QUALITY IMPROVEMENT OF ENDOSCOPY VIDEOS

Mirko Arnold∗, Stefan Ameling∗, Anarta Ghosh† and Gerard Lacey
School of Computer Science and Statistics
Trinity College
Dublin 2, Ireland
email: {arnoldma,stefan.ameling,aghosh,gerard.lacey}@cs.tcd.ie
∗Contributed Equally, †Corresponding Author.

ABSTRACT
We address two quality issues in endoscopy videos originating from: (a) Colour Channel Misalignment: In current endoscope systems, the colour channel images are most commonly acquired sequentially at different time instances. Whenever the camera moves significantly between acquisition instances, the colour channels get misaligned. This misalignment degrades the quality of the video due to the appearance of stroboscopic artefacts. (b) Specular Highlights: The surfaces of human organs, visualised in endoscopy videos, usually have a glossy appearance which causes specular highlights in the images. For many image analysis algorithms, these distinct and bright artefacts can become a significant source of error. In this paper, we propose two novel algorithms to remove the aforementioned artefacts. The colour channel misalignment artefacts are removed by estimating the camera motion in the time interval between the acquisition of two colour channel images. The issue of specular highlights is addressed using: (i) a segmentation method based on nonlinear filtering and colour image thresholding and (ii) a fast inpainting method. Both algorithms are evaluated using image sets extracted from colonoscopy videos. The colour channel realignment algorithm achieved a success rate of 86% and 78% in image sets with artificial and real misalignment, respectively. When specular highlights are a priori removed by the proposed algorithm, these success rates increase to 92% and 84%, respectively. The specular highlight removal algorithm achieved an accuracy of 91.99%.

KEY WORDS:
Endoscopy, Image Enhancement, Specular Highlight Removal, Colour Correction.

1 Introduction
Minimally invasive surgical (MIS) procedures have gained popularity in the medical community due to their ability to reduce patient recovery time as well as mortality. A key technological advancement that has contributed to the success of MIS procedures is video endoscopy. Endoscopy is the most commonly used imaging method in MIS procedures, e.g., colonoscopy, bronchoscopy, laparoscopy, rhinoscopy etc.. An endoscope is a bendable tube fitted with a camera at the tip. The tube is inserted through an orifice (natural or artificial) into the human body. The video captured by the camera is displayed in real-time on a monitor. Medical professionals perform diagnostic or therapeutic procedures using this visual information.

Evidently, the quality of the captured videos is crucial for accurate visualisation of the organ under consideration. Much research has been done towards improving the quality of endoscopic images and videos using computer vision techniques. Among them some of the notables are: navigation and virtual guidance systems [1], registration with other image modalities [2], 3D reconstruction [3] and image quality enhancement [4].

We address two quality issues in endoscopy images/videos originating from: (a) Colour Channel Misalignment: RGB sequential image acquisition system are very commonly used in endoscopy. In these systems the images corresponding to the red (R), the green (G) and the blue (B) colour channels are acquired at different time instances and merged to form the resulting video frame. However, an inherent technological shortcoming of such systems is: whenever the speed of the camera is high enough such that it moves significantly in the time interval between the acquisition instances of the images corresponding to two colour channels, they get misaligned in the resulting video frame, cf., Figures 1 and 2. This channel misalignment gives the images an unnatural, highly colourful, and stroboscopic appearance, which degrades the over-
The three colour channel images that are used to build up a frame are acquired by the camera at significantly different positions if the endoscope is moved with high speed. This causes misalignment of colour channels which are realigned by compensating for the camera movement.

![Figure 2: The three colour channel images that are used to build up a frame](image)

Examples of images from minimally invasive medical procedures showing specular highlights. (a) image showing a laparoscopic surgery procedure, (b) colonoscopic image showing a colonic polyp.

