21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Ulcer Detection in Wireless Capsule Endoscopy Images

Lecheng Yu
Sun Yat-sen University
Guangzhou, P. R. China
yulecheng@gmail.com

Pong C Yuen
Hong Kong Baptist University
Kowloon Tong, Hong Kong
pcyuen@comp.hkbu.edu.hk

Jianhuang Lai
Sun Yat-sen University
Guangzhou, P. R. China
stsljh@mail.sysu.edu.cn

Abstract

The invention of wireless capsule endoscopy greatly helps physician to view small intestine images without causing much pain to patients. It becomes very popular around the world for its usability and performance. However, physician requires a long time (around 45 minutes) to examine a capsule endoscopy video generated from each examination. In this paper, we propose a new image processing method using combination of local features for ulcer detection. The proposed method is developed based on bag-of-words model and feature fusion technique. Image patches are described by LBP and SIFT features. The pyramid bag-of-words is employed to model and represent the images, and SVM classifiers are trained. Finally feature fusion technique is employed to draw a final conclusion. Experimental results show that the proposed method outperforms single feature methods and existing methods.

1. Introduction

Millions of people all around the world suffer from digestive diseases, making a heavy loss to human health. The traditional endoscopy is still a main tool to diagnose digestive diseases. However conducting diagnosis in the small intestine is one of the major problems that endoscopy techniques have to be confronted in which it is hard to reach the small intestine. Besides, it brings uncomfortable to the patient and requires a skillful doctor to operate. Wireless capsule endoscopy (WCE) offers an alternative way to view the entire small intestine without causing any pain to the patient. It was invented by a group of researchers in 1989 and introduced by Given-Imaging since 2002 [1]. WCE is a pill-like small device compacted with a mini camera which takes four pictures per second. After the patient swallow the capsule, it is propelled by peristalsis and moves forward along digestive tract, taking pictures and sending them out. The whole process takes about eight hours. More than 100,000 images will be produced per examination. It requires much time and energy of doctors to analyze and make diagnosis. This drawback limits the world wide promotion of this newly emerging technique. So image processing techniques are proposed [1-6] to reduce the burden to medical doctors.

Abnormal regions in WCE images usually show more or less difference in appearance compared to its surroundings. Li and Meng [2] applied a patching scheme for ulcer detection. They divided WCE images into a number of small patches in order to deal with visibility problems and detect abnormalities more accurately. However, dividing images into small patches may isolate the abnormal regions from their surroundings. That may lose global image information, such as the global statistical features, the spatial relationship between local features. This motivates us to propose a new method which makes the advantage of the patch scheme as well as global spatial information. In this paper, we propose a methodology based on bag-of-words model and feature fusion technique for ulcer detection.

2. Bag-of-Words Model

In recent years, the bag-of-words (BOW) model has been successfully employed in computer vision to describe images with local features and have made impressive performance in natural scene categories [7] [8]. The BOW model of images consists of three major steps, namely (i) local region extraction and feature representation, (ii) vocabulary construction and (iii) image description based on constructed vocabulary. We follow the patching scheme and divide an image into a number of small patches and the block diagram is shown in Figure 1. Both LBP and SIFT features are computed for each patch. In the second step, k-means clustering is performed for all the patch features to partition these patches into k clusters which forming a vocabulary dictionary. Every patch belongs to the cluster with the nearest mean, that is represented by the word in the dictionary. Finally, bag-of-words histograms are obtained based on the vocabulary constructed in the second step. We employ SVM classifier for ulcer classification.

The traditional bag-of-words model has one major limitation. It disregards all information about the spatial layout of images and only focuses on the compositions of images. In [9], the authors used a spatial pyramid kernel to obtain spatial pyramid bag-of-words histograms to overcome this limitation. At the third step of generating bag-of-words model, we put a series of increasing coarser grid at resolutions 1, …, L over the image space, such that the grid at level l has 2^{l-1} cells along each dimension. For each cell, we obtain a cell based bag-of-words histogram, then concatenate all the cell histograms to form a pyramid bag-of-words histogram. The idea is shown in Figure 2. Then the distance between image X and image Y [9] is given by:

$$S^L(X,Y) = H^{L-1} + \sum_{l=0}^{L-2} \frac{1}{2^{L-1-l}} \left(H^l - H^{l+1} \right) \qquad (1)$$
$$= \frac{1}{2^{L-1}} H^0 + \sum_{l=1}^{L-1} \frac{1}{2^{L-l}} H^l$$

Here the H^l stands for the number of matches at level l, given by the histogram intersection function:

$$H^l(X,Y) = \sum_{i=1}^{D} \min\left(Hist_X^l(i), Hist_Y^l(i) \right) \qquad (2)$$

D is the number of histograms at level l and also includes all matches found at the finer level l+1. So match at level l is given by $H^l - H^{l+1}$. And match found at finer level should be more important, the formulation $\frac{1}{2^{L-1-l}}$ assigns the appropriate weight to matches found at each level.

