ORIGINAL PAPER

Multiple bleeding detection in wireless capsule endoscopy

Ouiem Bchir[1] · Mohamed Maher Ben Ismail[1] · Nada AlZahrani[1]

Abstract

Wireless capsule endoscopy (WCE) is an emerging technology that aims to detect pathology in the patient gastrointestinal tract. Physicians can use WCE to detect various gastrointestinal diseases at early stages. However, the diagnosis is tedious because it requires reviewing hundreds of frames extracted from the captured video. This tedious task has promoted researchers' efforts to propose automated diagnosis tools of WCE frames in order to detect symptoms of gastrointestinal diseases. In this paper, we propose an automatic multiple bleeding spots detection using WCE video. The proposed approach relies on two main components: (1) a feature extraction intended to capture the visual properties of the multiple bleeding spots, and (2) a supervised and unsupervised learning techniques which aim to accurately recognize multiple bleeding.

Keywords Wireless capsule endoscopy · Multiple bleeding spots · Recognition · Clustering · Classification

1 Introduction

Digestive diseases affect millions of people in the world. In 2009, studies found that 60 to 70 million US citizens suffer from such digestive diseases [1]. Different techniques have been used in the medical field to diagnose the gastrointestinal (GI) tract. The most common technique used to detect pathologies of the GI tract is endoscopy. However, reaching the main parts of the patient GI tract using conventional endoscopies such as intra-operative enteroscopy, colonoscopy and push enteroscopy is irritating and hard to accept by the patient. In fact, traditional endoscopies are uncomfortable and painful for them. In 2000, a new type of GI endoscopies known as wireless capsule endoscopy (WCE) [2] has been introduced. This new technology consists of a capsule that is swallowed by the patient, and allows viewing the whole area of the GI tract without pain. Simultaneously, a wireless transmitter sends frames from the inside GI tract to an outside receiver.

The resulting WCE video may consist of more than 55,000 frames captured at a frequency of two frames per second. These frames are then examined and analyzed by physicians in order to recognize disease symptoms. This task may last more than 2 hours per patient. This tedious diagnostic process along with the preciousness of physician time represents the main drawback of WCE and yields the development of computerized diagnostic systems based on machine learning techniques which aim to support physicians in detecting anomalies in WCE frames within a reasonable time [3, 4]. In fact, physicians need a reliable and robust approach to help them during the diagnosis phase, and reduce the examination time and effort. This goal becomes more challenging due to the variability of the different properties that characterize each disease. Also, for accurate diagnosis, various machine learning techniques should be implemented and tested to detect specific disease.

The three known GI diseases which can be detected using WCE are bleeding, tumor and ulcer. Each disease has its own symptoms and characteristics. The main focus of this paper is the multiple bleeding spots (MBS) detection using WCE video frames. MBS appears as small light spots as shown in Fig. 1. The detection of these symptoms is challenging because of its visual similarity to intestinal bubbles. Moreover, MBS symptoms may be captured in different areas of the GI tract with different color and texture characteristics, especially in the surrounding area. To detect these bleeding spots, visual feature descriptors should be extracted from each frame, and the obtained feature vectors would be fed into the classifier.

In this work, we propose an automated approach to detect MBS and discard useless frames. This detection process

✉ Mohamed Maher Ben Ismail
  mbenismail@ksu.edu.sa

1 Computer Science Department, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia
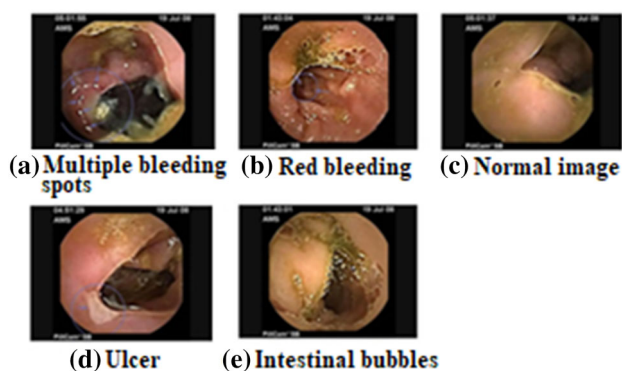
Fig. 1 WCE frame snapshots. **a** frame with multiple bleeding spots anomaly, **b** frame with red bleeding anomaly, **c** normal image, **d** frame with ulcer anomaly, **e** frame with intestinal bubbles

relies on an unsupervised learning technique to gather WCE frames sharing the same characteristics in homogeneous groups and a supervised learning technique that is intended to build a model able to distinguish between positive (containing MBS) and negative frames.

