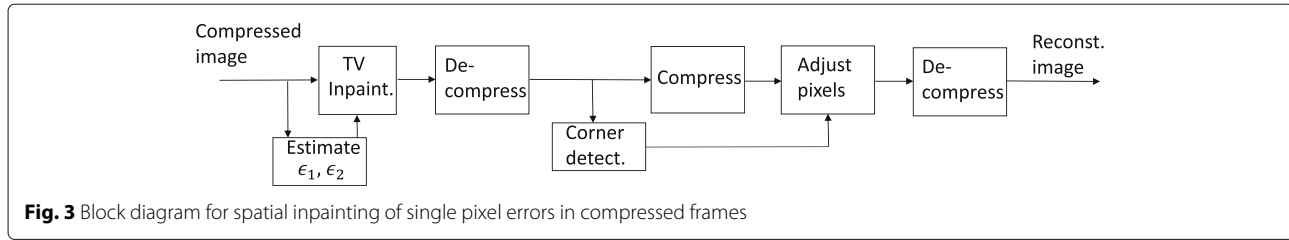
set of noisy pixels given in (2)), let v_0 denote the compressed noisy image on Ω_c. We seek the image v on Ω_c that is the minimizer of [23]

$$E[\,v|v_0, D_c] = \int_{\Omega_c} |\nabla v|\mathrm{d}\mathbf{x} + \frac{\lambda}{2} \int_{\Omega_c \setminus D_c} |v - v_0|^2 \mathrm{d}\mathbf{x}, \quad (1)$$

where λ controls the degree of noise reduction in v_0 outside the inpainting domain D_c, which is given by

$$D_c = \{\mathbf{x}|v_0(\mathbf{x}) \geq C_1 \vee v_0(\mathbf{x}) \leq C_2\}. \quad (2)$$

For salt and pepper noise $C_1 = \max(v_0)$ and $C_2 = \min(v_0)$. Since the noise resulting from PPM modulation is not exactly salt and pepper noise, we set $C_1 = \max(v_0) - \epsilon_1$ and $C_2 = \min(v_0) + \epsilon_2$, where ϵ_1 and ϵ_2 are determined for a relevant set of images. ϵ_1 and ϵ_2 cannot be chosen large enough for all noisy pixels to be contained within D_c without introducing blur in the compressed frame. This will be most problematic in very light or dark areas of the image. A "blob detection" algorithm (like "difference of Gaussians") [24] can be applied to detect what sets of pixels has the lightest and darkest values, then ϵ_1 and ϵ_2 can be adjusted from that. As the output from the DPCM coder has a Laplace-like distribution, this method works quite well, as we will see in Section 4.2.2 where more examples are provided. Since errors residing outside D_c are small and all blur introduced in the compressed frame leads to a bad reconstruction, a large λ should be chosen in Eq. (1). One may obtain further quality enhancement through the algorithm in [11] described above after TV inpanting. That is, by corner detection in the reconstructed image followed by pixel adjustment in the compression domain.

The result is shown in Fig. 4d. Although most of the prominent corners are removed and coarse details in the image are enhanced, there are still some false artifacts present due to smaller errors residing outside D_c in Eq. (2). These false artifacts are likely the reason why the SSIM is not larger than 0.87. We will provide a more thorough analysis of SSIM for this inpainting method in Section 4.2.2.
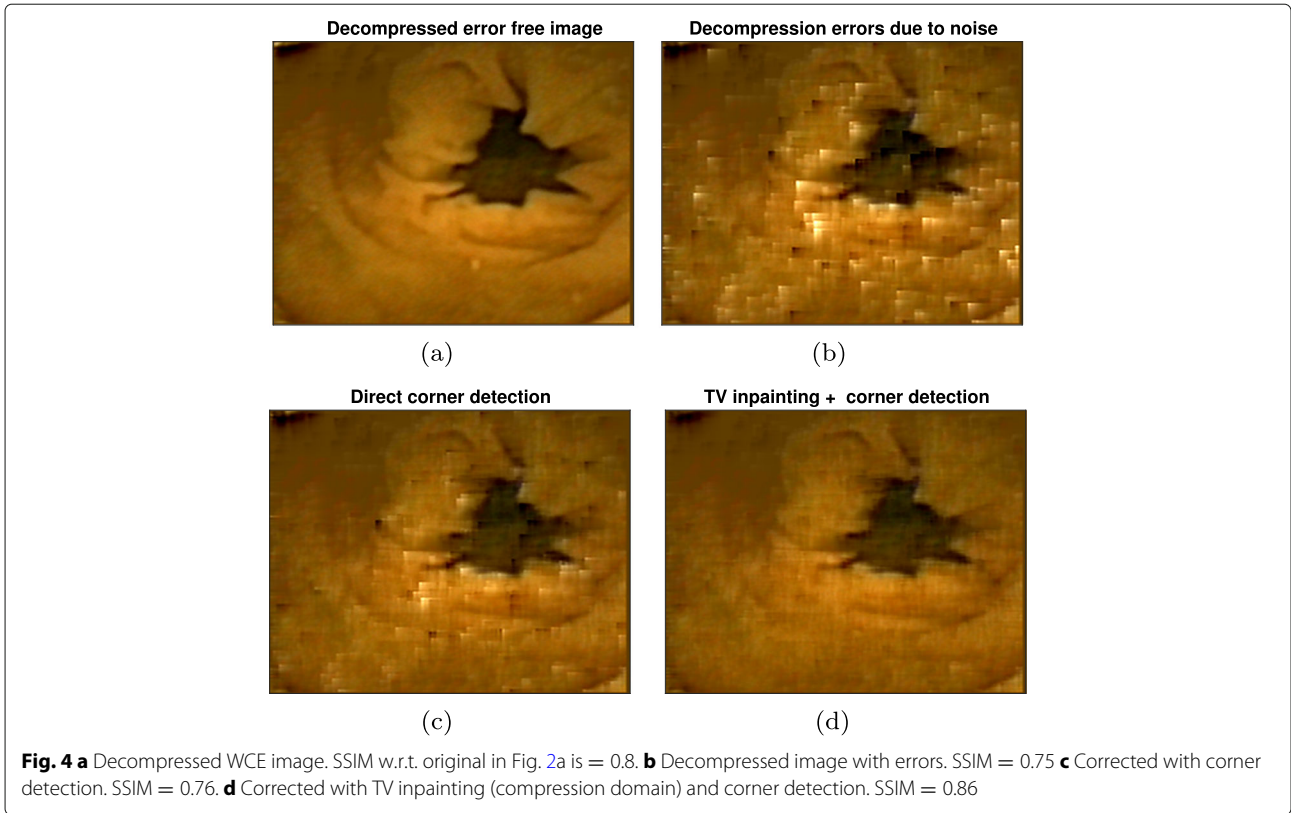
3.2 Scenario ii: temporal inpainting
The method proposed here is the same for compressed and uncompressed frames. We consider uncompressed frames.

If significant parts of a frame is missing, then large inpainting errors are unavoidable with spatial inpainting since the inpainting domain becomes too wide [25]. We utilize interframe correlation in a *temporal inpainting* strategy to cope with this situation: if neighboring frames are close enough content wise, then missing regions can be inserted from one of them. The advantage of this approach is that possible malign tissue that may become invisible due to an error block will become visible in the corrected frame, as information will be inserted from a neighboring frame. That is, information about malign tissue is not lost, and no false artifacts should be introduced.