3. Feature Fusion

Many computer vision and pattern recognition applications face the challenges of small inter-class variations and large intra-class variations. This is also the case in WCE images [10]. In this paper, we use Linear Classifier Dependency Modeling (LCDM) method proposed in [12] to conduct feature fusion. It is a classifier level fusion technique based on Bayesian theory, which explicitly model the dependency of classifiers to do feature fusion. LBP and SIFT operators are used here to extract features, then we build SVM classifiers using bag-of-words model. After that, LCDM is deployed to combine the different classifier results.

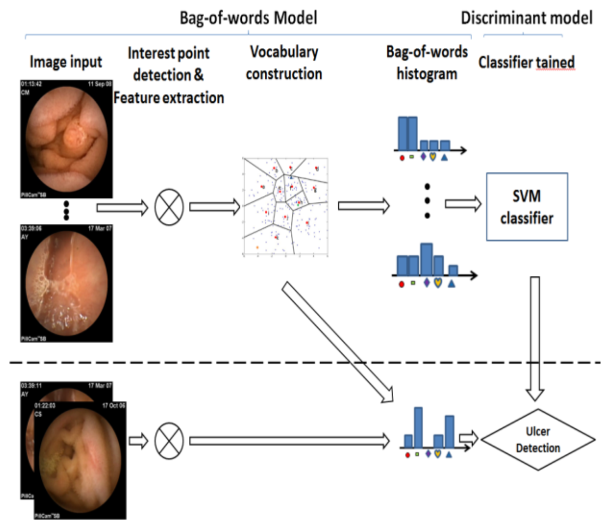

Fig. 1 Bag-of-words model for ulcer detection

4. Experimental Results

Experiments have been conducted to evaluate the performance of the proposed method and compare with existing methods. 344 endoscopic images were collected from 60 patients examination data for training, including 172 ulcer images and 172 normal images. For the testing, another 120 ulcer images and 120 normal images are selected. All the images are 576 x 576 pixels.

First, we evaluate the effectiveness of using traditional bag-of-words model for ulcer detection. We compare the performance with the existing method proposed in [2]. The experiment setting is the same as in [2]. We select 2256 patches from training image set as training patch set and 1080 patches from testing image set as testing patch set. Then we replace the curvelet LBP in the image patch feature extracting stage with LBP and SIFT, respectively to extract local features. SVM is trained for every single patch. We choose the rotation invariant uniform LBP to reduce the effects of illumination variations. Six statistical measurements of the feature histogram [13] are computing, namely standard deviation, skew, kurtosis, energy and mean. All experiments are performed under RGB space with each channel processed feature extraction independently.

Following the same criteria as in [2], accuracy, sensitivity and specificity are used. It can be observed from Table 1 that the proposed method outperforms the existing method. For LBP feature, the accuracy is improved by 11.76% while the accuracy is improved by 7.67% for SIFT feature.

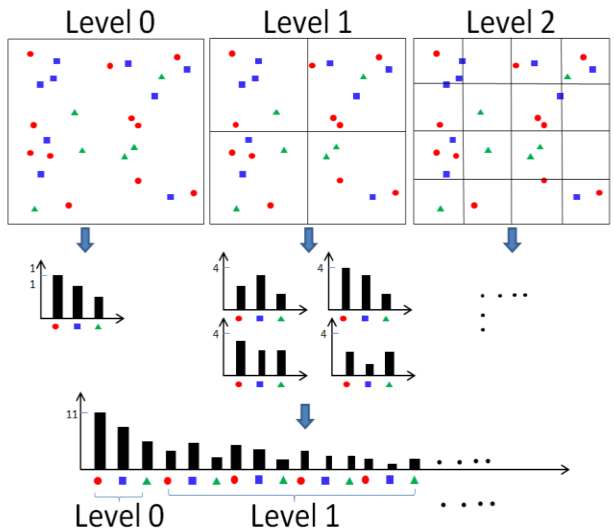

Fig. 2: Spatial pyramid bag-of-words histogram

Secondly, we evaluate three different ways in extracting local regions for constructing BOW model of images. (i) Evenly grid: Dividing an image into a number of small patches with the size 48 x 48. (ii) Harris affine region detector [14]: Detect corner points and the scales are the same as (i). (iii) LoG detector [15]: Extract blob regions of each image and the scales vary between 20 to 100 pixels. Also we use LBP and SIFT to extract features of detected regions.