2 Literature review

Wireless capsule endoscopy (WCE) is a new technology that assists physicians to detect various pathologies in the patient's gastrointestinal tract. Recently, multiple bleeding spot detection using WCE has gained researchers' attention. Many researches have contributed to provide different computer-based solutions to assist clinicians in their diagnosis. These solutions aim at improving the detection performance and reduce the investigation time. Some of these researches were introduced to diagnose a specific anomaly [3–5]. In particular, the authors in [5–7] used the deep learning techniques to automatically detect submucosal tumors, colorectal polyps and Hookworm, respectively. Other researchers intended to remove useless frames from the recorded WCE video in an unsupervised manner [8, 9] and enhance the frame quality [10]. In the following, several computer-based approaches, developed to detect multiple bleeding spots and summarize WCE video, are reviewed.

The informative frames that contain abnormalities of the gastrointestinal (GI) tract may appear in only two or three frames of the recorded video. Thus, some symptoms of digestive diseases captured by the capsule may not be distinguishable by naked eyes. Computer-based solutions are then needed to assist clinicians in their diagnosis. The recent solutions for automatic detection of multiple bleeding spots (MBS) rely on image processing techniques that are coupled with machine learning algorithms. In [11], a feature extraction method and various classification techniques for ulcer and bleeding detection in WCE frames were outlined. The

authors implemented their solution using color and texture features. Then, they used the support vector machines (SVM) [12], the decision tree [13] and the neural network classifiers [14]. Their results showed that *SVM* classifier yielded the best results. Similarly, the researchers in [15] adopted local binary histogram [16] as color feature and wavelet transform [17] as texture feature. Then, they classified the informative frames using SVM [12] and neural network [14] classifiers. The authors concluded that the SVM classifier along with color feature yields more accurate classification rate. On the other hand, selected histograms in hue, saturation and value (HSV) space and the dominant color feature as color descriptors were used in [18]. The authors found that HSV-based feature improved the recognition of useless frames containing black regions. Similarly to [11] and [15], an SVM classifier is trained to learn a classification model and detect frames containing bleeding spots. Besides, the researchers in [19] used HSI (hue, saturation and intensity) color domain and the region segmentation to extract the visual features. Their method consists of three phases; first, they represent the image in the HSI color domain. Then, they segment each image into bleeding and nonbleeding regions using JEDISON [20] region segmentation algorithm. Finally, they classify the segmented image into bleeding and nonbleeding areas using the color threshold technique. Another system is outlined in [21]. It combines the chrominance moments as color feature and the uniform local binary pattern as texture feature to distinguish between bleeding and normal regions. These features are conveyed to a multilayer perceptron neural network classifier [14] for bleeding detection in the WCE frames. In [22], the developed system starts by segmenting the WCE frames using superpixel segmentation approach. Then, superpixel classification using SVM classifier is performed. In [23], the authors introduced a four-step approach. First Karhunen–Loeve (K-L) transformation is applied on WCE frames. Then, smoothing and segmentation techniques in K-L color space are performed. After that, thresholding is used to detect blood-based regions. Finally, a local global graph is applied to segment blood-based regions. The authors in [24] reported an empirical comparison of visual descriptors to detect MBS in WCE videos. Their experiments show that the color moments feature in the HSV color space outperforms the other descriptors. In [25], the authors determined the dissimilarity between neighboring WCE frames using a frame-buffered versions of peak detection-based elimination algorithms. This yields the retention of the interesting/distinctive frames in the examination video. In [26], the researchers developed a new approach to reduce the number of frames in WCE video based on nonnegative matrix factorization [27, 28]. This algorithm extracts the representative frames from all WCE video frames. The combination of fuzzy C-means (FCM) [29] and nonnegative matrix factor-

ization was also adopted in [30] in order to discard irrelevant frames.