![Figure 3: Examples of images from minimally invasive medical procedures showing specular highlights](image)

all video quality of the MIS procedures. Moreover, in endoscopic images, the colour is an invariant characteristic for a given status of the organ [5]. Malignant tumors are usually inflated and inflamed. This inflammation is usually reddish and more severe in colour than the surrounding tissues. Benign tumors exhibit less intense colours. Hence, the colour is one of the important features used both in clinical and automated detection of lesions [6]. Consequently, removal of these artefacts is of high importance both from the clinical and the technical perspectives.

**Specular Highlights:** Images and videos from minimally invasive medical procedures largely show tissues of human organs, e.g., the mucosa of the gastrointestinal tract. These surfaces usually have a glossy appearance showing specular highlights due to reflection of light, cf., Figure 3. These highlights can negatively affect the perceived image quality. Furthermore, for many visual analysis algorithms, these distinct and bright features can become a significant source of error. Since the largest gradients can usually be found at the contours of specular highlights, they would interfere with all gradient based computer vision and image processing algorithms. Similarly, they will also affect texture based approaches. On the other hand, specular highlights hold important information about the surface orientation, if the relative locations of the camera and the illumination unit are known. Knowledge of the location of specular highlights may therefore improve the performance of 3D reconstruction algorithms. After detection, the removal of specular highlights can be achieved by interpolation using information from the neighbourhood pixels, which is generally referred to as image inpainting. For most applications in automated analysis of endoscopic videos, inpainting will not be necessary, since the localisation might suffice the need. However, a study by Vogt et al. [7], suggests that well inpainted endoscopic images are preferred by physicians to images showing specular highlights. Algorithms with the objective of visual enhancement may therefore benefit from a visually pleasing inpainting strategy.

We propose two novel algorithms to remove: (a) colour channel misalignment artefacts and (b) specular highlights in endoscopic videos. The proposed algorithms are evaluated using colonoscopy videos. Video Colonoscopy is an example of an endoscopic procedure, which is used for screening of patients for colorectal cancer, which in the last 20 years has been one of the leading causes of cancer related deaths all over the world [8]. In a video colonoscopy procedure an endoscope is inserted through the anus into the rectum and then gradually moved towards the most proximal part of the colon, viz., the terminal ileum, in order to detect abnormalities, e.g., the presence of polyps, lesions etc.. If these polyps, as depicted in Figure 1 and 3(b), remain undetected, they can grow to become malignant resulting in colorectal cancer. As shown in Figure 1 and 3(b) the presence of the aforementioned artefacts in colonoscopy videos is a major hindrance to efficient manual or automated detection and characterisation (malignant vs. benign) of polyps.

The paper is organised as follows: State of the art of present literature is briefly presented in Section 2. In Section 3 an elaborate description of the proposed algorithm to remove colour channel misalignment artefacts is presented. The algorithm to remove specular highlight artefacts is described in Section 4. The details of the experimental set up and an in-depth analysis of the achieved results are presented in Section 5. We conclude with a brief summary and possible future studies in Section 6.

## 2 Related Work

The problem of misaligned colour channels in endoscopic videos has only been tackled by Badiqué et al. [9] and Dahyot et al. [10]. Our new approach differs mainly in two ways: (a) the camera motion is estimated by the computation of the homography between two channels using correspondence of feature points, which is computationally more efficient and stable, (b) the proposed method is thoroughly evaluated using digitally captured and uncompressed high resolution colonoscopy video frames.

There exist a number of approaches to segment specular highlights in images, usually through detection of grey
scale intensity jumps [11] or sudden colour changes [12, 13, 14]. These approaches either can be applied to grey scale images only or assume aligned colour channels, for which specular highlights usually appear white. Our proposed method extends the state of the art through its ability to detect specular highlights in images with aligned and misaligned colour channels. Furthermore, it approximates the surface colour at the specular highlight regions and therefore allows for an efficient inpainting of the detected specular highlights. An overview of image inpainting algorithms can be found in [15] or, for video data, in [13].