The proposed scheme is depicted in Fig. 5. First corrupted parts of a frame is detected using the Harris detector. Due to capsule movement, the same features will seldom be located at the same coordinates and perspective on the screen in different frames. To *align* the two images so that their common features are located at the same set of coordinates, one can use a *homography* transform $\mathcal{H}$. That is, pixel coordinates of (past or future) frames I_{n+1} or I_{n-1}, denoted $\mathbf{x}$, are warped onto the coordinates of image I_n as $\tilde{\mathbf{x}} = \mathcal{H}\mathbf{x}$. Past frames can often cover the whole inpainting region at the cost of some blur as the WCE often moves closer to the background scene as it progresses through the digestive system. Future frames may not cover the whole inpainting region, but can be made as sharp as the original frame. We provide examples using past frames in the following.

$\mathcal{H}$ has to be estimated from the relevant frames. There are two main ways to do this: (I) *direct (pixel-based) method* which is described in [26]. (II) Estimate common features using the scale-invariant feature transform (SIFT) algorithm [27], then select the best matches (*inliers*) and find the best fit to $\mathcal{H}$ using the random sample consensus (RANSAC) algorithm [28]. I is likely the least complex method. However, we use method II here since it can determine an accurate $\mathcal{H}$ even from small overlapping regions of two images [26, pp. 15–33]. This implies that $\mathcal{H}$ can be found even when large parts of a frame is missing due to outage.

We applied the MATLAB implementation of SIFT, as well as other supporting functions, from the *VLFeat* library [29] in order to do the computations. Since certain artifacts due to compression and noise may be mistaken as features, it is important to make the SIFT algorithm

Fig. 4 a Decompressed WCE image. SSIM w.r.t. original in Fig. 2a is = 0.8. **b** Decompressed image with errors. SSIM = 0.75 **c** Corrected with corner detection. SSIM = 0.76. **d** Corrected with TV inpainting (compression domain) and corner detection. SSIM = 0.86

favor larger features. Therefore, we set a large "WindowSize" (variance of the Gaussian window), that is 4 *units of spatial bins* [29] (other parameters were set to default). Good matches were then found, as illustrated in Fig. 6.

Due to luminance differences between the original image and the inpainted part, edges may appear (see Fig. 7c). These can be removed through Poisson editing [30]: with Ω, the image domain, and D the inpainting domain with boundary ∂D, let u_0 denote the available image information on $\Omega - D$ and $\vec{v}$ be some "guiding" vector field on D. We seek the image u on D that is the minimizer of [30]

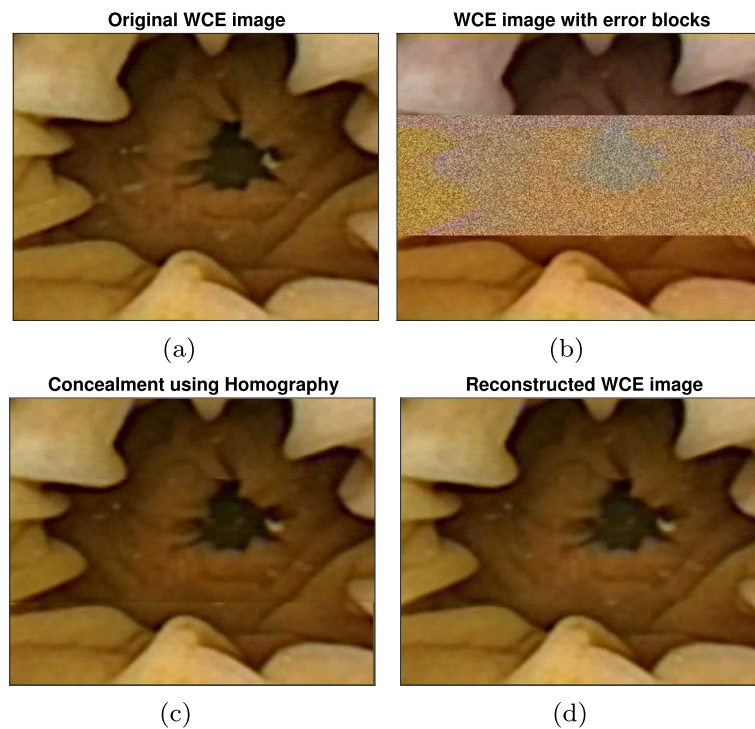$$\min_u \int_D |\nabla u - \vec{v}|^2 \mathrm{d}\mathbf{x}, \quad u_{|\partial D} = u_{0|\partial D}. \qquad (3)$$

The last condition ensures continuity over the boundary of D. Now let $f_D = \{I_{n-1}(\tilde{\mathbf{x}}) | \tilde{\mathbf{x}} = \mathcal{H}\mathbf{x} \in D\}$, i.e., the part inside D which is mapped from the neighboring image. Then, we can set $\vec{v} = \nabla f_D$.

Figure 7 shows the original image, an image with large error blocks in Y, U, and V channels as well as the reconstructed image. We have estimated the homography from the Y channel as depicted in Fig. 6. The SSIM for the noisy image is around 0.55, whereas the corrected image has a SSIM of = 0.93. In comparison, by applying (spatial) TV