Table 2 compares and contrasts the experimental results of the BOW model based on different local region detectors and representations. We can see from the Table 2 that evenly grid yields the best performance. So in the following experiments we use evenly grid method to extract local regions.

Next we will evaluate the effectiveness of using spatial pyramid kernel in the bag-of-words model for ulcer detection. For both LBP and SIFT features, we select the best three performance vocabulary size for the spatial pyramid kernel. Figures 3 and 4 show the accuracies of spatial pyramid kernel under different number of levels. Noted that when L=1 in the spatial pyramid kernel, it becomes a traditional bag-of-words model. So it can seen from both figures that using spatial pyramid kernel (L=2 and higher) improves the detection performance compared with traditional bag-of-words method (L=1) because the spatial pyramid kernel considers spatial information of patches in images. The best result often occurs at L=3, when the number of levels increased, the performance begin to deteriorate.

Table 1: Average classification results (%)

		Accuracy	Sensitivity	Specificity
LBP	PATCH [2]	71.76	64.23	79.51
	BOW	83.52	88.88	78.17
SIFT	PATCH [2]	66.48	64.42	68.61
	BOW	74.15	75.04	73.25

Table 2: Average classification results (%)

	Grid	Harris	LoG
LBP	83.52	75.42	76.10
SIFT	74.15	70.50	63.86

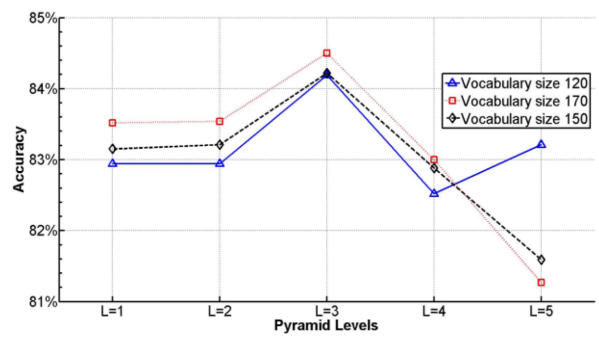

Fig. 3: Accuracy of spatial pyramid kernels (LBP)

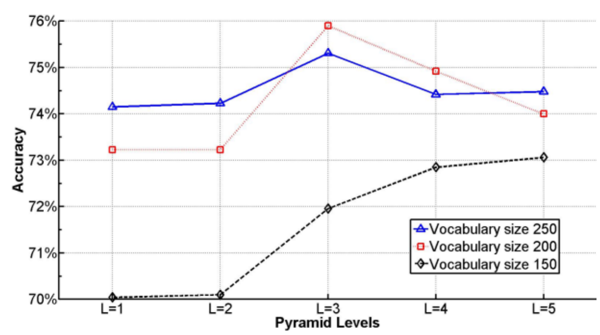

Fig. 4: Accuracy of spatial pyramid kernels (SIFT)

The last experiment is the feature fusion experiment. For both features, we select the best classifier on three vocabulary sizes used in the second experiment. Table 3 and Table 4 show the classifiers used in feature fusion and the fusion results. For LBP features in Table 2, the combined detection accuracy is 89.58% which is 2.91% higher than that of the single best classifier. Similar result can be seen in Table 4 for SIFT. Then we further combine all these three LBP and three SIFT features used in Tables 3 and 4. The ROC curves are shown in Figure 5.

Tab. 3 Feature fusion result of LBP (%)

	Accuracy	Sensitivity	Specificity
Size=120	86.67	92.50	80.83
Size=150	85.00	93.33	76.67
Size=170	85.83	91.67	80.00
Feature fusion	89.58	99.17	80.00

Tab. 4 Feature fusion result of SIFT (%)

	Accuracy	Sensitivity	Specificity
Size=150	72.08	79.17	65.00
Size=200	75.83	78.33	73.33
Size=250	77.50	80.00	75.00
Feature fusion	80.00	77.50	82.50

