

Uninformative Frame Detection in Colonoscopy Through Motion, Edge and Color Features

Mohammad Ali Armin^{1,2(✉)}, Girija Chetty¹, Fripp Jurgen²,
Hans De Visser², Cedric Dumas², Amir Fazlollahi²,
Florian Grimpen³, and Olivier Salvado²

¹ Department of Computer Science,
University of Canberra, Canberra, Australia
m.a.armin@gmail.com

² CSIRO Biomedical Informatics,
The Australian e-Health Research Centre Herston, Brisbane, Australia
Olivier.Salvado@csiro.au

³ Department of Gastroenterology and Hepatology,
Royal Brisbane and Women's Hospital, Herston, Australia

Abstract. Colonoscopy is performed by using a long endoscope inserted in the colon of patients to inspect the internal mucosa. During the intervention, clinicians observe the colon under bright light to diagnose pathology and guide intervention. We are developing a computer aided system to facilitate navigation and diagnosis. One essential step is to estimate the camera pose relative to the colon from video frames. However, within every colonoscopy video is a large number of frames that provide no structural information (e.g. blurry or out of focus frames or those close to the colon wall). This hampers our camera pose estimation algorithm. To distinguish uninformative frames from informative ones, we investigated several features computed from each frame: corner and edge features matched with the previous frame, the percentage of edge pixels, and the mean and standard deviation of intensity in hue-saturation-value color space. A Random Forest classifier was used for classification. The method was validated on four colonoscopy videos that were manually classified. The resulting classification had a sensitivity of 75 % and specificity of 97 % for detecting uninformative frames. The proposed features not only compared favorably to existing techniques for detecting uninformative frames, but they also can be utilized for the camera navigation purpose.

Keywords: Optical colonoscopy · Uninformative frames · Colonoscopy quality · Feature · Random Forest

1 Introduction

Colorectal cancer is the second leading cause of cancer related death after lung cancer in Australia, however detection and removal of polyps in early stages can increase the chance of survival by up to 90 % [1]. Optical colonoscopy is the gold standard in

inspecting and removing polyps. Each year 500,000 colonoscopies are performed in Australia [1]. One of the main issues for clinicians is to estimate the position of the endoscope inside the colon, and software solutions to help with navigation would be desirable [2–5]. As a whole, we aim to provide a technology to estimate camera pose during colonoscopy. However, colonoscopy video streams contain many frames with no or little clinical information such as the result of colon cleansing, a dirty lens, or close inspection of colon wall. In Fig. 1 some examples of colonoscopy frames are illustrated. We categorized colonoscopy frames as informative or uninformative. The informative frames include clear shot of the lumen Fig. 1(a–b) or wall (Fig. 1(c–d)). Uninformative frames are a result of blurriness (blurred), colon cleansing with water jet (water), lens contact to the colon wall with a various illumination (indistinct) or indistinct with big bubbles or a bubbles' colony that reduce clinical information in a frame (Fig. 1(e–h)). The uninformative frames decrease the quality of colon inspection by clinicians and may hamper our camera motion estimation algorithm. Some studies showed that uninformative frames can compromise up to 30–40 % of the entire video stream [6, 7]. Therefore, it is important to detect uninformative frames and remove them.

In recent years, several studies have reported automated identification of uninformative frames from endoscopy videos [6, 8, 9]. Oh et al. [6, 10] developed a method which is based on analyzing the gray level co-occurrence matrix (GLCM) texture of the discrete Fourier transform images and edge detection. Following that, Arnold et al. [8] proposed a Bayesian classification method to analyze the norm of the detail coefficients of wavelet decomposition to classify colonoscopy frames. They reported 92.3 % accuracy only in detecting indistinct frames similar to Fig. 1(g). Color features have also been used in Wireless Capsule Endoscopy (WCE) to identify useful frames prior to diagnosis [9, 11].