Fig. 5 Block diagram for suggested temporal inpainting scheme for error blocks

SIFT matches
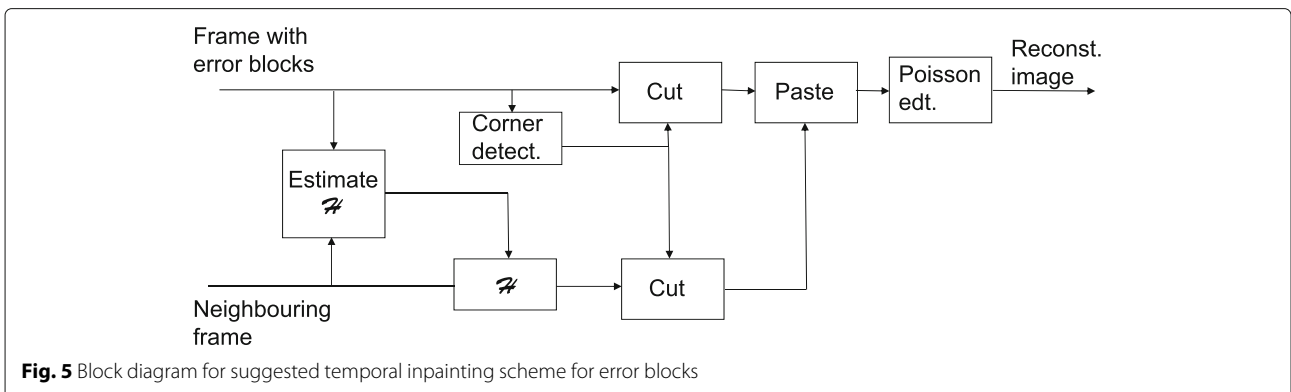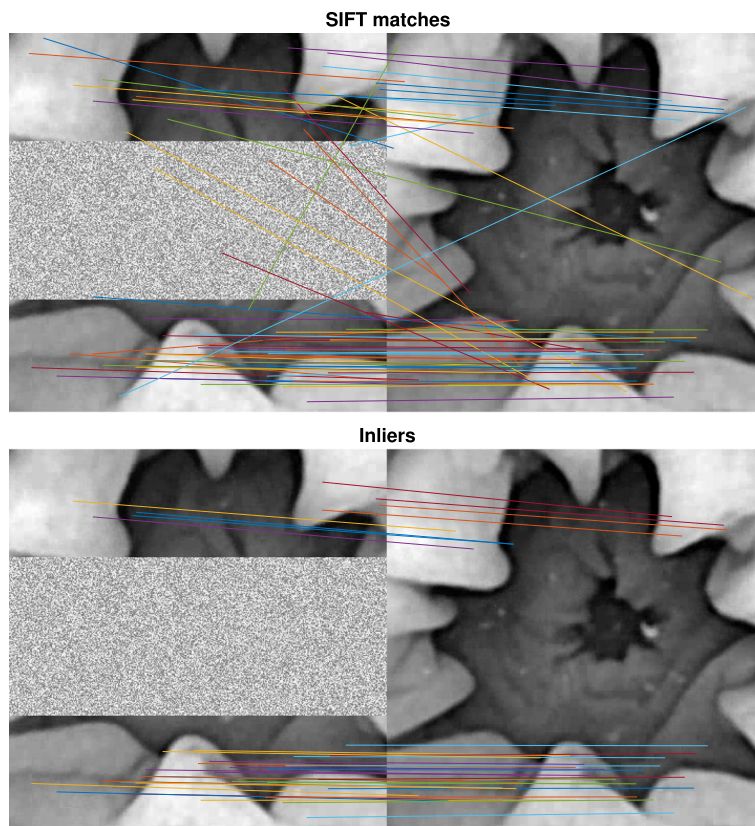
Inliers

Fig. 6 Estimating $\mathcal{H}$ for a highly corrupted frame. Upper: Matching SIFT features. Lower: inliers after RANSAC

Original WCE image

WCE image with error blocks

(a)

(b)

Concealment using Homography

Reconstructed WCE image

(c)

(d)

Fig. 7 Correction of error blocks. **a** Original WCE image. **b** WCE image with error blocks in *Y*, *U*, and *V* channels. SSIM = 0.55 **c** Corrected image using $\mathcal{H}$. SSIM = 0.91 **d** Corrected image using $\mathcal{H}$ and poisson editing. SSIM = 0.93

inpainting within the same noisy frame a SSIM value of 0.83 is obtained. These values are in line with the average SSIM presented in Section 4.2.

One may also apply the chrominance channels, *U* or *V* (from the RGB to YUV transform in Fig. 1), to estimate $\mathcal{H}$ if frame from the luminance channel *Y* is destroyed. This yields additional noise protection. However, since the energy in *U* or *V* is significantly lower, the accuracy of $\mathcal{H}$ may be less than that obtained with the *Y* channel.

It is important to note that $\mathcal{H}$ can only compensate for the WCE's movement, or rigid motion in general. When there are movements in the background due to muscle contractions etc., there will be distortions in the reconstructed frame. One may use *optical flow* [31] computed from neighboring frames to compensate for such motions, or techniques developed in so-called *non-rigid structure from motion* algorithms [32]. Still, it will be hard to obtain stable transforms among images if the correlation (i.e., similarity of image content) is too low, which will be the case when the WCE undergo rapid movements. However, it is likely that future WCE's will have higher framerate, making the above algorithm perform better in general.

3.3 Occurrence of single pixel errors and error blocks simultaneously

Single pixel errors and error blocks may both occur in the same image. There are two approaches to this problem: (i) deal with single pixel errors first and (ii) remove error blocks first.

Experiments clearly showed that approach (i) was the only functioning option: Although SIFT followed by RANSAC is very robust to noise in the images (as these are singled out as outliers through RANSAC) we get into trouble when we try to decide the area in the image that should replace error blocks. This since the corner/line detector becomes confused by the salt and pepper-like characteristic of the single pixel errors.

The result of approach i) is shown in Fig. 8. One can observe that the combined algorithm is capable of coping with both scenarios simultaneously. The SSIM is about the same as it was for block errors in isolation treated in the previous section. This implies that our approach is quite robust.

For compressed frames one would remove all corners in the image by using the method in Fig. 3 first, then remove block errors in the decompressed image. Then one will avoid that the DPCM decoder introduces new set of false artifacts due to the slight mismatch between original and temporally inpainted image.

4 Results and discussion
In this section we assess the performance and quality of the suggested post processing methods described in Section 3 and discuss the results.

We performed subjective tests for the temporal inpainting algorithm suggested in Section 3.2 as well as the spatial inpainting algorithms for uncompressed frames in Section 3.1.1. All algorithms are also evaluated objectively through the SSIM metric.

We did not perform subjective tests on the spatial inpainting algorithm for compressed frames in Section 3.1.2 due to the difficulty of obtaining raw frames. As explained in Section 3.1.2, we lack a good representative for original frame, and this will make the assessment of inpainting methods for compressed frames unfair. For this reason, we will make a more thorough assessment of this method objectively in Section 4.2.

4.1 Subjective testing
The experiment was conducted at "Innlandet Hospital Trust Gjøvik" (SI Gjøvik) with five gastroenterologists. Two are affiliated with SI Gjøvik. The three others were visiting from three other institutions in Norway (St Olavs Hospital Trondheim, Colosseumklinikken Medisinske Senter AS Oslo, and Telemark Hospital Skien), and this reduces possible biases due to "tradition" at a particular institution.

4.1.1 Description of experiment
Application and setup: the application was created using *MATLAB GUIDE* [33]. Three images, *original*, *noisy*, and *inpainted*, were displayed side by side horizontally in random order for each screen shot or *trial*. Thirty trials were done in total, with about 15 trials for each inpainting method. Among these, about 1/3 was with moderate noise, 1/3 with dense noise, and about 1/3 with very dense noise. Examples on dense and very dense noise are provided in Fig. 9.

The images were displayed on a Dell *ultrasharp* 24" monitor (U2412M) with aspect ratio 16:10 and 24-bit color resolution (approximately a sRGB gamut) over a *middle gray* (i.e., RGB values of [119,119,119]) background. The experiment was conducted in a room at SI Gjøvik with same type of lighting conditions as the room used for assessment of colonoscopy images, that is D65 lighting. The monitor was therefore calibrated for D65 lighting.