3 Proposed system

Multiple bleeding spots can be recognized as bright dots in the frames extracted from wireless capsule endoscopy (WCE) video. These dots may appear at any location of the intestinal track with different neighboring visual properties. In other words, categorization of frames showing multiple bleeding spots consists in assigning frames sharing the same visual properties (having small bright dots as a common pattern) to the same group. These pattern properties along with the small size of the region of interest and the diversity of the neighboring areas in the intestinal track affect the performance of most classifiers adopted to recognize multiple bleeding spots. In order to overcome this challenge, we propose to use an unsupervised machine learning technique to divide the main classification problem to multiple local classification subproblems. More specifically, clustering the training set is intended to partition the WCE frames into visually homogeneous subsets. The resulting cluster centroids are thus perceived as the output of this data summarization task.

Let $X = \{x_1, x_2, \ldots, x_n\}$ denote the feature vectors of a set of images that need to be partitioned into C clusters. Each vector $x_j = (f_1^j, f_2^j, \ldots, f_d^j)$ represents the d-dimensional visual features of the image j. The *FCM* algorithm [29] minimizes the objective function $J = \sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^m ||x_j - c_i^2$ where m, a real number larger than 1, controls the fuzziness degree, u_{ij} is the membership degree of x_j to the ith cluster, and c_i is the cluster center. The minimization of the objective function is conducted through an iterative optimization of the membership u_{ij} and the cluster centers c_i. The resulting membership degrees depend on the distance between the observed data and each cluster centroid.

Once the frame training set is clustered using *FCM*, the following step of the proposed approach is to assign each testing frame to the nearest cluster center, instead of comparing it with all the training frames as performed by the typical *KNN* classifier [31]. More specifically, the classification of each WCE frame is based on the nearest cluster center votes. Thus, given a test frame, the classifier assigns the class value with the most votes among the class values of the k nearest cluster centers. To determine the k nearest cluster centers, the Euclidean pairwise distances between the test frame x_j and the clusters centers c_i are calculated. The flowchart in Fig. 2 describes the proposed system. As it can be seen, it consists of three main phases. Namely, it relies on the feature extraction, the training set clustering and the testing set classification tasks.

In this pattern recognition problem, the training data consist of two groups of WCE frames. The first one is a set of
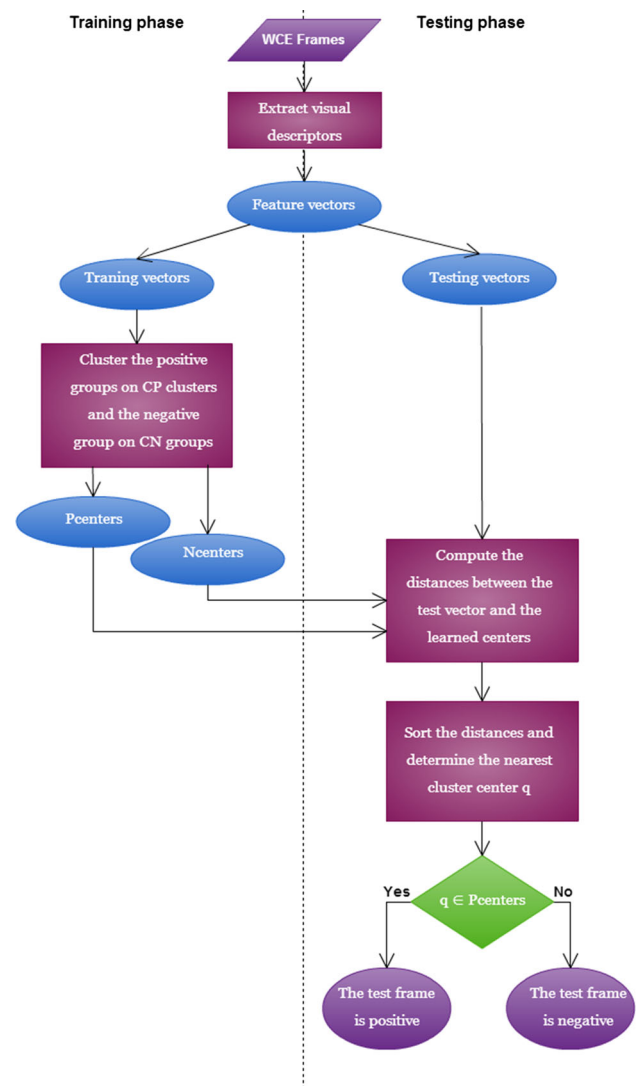

Fig. 2 Flowchart of the proposed approach

images containing multiple bleeding spots (positive group), and the second group is a set of normal frames (negative group). We cluster the positive frames into *CP* clusters and the negative frames into *CN* clusters using fuzzy *C*-means (*FCM*) clustering algorithm [29]. This clustering algorithm partitions the instances into a collection of *C* clusters and generates fuzzy membership values which reflect the degree of mutual dependency among the clusters. In fact, these fuzzy functions allow the distinction between the frames strongly belong to a particular cluster and those which exhibit marginal belongingness only with several clusters.