3 Removal of Colour Channel Misalignment Artefacts

Let $C_B, C_R, C_G$ be the three colour channels of a given endoscopy video frame. The proposed algorithm to remove the colour misalignment artefacts comprises the following key steps:

- Compute the Kullback-Leibler divergence, $d_{KL}$, between the intensity histograms of the colour channels, denoted as: $d_{KL}(h_C_i, h_C_j), i \neq j, \forall i, j \in \{R, G, B\}$. $h_C_i$ is the intensity histogram corresponding to colour channel $i$. Choose the colour channels $i$ and $j$, for which the $d_{KL}$ is minimum.

- Compute the homography ($H_{C,C}$) between the chosen colour channels $i$ and $j$, through feature matching. Assume linearity of motion and compute the homography between consecutive colour channels, $H_{C,C}, i,j \in \{R, G, B\}$.

- Align all the colour channels by using the inverse homography, $H_{C,C}^{-1}, i,j \in \{R, G, B\}$.

The aforementioned steps are illustrated schematically in Figure 4. In the following we discuss and justify the steps outlined above, especially the use of homography instead of full 3-D camera motion estimation.

In order to realign the colour channels it is imperative to estimate the camera motion in the interval between the acquisition of two consecutive colour channels and use that information to map the channels to a single reference frame (cf. Figure 2). The fundamental requirement of any extrinsic camera parameter estimation algorithm is to find corresponding points between multiple views of a single scene captured by the camera at different positions. We call these corresponding points feature points. Using the correspondence between the feature points, the camera extrinsic parameters are estimated by computing the Fundamental Matrix $F$, using for example the 8-point algorithm [16]. However, if the baseline, i.e., the distance between the positions of the cameras where the two images were acquired, is too small, computation of $F$ tends to be unstable. This is known as the small baseline problem in the field of structure from motion studies. As we estimate the camera motion between two colour channels, where the time interval is of the order of $\frac{1}{90}$ seconds, the baseline is very small in our case. However, in case of a small camera movement, homography can be used as a good approximation of the camera motion [17]. Generally, a homography is the mapping of points on one surface to another. In this context it is a mapping between points on two image planes that correspond to the same location on a planar object. However, due to the fact that the homography is computed between the channels of a single frame, which means the motion is small, this approximation was found to be acceptable and used in the literature [17]. Therefore, we estimate the camera motion by the homography $H$, which relates the positions of the points on a source image plane ($\hat{p}$) to the points on the destination image plane ($p$) by the following equations: $p = H\hat{p}$ and $\hat{p} = H^{-1}p$ with $p = \begin{pmatrix} x \\ y \end{pmatrix}$ and...
achieve the most accurate estimation and higher computation efficiency, correspondences are computed between one channel pair only, viz., the pair with the highest intensity distribution similarity. Kullback-Leibler divergence ($d_{KL}(\cdot, \cdot)$) is used as the dissimilarity metric. For two discrete probability mass functions ($pmf$) $P$ and $Q$, it is defined as: $d_{KL}(P,Q) = \sum_{k=1}^{n} P(k) \log \frac{P(k)}{Q(k)}$. In our application the pixel intensity distribution is used to compute the dissimilarity between the colour channels. The $pmf$ of the intensity distributions of two colour channels are estimated using normalised histograms. $d_{KL}(\cdot, \cdot)$ is computed between all three colour channels and the pair with minimum Kullback-Leibler divergence is chosen for direct computation of the homography. As it is necessary to estimate camera motion between two channel pairs (see Figure 2), the second homography is computed by using the first one. More precisely, a set of correspondences for the second channel pair is created based on the correspondences of the first pair. Let $(p, q)$ be a matched point pair for $C_B, C_G$ then a pair $(s, t)$ is created for $C_B, C_R$ with $s = p$ and

$$t_x = px + m \cdot (px - qx)$$

$$t_y = py + m \cdot (py - qy).$$

Here $m \in \mathbb{R}$ and $px, qx, py, qy$ are the x-coordinates and y-coordinates of the points $p, t$ and $q$, respectively. This new set of correspondences is used to estimate the camera motion between the channel pairs. Thus, a linear camera movement between $C_B, C_G$ is assumed. As the camera movement between two frames is very small, this approximation is well-justified.

