2019 IEEE 8th Data Driven Control and Learning Systems Conference
May 24-27, 2019, Dali, Yunnan Province, China

Automatic Classification Based on Features Fusion for Upper Gastrointestinal WCE Images

Min Yu[1], Sheng Li[1], Fengling Hu[2], Liping Chang[1], Ni Zhang[1], Qianru Jiang[1], Dongwei He[1], Xiongxiong He[1]
1. College of Information Engineering, Zhejiang University of Technology, Zhejiang 310023, P. R. China
E-mail: shengli@zjut.edu.cn
2. The First Affiliated Hospital, Zhejiang University, Zhejiang 310003, P. R. China

Abstract: Wireless Capsule Endoscopy (WCE) is an important clinical application which suffers a time-consuming review procedure. A WCE images automatic classification algorithm which fuses color and texture features is proposed to alleviate the burdensome task. A pre-process that aims to label *shadow* and *highlight* is implemented via a new automatic tuning algorithm based on superpixel-level. Hue-Saturation (HS) histograms and a new texture feature named colour scale invariant local ternary pattern (CSILTP) are extracted as local descriptors. A strategy of feature fusion that utilizes discrimination power analysis (DPA) is applied to reduce the dimension of features. The random forest (RF) classifier is then used to discriminate WCE images. The experiment results indicate a better performance of proposed method compared with some existing approaches.

Key Words: Wireless Capsule Endoscopy, Medical Image Processing, Automatic Organ Classification, Feature Fusion

1 Introduction

Traditional flexible endoscopy allows clinicians to explore the interior of the upper gastrointestinal (GI) tract which consists of esophagus, stomach, and duodenum (a part of intestine). However, this invasive procedure discomforts patients, and cannot examine the remainder of the intestine. Wireless Capsule Endoscopy (WCE) becoming popular recently which provides the ability to exam the whole GI tract with a non-invasive scheme [1]. Typically, WCE requires up to 8 hours to go through the GI tract and approximate 60000 images will be recorded in the form of a colour video sequence [2]. Huge amount of images makes the review procedure of WCE a time-consuming task. Fortunately, the computer aided diagnosis (CAD) can be employed to lighten the burden on clinicians. The first subsystem of CAD [3] is known as classification that is used to automatically segment the WCE videos into its corresponding organs. Coimbra *et al.*[4] utilize the *Scalable Color* and *Homogeneous Texture* descriptors of MPEG-7 with a Bayesian classifier to distinguish entrance, stomach, small intestine and colon. According to a useful cue that color information is the primary feature analyzed by clinicians, Berens *et al.* [5] automatically discriminate several meaningful parts (stomach, intestine and colon) by exploiting color features derived from hue saturation (HS) histograms. In [6], local binary patterns (LBP) and motion information are used to expand the works in [5]. Zhou *et al.* in [7] propose a novel color uniform local binary patterns (CULBP) descriptor and the Ada-SVM classifier which uses RBFSVMs as component of Adaboost classifier to elevate WCE images classification accuracy. In order to obtain a higher classification performance and efficiency, Li *et al.* utilize random forests and ferns for organ classification in [8]. In addition, Ma *et al.* present a new method of WCE organ classification by incorporating locality constraint based vector sparse coding (LCVSC) algorithm with the SVM classifier[9]. Convolutional neural network (CNN) is also used to address the classification problem of the digestive organs in [10]. Despite the powerful ability to learn layer-wise hierarchy models of CNN, huge amount of data acquisition with quality annotation is a big challenge for WCE image processing.

The research for WCE images automatic classification are mainly focused either on finding efficient classification algorithms such as in [8–10] or on exploring the robust feature descriptors like color features in [4, 5, 11] and texture features in [6, 12–14]. However, a single feature vector or stacking feature vectors cannot well characterize the WCE images comprehensively. In this paper, a scheme for feature fusion is developed, which improves classification accuracy by using a statistical-based analysis approach discrimination power analysis (DPA)[15]. In addition, WCE images suffer from illumination variations due to the non-ambient lighting, and thus, a new colour scale invariant local ternary pattern (CSILTP) algorithm is proposed to extract features of WCE images. CSILTP is an extension of SILTP[16] which is more robust to illumination variation. The main contributions can be summarized as follows.

