Improving the quality of colour colonoscopy videos

Rozenn Dahyot, Fernando Vilarino, and Gerard Lacey

[1]*School of Computer Science and Statistics, Trinity College Dublin, Ireland*

Correspondence should be addressed to Rozenn Dahyot (Rozenn.Dahyot@cs.tcd.ie)

ABSTRACT

Colonoscopy is currently one of the best method to detect colorectal cancer. Nowadays, one of the widely used colonoscope has a monochrome chipset recording successively at 60 Hz R, G and B components merged into one colour video stream. Misalignments of the channels occur each time the camera moves, and this artefact impedes both online visual inspection by doctors and offline computer analysis of the image data. We proposed to restore this artefact by first equalizing the colour channels and then performing a robust camera motion estimation and compensation.

1. INTRODUCTION

Colorectal cancer is the second leading cause of cancer death in the United States and colonoscopy, by removing polyps early, is currently one of the best method to reduce this fatality [12]. Colonoscopy is a minimally invasive endoscopic examination of the colon and the distal part of the small bowel with a fiber optic camera on a flexible tube. The video is inspected in realtime by the doctors to give a visual diagnosis (e.g. ulceration, polyps). This procedure also give the opportunity for biopsy of suspected lesions.

The quality of endoscopic screening is of significant concern in the medical community. Large inter-endoscopist variation in the number of polyps being missed has been measured in clinical studies [12]. Although no definitive cause for the high miss rates has been identified, the speed of camera movement has been suggested as a cause. Our research is within this context of identifying image quality artifacts that may be contributory factors to the high incidence of miss rates in endoscopy.

The inspection of colonoscopy videos can also be done offline, and computer aided methods are currently developed to assist medical doctors. For instance, in [5], a method is proposed to detect tumors in colonoscopy videos using colour wavelet covariance and linear Discriminant Analysis. In [4], the video is used to assess the endoscopist's skills by estimating the camera motion. In [8], edge detection and region growing are used to help the control of the colonoscope. In [14], an automatic labelling system for colonoscopy videos is presented using eye tracking of experts for training and indexing purposes. Labelled data is then used to feed a Support Vector Machine classifier to automatically detect tumors.

Endoscopes used in hospital use different imaging systems. Indeed some endoscopic systems use colour chipset cameras. However more recent endoscopes use monochrome chipsets with successive colour filters in order to improve spatio-temporal resolution of the videos. Those are now more commonly used in hospitals [10]. However one major problem occurres with monochrome chipset cameras: the three colour bands R, G and B composing each image are sometimes temporally desynchronized. This problem is illustrated by the image in figure 1. The current procedure used by doctors when they detect a potentially infected area of the colon, is to keep the camera steady the best they can while they visually inspect the images. Moreover this recurrent misalignment of colour channels in colonoscopy videos can impede any software using colour image processing techniques to assist doctors in their diagnosis.

In this article, we propose in section 2 to model the recording process of images by monochrome chipset endoscopes using successive colour filters. Following this modelling, a short review of related problems is given in paragraph 3. In paragraph 4, we present one possible solution to remove the colour misalignment and this is validated with experimental results in paragraph 5. Potential benefits of this work include facilitating the human and computer-aided visual inspection of colonoscopy videos

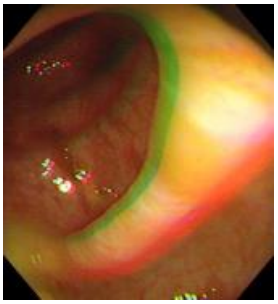


Fig. 1: The image I_{51} has misaligned colour channels.

performed online and offline.

2. COLONOSCOPY VIDEOS

The use of electronic imaging for endoscopy has been around for a long time [2]. The recordings from more recent cameras have better spatio-temporal resolution and work in a similar way as described in [2]: a monochrome image is produced by a black and white chip and is filtered by pulsed light to an RGB coloured system. This setting explains the artefact appearing in the recordings as illustrated in the figure 1. Because the colour channels of each image is not recorded at the same time and because the camera is most of the time moving, the RGB component of the images are misaligned in the videos.

Figure 2 illustrates the problem: the black oriented curve symbolised the camera trajectory. As the camera moves (at changing speed) on this trajectory, the bands $R_{t-\delta_R}$, G_t and $B_{t+\delta_B}$ are recorded at different times and are grouped to form the image I_t in the video. Due to the camera motion in between those recording times, the *RGB* bands in I_t are misaligned.