Data set: the images were captured with Pillcam®COLON [13] from GivenImaging with resolution of 576×576 pixels. The images contain a black frame surrounding the captured scene. The images were cropped to a rectangular shape of 361×361, effectively removing the surrounding frame. Thirty images taken from different parts of the colon was chosen, some normal and others with infected tissue. This is to illustrate a set of different images that would need to be restored in a realistic scenario. Examples are given in Fig. 9 (see also Fig. 7).
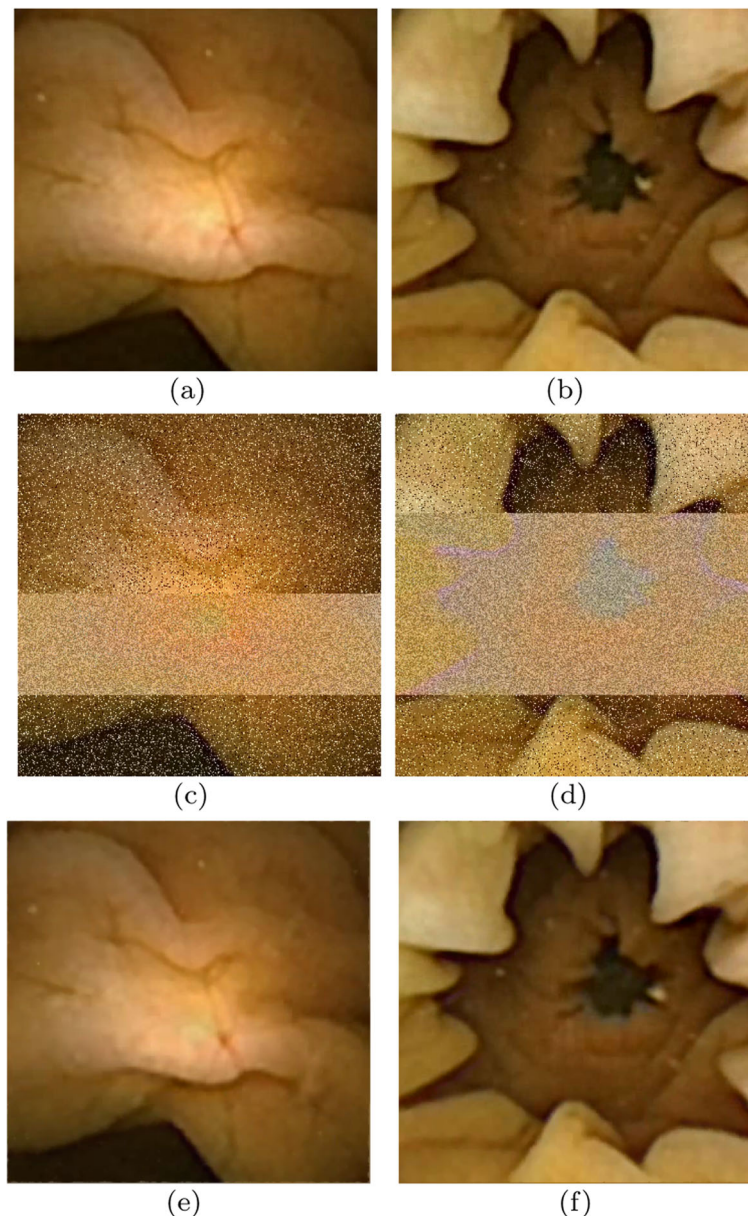
Fig. 8 Simultaneous correction of single pixel errors and error blocks. **a** Original image 1. **b** Original image 2. **c** Noisy image 1, SSIM = 0.005. **d** Noisy image 2, SSIM = 0.02 **e** Corrected image 1, SSIM = 0.91 **f** Corrected image 2, SSIM = 0.89

Assessment: Each candidate was asked to make the following assessments:

1) Image quality: the candidate was asked to categorize the images from A to D with, A being the highest quality, D being the lowest quality, corresponding to 4 to 1 points on a linear scale. Several images could be given the same score.

2) Usefulness: the candidate was asked to evaluate whether an image is useful for inspection or not. That is, whether the image is good enough to decide whether tissue is abnormal or not. As the original always appeared as one of the images, the candidate could determine if something artificial that could tamper the clinical evaluation was introduced into one of the images. The decision for usefulness was "yes" if (i) the image was clear enough to decide whether or not something was wrong, (ii) no significantly disturbing artificial artifacts was introduced into the image, and (iii) no important features were removed from the image. Otherwise, the candidate should click "no." There was a third option "irrelevant," which should be chosen if the given image had no clinical value in and off itself. This is to avoid setting a negative score on the
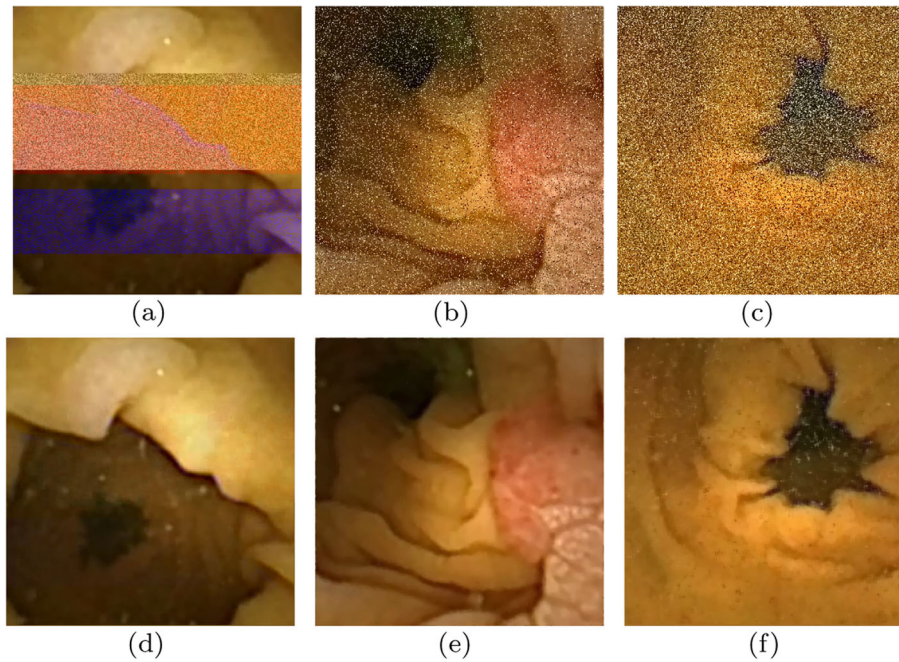
Fig. 9 Selection of noisy images and their inpainted version used during the subjective experiments. **a** Error blocks. **b** Dense noise. **c** Very dense noise. **d** Temporal inpainting with homography $\mathcal{H}$ and Poisson editing. **e** Inpainting with 5×5 MF. **f** Inpainting with 6×6 MF

inpainting algorithm when it is really the original image that is useless for clinical evaluation.

Information on candidates: The five gastroenterologist had somewhat different background and applied different parts of the scale for image quality score:
- Candidate 1: Gastroenterologist with long experience. Used scale B–D (3–1 points) consequently.
- Candidate 2: Gastroenterologist with long experience. Used scale A–C (4–2 points) consequently.
- Candidate 3: Young gastroenterologist with little experience. Used the whole scale A–D (4–1 points) consequently.
- Candidate 4: Gastroenterologist with long experience. Used scale B–D (3–1 points) consequently.
- Candidate 5: Gastroenterologist with long experience,

and with specialization in assessment of pillcam images. Used scale B–D (points 3–1) consequently.