1) An automatic tuning *shadow* and *highlight* detection algorithm based on superpixel-level is proposed.
2) A new feature descriptor CSILTP is extracted as a texture feature to improve the accuracy of WCE images classification.
3) An efficient scheme is developed to fuse multiple local features by employing DPA.

The rest of this paper is organized as follows. Section 2 and Section 3 introduces the proposed feature extraction and fusion approaches for WCE images automatic organ classification, respectively. The experiments and comparison results are illustrated and discussed in Section 4. Finally, the conclusions are drawn in Section 5.

2 Features Extraction

2.1 System Framework

The framework of the proposed WCE images automatic classification algorithm is shown in Fig. 1. Firstly, a WCE video is divided into images (frames). Secondly, a

This work was supported in part by the National Natural Science Foundation of China under Grant (61873239, 61503339, 61771430, 61675183), and in part by the Zhejiang Provincial Natural Science Foundation under Grant LY18F010023 and Grant LY14F050004.

978-1-7281-1454-5/19/$31.00 ©2019 IEEE

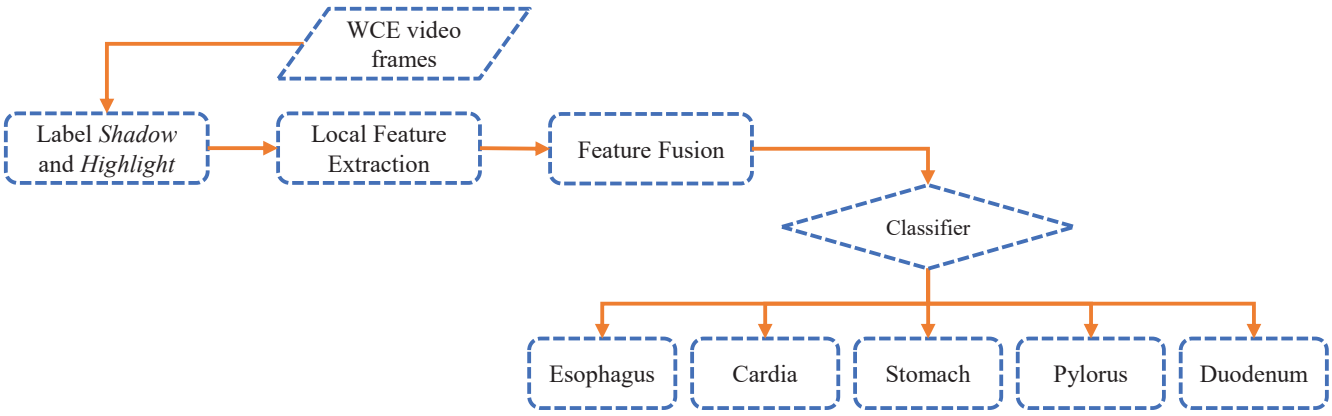


Fig. 1: Framework of the proposed method for organ classification

pre-process that aims to label *shadow* and *highlight* areas is developed via an automatic tuning algorithm based on superpixel-level. Then, color and texture features are extracted from WCE images, respectively. Thirdly, a strategy of feature fusion that utilizes discrimination power analysis (DPA) is applied to reduce the dimension of features. Finally, the random forest (RF) classifier is used to classify WCE images into organ such as esophagus, cardia, stomach, pylorus, and duodenum.

2.2 Label *Shadow* and *Highlight*

Since the tubular cavity structure of the GI tract, the lumen is expected to appear darker than surroundings in view. And the tissues closer to the light source are imaged with extreme bright due to the non-ambient lighting. Those regions are refer to as *shadow* and *highlight*, respectively [17]. The results of organ classification are unreliable if the whole regions in images are used to extract features [13]. Thus, the regions of *shadow* and *highlight* should be labeled and removed before feature extraction.