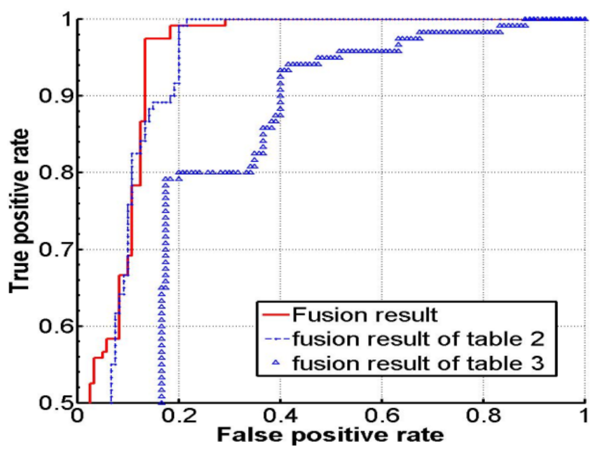

Fig. 5: ROC curves of feature fusion

From Figure 5, we can see that the detection accuracy of combining all six classifiers gives the best result. It means that combining more features can improve the detection accuracy.

5. Conclusions

In this paper, we have proposed to make use of multiple features for ulcer detection of WCE images. Based on the bag-of-words model, a new method is developed using spatial pyramid kernel and feature fusion technique for ulcer detection. Experimental results show that the method achieves promising results.

Acknowledgement

This project is supported by the Faculty Research Grant of Hong Kong Baptist University. The authors would like to thank Given Imaging Ltd for providing the endoscopic images and Dr. Grace Wong of Prince of Wales Hospital for her help in this preliminary study.

References

1. O. Lin, R. A. Kozarek, D. Schembre, et al. Blinded comparison of esophageal capsule endoscopy (ECE) versus conventional esophagogastroduodenoscopy (EGD) for identification of esophagitis and Barrett's seophagus in patients with chronic gastroesophaageal reflux disease (GERD). ICCE: AB87, 2005.
2. Baopu Li, Max Q.-H. Meng. Texture analysis for ulcer detection in capsule endoscopy images. Image and Vision Computing, 27(9): 1336-1342, 2009.
3. C.S. Lima, D. Barbosa, J. Ramos, A. Tavares, L. Monteiro, L. Carvalho. Classification of Endoscopic Cap-sule Images by using Color Wavelet Features, Higher order Statistics and Radial Basis Functions. EMBS, 1: 1242-1245, 2008.
4. M. Sendoh, K. Ishiyama, K.-I. Arai. Fabrication of magnetic actuator for use in a capsule endoscope. IEEE Transactions on Magnetics, 39(5): 3232-3234, 2003.
5. M. Mackiewicz, J. Berens, M. Fisher. Wireless capsule endoscopy color video segmentation. IEEE Transactions on Medical Imaging, 27(12): 1769-1781, 2008.
6. J.P. Silva Cunha, M. Coimbra, P. Campos, J.M. Soares. Automated topographic segmentation and transit time es-timation in endoscopic capsule exams. IEEE Transactions on Medical Imaging, 27(1):19-27, 2008.
7. G. Csurka, C. R. Dance, Lixin Fan, J. Willamowski, C. Bray. Visual Categorization with Bag of Keypoints. ECCV, 1:1-22, 2004.
8. J. Sivic, B. C. Russel, A. A. Efros, A. Zisserman, W. T. Freeman. Discovering object categories in image collec-tions. Computer Science and Artificial Intelligence Laboratory Report, 2005.
9. S. Lazebnik, C. Schmid, J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. CVPR, 2: 2169-2178, 2006.
10. Baopu Li, Lin Qi, Meng, M.Q.-H., Yichen Fan. Using Ensemble Classifier for Small Bowel Ulcer Detection in Wireless Capsule Endoscopy Images. ROBIO, 2326-2331, 2009.
11. A. Demiriz, K. P. Bennett, J. Shawe-Taylor. Linear programming boosting via column generation. Journal of Machine Learning Research, 46(1-3):225-254, 2002.
12. Andy J H Ma, Pong C Yuen. Linear Dependency Modeling for Feature Fusion. ICCV, 25: 2041-2048, 2011.
13. M. Boulougoura, E. Wadge. Intelligent systems for computer-assisted clinical endoscopic images analysis. Proceedings of Second International Conference on Biomedical Engineering, 405-408, 2004.
14. C. Harris, M. Stephens, A combined corner and edge detector, Proceedings of the 4th Alvey Vision Conference, 1999: 147—151.
15. T. Lindeberg, Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention, International Journal of Computer Vision, 1993, 11(3): 283--318.