Monochrome chip endoscopes give however better spatial resolution as a 3 chip camera or a bayer filter introduce approximations to the spatial/colour resolution. Also the LED lighting system can only produce white light through a combination of Red Green and Blue LEDs (there are no "white" LEDS). Thus sequential RGB delivers the best "static" image quality - which is important clinically.

Colonoscopy videos are recorded in a specific environment where several damaging events can occur and blur the images. As spotted in [4], out-of-focus frames usually originate from a too-close focus into the colon, or

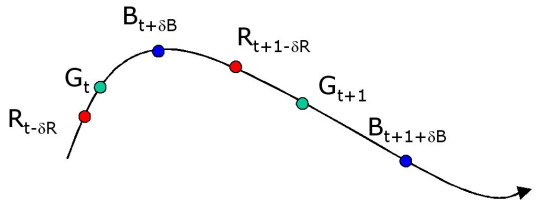


Fig. 2: Modelling the problem: R, G and B components of the images are recorded at different times and since the camera moves, at different positions.

because of substances (e.g. air bubbles) covering camera lens. Hwang et al. [4] propose to filter out those non-informative frames before performing any analysis. Using Fourier transform, they first classify non-informative frames (blurred) from informative ones.

Other artefacts occur in colonoscopy videos such as missing data. Indeed the nature of the colon and its humidity explain the occurrences of specular effects on its surface: the light projected from the colonoscope is entirely reflected in some areas of the colon surface. This creates saturated values (equal to 255) in the colour channels of the images. Figure 1 presents some specular regions (white spots). Figure 3 (top) shows the colour channels separately and the specular regions appear in each of them as white spots. Note that the position of those regions depends on the position and the direction of the light on the camera. Since the three colour channels have not been recorded at the same time and therefore are likely to not have been recorded at the same positions, those specular regions do not always appear as white (but also as reddish or greenish) in the original and restored frames (see figures 1 and 5). In those specular regions, some of the colour information has been lost.

3. RELATED WORKS

The misalignment of colour channels in images recorded by endoscopes has only been tackled by Badiqué et al. [1]. Taking the green channel as the reference frame, they proposed to match the red and the blue channel to it. Phase correlation is used to estimate locally the motion shift in between *R* and *G*, and *B* and *G*. The local shift map is then used to compensate the *R* and *B* to match *G*.

In [6], chromatic aberrations of lenses that provokes the colour channels to be mis-aligned are corrected. This

aberration is compensated by first calibrating the camera on a chessboard for each colour channel and then the displacement is estimated and compensated. The displacement in between RGB is the same for any image recorded by the same camera, so the calibration has to be performed once. The green channel is also chosen as the reference colour as it is midway within the visible spectrum [6]. Calibration cannot be used in our context since our misalignment is due to the motion of the camera that is changing and unpredictable.

In [15], multiplex fluorescence in situ hybridation (M-FISH), an imaging system to analyse chromosomes, shows mis-registrations in between the 6 channels recorded by the microscope which hampers the classification. The misalignment is generated from a combination of sources: lens distortion w.r.t. wavelength, and mechanical misalignment (e.g. vibrations) during the registration. An affine transformation is estimated using mutual information that is computationally expensive to optimize [16].

Motion estimation techniques can be classify into two categories [13]: frequency domain methods and spatial domain methods. The phase correlation method used in [1] belongs to the first category. It is not robust and limited by the displacement it can model. In the second category of methods, we propose to use the motion estimation proposed in [7] that has real-time potentials and is robust to outliers (e.g. specular areas).

4. A NEW RESTORING SCHEME

4.1. Overview

Considering an original frame I_t from a colonoscopy video, it is composed by the three colour channels $I_t = (R_{t-\delta R}, G_t, B_{t+\delta B})$ recorded at three different times. No prior hypotheses are assumed about the delays δR and δB (they can be different and negative). Our framework is therefore quite general and do not depend on the specification of the recording hardware used.

Our restoration method can be described in 3 steps:

1. **Colour channels equalization.** This first process transforms $R_{t-\delta R}$ and $B_{t+\delta B}$ into $\overline{R}_{t-\delta R}$ and $\overline{B}_{t+\delta B}$ respectively by histogram equalisation with G_t. This process is detailled in paragraph 4.2.