A novel automatic tuning *shadow* and *highlight* detection algorithm is proposed in this paper, which based on superpixel segmentation. Simple linear iterative clustering (SLIC) algorithm[18] is first utilized to obtain superpixels, as shown in Fig. 2(a). Every superpixel's intensity is assigned a value which equals to the average of all pixels in the surperpixel. Then, the color image is split into three intensity (gray scale) images (*i.e. hue*, *saturation* and *intensity* in HSI model). As shown in Fig. 2(b), the intensities of the *shadow* are relatively small (shown in dark gray). Thus, the *shadow* are where the intensities are less than a certain value (*i.e.* the threshold). However, threshold selection is a paradoxical process since a fixed threshold cannot be applied to all images. Thus, an automatic tuning threshold algorithm is proposed. Supposing the intensity of a surperpixel is the smallest in a set of surperpixels, then the surperpixel is called valley surperpixel (P_{VS}). A threshold statistical experiment of *shadow* detection is established on 100 samples. According to the experiment, if the intensity of a P_{VS} is small, the dark area around the P_{VS} may be large. Thus, we assume that the image with small P_{VS} intensity has a large gap between the threshold and the intensity of P_{VS}. A differential equation can be constructed:

$$(f(I_n) - I_n)' = \alpha * I_n + \beta \tag{1}$$

where I_n is intensity of P_{VS} in an arbitrary sample. $f(\cdot)$ is a function for automatic tuning threshold. α and β are the parameters of the linear function, respectively. The general solution of equation (1) is a quadratic function. The threshold T_n can be formulated

$$T_n = \begin{cases} a * I_n^2 + b * I_n + c, & I_n < T_{max} \\ T_{max}, & I_n \geq T_{max} \end{cases} \tag{2}$$

where a, b and c are the parameters of the quadratic function. I_s is intensity of P_{VS} in all training samples. T_{max} is the maximum threshold, and surperpixels whose intensities beyond T_{max} cannot be *shadow*. These conditions can be written as follows

$$\begin{cases} a * I_s^2 + b * I_s + c < T_{max} \\ a * T_{max}^2 + b * T_{max} + c < T_{max} \end{cases} \tag{3}$$

Besides, $T_n - I_n$ should be reduced along with the increase of I_n according to the assumption. So, this condition can be formulated as

$$\frac{\mathrm{d}(T_n - I_n)}{\mathrm{d}I_n} < 0, \quad I_n \in (I_s, T_{max}) \ . \tag{4}$$

The parameters of the quadratic function are obtained according to formulas (3) and (4):

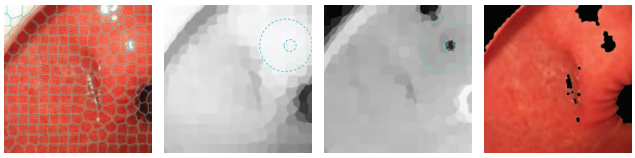
$$\begin{cases} 0 < a < 1/(T_{max} - I_s) \\ b = -a * (T_{max} + I_s) \\ c = T_{max} + a * T_{max} * I_s \end{cases} \tag{5}$$

where a determines the degree of the gap between the threshold and the intensity of P_{VS} along with the increase of I_n.

According to the methods in [13], the pixels in images with intensity values that satisfy $I(x, y) > threshold$ are supposed to be *highlight*. In fact, as shown in Fig. 2(a) and Fig. 2(b), there are some informative regions (areas within the cyan dotted circle), which also satisfy $I(x, y) > threshold$. So, it cannot detect the *highlight* by using *intensity* effectively. Fortunately, we find that the saturation values of *highlight* satisfy $S(x, y) < threshold$ as shown in Fig. 2(c). Therefore, the *highlight* can be detected in the same way as detection of *shadow*.

