Endoscopy Video Frame Classification Using Edge-based Information Analysis

Nicharee Rangseekajee, Sukanya Phongsuphap

Faculty of Information and Communication Technology, Mahidol University, Bangkok, Thailand

Abstract

Endoscopic surgery has brought about massive changes in cardiac and thoracic surgery. The therapeutic capabilities of endoscopy provide beneficial diagnosis and treatment for physician. Although, the endoscopic surgery shows a marked improvement in diagnosis and treatment, the endoscopy video still has some frames that contain no useful information. We need to distinguish the informative frames from the non-informative frames in order to utilize information from both of them for further operations for developing Computer Aided Diagnosis systems. This paper proposes a method based on edge information analysis to distinguish informative frames and non-informative frames by utilizing the Isolated Pixel Ratio Value (IPRV). The accuracy of the proposed method was remarkably higher than that of the other edge-based technique. Results indicate that the proposed method is an effective method for informative and non-informative thoracic endoscopy frame classification.

1. Introduction

Cardiac and thoracic endoscopy is the surgical procedures in the heart and chest cavity utilizing small incisions [1]. Examples of cardiac and thoracic endoscopy are cardiopulmonary endoscopy and thoracoscopy. The cardiopulmonary endoscopy technique is used to examine the cardiac chambers and the pulmonary arterial system [2]. Thoracoscopy can be performed of pleural-based abnormalities for diagnosis of diseases such as metastatic cancer, fungal, mesothelioma, etc [3]. The endoscope can help physician to determine the interpleural path in each vein graft that is very beneficial procedure to help physician for cardiac and thoracic's diagnosis and treatment [1][2][3]. Moreover, the thorocoscopy takes an advantage of maintaining minimal rib retraction [3].

In spite of the usage of endoscopy which is an information technology tool has a valuable remedy for the quality improvement of diagnostic procedures and technology of endoscopy shows a marked improvement in diagnosis and treatment, endoscopy video still contains some frames that have no any useful information for diagnosis [4]. In this paper, we call such a frame that contains useless contents as non-informative frame, and the one that contains some useful information as informative frame. We need to distinguish the informative frames from the non-informative frames in order to utilize information from both of them for further operations for developing automatic or semi-automatic Computer Aided Diagnosis systems. This classification also is benefit in reducing the number of images to be viewed by a physician. To distinguish informative frames and non-informative frames, we propose a method based on edge information analysis. The proposed method consists of three main processes: Preprocessing, Edge-based region description, and Frame classification.

The remainder of this paper is organized as follows. Firstly, the proposed technique is described in Section 2. Section 2.1 provides the technique of adaptive thresholding. Section 2.2 and Section 2.3 explain the techniques of edge-based information analysis, and connected-component labeling respectively. Section 2.4 describes the steps of frame classification. Section 3 shows the experimental results. Section 3.1 provides the data sets used in our experiments. Section 3.2 shows the results. Section 4 is the discussion and Section 5 is the conclusions.

2. Methodology

In this section, we explain the proposed method which consists of three main processes:

1) Pre-processing
2) Edge-based Information Analysis, and
3) Frame Classification.

The framework of our proposed method is shown in Figure 1. In the first step, the thoracoscopic frames are applied by adaptive thresholding technique to eliminate lighting effects, and then edge operators are performed on each frame to detect edge information. The last step is to classify frames into either informative frame or non-informative frame. Informative frames will be kept to be the useful information. In contrast, the non-informative ones will be removed from the video data to reduce time-consuming for physician's diagnosis and treatment.

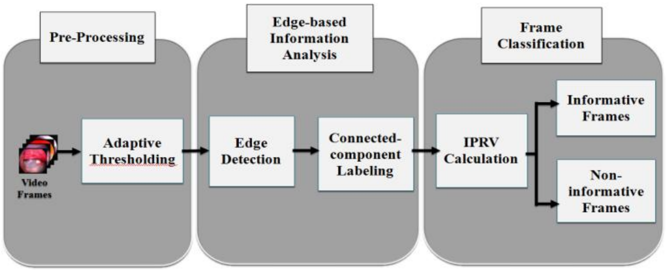


Figure 1. The framework of proposed method.