2. **Camera Motion estimation.** Considering the equalised frame $\overline{I}_t = (\overline{R}_{t-\delta R}, G_t, \overline{B}_{t+\delta B})$, the six parameter camera motion parameters, noted Θ_t^R and Θ_t^B, are estimated in between $(\overline{R}_{t-\delta R}, G_t)$ and $(\overline{B}_{t+\delta B}, G_t)$ respectively. Paragraph 4.3 presents the robust estimation scheme.

3. **Motion compensation.** The original image $I_t = (R_{t-\delta R}, G_t, B_{t+\delta B})$ is compensated and the restored image is noted $I_t^c = (R_{t-\delta R}^c, G_t, B_{t+\delta B}^c)$. $\overline{R}_{t-\delta R}$ and $\overline{B}_{t+\delta B}$ are compensated to align G_t using motion parameters Θ_t^R and Θ_t^B respectively.

4.2. Colour channels equalization

One major difficulty of our problem is to put in correspondence the R channel (respectively the B channel) with the G one. The grey-level content of each channel is different. We need to define a transformation so that the R values (respectively the B values) can be matched with the green ones. A similar problem arises when restoring flicker in videos. Flicker corresponds to random variations of brightness in the videos and several modelings have been proposed [9]. In particular, one modelling allows to simply compute the non-linear transformation from one cumulative histogram of grey levels to another. It is one of the simplest and earliest method to equalize the grey level dynamics of two images [3].

Considering the cumulative histograms C_R, C_G and C_B of each of the colour channel $(R_{t-\delta R}, G_t, B_{t+\delta B})$, the transfer functions f_R (respectively f_B) to transform the grey level values of $R_{t-\delta R}$ to match those of G_t (respectively to transform the grey level values of $B_{t+\delta B}$ to match those of G_t) is computed by [3]:

$$\begin{cases} f_R(v) = C_G^{-1} \circ C_R(v) \\ f_B(v) = C_G^{-1} \circ C_B(v) \end{cases} \quad (1)$$

f_R (respectively f_B) is applied to each values of $R_{t-\delta R}$ (respectively $B_{t+\delta B}$). Result of those transformations is shown in figure 3 (bottom). Grey level values in $\overline{R}_{51-\delta R}$ and $\overline{B}_{51+\delta B}$ are more similar to the one in G_{51}.

The effect of this equalisation can also be assessed by computing the histograms of the differences $\boldsymbol{\varepsilon} = B_{t+\delta B} - G_t$, $\boldsymbol{\varepsilon} = \overline{B}_{t+\delta B} - G_t$, $\boldsymbol{\varepsilon} = R_{t-\delta R} - G_t$ and $\boldsymbol{\varepsilon} = \overline{R}_{t-\delta R} - G_t$. Figure 4 presents those histograms for the frame I_{51}. We can notice that those histograms of differences after equalisation are centered on zero. This is a requirement

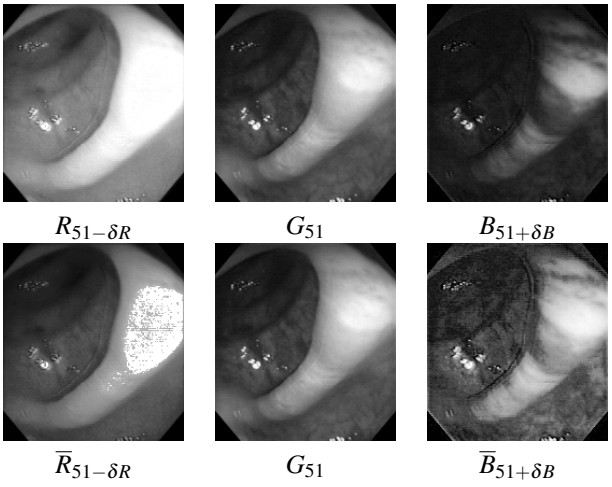
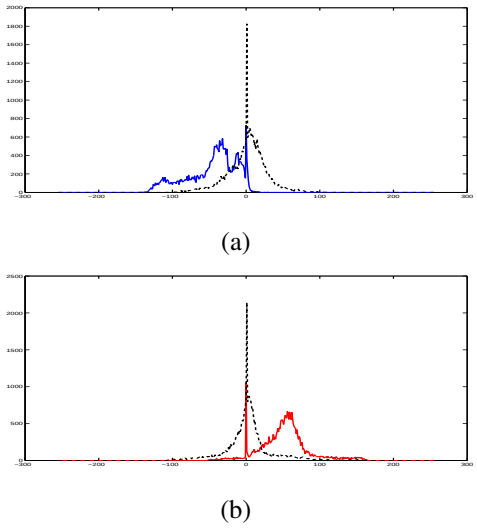


Fig. 3: Original colour channels of I_{51} and its equalized components $\overline{R}_{51-\delta R}$ and $\overline{B}_{51+\delta B}$.