2.3 Color Feature

Color information is the primary basis for analysis according to clinicians [6]. Meanwhile, color distribution is usually



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

Fig. 2: Illustration of detection of *shadow* and *highlight*. (a) Original image with 300 superpixels. (b) Average intensity superpixel map of original image. (c) Average saturation superpixel map of original image. (d) Image labeled valid regions (original is segmented by 1500 superpixels and the invalid regions are shown in black).

used to characterize images since its relatively invariants to scale changes, translation and rotation [19]. Hue-Saturation (HS) histograms are color-based features which are first applied to WCE images and reported a good results by Berens *et al.* in [5]. The process of HS feature extraction can be summarized as follows:

1) The WCE images are converted from the RGB to the HSI color space.
2) The hue and saturation are quantified to 32 bins respectively.
3) A 2-D color histogram is constructed via counting the number of pixels which satisfy corresponding hue and saturation bins simultaneously.
4) Calculate discrete cosine transform (DCT) coefficients of the 2-D color histogram.
5) HS histograms are formed by compressing DCT coefficients with PCA.

2.4 Texture Feature

The textures of inner walls of different GI organs appear distinguishable visually, such as the intestinal walls with folds and the stomach with smooth surfaces [5, 20, 21]. Local binary patterns (LBP) is often used in medical image analysis [6, 13, 22, 23]. However, the LBP operator is not robust to local image noises when neighboring pixels are similar. And in a case of illumination variation, the suddenly change of gray scale intensities of neighboring pixels simultaneously may cause different LBP forms. Thus, we introduce a scale invariant local ternary pattern (SILTP) operator [16] and extend it into color images by calculating norm values of three color channels in this work.

In LBP, local circular pixels in a window are thresholded by the central pixel, the process can be written as

$$S_\tau(I_c, I_k) = \begin{cases} 1, & I_k \geq I_c + \tau \\ 0, & I_k < I_c + \tau \end{cases} \tag{6}$$

where I_c, I_k are the intensity values of the center pixel and the neighbor pixels, respectively. τ is a scale factor indicating the comparing range and $\tau = 0$ in LBP.

Since the LBP operator is not robust to local image noises, therefore, Tan and Triggs proposed a local ternary pattern (LTP) operator [24] in order to handle the effect of noises. $S_\tau(I_c, I_k)$ in LTP is defined as

$$S_\tau(I_c, I_k) = \begin{cases} 1, & I_k \geq I_c + \tau \\ -1, & I_k \leq I_c - \tau \\ 0, & otherwise \end{cases} \tag{7}$$

They set τ as a constant and thus, there is a fault-tolerant range (*i.e.* local noises) between I_c and I_k. But the LTP descriptor cannot keep its invariance against gray scale changes. Thus, Liao *et al.* in [16] set τ automatically according to the value of I_c. The $S_\tau(I_c, I_k)$ can be redefined as

$$S_\tau(I_c, I_k) = \begin{cases} 01, & I_k \geq I_c(1 + \tau) \\ 10, & I_k \leq I_c(1 - \tau) \\ 00, & otherwise \end{cases} \tag{8}$$

Finally, given any pixel location (x_c, y_c), SILTP encodes it as

$$SILTP^{\tau}_{N,R}(x_c, y_c) = \bigoplus_{k=0}^{N-1} s_\tau(I_c, I_k) \tag{9}$$

where N is the amounts of neighborhood pixels which are equally spaced on a circle of radius R, and $\bigoplus$ denotes concatenation operator of binary strings.

Both LBP and SILTP extract texture features of gray images. There are colour LBP proposed in previous works for classification [25]. But colour LBP is usually obtained by concatenating the three LBP vectors of different color channels together, which can lead to loss of color information or correlation between channels of the original image. Thus, we employ a method that combines three different color channels via the usage of norm operation in [26]. The color norm value consists color information so that the classification performance becomes more effective and reliable than using only gray information. Supposing $I = [r\ g\ b]^T$ is a color image. Then a color norm value I_{norm} is defined as

$$I_{norm} = \|I\| = \sqrt{r^2 + g^2 + b^2} \tag{10}$$

Finally, a CSILTP feature is formed via calculating SILTP of I_{norm} .

3 Features Fusion

In general, classification accuracy can be improved by using multiple features of different properties, but stacking feature vectors will be very lengthy, which may lead to overfitting. Principal component analysis (PCA) is usually applied to reduce feature dimensionality in previous works. In this paper, we introduce a strategy to fuse several local features.