2.1. Pre-processing

In the pre-processing step, the lighting conditions are concerned because the light exposure can blur information in frames or there are some details covered by light reflections. For the thoracoscopic frames, the reflectivity causes disturbing the distinction of informative frame from non-informative one [5]. The accuracy and performance in frame classification step are reduced because of the reflectivity. To solve this problem, the adaptive thresholding technique is applied to each frame to reduce the effects of light. Moreover, the adaptive thresholding can be used to distinguish the ambiguity of image information in the thoracoscopic frames which is the main segmentation problem. The algorithm that is used in our method is Otsu's technique [6]. The advantage of this technique is to select a threshold automatically from a gray level histogram that has been derived from the discriminant analysis. The threshold value varies over the image by the discriminant criterion to maximize the separation in gray levels.

2.2. Edge-based information analysis

To distinguish the informative frames from the non-informative ones, we analyze the edge information based on a property of isolated pixels. We classify them into two types: edges and noises. We define edge as the connected edge pixel (CP), and noise as the isolated edge pixel. The edge detection operators that are used here are Canny, Prewitt, Laplacian and Sobel. Each frame will be tested with the four operators to compare performance. The Sobel and Prewitt operators are the first derivative edge detection operators. For the Laplacian operator, it is approximating the second derivative. The Canny algorithm represents the complicated process to detect edge information, and it can detect fine details [7].

2.3. Connected-component labeling

After we apply an edge operator to each frame, the edge information is detected. We construct the connected edge pixel by using two-dimensional connectivity: 8-connectivity component labeling. The pixel is said to have 8-connectivity when there is the adjacent pixel touches any of the 4 edges and 4 corners of pixel. We used the 8-connectivity to label connected pixel in vertical, horizontal, left and right diagonal directions, and it is labeled as connected edge pixel (CP). If any pixel has no neighborhood of pixel in every direction, it is labeled as isolated edge pixel (IP). The CP and IP are used to calculate the Isolated Edge Pixel Ratio Value (IPRV) as defined in Eq (1).

$$IPRV = \frac{IP}{IP+CP} \times 100\% \qquad (1)$$

where IP is Isolated Pixel Value, CP is Connected Edge Pixel Value and TP is Total Edge Pixel (TP = CP + IP).

2.4. Frame classification

To classify frames into two types: informative frame and non-informative frame, we define a specified threshold value based on the IPRV from a training data set. The steps of finding the optimal threshold value are described as follows:

(1) Define the initial threshold value T.
(2) Assign each frame from the training set into one of two categories:
Category 1: IPRV > T
Category 2: IPRV ≤ T
(3) Find the average of IPRV of each category, we get μ_1 and μ_2.
(4) Calculate the new threshold value T_n defined as follows:

$$T_n = \frac{(\mu_1+\mu_2)}{2} \qquad (2)$$

(5) Calculate the difference between T_n and T_{n-1} as follows:

$$D = T_n - T_{n-1} \qquad (3)$$

(6) If D is close to zero, then stop. Otherwise we set $T = T_n$, then repeats step (2) to (5) until D is close to zero.

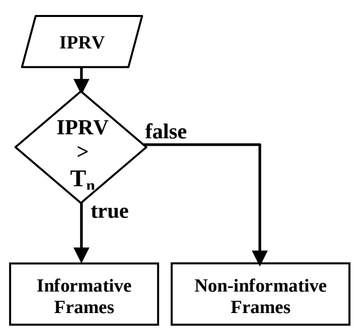


Figure 2. The diagram of frame classification step.

The specified threshold is used to classify frame. The diagram for classifying the frames is shown in Figure 2. From the diagram, the IPRV of each frame is checked with the specified threshold value. If the IPRV of the frame is greater than the specified threshold value, then the frame is classified as informative frame. In contrast, if the IPRV of a frame is less than or equal to the specified threshold value, then the frame is classified as non-informative frame.