(a)

(b)

Fig. 4: (a) Histograms of the differences $\boldsymbol{\varepsilon} = B_{51+\delta B} - G_{51}$ (blue continuous) and $\boldsymbol{\varepsilon} = \overline{B}_{51+\delta B} - G_{51}$ (black dots). (b) Histograms of the differences $\boldsymbol{\varepsilon} = R_{51-\delta R} - G_{51}$ (red continuous) and $\boldsymbol{\varepsilon} = \overline{R}_{51-\delta R} - G_{51}$(black dots).

to apply the motion estimation as explained in the next paragraph.

4.3. Camera Motion Estimation

We use a 6 parameter affine camera motion instead of 2 used by [1], as it is better suited to the zooming effect created in colonoscopy videos when the camera is moving backward and forward. The frame rate of the endoscope used is 60 fps meaning that in between the recording of the R component and the successive G, only 0.0167s has passed. The 6 parameter motion model is then expected to be sufficient. It is a good trade off between complexity and representativeness [7].

We only present here the estimation of the displacement in between $R_{t-\delta R}$ and G_t. It is the same process for matching $B_{t+\delta B}$ to G_t. In the following, we simplify the notation replacing Θ_t^R by Θ.

The displacement to apply to a pixel at position $\mathbf{x} = (x, y)$ in the image $R_{t-\delta R}$ to match G_t is expressed by:

$$F(\mathbf{x}, \Theta) = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d_x \\ d_y \end{pmatrix} \quad (2)$$

where the camera motion parameter to estimate is $\Theta = (a_1, a_2, a_3, a_4, d_x, d_y)$. Following [7], Θ is estimated by maximizing a probability of the form:

$$\hat{\Theta} = \arg\max_{\Theta} \left\{ \mathscr{P}(\boldsymbol{\varepsilon}) \propto \exp\left[-\frac{1}{2} \sum_{\mathbf{x}} \rho\left(\frac{\varepsilon(\mathbf{x}, \Theta)}{\sigma_\rho} \right) \right] \right\} \quad (3)$$

where $\varepsilon(\mathbf{x}, \Theta) \simeq G_t(\mathbf{x}) - R_{t-\delta R}(F(\mathbf{x}, \Theta))$, ρ is a robust function and σ_ρ is its scale parameter that controls the rejection of outliers in the estimation. More detail on the estimation process can be found in [7]. A robust procedure is preferred to not be sensitive to outliers that arise when the content in the two images to match has changed, or when artefacts occur (e.g. specular areas).

The function ρ is basically reproducing the behaviour of a centered gaussian distribution when the difference $\varepsilon(\mathbf{x}, \Theta)$ is inferior to σ_ρ. On the contrary, when the difference $\varepsilon(\mathbf{x}, \Theta)$ is much larger than σ_ρ, the term is penalized so that its contribution in the estimation is decreased. We have chosen a monotone robust function [11]:

$$\rho(\varepsilon) = 2\sqrt{1 + \varepsilon^2} - 2 \quad (4)$$

This allows to not penalize too strongly pixels that are not perfectly matched after the equalization process. Similarly as in [7], the scale parameter is automatically computed and is proportional to the Median Absolute Deviation (MAD).

