Hindawi Publishing Corporation
EURASIP Journal on Image and Video Processing
Volume 2008, Article ID 139429, 7 pages
doi:10.1155/2008/139429

Research Article

Improving the Quality of Color Colonoscopy Videos

Rozenn Dahyot, Fernando Vilariño, and Gerard Lacey

Department of Computer Science, School of Computer Science and Statistics, Trinity College Dublin, College Green, Dublin 2, Ireland

Correspondence should be addressed to Rozenn Dahyot, rozenn.dahyot@cs.tcd.ie

Received 1 August 2007; Revised 20 November 2007; Accepted 22 January 2008

Recommended by Shoji Tominaga

Colonoscopy is currently one of the best methods to detect colorectal cancer. Nowadays, one of the widely used colonoscopes has a monochrome chipset recording successively at 60 Hz R, G, and B components merged into one color 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 propose to restore this artefact by first equalizing the color 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 methods to reduce this fatality [1]. 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 gives the opportunity for biopsy of suspected lesions.

The quality of endoscopic screening is of significant concern in the medical community. Large interendoscopist variation in the number of polyps being missed has been measured in clinical studies [1]. 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 artefacts 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 [2], a method is proposed to detect tumors in colonoscopy videos using color wavelet covariance and linear discriminant analysis. In [3], the video is used to assess the endoscopist's skills by estimating the camera motion. In [4], edge detection and region growing are used to help the control of the colonoscope. In [5], an automatic labeling system for colonoscopy videos is presented using eye tracking of experts for training and indexing purposes. Labeled 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 color chipset cameras. However, more recent endoscopes use monochrome chipsets with successive color filters in order to improve spatiotemporal resolution of the videos. Those are now more commonly used in hospitals [6]. However, one major problem occurs with monochrome chipset cameras: the three color 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 color channels in colonoscopy videos can impede any software using color 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 color filters. Following this modeling, a short review of related problems is given in Section 3.

Figure 1: The image I_{51} has misaligned color channels.

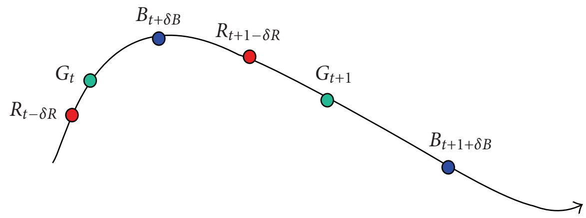


Figure 2: Modeling the problem: R, G, and B components of the images are recorded at different times and the camera moves at different positions.

In Section 4, we present one possible solution to remove the color misalignment and this is validated with experimental results in Section 5. Finally, a conclusion is drawn in Section 6. Potential benefits of this work include facilitating the human and computer-aided visual inspection of colonoscopy videos performed online and offline.

2. COLONOSCOPY VIDEOS

The use of electronic imaging for endoscopy has been around for a long time [7]. The recordings from more recent cameras have better spatiotemporal resolution and work in a similar way as described in [7]: a monochrome image is produced by a black and white chip and is filtered by pulsed light to an *RGB* colored system. This setting explains the artefact appearing in the recordings as illustrated in Figure 1. Because the color channels of each image are not recorded at the same time and because the camera is most of the time moving, the *RGB* components of the images are misaligned in the videos.

Figure 2 illustrates the problem: the black oriented curve symbolized 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. The index t actually corresponds to the frame number of the color frame I_t in the color video, and also indexes the corresponding band G. The variables δ_R and δ_B used with t to index R and B emphasize the fact that those are not recorded at the exact same time as G_t. Due to the camera motion in between those recording times, the *RGB* bands in I_t are misaligned.

Monochrome chip endoscopes give, however, a better spatial resolution as a 3-chip camera or a bayer filter, and introduce approximations to the spatial/color 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 [3], out-of-focus frames usually originate from a too-close focus into the colon, or because of substances (e.g., air bubbles) covering the camera lens. Hwang et al. [3] propose to filter out those noninformative frames before performing any analysis. Using Fourier transform, they first classify noninformative 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 color channels of the images. Figure 1 presents some specular regions (white spots). Figure 3(top) shows the color 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 color 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 color information has been lost.

3. RELATED WORKS

The misalignment of color channels in images recorded by endoscopes has only been tackled by Badiqué et al. [8]. Taking the green channel as the reference frame, they proposed to match the red and the blue channels 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 [9], chromatic aberrations of lenses that provoke the color channels to be misaligned are corrected. This aberration is compensated by first calibrating the camera on a chessboard for each color 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 color as it is midway within the visible spectrum [9]. Calibration cannot be used in our context since our misalignment is due to the motion of the camera that is changing and unpredictable.