3. Experimental results

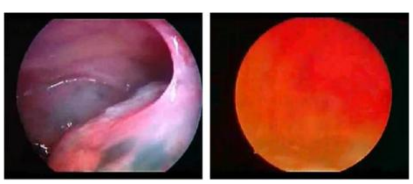
In this section, we explain about our data sets and results from the experiments. Firstly, the data sets are described in Section 3.1 then, the results are shown in Section 3.2.

3.1. Data sets

Our data consists of 387 thoracoscopic frames: 347 informative images and 40 non-informative images from two data sets. The data set 1 consists of 117 informative frames and 23 non-informative frames. The data set 2 consists of 230 informative frames and 17 non-informative frames. The size of a frame is 480×360 pixels. The details of data sets are shown in Table 1. Examples of data are shown in Figure 3.

Table 1. Data sets of thoracoscopic frames used in experiments.

Data Set	Informative	Non-informative	Total
1	117	23	140
2	230	17	247
Total	347	40	387



(a) (b)

Figure 3. (a) Informative frame, (b) Non-informative frame.

3.2. Results

We had evaluated performances of thoracic endoscopy video frame classification based on four measures: precision, sensitivity, specificity, and accuracy [8]. Our proposed method provided more than 90% for precision, specificity and accuracy. The results showed that Sobel operator gave the remarkable high performance for precision, specificity, and accuracy for both data sets. For Prewitt operator, it showed the high performance in precision, specificity, and accuracy for data set 2. For Laplacian and Canny operators, they showed high precision results. The details are shown in Table 2.

The experiment indicates that Sobel operator is the best operator for edge information detection in the thoracoscopic frames comparing with other edge detectors. We had compared performance of our method with that of Oh et al.'s method [5]. The results are shown in Table 3. We found that our proposed method provide the better performance than Oh et al.'s method for all measures.

Table 2. Comparison of four edge operators performance.

Edge operator	Data Set	Performance (%)			
		Precision	Sensitivity	Specificity	Accuracy
Sobel	1	95.7	95.7	78.3	92.9
	2	97.8	97.0	70.6	95.1
Prewitt	1	95.8	79.3	82.6	78.6
	2	97.8	94.8	70.6	93.1
Laplacian	1	85.7	66.7	43.5	62.9
	2	90.0	43.0	35.3	42.5
Canny	1	89.5	58.1	65.2	59.3
	2	94.3	71.3	41.2	69.2

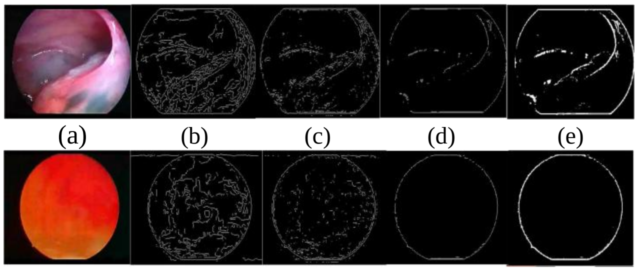
Table 3. Comparison of our proposed method and Oh et al.'s method.

Method	Data Set	Performance (%)			
		Precision	Sensitivity	Specificity	Accuracy
Proposed method	1	95.7	95.7	78.3	92.9
	2	97.8	97.0	70.6	95.1
Oh et al.'s method	1	84.9	53.0	52.2	52.9
	2	94.6	69.5	47.0	68.0

4. Discussions

This study demonstrates that the adaptive thresholding technique can result in improving the accuracy and performance in frame classification. The reflectivity and lighting conditions that cause misclassification between informative frames and non-informative frames are reduced. Comparing with Oh et al.'s method, they did not have the pre-processing step for reducing the effect of the lighting conditions.

After investing we found that the Sobel operator shows the better results in edge-based information analysis. The Sobel operator can highlight the edge strength of pixel information comparing with other operators. The useful pixel information in each frame is detected accurately. The results shown in Figure 4 present the thoracoscopic frames that are applied by four edge operators in which Figure 4(a) is an informative frame, and (b) to (e) are the edge information images generated from (a) by applying the Laplacian, Canny, Prewitt and Sobel operators respectively. Figure 4(f) is a non-informative frame, and (g) to (j) show the images generated from (f) by using the same edge operators as applying in Figure 4(b) to (e).