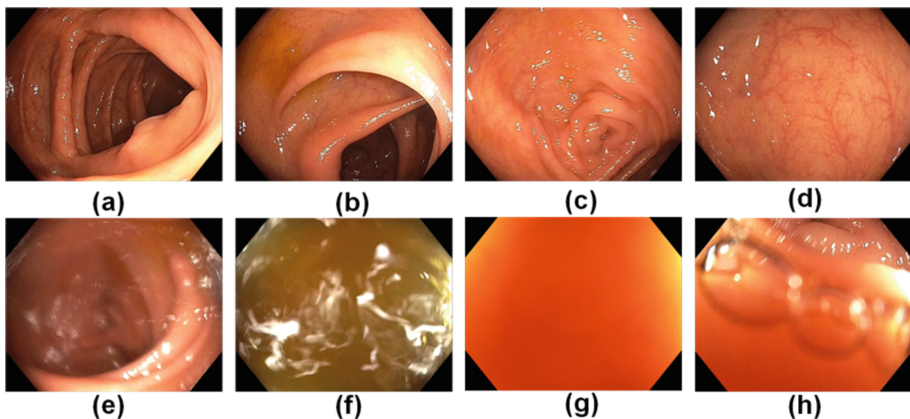


Fig. 1. First row is an example of informative frames, lumen view (a and b), wall view (c–d), and second row represent uninformative frames: blurred (e), water (f), indistinct (g), indistinct with bubble (h).

2 Method

The outline of the proposed method to classify colonoscopy frames is shown in Fig. 2. In this study, we investigate several features and use a Random Forest (RF) [12] classifier to detect uninformative frames. As the first step, all image frames were converted to the Hue-Saturation-Value (HSV) color space and smoothed using a Gaussian filter (By applying a Gaussian filter we aim at removing the noise, as well as moving mildly blurred frames to the blurred category). Subsequently, three shape-feature descriptors were investigated based on the following assumptions: (i) Consecutive uninformative frames results in a lower number of features detected by motion flow. For this, we computed the number of features detected by the Kanade Lucas Tomasi (KLT) tracker [13]. (ii) Uninformative frames such as Fig. 1(f–h) appear with a uniform color distribution. Therefore, to further emphasize on the color aspect, HSV color space was considered for computing the mean and standard deviation (STD) as features. (iii) Those uninformative frames which are blurred or mildly blurred have fewer sharp edges than a typical good quality colonoscopy image; for this we computed the percentage of edge pixels. The motivation of using these features is to utilize features currently computed for camera motion estimation to classify colonoscopy frames. This can also reduce the complexity of uninformative frame detection.

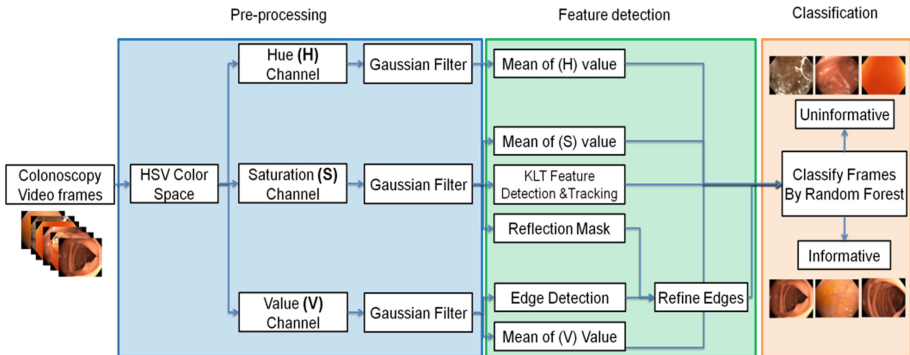


Fig. 2. The diagram of the proposed method for classification of colonoscopy frames.

2.1 Dataset

The data used for preparing this study were collected from four colonoscopy videos of different parts of the colon from different patients. Videos were captured by a 190HD Olympus colonoscope, with 50 frame/sec with a frame size of 1856×1044 pixels. A medical expert manually marked videos for uninformative frames. The details of our experimental videos are shown in Table 1.