In [10], multiplex fluorescence in situ hybridation (M-FISH), an imaging system to analyze chromosomes, shows misregistrations in between the 6 channels recorded by the microscope which hampers the classification. The misalignment is generated from a combination of sources: lens distortion with respect to wavelength, and mechanical misalignment (e.g., vibrations) during the registration. An affine transformation is estimated using mutual information that is computationally expensive to optimize [11].

Motion estimation techniques can be classified into two categories [12]: frequency domain methods and spatial domain methods. The phase correlation method used in [8] 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 [13] 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 color 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 does not depend on the specification of the recording hardware used.

Our restoration method can be described in the following three steps.

(1) *Color 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 equalization with G_t. This process is detailed in Section 4.2.

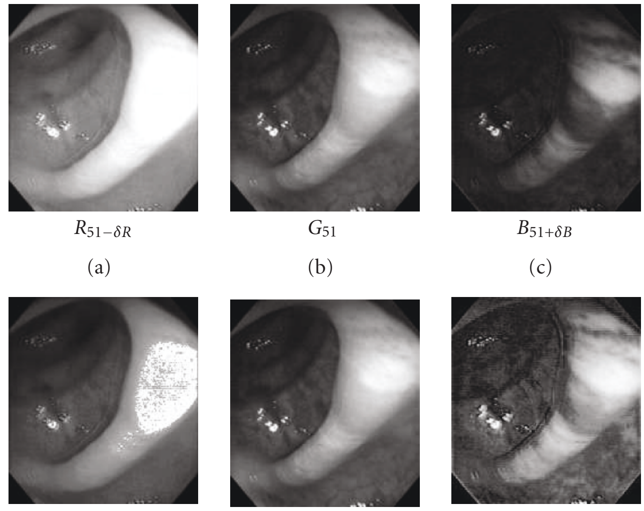
(2) *Camera motion estimation.* Considering the equalized frame $\overline{I}_t = (\overline{R}_{t-\delta R}, G_t, \overline{B}_{t+\delta B})$, the six camera motion parameters, noted by Θ_t^R and Θ_t^B, are estimated in between $(\overline{R}_{t-\delta R}, G_t)$ and $(\overline{B}_{t+\delta B}, G_t)$, respectively. Section 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 by $I_t^c = (R_{t-\delta R}^c, G_t, B_{t+\delta B}^c)$. $\overline{R}_{t-\delta R}$ 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. Color channels equalization

One major difficulty of our problem is to put in correspondence the R channel (resp., the B channel) with the G one. The gray-level content of each channel is different. We need to define a transformation so that the R values (resp., 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 [14]. In particular, one modeling allows to simply compute the nonlinear transformation from one cumulative histogram of gray levels to another. It is one of the simplest and earliest method to equalize the gray-level dynamics of two images [15].

Considering the cumulative histograms C_R, C_G, and C_B of each of the color channels $(R_{t-\delta R}, G_t, B_{t+\delta B})$, the transfer



$R_{51-\delta R}$ (a) G_{51} (b) $B_{51+\delta B}$ (c)

$\overline{R}_{51-\delta R}$ (d) G_{51} (e) $\overline{B}_{51+\delta B}$ (f)

Figure 3: Original color channels of I_{51} and their equalized components $\overline{R}_{51-\delta R}$ and $\overline{B}_{51+\delta B}$.

functions f_R (resp., f_B) to transform the gray-level values of $R_{t-\delta R}$ to match those of G_t (resp., to transform the gray-level values of $B_{t+\delta B}$ to match those of G_t) are computed by [15]

$$f_R(v) = C_G^{-1} \circ C_R(v),$$
$$f_B(v) = C_G^{-1} \circ C_B(v). \quad (1)$$