4.1.2 Results

Table 1 summarizes the image quality assessment for each candidate as well as the average over all candidates. Scores are in terms of average *Mean Opinion Score* (MOS) and average *standard Z* score. The MOS is computed as the (arithmetic) mean of all ratings corresponding to the grades A–D (that is, rating 4–1 on a linear scale). We consider a 95% confidence interval computed in the standard way assuming that the variation in the mean is normally distributed [34]. The individual standard deviations are estimated from the data. Average Z score is computed according to Montag's method [35] assuming normal distribution and equal variance for all cases. Due to differing

Table 1 Quality assessment

	Δ_{OI}	Δ_{ON}	Z_O	Z_N	Z_I
Cand. 1	0.26	1.63	0.235	-0.501	0.266
Cand. 2	0.13	1.70	0.204	-0.510	0.306
Cand. 3	0.33	2.47	0.242	-0.500	0.258
Cand. 4	0.06	1.50	0.151	-0.530	0.379
Cand. 5	0.20	2.00	0.187	-0.517	0.330
All	0.20	1.86	0.219	-0.508	0.289

Δ_{OI} and Δ_{ON} denotes average difference in score between original and inpainted image and original and noisy image respectively. The last three columns contain average Z-score for original, noisy and inpainted image

Table 2 Usefulness assessment

Useful I/N:	True yes %	True no %	Irrel. %
Cand. 1	100/33.3	0/60.7	10
Cand. 2	100/13.3	0/86.7	0
Cand. 3	92.3/3.84	7.7/96.2	13.3
Cand. 4	100/17.3	0/82.7	3.33
Cand. 5	86.7/0	13.3/100	0
All	95.8/13.4	4.2/86.6	5.33

Inpainted/noisy (I/N) percentages are shown. "True" means that irrelevant images (percentage is shown in the last column) are taken out during percentage computation

use of the scale among candidates, we have listed the difference in MOS. That is, Δ_{OI} denotes the difference in average MOS between original and inpainted image, whereas Δ_{ON} denotes the difference between average MOS for original and noisy images.

Table 2 summarizes the usefulness assessment. Scores in percent is shown for inpainted/noisy in the first two columns. The percentage values have been computed by removing irrelevant images from the total (thereby the name "true"). The percentage of irrelevant images is listed in the last column.

Figures 10 and 11 show the statistics for MOS as well as histograms for usefulness test (with no compensation for irrelevant images) for the strictest candidate (candidate 3) as well as the total result including all candidates.

Notes on the results: the strictest evaluation overall was done by candidate 3 (the candidate with least experience), whereas candidate 5 (with background in pillcam image assessment) had the strictest judgement on usefulness.

In total, 4 out of 30 inpainted images was deemed useless by candidate 5, three of which was very dense noise corrected with 5×5 or 6×6 median filter (MF). The last one was a temporally inpainted image and was discarded due to color changes in certain areas (however, such errors could easily be eliminated by another inpainting in one of the chromaticity channels).

All in all, the inpainted images have a good score being only 0.1 to 0.3 points away from the original on average. This compared to the noisy images which has a score of about 2 points lower than the original on average. It is more surprising that most of the inpainted images were rated as useful for 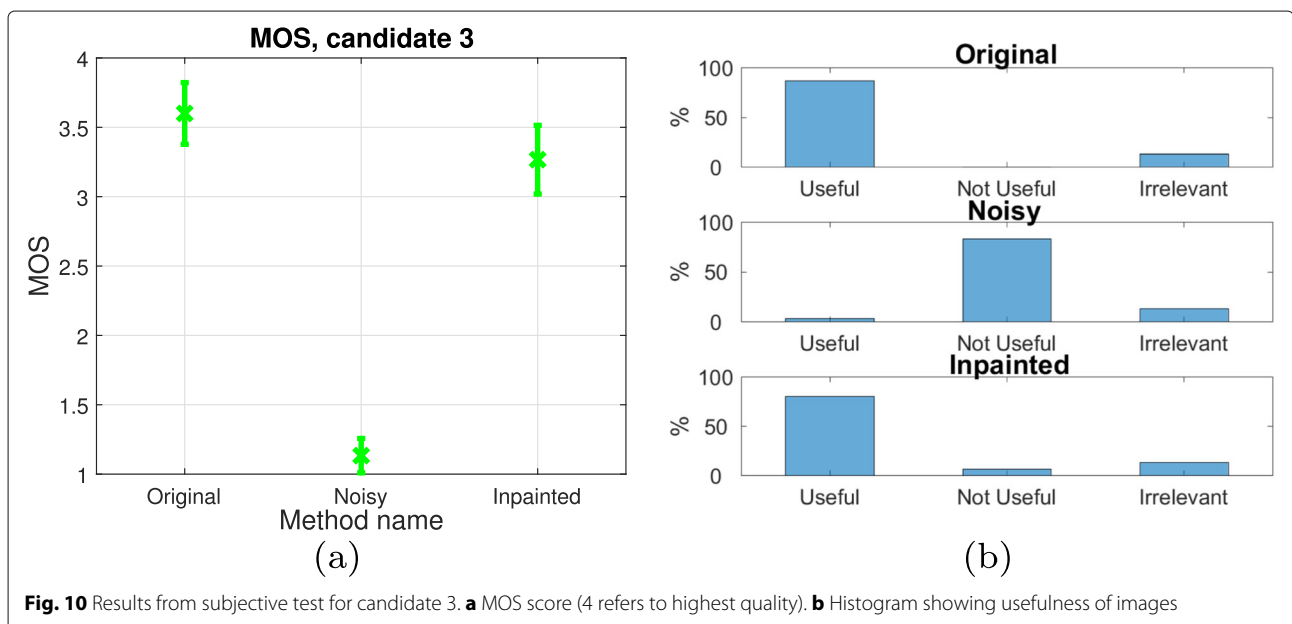inspection, implying that they can actually function as a substitute for the original image whenever noise has corrupted it heavily.
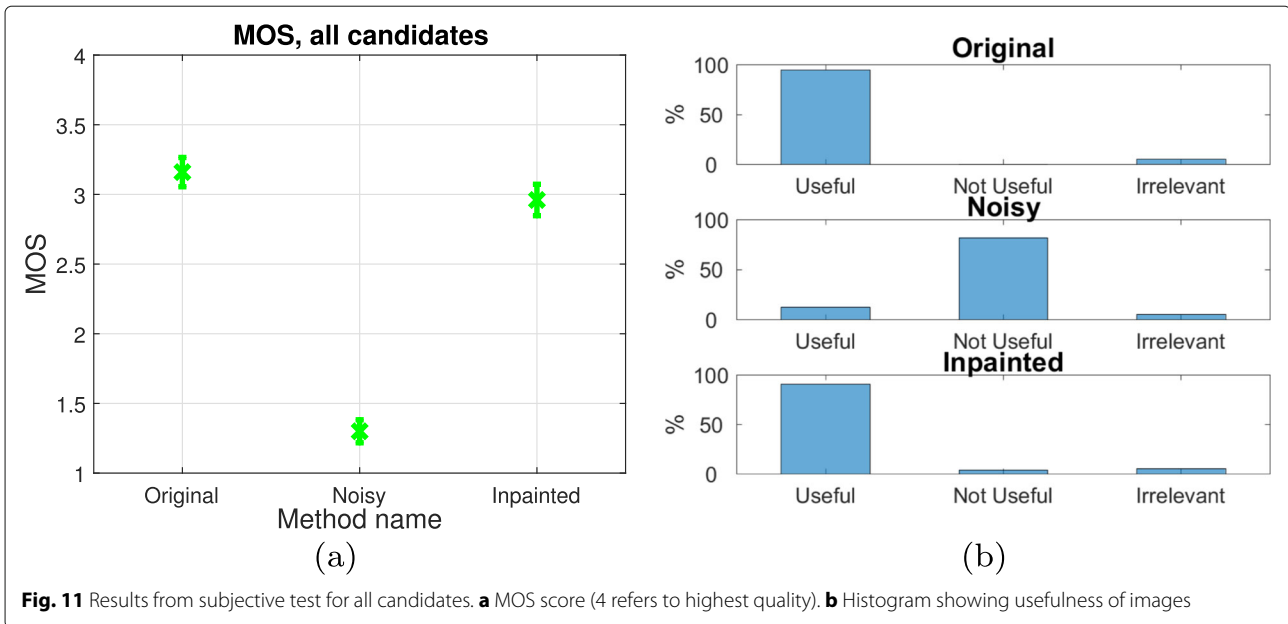
Notes on assessments: candidates 1, 2, 4, and 5 went through the test quite fast, gazing relatively quickly at each image. Candidate 5, with long experience in pillcam image assessment, was generally more critical to any sort of noise artifact. In fact, this candidate consequently gave the highest score to the original image. Candidate 3, with less experience than the other candidates, studied each image more carefully in order to categorize each image differently. This was a valuable contribution w.r.t. quality in that each image was studied more thoroughly before a decision was made. It is interesting to note that candidates 3 and 5 rated the inpainted images with very similar score as the other candidates (taking the differing use of the scale A-D into account), indicating a consistency for the chosen inpainting methods. It is also interesting that candidates 1, 2, and 4, which have mainly evaluated high quality colonoscopy images, judged the usefulness of the inpainted images as being the same as the original ones.