A statistic-based analysis approach discrimination power analysis (DPA) is proposed by Dabbaghchian *et al.* to select coefficients of discrete cosine transform (DCT) [15]. The basic idea of DPA is to select components which have large variations between the classes and small variations within the class (it means large discrimination power). Thus, it is expected that the components with large DP in feature vectors can be formed a low dimensional feature vector for classification. And this idea is the basis of the proposed features fusion.

A M-dimensional feature vector extracted from an arbitrary image can be written as $F = [x_1, x_2, ..., x_M]^T \in \mathbb{R}^M$. DP of each component $x_i (i = 1, 2, ..., M)$ can be estimated as follows:

1) Construct the training set matrix A_i with the ith dimensional component of feature vectors for all categories and all training images:

$$A_i = \begin{bmatrix} x_i(1,1) & x_i(1,2) & \cdots & x_i(1,C) \\ x_i(2,1) & x_i(2,2) & \cdots & x_i(2,C) \\ \vdots & \vdots & \ddots & \vdots \\ x_i(S,1) & x_i(S,2) & \cdots & x_i(S,C) \end{bmatrix}_{S \times C} \tag{11}$$

where $x_i(S,C)$ is the ith dimensional component of the Sth sample of the Cth category of images.

2) Calculate the mean value of each category:

$$M_i^c = \frac{1}{S} \sum_{s=1}^{S} A_i(s,c), c = 1,2,...,C \tag{12}$$

where c and s denote the numbers of the category and the training image for each category respectively.

3) Calculate variance of each category:

$$V_i^c = \sum_{s=1}^{S} (A_i(s,c) - M_i^c)^2, c = 1,2,...,C \tag{13}$$

4) Averaging the variance of all the categories:

$$V_i^W = \frac{1}{C} \sum_{c=1}^{C} V_i^c \tag{14}$$

5) Calculate the mean of all training samples:

$$M_i = \frac{1}{S \times C} \sum_{c=1}^{C} \sum_{s=1}^{S} A_i(s,c) \tag{15}$$

6) Calculate the variance of all training samples:

$$V_i^B = \sum_{c=1}^{C} \sum_{s=1}^{S} (A_i(s,c) - M_i)^2 \tag{16}$$

7) Estimate the DP on the ith dimensional component of feature vectors:

$$D_i = \frac{V_i^B}{V_i^W}, 1 \le i \le M \tag{17}$$

DP of each component of a feature vector is obtained by this way. Feature dimensions associated with several feature vectors in testing images can be reduced by selecting components with large DP in training images. In addition, All selected feature components are weighted in order to further improve the rate of recognition. The weight of the ith component can be determined by the ratio of its DP value to the total DP values of all selected components in the vector presented by Leng *et al.* in [27].

$$\omega_i = \frac{D_i}{D_1 + D_2 + ... + D_M}, i = 1,2,...,M \tag{18}$$

4 Experiments

4.1 Performance Criterion

For an arbitrary WCE frame, its organ category in GI tract is predicted through a Random forests (RF) multi-label classifier. RF has been shown to outperform other widely-used classifiers such as AdaBoost and SVM in medical image processing [28]. The performance of single organ classification are measured precision and recall.

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

Fig. 3: Examples of various organs of the upper GI tract. (a) Esophagus. (b) Cardia. (c) Stomach. (d) Pylorus. (e) Duodenum.

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

where TP is the number of true positive images, FP is the number of false positive images, TN is the number of true negative images, and FN is the number of false negative images.

The performance of all organs classification is measured by accuracy and macro-averaging of precision and recall:

$$Accuracy = \frac{Number\ of\ correct\ detections}{Number\ of\ total\ samples} \tag{21}$$

$$Macro\_Pre = \frac{1}{n} \sum_{i=1}^{n} Precision(i) \tag{22}$$

$$Macro\_Rec = \frac{1}{n} \sum_{i=1}^{n} Recall(i) \tag{23}$$

where n is the class number.

4.2 Dataset and Experimental Results

A total of 11 WCE video clips within esophagus, cardia, stomach, pylorus and duodenum are obtained from the Internet [29–31]. Examples of various organs of the upper GI tract are shown in Fig. 3. A total of 3680 frames (each organ consists 736 frames) are screened as experimental dataset under the guidance of an experienced clinician. The details are shown in Table 1.