## 4 Removal of Specular Highlights Artefacts

### Detection of Specular Highlights

In the first step of our approach, we use a global threshold $T_1$ on the green channel ($C_G$), the blue channel ($C_B$) and a grey scale image representation obtained from RGB through $E = 0.2989 \cdot C_R + 0.5870 \cdot C_G + 0.1140 \cdot C_B$. Any given pixel, say $p$, is marked as a possible specular highlight when the following condition is met: $C_G(p) > T_1$ $\lor$ $C_B(p) > T_1$ $\lor$ $E(p) > T_1$. The red channel is not used in this first step, because intense reddish colours are very common in colonoscopic videos and therefore saturated red occurs due to reasons other than being specular highlights. The second step detects pixels that can be described as positive outliers when compared to the surrounding tissue surface. Other than segmenting the image into tissue regions of similar colour (as, e.g., in [14]) we compute this neighbourhood for each pixel using a modified median filter approach. Using a square kernel of size $w \times w$, with $w$ on the order of 1/5 of the image width, we compute a median filtered image. The possible specular highlight regions that were detected in the first step enter the median filter with the centroid of the colours of a region of pixels adjacent to the specular region’s outline. This prevents the filtered image to appear too bright in regions where specular highlights cover a large area. Smaller specular highlights are effectively removed due to the relatively large window size. The pixel value at

\[\hat{p} = \left(\hat{x}, \hat{y}\right),\] where $(x, y)$ and $(\hat{x}, \hat{y})$ are the coordinates of $p$ and $\hat{p}$. As $H$ has eight independent elements [17], it can be computed using four point correspondences. However, the use of more points is beneficial, because in practice there will be noise and other inconsistencies in the feature point correspondences. Hence, in practice, the homography is computed through minimisation of the so called back-projection error using a set of point correspondences:

\[\Phi_H = \sum_i \left(\hat{x}_i - \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}}\right)^2 + \sum_i \left(\hat{y}_i - \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}}\right)^2.\] (1)

Here $h_{mn}$ is the $m^{th}$ row and $n^{th}$ column entry of the homography matrix $H$ and $(x_i, y_i), (\hat{x}_i, \hat{y}_i)$ are the coordinates of the $i^{th}$ feature point in two views respectively. As not all of the point pairs may fit the transformation, the estimation of the correct transformation is done using robust statistical methods, e.g., Random Sample Consensus (RANSAC) [18].

The computed homography is then used to map the channels by transforming the pixels. By merging the transformed colour component images to a colour frame, the camera shift between the channels is compensated. To achieve the most accurate estimation and higher computational efficiency, correspondences are computed between one channel pair only, viz., the pair with the highest intensity distribution similarity. Kullback-Leibler divergence ($d_{KL}(\cdot, \cdot)$) is used as the dissimilarity metric. For two discrete probability mass functions ($pmf$) $P$ and $Q$, it is defined as: $d_{KL}(P,Q) = \sum_{k=1}^{n} P(k) \log \frac{P(k)}{Q(k)}$. In our application the pixel intensity distribution is used to compute the dissimilarity between the colour channels. The $pmf$ of the intensity distributions of two colour channels are estimated using normalised histograms. $d_{KL}(\cdot, \cdot)$ is computed between all three colour channels and the pair with minimum Kullback-Leibler divergence is chosen for direct computation of the homography. As it is necessary to estimate camera motion between two channel pairs (see Figure 2), the second homography is computed by using the first one. More precisely, a set of correspondences for the second channel pair is created based on the correspondences of the first pair. Let $(p, q)$ be a matched point pair for $C_B, C_G$ then a pair $(s, t)$ is created for $C_B, C_R$ with $s = p$ and

\[t_x = px + m \cdot (px - qx)\]

\[t_y = py + m \cdot (py - qy).\]

Here $m \in \mathbb{R}$ and $px, qx, py, qy$ are the x-coordinates and y-coordinates of the points $p, t$ and $q$, respectively. This new set of correspondences is used to estimate the camera motion between the channel pairs. Thus, a linear camera movement between $C_B, C_G$ is assumed. As the camera movement between two frames is very small, this approximation is well-justified.