Comment on number of candidates: ideally, a large number of candidates should have performed the subjective test. However, it is difficult to gather enough gastroenterologists over a limited time period due to availability of such qualified personnel. Since the experiment showed a clear consistency after the five candidates, we were able to get hold of, we chose to conclude the experiment.

4.2 Objective tests through SSIM metric

We evaluate all suggested methods here. However, we treat the method in Section 3.1.2 more thoroughly since

Fig. 10 Results from subjective test for candidate 3. **a** MOS score (4 refers to highest quality). **b** Histogram showing usefulness of images

Fig. 11 Results from subjective test for all candidates. **a** MOS score (4 refers to highest quality). **b** Histogram showing usefulness of images

subjective tests were not performed for this case. The other two methods are treated mostly to show correspondence between subjective and objective assessments.

The average SSIM scores with corresponding standard deviation are provided for all inpainting methods along with the corresponding noisy images in Table 3. The following abbreviations are used in the table: "high density single pixel errors" (HSPE), "error blocks" (EB), "corner detection" (CD), "homography with Poisson editing" (HP), and total variation (TV).

4.2.1 *Temporal inpainting and spatial inpainting for uncompressed frames*

To compute the values in Table 3 for uncompressed frames we use the same set of images as in the subjective experiment in Section 4.1.

Consider median filter (MF) and high density single pixel errors (HSPE): an average SSIM of 0.9510 was obtained for MF-inpainted images with a standard deviation of 0.0481, whereas the noisy images with HSPE had a SSIM of 0.1362 on average with standard deviation 0.1165. This is about what one would expect given the results of the subjective test in Section 4.1

Consider homography with Poisson editing (HP) and error blocks (EB): an average SSIM of 0.9093 was obtained with HP-inpainted images with a standard deviation of 0.0214, whereas the noisy images with EB had a SSIM of 0.6636 on average with standard deviation 0.0718. The SSIM for inpainted images is about what one can expect from the results of the subjective test in Section 4.1. The images with EB, on the other hand, seems to have a rather high SSIM value. One likely reason is that the parts of the image not affected by noise is exactly equal to the original.

The SSIM does not seem to account fully for the visual disturbance caused by blocks of random pixels (as seen in Fig. 9a), which was one of the reasons why the subjective test resulted in a low score for images with EB.

4.2.2 *Spatial inpainting for compressed frames*

From the SSIM values in Table 3, one can see that an average SSIM of 0.8538 is obtained for the suggested inpainting method: Total variation inpainting followed by corner detection (TV + CD). This compared to 0.8257 for direct CD from [11] and 0.7828 for noisy images. From these numbers, it appears that only small gains are obtained through both inpainting methods, and that the difference between the suggested inpainting method (TV + CD) and direct CD is quite insignificant. This does not correspond well with reality when inspecting the resulting images. We provide some examples here, in order to

Table 3 Average SSIM scores with corresponding standard deviation for all inpainting methods along with the corresponding noisy frames

	Average SSIM	Std. dev.
Uncomp. HSPE	0.136	0.117
Uncomp. MF inp.	0.951	0.048
Uncomp. EB	0.664	0.072
Uncomp. HP. inp.	0.909	0.021
Compr. HSPE	0.783	0.042
Compr. direct CD	0.826	0.055
Compr. TV + CD	0.854	0.053

Abbreviations are as follows: *HSPE* high-density single pixel errors, *MF* median filter, *EB* error blocks, *CD* corner detection, *HP* homography with Poisson editing, *TV* total variation, *Uncompr* uncompressed, *Compr.* compressed

show that the SSIM does not capture subjective reality very well in this case. Figure 12 shows three examples of compressed image, noisy image, direct CD, and TV + CD. Take for example image 12 (i) versus image 12(l). One would expect the difference in SSIM between these two images to be larger than 0.04.