4.4. Restoring the colour frame

Once the displacement Θ_t^R and Θ_t^B have been estimated, the compensated frames $R^c_{t-\delta R}$ and $B^c_{t+\delta B}$ are computed from the original frames $R_{t-\delta R}$ and $B_{t+\delta B}$, and then rearranged in the restored colour image $I_t^c = (R^c_{t-\delta R}, G_t, B^c_{t+\delta B})$. Figure 5 shows the result of the restoration for the image I_{51} (cf. fig. 1). Note that the misalignment in this case was quite important, but is however properly restored. Missing data in $R^c_{t-\delta R}$ and $B^c_{t+\delta B}$ may appear on the edge of the restored frame depending on the motion compensation. This effect appears in figure 5 where the bottom and right areas appear green. This is because the red component has been properly aligned with the green but there is no knowledge on the red values on those (bottom and right) areas from the original frame $R_{t-\delta R}$. Those missing values are filled with zeros. One way to improve the visualisation is to crop the restored frame. Alternatively, we are currently investigating inpainting methods to resolve this. Results shown in this article do present those missing data which allow to appreciate the important displacements that sometimes arise in colonoscopy videos.

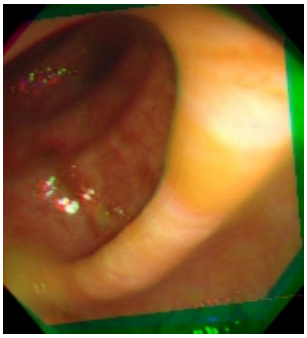


Fig. 5: Restored frame I^c_{51} of I_{51}.

5. EXPERIMENTAL RESULTS

We have collected several hours of colonoscopy in DV compressed format. The assessment shown here is done qualitatively by visual inspection on more than 200 images coming from different sequences. Some restored videos can be seen at `https://www.cs.tcd.ie/Rozenn.Dahyot/Demos/DemosColonoscopy.html`.

Example of successful restoration are reported in the figure 6. For the image I_{12}, the red and green colour channels are misaligned in the original image (right). The misalignment is corrected in the restored image (left). The motion compensation between colour channels implies that missing data may appear on the sides of the restored image. As already explained for the result in figure 5, the missing data are filled with zeros. The result of the restoration process is therefore better appreciated looking at the center of the images and in particular near the strong edges of the lumen.

It is difficult to assess quantitatively the restoration as we do not know what is the ground truth in our videos. We define a failed restoration when the restored image I_t^c is worse than the original one. Figure 7 shows two examples: the compensated image I_{76} is not worse than the original and is not counted as a failure, but image I_{134} is. We assessed that about 10% of the restored frames are worse than the originals.

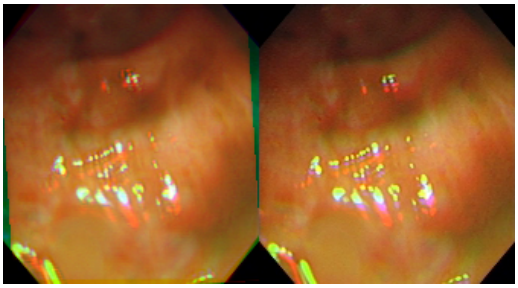
Most of those failed restorations are explained by the really low quality of the original images. Those images are blurred with low edge content, or present really weird colour dynamics (e.g. image I_{134} in figure 7). It is understood that most of those frames would have been classified as non-informative in the system presented by Hwang et al. [4].

Another source of error comes from specular areas which create strong edges on which most motion estimators (including ours) relies heavily. As explained earlier, those specular areas may not be aligned in the R, G and B frames since they appear at different location due to the different orientation and position of the camera at the time of their recordings.

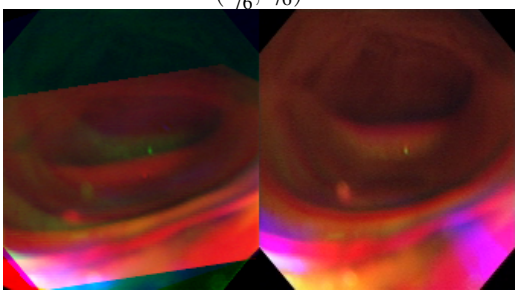
At last, DV uses chroma subsampling that creates artefacts in the R,G and B frames. It means that when decoding the frame in DV, we cannot recover clean R, G, B channels has recorded by the endoscope. It would be difficult to try to recover clean R, G and B frames from the DV files using a software solution. Instead our current work investigates the use of dedicated hardware to acquire high definition raw colour frames in real-time. It is expected that our method to re-align colour channels will then achieve even better performances on cleaner data.