## 4 Removal of Specular Highlights Artefacts

### Detection of Specular Highlights

In the first step of our approach, we use a global threshold $T_1$ on the green channel ($C_G$), the blue channel ($C_B$) and a grey scale image representation obtained from RGB through $E = 0.2989 \cdot C_R + 0.5870 \cdot C_G + 0.1140 \cdot C_B$. Any given pixel, say $p$, is marked as a possible specular highlight when the following condition is met: $C_G(p) > T_1$ $\lor$ $C_B(p) > T_1$ $\lor$ $E(p) > T_1$. The red channel is not used in this first step, because intense reddish colours are very common in colonoscopic videos and therefore saturated red occurs due to reasons other than being specular highlights. The second step detects pixels that can be described as positive outliers when compared to the surrounding tissue surface. Other than segmenting the image into tissue regions of similar colour (as, e.g., in [14]) we compute this neighbourhood for each pixel using a modified median filter approach. Using a square kernel of size $w \times w$, with $w$ on the order of 1/5 of the image width, we compute a median filtered image. The possible specular highlight regions that were detected in the first step enter the median filter with the centroid of the colours of a region of pixels adjacent to the specular region’s outline. This prevents the filtered image to appear too bright in regions where specular highlights cover a large area. Smaller specular highlights are effectively removed due to the relatively large window size. The pixel value at

\[\hat{p} = \left(\hat{x}, \hat{y}\right),\] where $(x, y)$ and $(\hat{x}, \hat{y})$ are the coordinates of $p$ and $\hat{p}$. As $H$ has eight independent elements [17], it can be computed using four point correspondences. However, the use of more points is beneficial, because in practice there will be noise and other inconsistencies in the feature point correspondences. Hence, in practice, the homography is computed through minimisation of the so called back-projection error using a set of point correspondences:

\[\Phi_H = \sum_i \left(\hat{x}_i - \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}}\right)^2 + \sum_i \left(\hat{y}_i - \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}}\right)^2.\] (1)

Here $h_{mn}$ is the $m^{th}$ row and $n^{th}$ column entry of the homography matrix $H$ and $(x_i, y_i), (\hat{x}_i, \hat{y}_i)$ are the coordinates of the $i^{th}$ feature point in two views respectively. As not all of the point pairs may fit the transformation, the estimation of the correct transformation is done using robust statistical methods, e.g., Random Sample Consensus (RANSAC) [18].

The computed homography is then used to map the channels by transforming the pixels. By merging the transformed colour component images to a colour frame, the camera shift between the channels is compensated. To achieve the most accurate estimation and higher computational efficiency, correspondences are computed between one channel pair only, viz., the pair with the highest intensity distribution similarity. Kullback-Leibler divergence ($d_{KL}(\cdot, \cdot)$) is used as the dissimilarity metric. For two discrete probability mass functions ($pmf$) $P$ and $Q$, it is defined as: $d_{KL}(P,Q) = \sum_{k=1}^{n} P(k) \log \frac{P(k)}{Q(k)}$. In our application the pixel intensity distribution is used to compute the dissimilarity between the colour channels. The $pmf$ of the intensity distributions of two colour channels are estimated using normalised histograms. $d_{KL}(\cdot, \cdot)$ is computed between all three colour channels and the pair with minimum Kullback-Leibler divergence is chosen for direct computation of the homography. As it is necessary to estimate camera motion between two channel pairs (see Figure 2), the second homography is computed by using the first one. More precisely, a set of correspondences for the second channel pair is created based on the correspondences of the first pair. Let $(p, q)$ be a matched point pair for $C_B, C_G$ then a pair $(s, t)$ is created for $C_B, C_R$ with $s = p$ and

\[t_x = px + m \cdot (px - qx)\]

\[t_y = py + m \cdot (py - qy).\] (2)