Table 1. Dataset used in our experiment to detect uninformative frames

Dataset	Uninformative frames	Informative frames	Total frames	Informative and uninformative sequences
Patient 1	1205	1295	2500	2×30
Patient 2	112	1888	2000	2×14
Patient 3	201	1498	1699	2×11
Patient 4	702	2368	3070	2×40
Total	2220	7049	9269	190

2.2 Feature Detection

Number-of-features Descriptor Computed by KLT from Saturation Channel. The saturation color channel of HSV was used to extract and track features by the KLT method. This channel was used because our camera estimation parameters empirically obtained a better performance in feature detection. The KLT method detects corner like features with high contrast by measuring the minimum eigenvalue of each 2×2 gradient matrix in a frame. The displacement of selected features between consecutive frames was estimated by using an optimizer to minimize the difference between two feature windows for the image intensity. To address the large displacements, a pyramid based approach was used to track features.

Based on our assumption, frames with low numbers of features should have inadequate information to be used for camera motion estimation, and should be classified as uninformative frame. The number-of-features detected on a set of informative and uninformative frames are shown in Fig. 3.

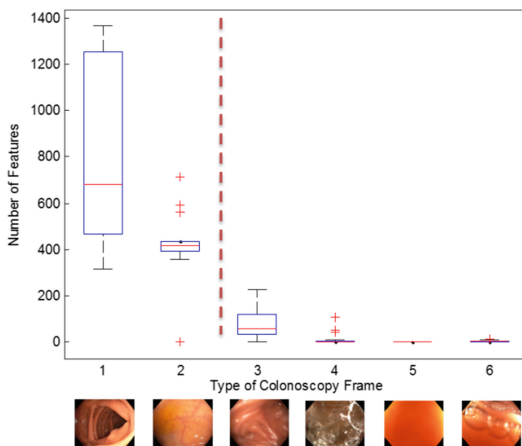


Fig. 3. Number-of-feature detected by the Kanade Lucas Tomasi (KLT) for several informative (1–2) and uninformative (3–6) frames using the Saturation color space.

Color Features from H-, S- and V-channel. Colonoscopy images are commonly presented in RGB color space. However, the HSV color space has shown a better ability in dichotomizing chromaticity (hue and saturation) from luminance [11]. Frames with no information such as the ones captured from a close inspection of the colon wall (Fig. 1(g–h)) or during colon cleansing have distinct signal from informative frames. Such distinction can be estimated by computing the STD and mean of the three HSV channels. The STD of hue, saturation and value for a set of informative and uninformative frames are presented in Fig. 4.