6. CONCLUSION

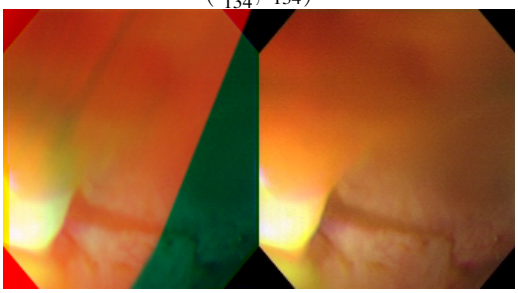
We have presented a new method to restore frames from colonoscopy videos that present a misalignment in their colour channels. This artefact is due to a delay in between the recordings of the different channels and the



(I^c_{76}, I_{76})



(I^c_{134}, I_{134})



(I^c_{20}, I_{20})

Fig. 7: The restoration of the image I_{76} does not improve the original image. The restored images I^c_{134} and I^c_{20} are worse than the originals and are counted as failed restorations.

camera motion inside the colon creates the misalignments. Experimental results show that our method works well and mainly fails when the quality of the images is very low. Current development of our system aims at first improving the quality of the videos using dedicated hardware. It is then believed that any computer-aided analysis of colonoscopy videos would benefit from this restoration performed at an early stage.

7. REFERENCES

[1] E. Badiqué, N. Ohyama, M. Yachida, T. Honda, and J. Tsujiuchi. Compensation of motion related blur in ccd color endoscope images. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1986.

[2] G. Berci and M. Paz-Partlow. Electronic imaging in endoscopy. *Surgical Endoscopy*, pages 227–233, 1988.

[3] R. C. Gonzales and P. Wintz. *Digital image processing (2nd ed.)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1987.

[4] S. Hwang, J. Oh, J. Lee, Y. Cao, W. Tavanapong, D. Liu, J. Wong, and P. C. de Groen. Automatic measurement of quality metrics for colonoscopy videos. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 912 – 921, Singapore, 2005.

[5] S. A. Karkani, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras. Computer-aided tumor detection in endoscopic video using colour wavelet features. *IEEE Transactions on Information Technology in Biomedicine*, 7(3), September 2003.

[6] J. Mallon and P. F. Whelan. Calibration and removal of lateral chromatic aberration in images. *Pattern Recognition Letters*, 28(1):125–135, January 2007.

[7] J.-M. Odobez and P. Bouthémy. Robust multiresolution estimation of parametric motion models. *Journal of visual communication and image representation*, 6:348–365, 1995.

[8] S. Phee, W. Ng, I. Chen, F. Seow-Choen, and B. Davies. Automation of colonoscopy part ii: Visual-control aspects. interpreting images with a computer to automatically maneuver the colonoscope. *IEEE Engineering in Medicine and Biology*, pages 81–88, May, June 1998.

[9] F. Pitié, R. Dahyot, F. Kelly, and A. Kokaram. A new robust technique for stabilizing brightness fluctuations in image sequences. In *Workshop on Statistical Methods in Video Processing (in conjunction to ECCV)*, Prague, Czech Republic, May 2004. Springer, ISBN: 3-540-23989-8.

[10] J. Simpson. *Manual of Canine and Feline Gastroenterology*, chapter Chapter 4. Gastrointestinal endoscopy, pages 34–49. British Small Animal Veterinary Association, 2nd edition, 2005.

[11] C. Stewart. Bias in robust estimation caused by discontinuities and multiple structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(8):818 – 833, August 1997.