Here $m \in \mathbb{R}$ and $px, qx, py, qy$ are the x-coordinates and y-coordinates of the points $p, t$ and $q$, respectively. This new set of correspondences is used to estimate the camera motion between the channel pairs. Thus, a linear camera movement between $C_B, C_G$ is assumed. As the camera movement between two frames is very small, this approximation is well-justified.
Figure 6: Overview of the specular highlight detection algorithm.

Each position in the resulting image is therefore an estimate of the colour of the surrounding tissue surface.

Specular highlights are then found as positive colour outliers by comparing the pixel values in the input and the median filtered image. For such comparison, several distance measures and ratios are possible. During evaluation we found that the maximal ratio of the three colour channel intensities in the original image and the median filtered image produces optimal results. For each pixel location $p$, this ratio $\epsilon_{\text{max}}$ is computed as

$$\epsilon_{\text{max}}(p) = \sup \left\{ \frac{C_R(p)}{C_R^*(p)}, \frac{C_C(p)}{C_C^*(p)}, \frac{C_B(p)}{C_B^*(p)} \right\},$$

(3)

with $C_R^*(p), C_C^*(p)$ and $C_B^*(p)$ being the intensities of the red, green and blue colour channel in the median filtered image, respectively. Using a second threshold $T_2$, the pixel at location $p$ is then classified as a specular highlight pixel, if $\epsilon_{\text{max}}(p) > T_2$.

During initial tests we noticed that some bright regions in the image are mistaken for specular highlights by the algorithm presented so far. In particular, the mucosal surface in the close vicinity of the camera can appear saturated without showing specular reflection and may therefore be picked up by the detection algorithm. To address this problem, we made use of the property, that the outline of specular highlights generally consists of strong edges. Therefore, we test the area around the outline for its gradient to detect false positives by computing the mean of the gradient magnitude in a stripe-like area around the detected outlines of the specular regions. Figure 5 illustrates this idea. A morphological dilation and erosion is separately applied to a binary mask describing the areas of specular highlights, yielding two different output masks $S_d$ and $S_e$. Logical exclusive disjunction of $S_d$ and $S_e$ results in the desired outline areas.

Figure 7: The results of the proposed specular highlights detection algorithm. In the left and the right columns, the original and the resulting images after specular highlights being removed are shown respectively.

Following this, only those specular regions are retained, whose corresponding outline regions meet the condition

$$\frac{1}{N} \sum_{n=1}^{N} |\text{grad}(E_n)| > T_3,$$

(4)

with $|\text{grad}(E_n)|$ being the grey scale gradient magnitude of the $n$-th out of $N$ pixels of outline area corresponding to a given possible specular region. The gradient is approximated by vertical and horizontal differences of directly neighbouring pixels. Using this approach, bright, non-specular regions such as the one in the bottom left corner of Figure 5(a), can be identified as false detections.

Figure 6 shows a condensed overview of the presented detection algorithm.

**Inpainting of Specular Highlights:** For an efficient inpainting, the binary mask marking the specular regions in the image is first converted to a smooth weighting mask. This mask is then used to fill the specular regions with the content from the median filtered image obtained during the detection phase. More specifically, the smoothing of the binary mask is performed by adding a logistic decay to the outlines of the specular regions. The resulting weighting mask $M(p)$ is used to blend between the original image $I(p)$ and the median filtered image $I_{\text{med}}(p)$. The inpainted image $I_{\text{inp}}$ is computed for all pixel locations $p$ using

$$I_{\text{inp}}(p) = M(p) \cdot I_{\text{med}}(p) + (1 - M(p)) \cdot I(p)$$

with $M(p) \in [0, 1]$ for all pixel locations $p$. 