Three experiments are designed to demonstrate the reliability and efficiency of the proposed algorithm. The first experiment presents the effect of the proposed feature descriptor. In the experiment, the organs classification performance of CSILTP and other texture features are contrasted by using a RF classifier. As shown in Fig. 4, The SILTP feature significantly improves the classification performance of LBP since it further reduces the affect of illumination, as demonstrated in [16]. The results of classification using CULBP are improved over ULBP in accuracy, precision and recall respectively due to CULBP feature combines texture information of three color channels, which is more robust to the variation of illumination and more discriminative for classification [7]. Similarly, the classification of CSILTP feature reaches an average of 97.51% accuracy, which outperforms

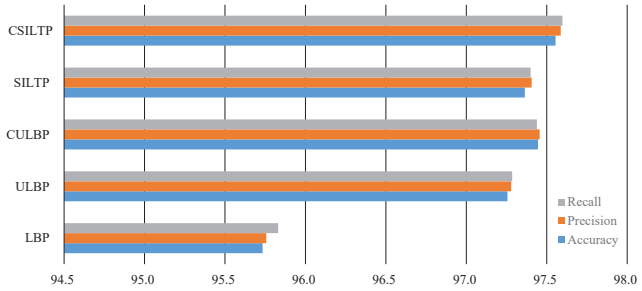


Fig. 4: The comparison of our presented CSILTP and other texture features.

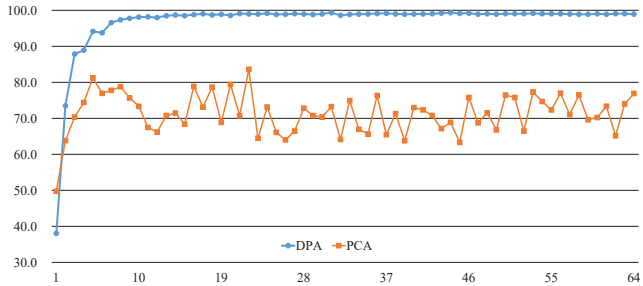


Fig. 5: Accuracies of methods used PCA and DPA in different dimension.

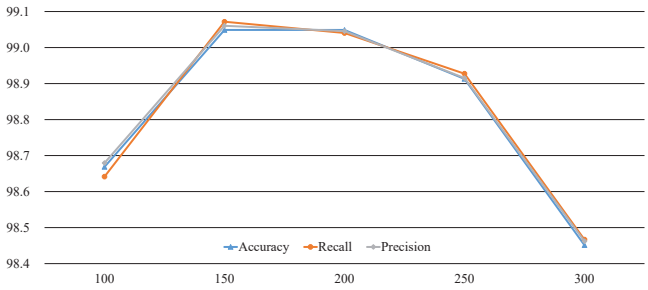


Fig. 6: Performance of classification used RF in different grown trees.

SILTP with an average of 97.36% accuracy. As a result, The CSILTP is more robust for classification.

In the second experiment, the effect of feature fusion by utilizing DPA is compared with PCA. The HS color features and CSILTP texture features are stacked firstly. Then DPA and PCA are used to reduce the dimension of stacking feature vectors, respectively. Accuracies of the two methods in different dimension are presented. It can be seen from Fig. 5 that the classification accuracy using DPA reaches a good result when the feature dimension equals to 24 approximately. However, the results using PCA fluctuate greatly as the change of feature dimensions, and the accuracies are lower than DPA. The reason is that DPA makes full use of the statistical information of training data to further the discrimination of fused features.

The last experiment evaluates the capability of single image classification based on a 10-fold cross-validation test performed on the 3680 frames dataset. In the experiment, HS and CSILTP features are fused by DPA, and fusion features are used to train RF classifier. The dimension of fusion features is 24, which is given in the second experiment. As shown in Fig. 6, RF classifier with 150 grown trees has better performance in organs classification. Thus, the number of grown trees in RF classifier is set to 150 in the last experiment. Table 2 and Table 3 shows the performance of organs classification which are predicted by RF classifier with 150 decision trees, respectively. The results show that the proposed method is better than the existing methods[5–7]. And the averages of total accuracy, micro-average precision and macro-average recall are 98.99%, 99.00% and 99.01%, respectively.