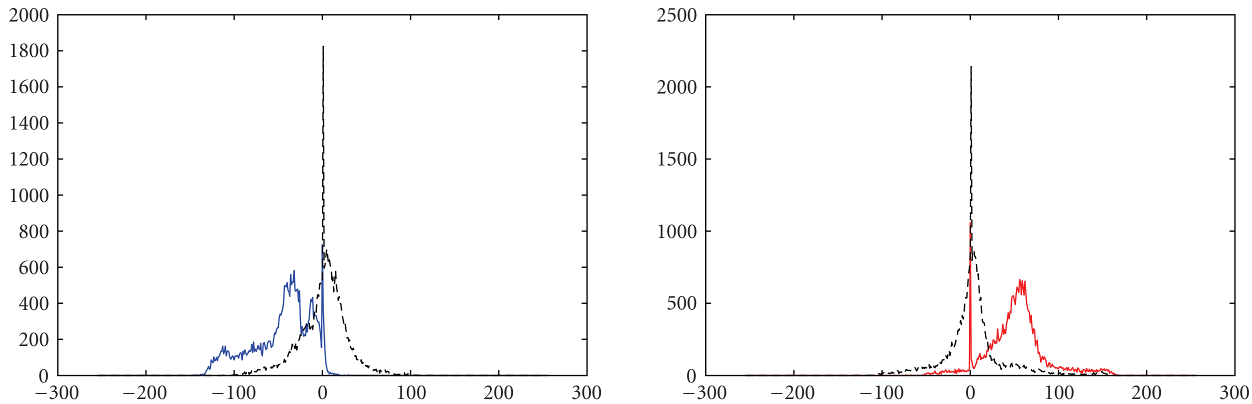
f_R (resp., f_B) is applied to each value of $R_{t-\delta R}$ (resp., $B_{t+\delta B}$). The result of those transformations is shown in Figure 3(bottom). Gray-level values in $\overline{R}_{51-\delta R}$ and $\overline{B}_{51+\delta B}$ are more similar to those in G_{51}.

The effect of this equalization 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 equalization are centered on zero. This is a requirement to apply the motion estimation as explained in the next section.

4.3. Camera motion estimation

We use a 6-parameter affine camera motion instead of 2 used by [8], 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 0.0167 s has passed. The 6-parameter motion model is then expected to be sufficient. It is a good tradeoff between complexity and representativeness [13].

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.



(a) (b)

FIGURE 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).

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 [13], Θ is estimated by maximizing a probability of the form

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

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 details on the estimation process can be found in [13]. A robust procedure is preferred not to 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 behavior 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 [16]

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

This allows to not penalize too strongly pixels that are not perfectly matched after the equalization process. Similarly as in [13], the scale parameter is automatically computed and is proportional to the median absolute deviation (MAD).