360
Figure 8: The results of colour channel realignment algorithm in Datasets 1 (first row) and 2 (second row). Left Column: the original images. Right Column: the resulting images after the colour channel misalignment artefacts are removed.

Figure 9: The results of colour channel realignment algorithm in Datasets 3 (first row) and 4 (second row). Left Column: the original images. Right Column: the resulting images after the colour channel misalignment artefacts are removed.

Figure 7 shows examples of images before and after inpainting. It can be seen that the inpainting method produces only minor artefacts for small specular highlights.

5 Experiments, Results and Discussion

5.1 Colour Channel Misalignment Artefact Removal

As described above the most important aspect of the proposed algorithm is to find efficient and robust point correspondences between the colour channels of a given endoscopy video frame in order to compute the homography. The correspondences can be found using two different paradigms: (a) feature point matching and (b) feature point tracking. In both paradigms the feature points need to be detected first. In order to find the optimal feature detector and descriptor we chose 50 colonoscopy video frames and induced artificial colour misalignment, cf. Figure 8. We exhaustively evaluated the performance of the following feature detectors, descriptors and trackers in finding point correspondences between colour channels [19]: (a) Feature Detectors: Harris Corner Detector, Hessian Detector, Hessian-Laplace, Harris-Laplace, Fast-Hessian, Difference of Gaussian, Maximally Stable Extremal Regions, Features from Accelerated Segment Test, Center Surround Extrema (CenSurE). (b) Feature Descriptors: SIFT, SURF, SPIN Image, Shape Context. (c) Tracker: Lucas-Kanade method. The combination of the Harris corner detector and the SIFT descriptor was found to yield optimal results. All the experiments, mentioned below, are done using this combination. For brevity a detailed discussion on the choice of the feature detectors and descriptors and original references corresponding to the aforementioned methods are not included in this paper. However, more details can be found in [19]. The parameter $m$ in (2) was experimentally determined to be 0.5 and kept constant throughout the evaluation process.

The performance of the proposed algorithm is evaluated using the following four test datasets: Dataset 1: 50 colonoscopy video frames with artificially induced colour channel misalignment, cf. Figure 8. Dataset 2: 50 colonoscopy video frames with artificially induced colour channel misalignment where specular highlights were removed using the proposed algorithm, cf. Figure 8. Dataset 3: 50 colonoscopy video frames with natural colour channel misalignment, cf. Figure 9. Dataset 4: 50 colonoscopy video frames with natural colour channel misalignment, where specular highlights were removed by the proposed algorithm, cf. Figure 9.

The performance of the proposed algorithm is evaluated using the following measures: (a) percentage of images where colour channels were successfully realigned (SR), (b) percentage of images where colour channels were not successfully realigned but they were not distorted either (USRND), (c) percentage of images where colour channels were not successfully realigned moreover they were also distorted (USRD). Successful realignment and distortion of the images were evaluated using visual inspection. Moreover, the effect of the algorithm on the video frames where no colour channel misalignment is present, was tested using 5 images, cf. Figure 10.

The quantitative performance of the proposed algorithm is depicted in Table 1. As expected, the performance of the algorithm is better in images where the colour channels were misaligned artificially. However, in the real images the performance was also found to be very promising.
Moreover, it was evident that when the specular highlights are removed by the proposed algorithm, the performance of the colour channel misalignment artefact removal is greatly enhanced. We found that the positions of the specular highlights remain unchanged between colour channels even if they are misaligned. This phenomenon can be empirically explained by the time taken for saturation and desaturation of the sensor. This invariance of specular highlight positions greatly deteriorates the feature matching process, as it is highly likely, due to the presence of strong edges and corners, that a considerable number of points from these regions are detected as features. This performance enhancement emphasizes the usefulness of the proposed specular highlight removal algorithm. In Figures 8, 9, 10 the performance of the proposed colour channel misalignment artefact removal algorithm is illustrated. In Figure 10 we see that when colour channels are not misaligned the proposed algorithm does not (visibly) distort the image.