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Figure 4. (a) Informative frame, (b) - (e) edge detection from (a) by using Laplacian, Canny, Prewitt and Sobel operators respectively; (f) Non-informative frame, and (g) - (j) edge detection from (f) by using Laplacian, Canny, Prewitt and Sobel operators respectively.

Oh et al. presented the edge-based frame classification by using the Canny operator. When we applied their method with our data set, we found that the Canny operator shows the low performance in sensitivity, specificity and accuracy because the Canny operator detect fine details including noise information in the frames and those pixels are not related to the real information in the frames as shown in Figure 5(b). The image result that is detected by the Sobel operator shows no edge pixel information as shown in Figure 5(c). The non-informative frames that are applied the Sobel operator are correctly classified because the Sobel operator detects only the strong edge pixels.

(a) (b) (c)

Figure 5. (a) The non-informative frame, (b) The result of (a) applied by the Canny operator and (c) The result of (a) applied by the Sobel operator.

As we mentioned before, the reflectivity is a problem to detect incorrect edge pixel. This effect causes misclassification of the informative frames and the non-informative frames. Although, adaptive thresholding technique can reduce some effects of reflectivity, there are some frames that still are classified incorrectly, which need further investigations.

5. Concluding remarks

In conclusion, in this paper we have proposed a method for thoracoscopic frame classification. The method consists of three main steps consisting of pre-processing, edge-based information analysis and frame classification. In the pre-processing step, we apply adaptive thresholding technique to reduce the effect of lighting conditions for reducing frame misclassification. Edge-based information analysis step uses the Sobel operator to detect the pixel information in the frames. In the frame classification, we classify frames based on Isolated Pixel Ratio Value (IPRV). The last results are two types of frames: informative frame and non-informative frame. The non-informative frames that have no useful information will be removed from the thoracoscopy video to reduce the time consuming in physician's diagnosis and treatment.

Our proposed method can achieve high accuracy (over 92%) while also providing high precision (over 95%), sensitivity (over 95%), and specificity (over 70%). Our on-going work is to find the more effective method for eliminating reflection effect which is one of the major causes of misclassification.

References

[1] Szpala S, Guiraudon G, Peters T. Cardiac Endoscopy Enhanced by Dynamic Organ Modeling for Minimally-Invasive Surgery Guidance. Medical Image Computing and Computer Assisted Intervention 2003;2878:499-506.

[2] Ngaage DL, Shah R, Sanjay SP, Cale ARJ. Cardiopulmonary Endoscopy: An Effective and Low Risk Method of Examining the Cardiopulmonary System during Cardiac Surgery. European Journal of Cardio-thoracic Surgery 2001;19:152-155.

[3] Ohtsuka T, Kubota H, Motomura N, Takamoto S. Thoracoscopy for Minimally Invasive Axillo-Coronary Artery Bypass. European Journal of Cardio-Thoracic Surgery 2001;20:856-857.

[4] Jozwiak R, Przelaskowski A, Duplaga M. Diagnostically Useful Video Content Extraction for Integrated Computer-Aided Bronchoscopy Examination System. AISC 2009;57:151-158.

[5] Oh J, Hwang S, Lee J, Tavanapong W, Wong J, de Groen PC. Informative Frame Classification for Endoscopy Video. Medical Image Analysis 2007;11:110-127.

[6] Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. System, Man and Cybernetics 1979;9:62-66.

[7] Sonka M, Hlavac V, Boyle R. Image Processing, Analysis, and Machine Vision. PWS Pub., 1999.

[8] Baldock J, Graham J. Image Processing and Analysis: A practical approach. Oxford University Press, 2000.

Address for correspondence:

Sukanya Phongsuphap
Faculty of Information and Communication Technology
Mahidol University, 999 Phuttamonthon 4 road
Salaya, Nakhon Pathom 73170, Thailand
E-mail: itsps@mahidol.ac.th