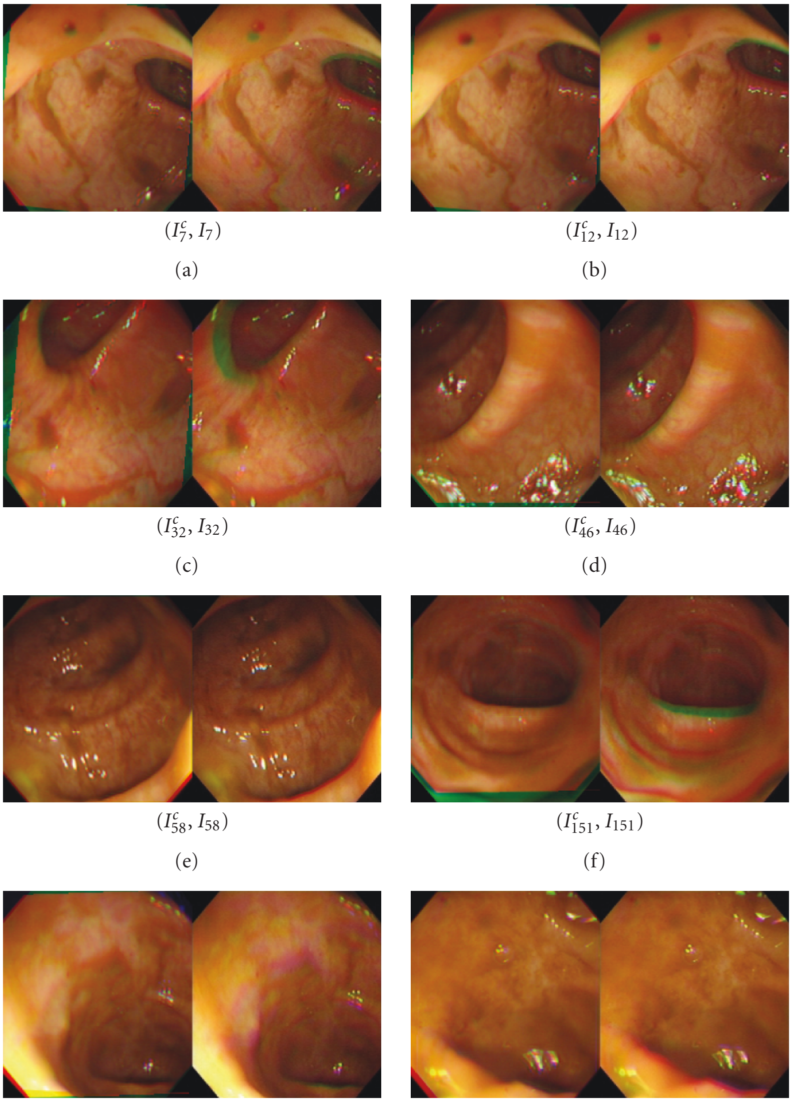
4.4. *Restoring the color frame*

Once the displacements Θ_t^R and Θ_t^B have been estimated, the compensated frames $R_{t-\delta R}^c$ and $B_{t+\delta B}^c$ are computed from the original frames $R_{t-\delta R}$ and $B_{t+\delta B}$, and then rearranged in the restored color image $I_t^c = (R_{t-\delta R}^c, G_t, B_{t+\delta B}^c)$. Figure 5 shows the result of the restoration for the image I_{51} (cf. Figure 1). Note that the misalignment in this case was quite important, but is, however, properly restored. Missing data in $R_{t-\delta R}^c$ and $B_{t+\delta B}^c$ 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 visualization 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. 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.

FIGURE 5: Restored frame I_{51}^c of I_{51}.

5. EXPERIMENTAL RESULTS

We have collected several hours of colonoscopy in DV compressed format. The assessment shown here is done



(I^c_7, I_7)

(a)

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

(b)

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

(c)

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

(d)

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

(e)

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

(f)

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

(g)

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

(h)

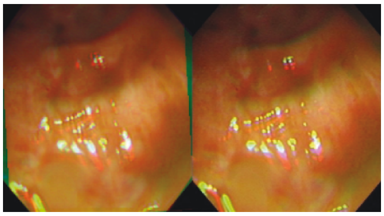
Figure 6: Successful restorations: the left images are the restored frames and the right ones are the originals.

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.

Examples of successful restorations are reported in Figure 6. For the image I_{12}, the red and green color channels are misaligned in the original image (right). The misalignment is corrected in the restored image (left). Successful restorations: the left images are the restored frames and the right ones are the originals.

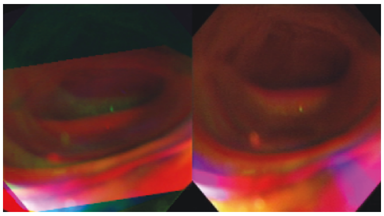
It is difficult to assess quantitatively the restoration as we do not know what is the groundtruth in our videos. We define a failed restoration when the restored image I^c_t is worse than the original one. Figure 7 shows two examples: the compensated image I_{76} is not worse than the original and is 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 color dynamics (e.g., image I_{134} in Figure 7). It is understood that most of those frames would have been classified as noninformative in the system presented by Hwang et al. [3].

In conjunction with blurredness, a possible additional source of error comes from specular areas which create strong edges on which most motion estimators (including ours) rely heavily in some particular situation. As explained earlier, those specular areas may not be aligned in the R, G, and B frames since they appear at different locations due to the different orientations and positions of the camera at the time of their recordings. When no other edge information appears in the image than the specular areas, for instance in



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

(a)



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

(b)

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

(c)

Figure 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.

blurred and uniform color images, it is then likely that our robust estimation process will compensate for the local motion of those specular areas instead of the global motion of the camera. Those specular areas can be detected by searching for saturated pixels (e.g., which values are close to 255) and can be weighted down in our robust estimation scheme.

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, and B channels as recorded by the endoscope.

Our current and future efforts for improving the restoration aim at the following.

(i) *Improving the quality of the images by avoiding compression that creates artefacts.* 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 uncompressed high definition color frames in real-time. It is expected that our method to realign color channels will then achieve even better performances on cleaner data.

(ii) *Detecting and reducing the failed restoration.* We assessed that 10% of the frames are not properly restored and can be even worse than the originals. This can be corrected by one of the following approaches.