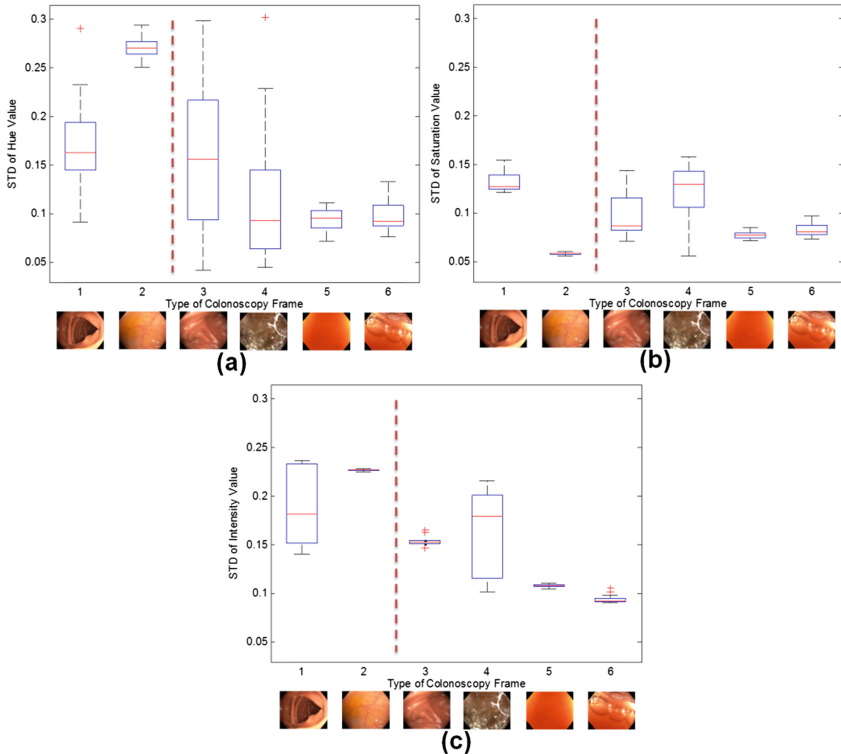


Fig. 4. The STD of Hue (a), Saturation (b), and Value (b) for several informative (1–2) and uninformative (3–6) frames.

Percentage-of-edge-pixel Feature Estimated from Value Channel. To detect uninformative frames, we analyzed the percentage of the edge pixels as the number of the edge pixels to all pixels in a frame. The edges were detected by using the Canny edge detector [14] from the Value channel. The percentage of isolated pixels introduced by Oh et al. [6] was also estimated for comparison.

Based on our experiments, frames with a higher percentage of edge pixels were informative whereas uninformative frames (including blurred, mild blurred and indistinct) had a lower percentage. Reflections can increase the number of edges,

especially when there are bubbles or a water jet for cleaning the colon. The reflection effect was removed by generating a mask using an automatic Otsu thresholding from a frame in the Saturation channel. The percentage of edge pixels on a set of informative and uninformative frames is shown in Fig. 5.

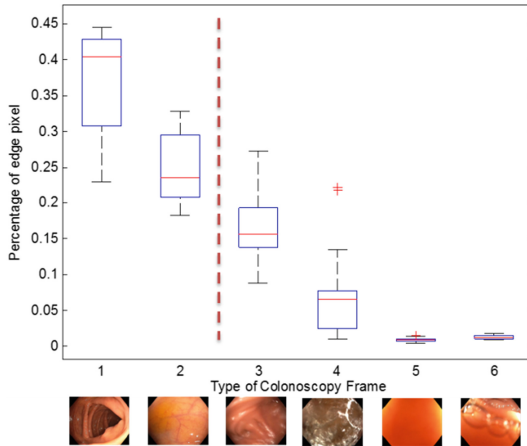


Fig. 5. Percentages-of-edge-pixel to all pixels for different types of colonoscopy frames (1 and 2 represent informative and 3 to 6 represent uninformative frames) using the Value channel and Canny edge detector.

2.3 Random Forest Classification

To classify frames into informative and uninformative classes, a binary Random Forest (RF) classifier [12] was used. On all available frames, feature metrics, including number of motion features, mean and STD of each color channel, and percentage-of-edge-pixel were calculated. In a colonoscopy video, consecutive frames may provide similar information which reduces the efficiency of the RF classifier if selected together. Therefore, prior to classification, all the informative and uninformative sequences were divided into half. We used the first half for training and second half for testing. The parameters used for RF training were: 100 trees, sample selection without replacement, and a node size of maximum 2.

2.4 Experimental Evaluation

To evaluate the performance of the proposed detection technique, sensitivity, precision, specificity and accuracy were considered. To compare the effectiveness of the proposed feature descriptors with similar studies, the gray level co-occurrence matrix (GLCM) and percentage-of-isolated-pixel (IPR) [6] were also included.

3 Results

Two representative examples of the KLT, edge and color features computed on informative and uninformative frames are illustrated in Figs. 6 and 7. A high number of motion vectors and edge pixels were identified for informative frames which demonstrate the potentials of the proposed features.