Table 1: A total of 11 WCE video fragments within 5 upper gastrointestinal organs.

S/N	Clips	Frame rate(fps)	Duration(sec)	Total(f)
1	esophagus-stomach	25	45	1125
2	cardia	25	45	1125
3	esophagus-cardia	25	21	537
4	esophagus-cardia	25	20	514
5	esophagus-cardia	25	51	1284
6	stomach	30	112	3916
7	stomach-pylorus	30	127	3202
8	pylorus	25	39	980
9	pylorus	25	43	1087
10	duodenum	25	36	906
11	duodenum	30	514	17

Table 2: Single organ classification performance comparison by precision and recall.

Organs	Precision		Recall	
	Mean	Variance	Mean	Variance
Esophagus	100.00	0.00	99.61	0.82
Cardia	99.46	0.85	99.17	0.52
Stomach	98.59	1.00	97.52	3.40
Pylorus	99.34	0.85	99.05	0.44
Duodenum	97.62	2.99	99.70	0.40

Table 3: WCE images classification performance comparison by three evaluation metrics.

Methods	Accuracy		Precision		Recall	
	Mean	Variance	Mean	Variance	Mean	Variance
HS[5]	98.04	0.69	98.02	2.98	98.04	2.65
HS+LBP[6]	98.29	1.28	98.27	3.59	98.29	3.18
HS+CULBP[7]	98.32	1.33	98.37	6.05	98.35	4.82
Proposed method	**98.99**	**0.15**	**99.00**	**1.14**	**99.01**	**1.12**

5 Conclusions

We present a method to automatically discriminate the frames of WCE videos in the upper gastrointestinal tract based on color and texture features. Besides, a statistical-based DPA scheme is employed to reduce fused feature dimensions. The result of organs classification reaches a total accuracy of approximately 98.99% in average. In the future, we will extend the classification of WCE images to the entire GI tract. And a potential application of our method is automatic segmentation and annotation of WCE videos.

References

[1] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless capsule endoscopy," *Nature*, vol. 405, pp. 725–729, May 2000.

[2] "Given imaging." [Online]: https://www.medtronic.com/us-en/index.html/. [Accessed]: Dec. 4, 2018.

[3] M. Riegler, K. Pogorelov, P. Halvorsen, T. de Lange, C. Griwodz, P. T. Schmidt, S. L. Eskeland, and D. Johansen, "Eir - efficient computer aided diagnosis framework for gastrointestinal endoscopies," in *2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 1–6, Jun. 2016.

[4] M. Coimbra, P. Campos, and J. P. S. Cunha, "Extracting clinical information from endoscopic capsule exams using mpeg-7 visual descriptors," in *The 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, 2005. EWIMT 2005. (Ref. No. 2005/11099)*, pp. 105–110, Nov. 2005.

[5] J. Berens, M. Mackiewicz, and D. Bell, "Stomach, intestine, and colon tissue discriminators for wireless capsule endoscopy images," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5747, pp. 283–290, Apr. 2005.

[6] M. Mackiewicz, J. Berens, and M. Fisher, "Wireless capsule endoscopy color video segmentation," *IEEE Trans. on Medical Imaging*, vol. 27, pp. 1769–1781, Dec. 2008.

[7] R. Zhou, B. Li, H. Zhu, and M. Q.-H. Meng, "A novel method for capsule endoscopy video automatic segmentation," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3096–3101, Nov. 2013.

[8] B. Li, R. Zhou, C. Yang, M. Q.-H. Meng, G. Xu, and C. Hu, "Capsule endoscopy images classification by random forests and ferns," in *2014 4th IEEE International Conference on Information Science and Technology*, pp. 414–417, Apr. 2014.

[9] T. Ma, Y. Zou, Z. Xiang, L. Li, and Y. Li, "Wireless capsule endoscopy image classification based on vector sparse coding," in *2014 IEEE China Summit International Conference on Signal and Information Processing (ChinaSIP)*, pp. 582–586, Jul. 2014.