The obtained set of cluster centers (*Pcenters and Ncenters*) summarizes the dataset and will be used for the supervised learning task. Finally, the distances between a given test frame and the learned centers *Pcenters* and *Ncenters* are computed and sorted to determine the nearest center q to the test frame. If the nearest center q is the learned proto-
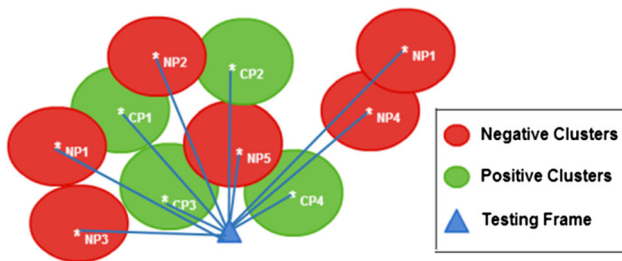
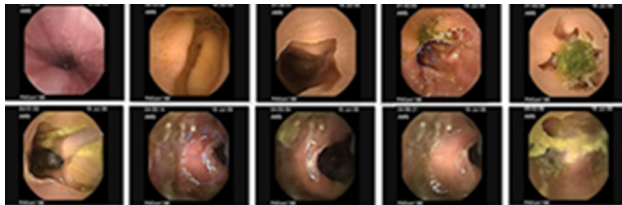**Fig. 3** Illustrative example for the proposed classification approach

Fig. 4 Sample video frames showing various anomalies

Fig. 5 Accuracy obtained using various visual features and different *P* and *Q* values

type of a positive cluster, the test frame is labeled as positive (contains multiple bleeding spots). On the other hand, if the nearest center is the prototype of a negative cluster, the test frame is labeled negative (normal frame). This is the same concept as in *K* nearest neighbor classifier *KNN* [27] with the number of neighbor parameter equal to 1, applied on the learned centers instead of the whole training set. Figure 3 shows an illustrative example of multiple instance-based classification. As it can be seen, *CP3*, which is a center of a positive cluster, is the nearest cluster to the testing vector. Thus, it is labeled as positive.

4 Experiments

In order to assess the performance of the proposed system, we first investigate the impact of applying a different number of clusters on the accuracy of the multiple bleeding spots detection. Then, we evaluate the performance of the proposed system and compared it to the typical KNN- and *SVM*-based detection systems. We use a real WCE videos collection that yields 1275 frames representing the different parts of the gastrointestinal tract. These frames were captured using Imaging PillCam® SB [26]. The size of each frame is 480×360. Sample video frames with different pathology symptoms are shown in Fig. 4. Each frame is labeled according to the anomaly it contains. The dataset consists of 915 "normal" frames and 360 "multiple bleeding spots" frames.

In these experiments, we consider a set of MPEG-7 features [27]. Namely, we use the color moments, the color histogram, the local color moments, the Gabor filter, the discrete wavelet transform (DWT) and the local binary pattern (LBP) features [24]. Besides, we adopt a 10-fold cross-
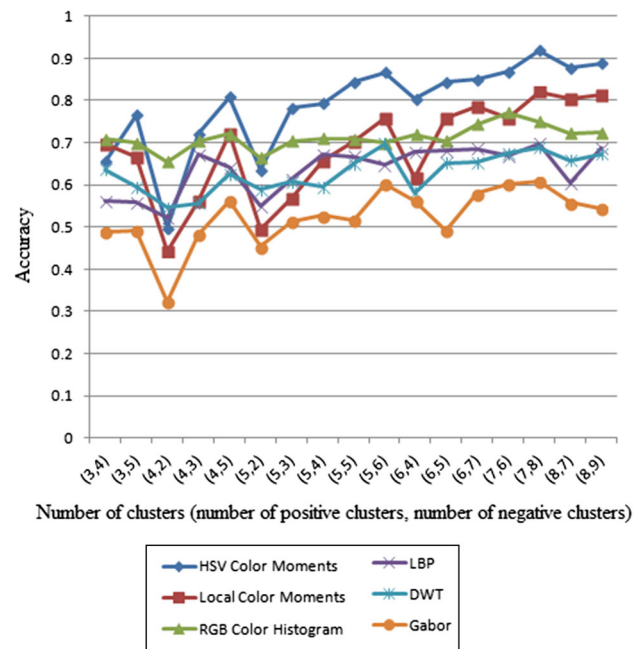
validation to divide the dataset into training and testing. In each fold, the part of dataset that we categorized as normal is clustered into *P* clusters. On the other hand, the frames which include MBS are clustered into *Q* clusters. As shown in Fig. 5, we deployed 18 scenarios by setting different combinations for the number of clusters (*P* and *Q values*). As it can be seen in Fig. 5, partitioning "normal" frames and "multiple bleeding spots" frames into 7 and 8 clusters, respectively, yields the highest accuracy for the HSV color moments, the local color moments, local binary pattern and Gabor filter. On the other hand, the RGB color histogram and the discrete wavelet transform yield higher performance with a smaller number of clusters.