- (a) *Not restoring noninformative images (i.e., images that are too blurry).* The detection of such blurry frames is performed by Hwang et al. [3].
- (b) *The second possible approach is to include prior information on the possible motions in the colonoscopy videos.* Some estimated parameters are not coherent with respect to previous and future estimated parameters. Kalman filtering encapsulating priors could be used. Also the displacement of the endoscope manually controlled by medical doctors, in the temporal window of 1/60 seconds, is bounded in the motion parameter space. As can be seen in Figure 7, the failed restorations (frames 134 and 20) involve unrealistic displacement. Current works aim at including more prior information to constrain better the restoration.

(iii) *Filling missing data using inpainting methods.* This can be used to improve further the quality of the images by both correcting the borders of the images after color channel realignment and also filling in specular areas.

6. CONCLUSION

We have presented a new method to restore frames from colonoscopy videos that present a misalignment in their color channels. This artefact is due to a delay in between the recordings of the different channels and the 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. It is believed that any computer-aided analysis of colonoscopy videos would benefit from this restoration performed at an early stage.

ACKNOWLEDGMENTS

This work has been partly funded by the Enterprise Ireland Project PC-2006-038 *Endoview* and the European Network of Excellence on *Multimedia Understanding through Semantics, Computation and Learning* (MUSCLE) FP6-5077-52, available at http://www.muscle-noe.org.

REFERENCES

[1] 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*, vol. 101, no. 2, pp. 343–350, 2006.

[2] S. A. Karkani, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras, "Computer-aided tumor detection in endoscopic video using color wavelet features," *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 3, pp. 141–152, 2003.

[3] S. Hwang, J. Oh, J. Lee, et al., "Automatic measurement of quality metrics for colonoscopy videos," in *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*, pp. 912–921, Singapore, November 2005.

[4] S. J. Phee, W. S. Ng, I. M. Chen, F. Seow-Choen, and B. L. Davies, “Automation of colonoscopy. II. Visual-control aspects,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 17, no. 3, pp. 81–88, 1998.

[5] F. Vilariño, G. Lacey, J. Zhou, H. Mulcahy, and S. Patchett, “Automatic labeling of colonoscopy video for cancer detection,” in *Proceedings of the 3rd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA ’07)*, J. Mart, J.-M. Bened, A. M. Mendona, and J. Serrat, Eds., vol. 4477 of *Lecture Notes in Computer Science*, pp. 290–297, Springer, Girona, Spain, June 2007.

[6] J. Simpson, “Manual of canine and feline gastroenterology,” in *Gastrointestinal Endoscopy*, pp. 34–49, chapter 4, British Small Animal Veterinary Association, Gloucester, UK, 2nd edition, 2005.

[7] G. Berci and M. Paz-Partlow, “Electronic imaging in endoscopy,” *Surgical Endoscopy*, vol. 2, no. 4, pp. 227–233, 1988.

[8] E. Badiqué, N. Ohyama, M. Yachida, T. Honda, and J. Tsujiuchi, “Compensation of motion related blur in ccd color endoscope image,” in *Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP ’86)*, vol. 11, pp. 1785–1788, Tokyo, Japan, April 1986.

[9] J. Mallon and P. F. Whelan, “Calibration and removal of lateral chromatic aberration in images,” *Pattern Recognition Letters*, vol. 28, no. 1, pp. 125–135, 2007.

[10] Y.-P. Wang, “M-FISH image registration and classification,” in *Proceedings of the 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano (ISBI ’04)*, vol. 1, pp. 57–60, Arlington, Va, USA, April 2004.

[11] G. Wollny, “Analysis of changes in temporal series of medical images,” Ph.D. thesis, University of Leipzig, Germany, 2003.

[12] 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*, vol. 2006, Article ID 71459, 14 pages, 2006.

[13] J. M. Odobez and P. Bouthemy, “Robust multiresolution estimation of parametric motion models,” *Journal of Visual Communication and Image Representation*, vol. 6, no. 4, pp. 348–365, 1995.

[14] F. Pitié, R. Dahyot, F. Kelly, and A. Kokaram, “A new robust technique for stabilizing brightness fluctuations in image sequences,” in *Proceedings of the 2nd Statistical Methods for Video Processing Workshop, in conjunction with the European Conference on Computer Vision*, vol. 3247, Springer, Prague, Czech Republic, May 2004.

[15] R. C. Gonzales and P. Wintz, *Digital Image Processing*, Addison-Wesley, Boston, Mass, USA, 2nd edition, 1987.

[16] C. V. Stewart, “Bias in robust estimation caused by discontinuities and multiple structures,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 8, pp. 818–833, 1997.