The SSIM for the suggested inpainting algorithm (TV+DC) may not be too far off. Its mainly the difference in SSIM to the other images that seems unrealistic. A possible reason why the SSIM fails to capture distortions in these images may be that many features in the image are mostly present. However, the fact that the errors perceived
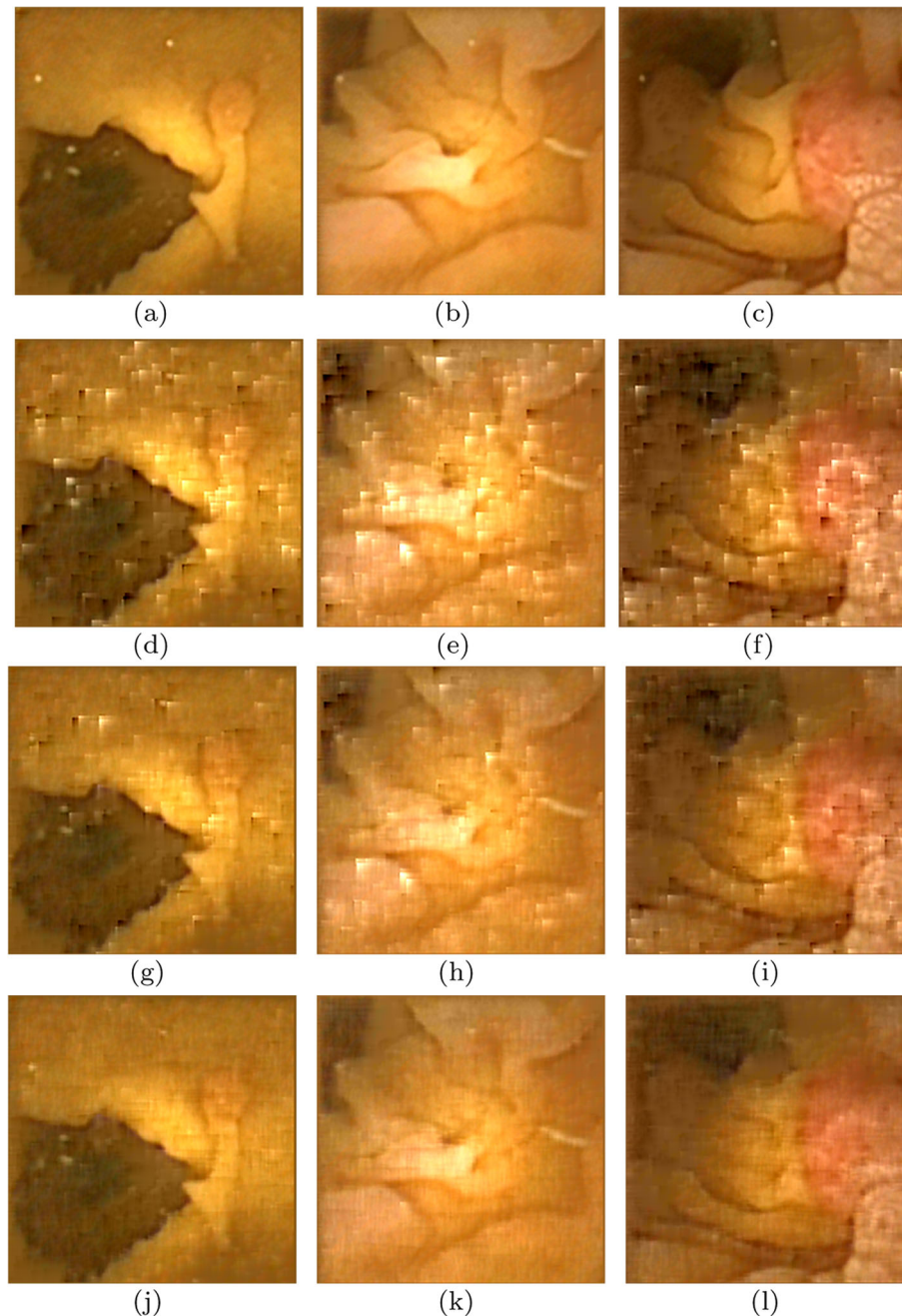
Fig. 12 Selection of compressed images used during SSIM calculation. **a** Compressed image 1. **b** Compressed image 2. **c** Compressed image 3. **d** Noisy image 1, SSIM = 0.745. **e** Noisy image 2, SSIM = 0.794 **f** Noisy image 3, SSIM = 0.751. **g** CD image 1, SSIM = 0.796. **h** CD image 2, SSIM = 0.827. **i** CD image 3, SSIM = 0.790. **j** TV + CD image 1, SSIM = 0.848. **k** TV + CD image 2, SSIM = 0.846. **l** TV + CD image 3, SSIM = 0.828

by our eyes is very disturbing, is seemingly not taken into account by the SSIM here.

5   Conclusion and summary

We have illustrated that post processing at the receiver can successively conceal a high density of errors as well as large missing parts in WCE images. This may be utilized, together with channel coding, to provide a more robust error correction protocol without any additional processing in the pill itself.

Subjective tests show that inpainted images obtained using techniques suggested for temporal inpainting as well as spatial inpainting for uncompressed frames has good quality, and more surprisingly, that they can be applied for clinical assessment. The quality obtained by the suggested techniques are also confirmed through objective tests using the SSIM metric.

When it comes to spatial inpainting on compressed frames, it is not yet possible to draw firm conclusions. The reason is that it is difficult to get hold of raw frames from pillcameras (the ones available are already compressed). Therefore we lack a good representative for "original", that is, noise free compressed frame, when performing subjective tests. For this reason, only objective tests with SSIM metric was performed. Although the suggested inpainting method clearly improve the image quality, as seen through visual inspection, the SSIM does not seem to capture this properly. Although the method seems promising, the results are still inconclusive.

There are cases that cannot be tackled by the proposed methods, like lengthy outage periods and lack of correlation within and among frames. To cope with longer outage periods one may use techniques like optical flow to interpolate frames that have been destroyed. The accuracy will again depend on the correlation with the nearest neighboring frames. However, in future WCE prototypes, one can expect that the framerate will increase, enhancing the performance of the suggested approach. Also, by using strong error correcting codes in combination with post processing algorithms, it is less likely that both methods break down simultaneously.

The method proposed in this paper has at this stage only been implemented through simulation. It is likely that other problems may arise in real clinical application. In the future, we therefore seek to verify the suggested algorithms in real scenarios through clinical trials.

Abbreviations
WCE: Wireless capsule endoscope; SSIM: Structural similarity; UWB: Ultra wide band; HD: High definition; DPCM: Differential pulse-coded modulation; PPM: Pulse-position modulation; SNR: Signal-to-noise ratio; MF: Medial filtering; TV: Total variation; MOS: Mean opinion score; HSPE: High density single pixel errors; EB: Error blocks; CD: Corner detection; HP: Homography with Poisson editing

Author's contributions
The work presented in this paper was carried out in collaboration among all authors. PAF carried out the main research and wrote the manuscript. IF contributed to the theoretical parts of the paper as well as all methods applied. MP contributed to the objective evaluation as well as the setup of the subjective experiment. ØH contributed to the subjective evaluation of the result giving guidelines and recommendations, providing facilities as well as setting up meetings with qualified personnel. All authors have read and approved the manuscript.

Authors' information
Pål Anders Floor received the B.Sc. degree from Gjøvik University College, Gjøvik, Norway, in 2001, and the M.Sc. and Ph.D. degrees from the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway, in 2003 and 2008, respectively, all in electrical engineering. He held a post-doctoral position with the Intervention Center, Oslo University Hospital, and the Institute of Clinical Medicine, University of Oslo, Oslo, Norway, and the Department of Electronics and Telecommunications, NTNU, from 2008 to 2015. He currently holds a post-doctoral position with the Norwegian Color and Visual Computing Laboratory, Department of Computer Science, NTNU, Gjøvik. His current research interests include joint source-channel coding, information theory and signal processing applied on point-to-point links, in small and large networks, and in neuroscience as well as lightweight cryptography for low complexity devices. He is currently doing research on communication and image enhancement for capsule endoscopy.
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
Marius Pedersen received his B.Sc. in Computer Engineering in 2006, and M.Sc. 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 a full 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.
Øistein Hovde received his Doctor of Medicine (MD) from the University of Oslo in 1984. He is currently associate professor at Institute of Clinical Medicine, University of Oslo. He is also a senior consultant (gastroenterologist) at Innlandet Hospital Trust, Gjøvik, Norway, and he is the Head of the gastroenterological division in the hospital. His PhD focused on epidemiology in inflammatory bowel diseases (IBD: Crohn's disease and ulcerative colitis).His clinical work is focused on inflammatory bowel diseases, endoscopy and therapeutic endoscopy. For many years he has had different positions in national and European gastrointestinal organizations (Norwegian Gastroenterological Association and United European gastroenterological organization (ueg)).