The performance of the above mentioned features in detecting uninformative frames using RF classifier is shown in Table 2. The collective performance of the proposed features, with accuracy of 94 % and specificity of 97 %, compares favorably to GLCM + IPR features, with accuracy of 92 % and specificity of 96 %. The calculation

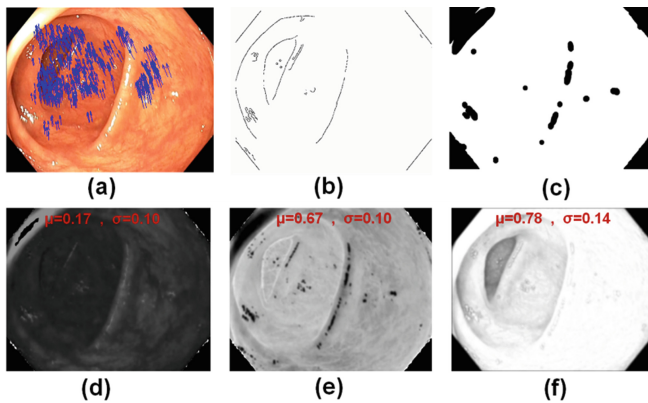


Fig. 6. The proposed motion (a) and edge features (b) computed on a representative informative frame along with the reflection mask (c) and three HSV channels (d–f). The mean (μ) and STD (σ) features are also shown on each color space.

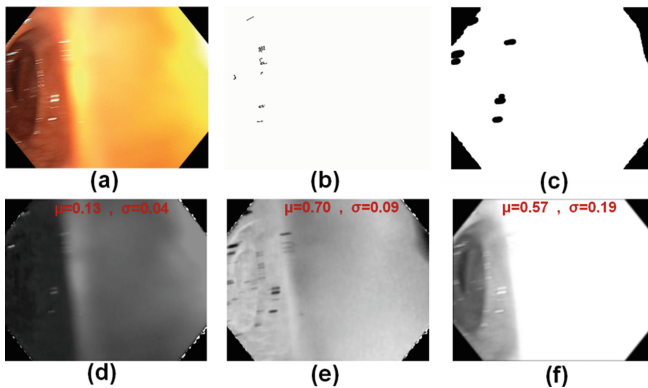


Fig. 7. The proposed motion (a) and edge features (b) computed on a representative uninformative frame along with the reflection mask (c) and three HSV channels (d–f). The mean (μ) and STD (σ) features are also shown on each color space.

Table 2. The performance of different feature descriptors on identifying uninformative frames

Feature	Classification results			
	Precision	Sensitivity	Specificity	Accuracy
KLT	0.83	0.72	0.97	0.93
STD color	0.62	0.36	0.96	0.85
Mean color	0.76	0.55	0.96	0.90
Percentage of edge pixels	0.75	0.51	0.97	0.89
All proposed features	0.86	0.75	0.97	0.94
GLCM	0.76	0.64	0.96	0.91
IPR	0.49	0.32	0.90	0.75
GLCM + IPR	0.79	0.67	0.96	0.92

time on average for KLT features computation was 0.16 s/frame whereas 0.072 s/frame was spent for GLCM feature calculation by using a standard PC, MATLAB, and non-optimized scripts.

4 Discussion

This paper proposes a method based on the KLT motion, color and edge features for detecting uninformative frames as the initial stage of our main pipeline for camera motion estimation algorithm. The proposed features were evaluated using a binary Random Forest classifier and obtained 86 % precision, 75 % sensitivity, 97 % specificity, and 94 % accuracy.

In the present work, the KLT motion features were proposed as a metric for identifying uninformative frames. To increase the number of these features in each frame, the HSV color space was found more suitable. More importantly, motion features were already available for estimating camera motion in our algorithm which will reduce the computational complexity. Besides, there are some frames, e.g. wall view, with fewer textures compared to lumen which will result in less motion features. To identify these frames, color information in HSV color space was calculated.

Adding more features such as GLCM, IPR or wavelet as used in literature might improve our method in classifying more complicated colonoscopy frames. For instance, these features might be useful when color features show a partial overlapping between a subset of uninformative and informative frames. However, the aim of this study was to investigate the feasibility of using KLT features which were concurrently computed for camera pose estimation. Furthermore, this approach can be used in endoscopy videos such as bronchoscopy and wireless capsule endoscopy to remove uninformative frames during camera motion estimation.