Figure 6 depicts the mean and standard deviation of the accuracy obtained using all visual features and different values of *P* and *Q*. As it can be noticed, the HSV color moments feature yields the highest average accuracy. Moreover, as shown in Fig. 6, its standard deviation proves that the lowest accuracy obtained using this feature exceeds the average accuracy reached by the other visual features. Besides, based on the obtained results presented in Figs. 5 and 6, we can conclude that the best and most discriminative color feature is the HSV color moments. In fact, MBS appear as white dots in WCE frames. These white dots are represented using a [360, 0, 100] vector in HSV color space, while the rest of the image is represented toward red color [360, 100, 100]. These white spots affect the mean and standard deviation of the color moments. Figure 7 shows the ROC plots obtained based on clustering the WCE frames with $P=7$ and $Q=8$ using the obtained visual features. We can notice that these
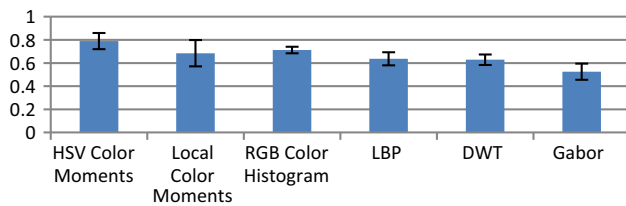
Fig. 6 Average accuracy and standard deviation obtained using various visual features and different P and Q values
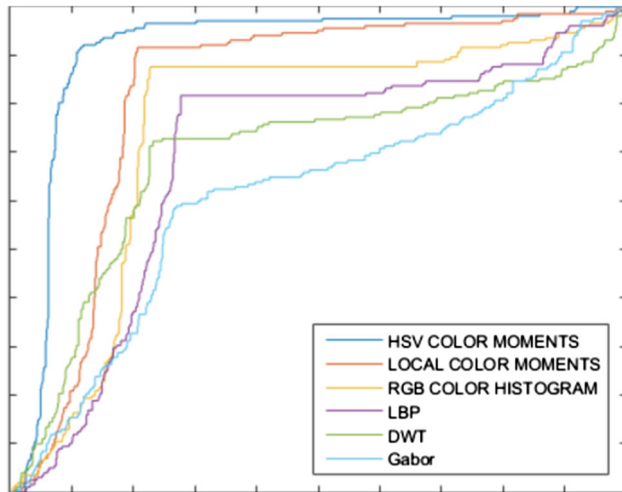

Fig. 7 ROC curves for the classification results obtained using $P = 7$ and $Q = 8$

results confirm the results in Fig. 5. The HSV color moments feature outperforms the other features and yields the best performance.

In the following, we compare the classification performance of the proposed system to those obtained using the widely used *SVM* classifier. In addition, we compare our approach to the typical *KNN* where no clustering is conducted. The number of neighbors in the *KNN* classifier is set to 3. Thus, we deploy the typical *KNN* and *SVM* classifiers using all visual features on the same dataset, and we report the obtained accuracy values in Table 1. As one can notice, the proposed system outperforms *KNN*- and *SVM*-based classification for all visual features. Besides, the *HSV* color moment yields the best performance for all classifiers.