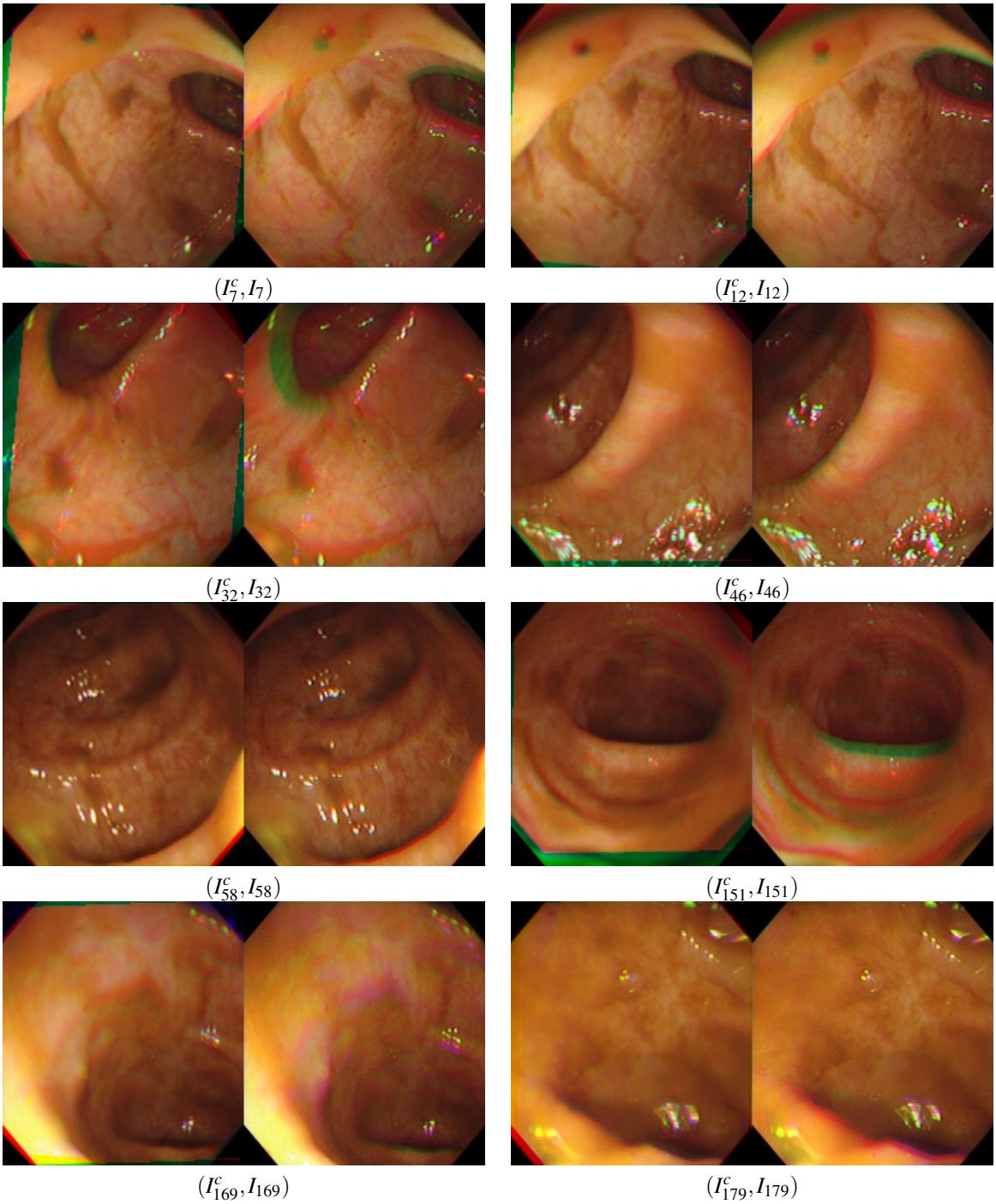
[12] J. C. van Rijn, J. B. Reitsma, J. Stoker, P. M. Bossuyt, S. J. van Deventer, and E. Dekker. Polyp miss rate determined by tandem colonoscopy: A systematic review. *American Journal of Gastroenterology*, pages 343–350, 2006.

[13] P. Vandewalle, S. Süsstrunk, and M. Vetterli. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP Journal on Applied Signal Processing*, pages 1–14, 2006.

[14] F. Vilario, G. Lacey, J. Zhou, H. Mulcahy, and S. Patchett. Automatic labeling of colonoscopy video for cancer detection. In J. Mart, J.-M. Bened, A. M. Mendona, and J. Serrat, editors, *IbPRIA (1)*, volume 4477 of *Lecture Notes in Computer Science*, pages 290–297. Springer, 2007.

[15] Y.-P. Wang. M-fish image registration and classification. In *IEEE International Symposium on Biomedical Imaging: Macro to Nano*, volume 1, pages 57–60, April 2004.

[16] G. Wollny. *Analysis of Changes in Temporal Series of Medical Images*. PhD thesis, University of Leipzig, Germany, 2003.



(I^c_7, I_7) (I^c_{12}, I_{12})

(I^c_{32}, I_{32}) (I^c_{46}, I_{46})

(I^c_{58}, I_{58}) (I^c_{151}, I_{151})

(I^c_{169}, I_{169}) (I^c_{179}, I_{179})

Fig. 6: Successful restorations: the left images are the restored frames and the right ones are the originals.