The main limitation of the current study is the small dataset size, increasing the number of frames might slightly change the reported performance. In future work, we aim to validate our method on a bigger dataset acquired from different colonoscopes with different field of view and resolution. A diverse colonoscopy video datasets from different patients will allow us to validate the proposed features with other training

approaches for RF classifier such as one-video-leave-out approach and clustering based methods such as K-mean clustering. Furthermore, other feature descriptors such as scale invariant feature transform (SIFT) or speeded up robust features (SURF) will be investigated.

5 Conclusion

This study demonstrated that KLT motion, color and edge features can together provide effective detection of uninformative colonoscopy frames. The proposed method can be performed simultaneously with camera pose estimation. This would reduce the computational burden and necessity to compute other complex features for uninformative frame detection.

References

1. Australian Institute of Health and Welfare. <http://www.aihw.gov.au/>
2. Liu, J., Subramanian, K.R., Yoo, T.S.: A robust method to track colonoscopy videos with non-informative images. *Int. J. Comput. Assist. Radiol. Surg.* **8**, 575–592 (2013)
3. Puerto-Souza, G.A., Staranowicz, A.N., Bell, C.S., Valdastrì, P., Mariottini, G.-L.: A comparative study of ego-motion estimation algorithms for teleoperated robotic endoscopes. In: Luo, X., Reich, T., Mirota, D., Soper, T. (eds.) CARE 2014. LNCS, vol. 8899, pp. 64–76. Springer, Heidelberg (2014)
4. Mori, K., Deguchi, D., Sugiyama, J., Suenaga, Y., Toriwaki, J., Maurer, C.R., Takabatake, H., Natori, H.: Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. A preliminary version of this paper was presented at the Medical Image Computing and Computer-Assisted Intervention (MICCAI) Conference, Utrecht, The Netherlands (Mori et al. 2001). *Med. Image Anal.* **6**, 321–336 (2002)
5. Rai, L., Helferty, J.P., Higgins, W.E.: Combined video tracking and image-video registration for continuous bronchoscopic guidance. *Int. J. Comput. Assist. Radiol. Surg.* **3**, 315–329 (2008)
6. Oh, J., Hwang, S., Lee, J., Tavanapong, W., Wong, J., de Groen, P.C.: Informative frame classification for endoscopy video. *Med. Image Anal.* **11**, 110–127 (2007)
7. Oh, J., Hwang, S., Cao, Y., Tavanapong, W., Liu, D., Wong, J., de Groen, P.C.: Measuring objective quality of colonoscopy. *IEEE Trans. Biomed. Eng.* **56**, 2190–2196 (2009)
8. Arnold, M., Ghosh, A., Lacey, G., Patchett, S., Mulcahy, H.: Indistinct frame detection in colonoscopy videos. In: 13th International Machine Vision and Image Processing Conference (IMVIP), pp. 47–52 (2009)
9. Mackiewicz, M., Berens, J., Fisher, M.: Wireless capsule endoscopy color video segmentation. *IEEE Trans. Med. Imaging* **27**, 1769–1781 (2008)
10. Oh, J., Hwang, S., Tavanapong, W., de Groen, P.C., Wong, J.: Blurry-frame detection and shot segmentation in colonoscopy videos. In: Proceedings of SPIE, pp. 531–542 (2003)
11. Bashar, M.K., Mori, K., Suenaga, Y., Kitasaka, T., Mekada, Y.: Detecting informative frames from wireless capsule endoscopic video using color and texture features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 603–610. Springer, Heidelberg (2008)

12. Random Forest (Regression, Classification and Clustering) implementation for MATLAB. <https://code.google.com/p/randomforest-matlab/>
13. Shi, J., Carlo, T.: Good features to track. In: CVPR, pp.593–600 (1994)
14. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-8**, 679–698 (1986)