5.2 Specular Highlights Artefact Removal

To evaluate the performance of the presented algorithm, we follow an image patch based approach. The images were divided into smaller patches and the algorithm had to assess whether these patches contained specular highlights or not. Ground truth data was obtained by dividing the images into 9 non-overlapping, rectangular image patches, which were then manually labelled.

The parameters of the algorithm were optimised using a training set of image patches obtained from 28 images of one colonoscopy video. The optimised algorithm was then tested on a different set of image patches obtained from 68 images from 2 different colonoscopy videos.

The training set contained 184 patches with specular highlights and 68 without, and the test set contained 446 patches with specular highlights and 166 without. The thresholds \( T_1 = 200, T_2 = 1.87 \) and \( T_3 = 2.72 \) were found to be optimal for the training set after a direct search optimisation initialised by a large range of parameter combinations. We measured the performance of our algorithm using the following commonly used measures: accuracy, precision, sensitivity and specificity [20]. The results of the test are shown in Table 2. The method achieved an overall accuracy of 91.99%.

![Figure 10: The effect of the proposed colour channel realignment algorithm when no misalignment is present. Left Column: original image. Right Column: The result of the application of the colour realignment algorithm. The original does not get distorted when there is no misalignment.](image)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Performance [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>91.99</td>
</tr>
<tr>
<td>Precision</td>
<td>96.49</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>92.38</td>
</tr>
<tr>
<td>Specificity</td>
<td>90.96</td>
</tr>
</tbody>
</table>

6 Conclusion

We have proposed two novel algorithms to remove colour channel misalignment and specular highlight artefacts in endoscopy videos. The algorithms were extensively evaluated using colonoscopy video frames and the performance was found to be promising. The colour channel realignment algorithm was evaluated more rigorously than the state of the art. However, the proposed algorithm can be further improved by the use of a more objective and quantitative evaluation criterion. Furthermore, the edges of the images get distorted due to the realignment procedure. Development of an adaptive inpainting algorithm to remove these distortions will be pursued in future. The specular highlight detection algorithm involves less parameters than Oh’s approach [14] and does not require an initial segmentation. Furthermore, in contrast to other approaches, our detection method is applicable to the widely used sequential RGB image acquisition systems. One shortcoming of the algorithm is that large specular regions appear strongly blurred, which is an obvious consequence of using the median filtering approach. For more visually pleasing results for very large specular areas, it would be necessary to use

<table>
<thead>
<tr>
<th>Dataset</th>
<th>SR (%)</th>
<th>USRND (%)</th>
<th>USRD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset 1</td>
<td>86</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>92</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Dataset 3</td>
<td>78</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Dataset 4</td>
<td>84</td>
<td>14</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1: Performance of the proposed colour channel misalignment artefact removal algorithm. **SR:** percentage of images where the colour channels were successfully realigned. **USRND:** percentage of images where the colour channels were not successfully realigned, however they were not distorted. **USRD:** percentage of images where the colour channels were not successfully realigned and they were also distorted. **Dataset 1:** 50 colonoscopy video frames with artificially induced colour channel misalignment. **Dataset 2:** Dataset 1 with specular highlights removed using the proposed algorithm. **Dataset 3:** 50 colonoscopy video frame with natural colour channel misalignment. (iv) **Dataset 4:** Dataset 3 with specular highlights are removed by the proposed algorithm.
additional features of the surrounding regions, such as texture or visible contours. Though such large specular regions are rare in clear colonoscopic images and errors arising from them can therefore usually be neglected, yet this issue will be dealt with in our future studies. Future work will also include a more rigorous evaluation of the presented methods by comparing their performance directly with implementations of other relevant methods.

Acknowledgment

This work has been supported by the Enterprise Ireland Endoview project CFTD-2008-204 under the National Development Plan.

References