Figure 8 displays the ROCs for typical *KNN* and *SVM* classifiers, and the proposed approach obtained using the color moment descriptor. On the other hand, Table 2 shows the corresponding accuracy results. As it can be seen, the typical *KNN*-based detection gives the worst results. This is due to the fact that MBS appear in any place in the gastrointestinal track, so the surrounding areas show high variance. On the other hand, the proposed approach partitions the frames, such as frames with the same surrounding area are grouped into one cluster. This clustering phase improved the accuracy of detecting MBS in the test frames. Moreover, typical *SVM*-

Table 1 Accuracy obtained using KNN, SVM classifiers and our proposed system with different visual features

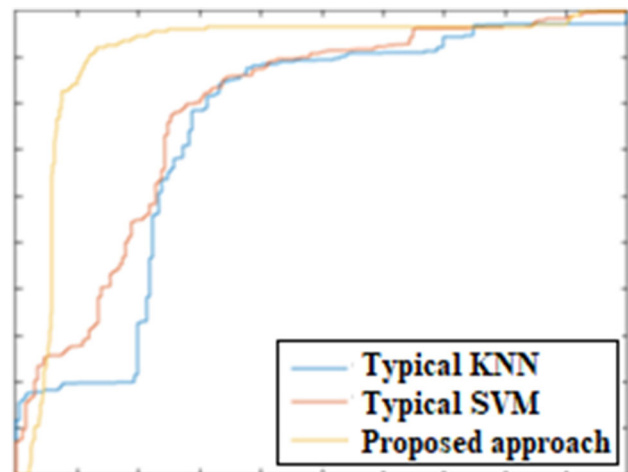| Feature	KNN	SVM	Proposed approach
HSV color moments	0.7341	0.8133	0.9092
Local color moments	0.4810	0.4716	0.8213
RGB color histogram	0.6332	0.5938	0.7496
LBP	0.4769	0.4543	0.6887
DWT	0.6223	0.6605	0.6889
Gabor	0.4356	0.3725	0.6081

Fig. 8 ROC curves obtained KNN, SVM classifiers and the proposed approach along with the color moments feature

Table 2 Performance results obtained using KNN, SVM classifiers and the proposed approach along with the color moments feature

	KNN	SVM	Proposed approach
Accuracy	0.7341	0.8133	0.9092

based detection gives lower performance than our approach because *SVM* is mainly a binary classifier, and although the MBS detection problem seems to be binary with "normal" and "MBS" classes, the actual visual properties of the multiple bleeding spots show high variance depending on their location in the gastrointestinal tract. In other words, the "MBS" class contains subcategories corresponding to the location with respect to the gastrointestinal tract.

In order to validate the results obtained in the experiments conducted previously, we performed a statistical test to compare the detection results obtained using the three classifiers. We compared the accuracy of our proposed approach using the *t* test method [32] twice. First, we test the accuracy of our

approach compared to the accuracy of the typical *KNN* classifier accuracy. Then, we compared it to the accuracy of the typical *SVM*-based classification. Both cases gave decision result of 1, when the confidence level is set to 0,05. In other words, the results, obtained using our proposed approach, are statistically significant compared to the typical *KNN-* and *SVM*-based results.

5 Conclusions

In this paper, we proposed a new multiple bleeding spots detection approach. The first stage of the proposed multiple bleeding detection system consists in extracting several visual descriptors from each WCE frame in order to capture the visual properties of the frames containing multiple bleeding spots. Then, an unsupervised learning technique is applied to reduce the similarity computation cost and shrink the search space when we classify new WCE frames. We evaluated the proposed system using real WCE frames captured using Imaging PillCam® SB [23]. We investigated different color and texture feature extraction techniques and applied them to "normal" and "multiple bleeding spots" frames. The results showed that the HSV color moments descriptor represents the best visual characteristic of the multiple bleeding spots. In addition, we compared the performance of the proposed multiple bleeding detection system with the typical *KNN* and *SVM* classifiers-based detection. The results were promising and showed that the proposed system outperforms these state-of-the-art methods.