[10] Y. Zou, L. Li, Y. Wang, J. Yu, Y. Li, and W. J. Deng, "Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*, pp. 1274–1278, Jul. 2015.

[11] C. Wang, Z. Luo, X. Liu, J. Bai, and G. Liao, "Organic boundary location based on color-texture of visual perception in wireless capsule endoscopy video," *Journal of Healthcare Engineering,2018,(2018-1-10)*, vol. 2018, pp. 1–11, Jan. 2018.

[12] M. Mackiewicz, J. Berens, M. Fisher, and D. Bell, "Colour and texture based gastrointestinal tissue discrimination," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 2, pp. II–II, May 2006.

[13] R. Zhou, B. Li, Z. Sun, C. Hu, and M. Q.-H. Meng, "Wireless capsule endoscopy video automatic segmentation," in *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 825–830, Dec. 2012.

[14] B. Li, G. Xu, R. Zhou, and T. Wang, "Computer aided wireless capsule endoscopy video segmentation," *Medical Physics*, vol. 42, pp. 645–652, Jan. 2015.

[15] S. Dabbaghchian, M. P. Ghaemmaghami, and A. Aghagolzadeh, "Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology," *Pattern Recognition*, vol. 43, pp. 1431–1440, Apr. 2010.

[16] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikinen, and S. Z. Li, "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1301–1306, Jun. 2010.

[17] X. Zabulis, A. A. Argyros, and D. P. Tsakiris, "Lumen detection for capsule endoscopy," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3921–3926, Sep. 2008.

[18] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 2274–2282, Nov. 2012.

[19] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, pp. 11–32, Nov. 1991.

[20] J. P. S. Cunha, M. Coimbra, P. Campos, and J. M. Soares, "Automated topographic segmentation and transit time estimation in endoscopic capsule exams," *IEEE Trans. on Medical Imaging*, vol. 27, pp. 19–27, Jan. 2008.

[21] L. Igual, J. Vitrià, F. Vilariño, S. Seguí, C. Malagelada, F. Azpiroz, and P. Radeva, "Automatic discrimination of duodenum in wireless capsule video endoscopy," in *4th European Conference of the International Federation for Medical and Biological Engineering*, vol. 22, pp. 1536–1539, 2009.

[22] P. Wang, S. M. Krishnan, C. Kugean, and M. P. Tjoa, "Classification of endoscopic images based on texture and neural network," in *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 4, pp. 3691–3695, Oct. 2001.

[23] C. Azzopardi, Y. A. Hicks, and K. P. Camilleri, "Exploiting gastrointestinal anatomy for organ classification in capsule endoscopy using locality preserving projections," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3654–3657, Jul. 2013.

[24] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. on Image Processing*, vol. 19, pp. 1635–1650, Jun. 2010.

[25] J. Y. Choi, K. N. Plataniotis, and Y. M. Ro, "Using colour local binary pattern features for face recognition," in *2010 IEEE International Conference on Image Processing*, pp. 4541–4544, Sep. 2010.

[26] S. H. Lee, J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Local color vector binary patterns from multichannel face images for face recognition," *IEEE Trans. on Image Processing*, vol. 21, pp. 2347–2353, Apr. 2012.

[27] L. Leng, J. Zhang, J. Xu, M. K. Khan, and K. Alghathbar, "Dynamic weighted discrimination power analysis in dct domain for face and palmprint recognition," in *2010 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 467–471, Nov. 2010.

[28] A. Criminisi and J. Shotton, *Decision Forests for Computer Vision and Medical Image Analysis*, pp. 273–293. Springer: London, 2013.

[29] "Science photo library." [Online]: https://www.sciencephoto.com/. [Accessed]: Dec. 4, 2018.

[30] "Ei salvador atlas of gastrointestinal video endoscopy." [Online]: http://www.gastrointestinalatlas.com/. [Accessed]: Dec. 4, 2018.

[31] "Gastrolab - the gastrointestinal image site." [Online]: http://www.gastrolab.net/index.htm/. [Accessed]: Dec. 4, 2018.