Availability of data and materials
The dataset applied in this paper is available in the [36] repository. Please contact the corresponding author for any further data requests.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway. [2]Department of Gastroenterology, Innlandet Hospital Trust, Gjøvik, Norway.

References

1. Colorectal cancer. http://www.euro.who.int/en/health-topics/noncommunicable-diseases/cancer/news/news/2012/2/early-detection-of-common-cancers/colorectal-cancer. Accessed 14 Dec 2018
2. M. Bugajski, P. Wieszczy, G. Hoff, M. Rupinski, J. Regula, M. F. Kaminski, Modifiable factors associated with patient-reported pain during and after screening colonoscopy. Gut. **67**(11), 1958–1964 (2018)
3. P. Swain, The future of wireless capsule endoscopy. World J. Gastroenterology. **14**(26), 4142–4145 (2008)
4. A. Kim, T. A. Ramstad, I. Balasingham, in *4th Int. Symp. on Appl. Sci. in Biomed. and Comm. Technol. (ISABEL)*. Very low complexity low rate image coding for the wireless endoscope (ACM, Barcelona, Spain, 2011), pp. 1–5
5. P. A. Floor, R. Chàvez-Santiago, A. N. Kim, K. Kansanen, T. A. Ramstad, I. Balasingham, Communication aspects for a measurement based uwb in-body to on-body channel. IEEE Access. **7**, 29425–29440 (2019)
6. R. Chàvez-Santiago, K. Sayrafian-Pour, A. Khaleghi, J. Takizawa, J. Wang, I. Balasingham, H.-B. Li, Propagation models for IEEE 802.15.6 standardization of implant communication in body area networks. IEEE Comm. Mag. **51**(8), 80–7 (2013)
7. S. Støa, R. Chàvez-Santiago, I. Balasingham, in *GLOBECOM*. An ultra wideband communication channel for the human abdominal region (IEEE, Miami, FL, USA, 2010), pp. 246–250
8. D. Anzai, S. Aoyama, J. Wang, Specific absorption rate reduction based on outage probability analysis for wireless capsule endoscope with spatial receive diversity. IET Microwaves Antennas Propag. **8**(10), 695–700 (2014)
9. S. Støa, R. Chàvez-Santiago, I. Balasingham, in *3rd Int. Symp. on Appl. Sci. in Biomed. and Comm. Technol. (ISABEL)*. An ultra wideband communication channel model for capsule endoscopy (IEEE, Rome, Italy, 2010), pp. 1–5
10. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
11. A. Kim, E. J. Daling, T. A. Ramstad, I. Balasingham, in *7th Int. Conf. on Body Area Netw. (BODYNETS)*. Error concealment and post processing for the capsule endoscope (ACM, Oslo, Norway, 2012), pp. 149–152
12. R. Chàvez-Santiago, A. Khaleghi, I. Balasingham, T. A. Ramstad, in *2nd Int. Symp. on Appl. Sci. in Biomed. and Comm. Technol. (ISABEL)*. Architecture of an ultra wideband wireless body area network for medical applications (IEEE, Bratislava, Slovakia, 2009), pp. 1–6
13. A colon exam in a capsule. https://www.medtronic.com/covidien/en-us/products/capsule-endoscopy/pillcam-colon-2-system.html. Accessed 14 June 2019
14. D. Turgis, R. Puers, Image compression in video radio transmission for capsule endoscopy. Elsevier J. Sensors Actuators. **123-124**, 129–136 (2005)
15. S. Haykin, *Communication Systems. 3rd edn.* (Wiley, New York, USA, 1994)
16. A. N. Kim, P. A. Floor, T. A. Ramstad, I. Balasingham, in *Proc. 7th Int. Conf. on Body Area Netw. (BODYNETS)*. Communication using ultra wide-band pulse position modulation for in-body sensors (ACM, Oslo, Norway, 2012), pp. 159–165
17. J.-C. Brumm, G. Bauch, On the placement of on-body antennas for ultra wideband capsule endoscopy. IEEE Access. **5**, 10141–10149 (2017)
18. J. D. Parsons, *The Mobile Radio Propagation Channel. 2nd edn.* (Wiley, Chichester, UK, 2000)
19. J. M. Wozencraft, I. M. Jacobs, *Principles of Communication Engineering*. (New York: John Wiley & Sons, Inc, Long Grove, IL, USA, 1965)
20. T. Huang, G. Yang, G. Tang, A fast two-dimensional median filtering algorithm. IEEE Trans. Acoust., Speech, Signal Proc. **27**(1), 13–18 (1979)
21. T. Chan, J. Shen, *Image Processing and Analysis*. (Soc. for Ind. and Appl. Math., Philadelphia, PA, USA, 2005)
22. C. Harris, M. Stephens, in *4th Alvey Vision Conf.* A combined corner and edge detector (BMVA, Univ. of Manchester, UK, 1988), pp. 147–151
23. J. Shen, T. F. Chan, Mathematical models for local nontexture inpaintings. SIAM J. Appl. Math. **62**(3), 1019–1043 (2002)
24. T. Lindeberg, Image matching using generalized scale-space interest points. J. Math. Imaging Vis. **52**(1), 3–36 (2015)
25. T. F. Chan, S. H. Kang, Error analysis for image inpainting. J. Math. Imag. Vis. **26**(1), 85–103 (2006)
26. R. Szeliski, Image alignment and stitching: a tutorial. Technical report (2004). https://www.microsoft.com/en-us/research/publication/image-alignment-and-stitching-a-tutorial/. Accessed 17 Jan 2017
27. D. G. Lowe, in *7th IEEE Int. Conf. Comput. Vision (ICCV)*. Object recognition from local scale-invariant keypoints (IEEE, Kerkyra, Greece, 1999), pp. 1150–1157
28. M. A. Fischler, R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM. **24**, 381–395 (1981)
29. A. Vedaldi, B. Fulkerson, VLFeat: An open and portable library of computer vision algorithms (2008). http://www.vlfeat.org/. Accessed 21 Feb 2017
30. P. Pérez, M. Gangnet, A. Blake, Poisson image editing. ACM Trans. Graph. **22**(3), 313–318 (2003)
31. J. J. Gibson, The perception of the visual world. Science. **113**(2940), 535–535 (1951)
32. L. Torresani, D. B. Yang, E. J. Alexander, C. Bregler, in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Tracking and modeling non-rigid objects with rank constraints (IEEE, Kauai, Hawaii, USA, 2001)
33. Create apps with graphical user interfaces in MATLAB. https://se.mathworks.com/discovery/matlab-gui.html. Accessed 15 Dec 2018
34. D. R. Bull, in *Communicating Pictures*, ed. by D. R. Bull. Chapter 10 - measuring and managing picture quality (Academic Press, Oxford, 2014), pp. 317–360. http://www.sciencedirect.com/science/article/pii/B9780124059061000106. Accessed 26 Jan 2020
35. E. Montag, Empirical formula for creating error bars for the method of paired comparison. J. Electron. Imaging. **15**, 9–11 (2006)
36. GivenImaging: Capsule Video Endoscopy: Atlas 2016 (2016). https://www.medtronic.com/covidien/en-us/products/capsuleendoscopy.html. Accessed 30 Jan 2017

Publisher's Note