References

1. US Department of Health and Human Services: Opportunities and Challenges in Digestive Diseases Research: Recommendations of the National Commission on Digestive Diseases. National Institutes of Health, Bethesda (2009)
2. Iddan, G., et al.: Wireless capsule endoscopy. Nature **405**, 417 (2000)
3. Serej, N.D., et al.: A robust keypoint extraction and matching algorithm based on wavelet transform and information theory for point-based registration in endoscopic sinus cavity data. Signal Image Video Process. **10**(5), 983–991 (2016)
4. Martinez-Herrera, S.E., et al.: Identification of precancerous lesions by multispectral gastroendoscopy. Signal Image Video Process. **10**(3), 455–462 (2016)
5. Xiao, Wu, et al.: Automatic hookworm detection in WCE images. IEEE Trans. Med. Imaging **35**(7), 1741–1752 (2016)
6. Winstone, B., et al.: Toward bio-inspired tactile sensing capsule endoscopy for detection of submucosal tumors. IEEE Sens. J. **17**(3), 848–857 (2017)
7. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**, 788–791 (1999)
8. Bashar, M., et al.: Automatic detection of informative frames from WCE images. Med. Image Anal. **14**, 449–470 (2010)
9. Suykens, J.A., Vandewalle, J.: Least squares support vector machine classifiers. Neural Process. Lett. **9**, 293–300 (1999)
10. Li, B., Meng, M.-H.: WCE images enhancement using contrast driven forward and backward anisotropic diffusion. In: Proceedings IEEE International Conference on Image Processing, pp. 437–440 (2007)
11. Yeh, J.Y., et al.: Bleeding and ulcer detection using WCE images. J. Softw. Eng. Appl. **7**, 422 (2014)
12. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**, 273–297 (1995)
13. Quinlan, J.R.: Simplifying decision trees. Int. J. Man Mach. Stud. **27**, 221–234 (1987)
14. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. **5**, 115–133 (1943)
15. Khun, P.C., et al.: Feature selection and classification for wireless capsule endoscopic frames. In: Proceedings of International Conference on Biomedical and Pharmaceutical Engineering, pp. 1–6 (2009)
16. He, D.-C., Wang, L.: Texture unit, texture spectrum, and texture analysis. IEEE Trans. Geosci. Remote Sens. **28**, 509–512 (1990)
17. Meyer, Y.: Wavelets and Operators, vol. 1. Cambridge University Press, Cambridge (1995)
18. Giritharan, B., et al.: Bleeding detection from capsule endoscopy videos. In: Proceedings of 30th International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4780–4783 (2008)
19. Shah, S.K., et al.: Classification of bleeding images in WCE using HSI color domain and region segmentation. In: Conference URINE ASEE (2007)
20. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 603–619 (2002)
21. Li, B., et al.: Computer-aided detection of bleeding regions for capsule endoscopy images. IEEE Trans. Biomed. Eng. **56**, 1032–1039 (2009)
22. Fu, Y., et al.: Computer-aided bleeding detection in WCE video. IEEE J. Biomed. Health Inform. **18**, 636–642 (2014)
23. Karargyris, A., Bourbakis, N.: A methodology for detecting blood-based abnormalities in WCE videos. In: IEEE International Conference on BioInformatics and BioEngineering, pp. 1–6 (2008)
24. Alotaibi, S., Qasim, S., Bchir, O., Ismail, M.M.B.: Empirical comparison of visual descriptors for multiple bleeding spots recognition in wireless capsule endoscopy video. In: Computer Analysis of Images and Patterns, pp. 402–407 (2013)
25. Li, C., et al.: Online redundant image elimination and its application to wireless capsule endoscopy. Signal Image Video Process. **8**(8), 1497–1506 (2014)
26. Dr Khoroo's Medical Clinic/Trust. online: http://www.drkhuroo.in/#, [Sep. 6, 2015]
27. Manjunath, B., Salembier, P., Sikora, T.: Introduction to MPEG 7: Multimedia Content Description Language. Wiley, Hoboken (2002)
28. Iakovidis, D.K., Tsevas, S., Polydorou, A.: Reduction of capsule endoscopy reading times by unsupervised image mining. Comput. Med. Imaging Graph. **34**, 471–478 (2010)
29. Bezdek, J.C., Ehrlich, R., Full, W.: FCM: the fuzzy c-means clustering algorithm. Comput. Geosci. **10**, 191–203 (1984)
30. Okun, O., Priisalu, H.: Unsupervised data reduction. Signal Process. **87**, 2260–2267 (2007)
31. Mitchell, T.: Machine Learning. McGraw Hill, New York (1997)
32. Davis, J.C., Sampson, R.J.: Statistics and data analysis in geology, vol. 646. Wiley, Hoboken (1986)