CE Video Summarization Using Relational Motion Histogram Descriptor

Mohamed Maher Ben Ismail and Ouiem Bchir
College of Computer and Information Sciences, Computer Science department, King Saud University, Riyadh, KSA
Email: {mbenismail, obchir}@ksu.edu.sa

***Abstract*—We propose a capsule endoscopy summarization system relying on two main components. The first one consists of a semi-supervised Clustering and Local Scale Learning (SS-LSL) algorithm which is used to group video frames into prototypical clusters that summarize the video scene. The second component of the system relies on a novel Relational Motion Histogram (RMH) descriptor that is designed to represent local motion distribution between two contiguous frames. The main idea is to identify "highlight" frames which contain typical variations within the frame collection. These variations are due to different pathologies, small tumors and other subtle abnormalities of the small intestine, etc. The proposed video summarization system is trained, field-tested, evaluated, and compared through a large-scale cross-validation experiment.**

***Index Terms*—capsule endoscopy, semi-supervised clustering, relational clustering, motion descriptor**

I. Introduction

Capsule endoscopy (CE) emerged as the latest imaging modality for screening the small intestine in order to detect pathologies such as Colorectal cancer (CRC) [1]. The patient swallows the endoscopy capsule which starts capturing consecutive views of the digestive system. The video is real-time-saved into a portable device placed on the patient himself. CE procedure can produce up to 60,000 images for each examination, which cause a time consuming and attention intensive task for physicians. The earliest efforts in this area were directed towards enhancing the performance of the endoscopy capsule itself [2]. The CE has been an extremely valuable addition to the diagnostic armamentarium available for the evaluation and management of CRC pathologies.

The main goal of endoscopy video summarization is the development of learning techniques that capture the visual descriptors such as color, texture, and shape present in the extracted frames, and group them into homogeneous categories. In [3], the authors proposed a video summarization approach based on detecting video boundary among the video sequence through finding local maximal value along the dissimilarity curve of the CE video. Then, they used a simple k-means algorithm [4] to obtain key frames from different segments. However, k-means algorithm is a hard-partitioning clustering algorithm, and does not handle overlapping clusters. The proposed method in [5] takes into consideration overlapping clusters by using the Fuzzy c-means (FCM) [6] algorithm to cluster video frames. The obtained groups of video frames and their respective representatives summarize the training data and can be used as the basis for digestive pathologies detection. However, the clustering problem in this application is not trivial as it involves high dimensional and possibly multi-modal features. In [7], the researchers summarize the endoscopy video using non-negative matrix factorization (NMF) [8] technique by keeping the most representative. The authors in [9] proposed a two-stage preprocessing approach in order to remove irrelevant frames in CE videos. First, frames of gastric juice are eliminated using local HS histogram features. Then, the bubbles frames in the CE video are removed by combining Color Local Binary Patterns (CLBP) algorithm with Discrete Cosine Transform (DCT). The K Nearest Neighbor (KNN) classifier is used in both stages of the approach. Lately, this work was extended in [10], and the authors proposed the color uniform local binary pattern (CULBP) feature which includes two kinds of patterns. Namely, the color norm patterns and the color angle patterns. Also the authors proposed the Ada-SVM classifier in order to improve the system accuracy. In [11], the authors proposed a CE video segmentation approach which relies on an unsupervised learning approach. Namely, they used the probabilistic latent semantic analysis (pLSA) model as clustering approach, along with the Scale Invariant Feature Transform (SIFT) as low-level features. The approach showed promising results, however the clustering process suffers from the curse of dimensionality and the sensitivity to noise samples. Similarly, in [12], a framework for CE video segmentation is proposed. It starts by representing each frame using one feature vector by combining color, texture, and motion information. Then, the video is segmented using supervised learning algorithms. Also the authors in [13] outlined a novel scheme to categorize CE video sequences with respect to abnormalities. The aim of their work was to assist physicians in their diagnosis, and save their time. The approach relies on multi-feature extraction and fusion technique to detect key-frames. However the proposed framework requires prior knowledge including videos of patients containing normal and abnormal symptoms. We believe that the independence assumption between frames in the state-of-the-art approaches does not hold. Thus, we propose a

Manuscript received February 2, 2015; revised June 12, 2015.

doi: 10.18178/joig.3.1.34-39

novel relational descriptor to capture the motion flow between successive frames. Moreover, these relational descriptors are provided as input to a new relational semi-supervised clustering algorithm where side information is used not only to reward or penalize a traditional relational fuzzy clustering algorithm decision, but is also used to affect the computation of the local distance measure.

In this paper, we propose an endoscopy video summarization system based on novel relational motion histogram descriptor (RMH) that is designed to represent the local motion distribution between two contiguous frames, and a Semi-supervised Clustering and Local Scale Learning algorithm. Every video frame is represented using its RMH descriptor. Then, the learning algorithm partitions the video frames into typical and non-typical frames in a semi-supervised manner based on their Relational Motion Histogram descriptor, and generates "highlight" frame collection as summary video.

II. Semi-Supervised Relational Clustering with Local Scaling Parameter

Let $x_1, \ldots, x_N$ be a set of N data points and let $R=[r_{jk}]$ be a relational matrix where r_{jk} represents the distance between x_j and x_k. We assume that some partial information is available and let SL be the indicator matrix for the set of "Should-Link" pairs of constraints such that $SL(j,k)=1$ means that x_j and x_j should be assigned to the same cluster and 0 otherwise. Similarly, let SNL be the indicator matrix for the set of "Should not-Link" pairs such that $SNL(j,k)=1$ means that x_j and x_k should not be assigned to the same cluster and 0 otherwise. The Semi-Supervised Local Scaling Learning (SS-LSL) minimizes the following multi-term objective function:

$$\begin{aligned} J &= \sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{k=1}^{N} u_{ij}^m u_{ik}^m \left(1-\exp\left(-\frac{r_{jk}}{\sigma_i}\right)\right) \\ &-w\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{k=1}^{N} u_{ij}^m u_{ik}^m \left(SL(\mathrm{j,k})-\mathrm{SNL(j,k)}\right)-\sum_{i=1}^{C}\frac{K_1}{\sigma_i^2} \end{aligned} \tag{1}$$

Subject to $0 \le u_{ij} \le 1$ and $\sum_{i=1}^{C} u_{ij} = 1 \text{for } j \in \{1,\ldots,N\}$.

It is an extension of the LSPL algorithm [key-14] that incorporates partial supervision. As in the LSPL [14] objective function, the first term in (1) seeks compact clusters, and the last term is a regularization term to avoid the trivial solution where all σ_i are infinitely large. The second term in (1) is a reward term for satisfying "Should-Link" constraints. It is constructed in such a way that the reward between nearby "Should-Link" points is higher than that between distant ones. The third term is a penalty for violating "Should Not-Link" constraints. It is constructed in such a way that the penalty between nearby "Should not-Link" points is higher than distant points one. In (1), the weight $\mathrm{w} \in (0,1)$provides a way of specifying the relative importance of the "Should-Link" and "Should Not-Link" constraints compared to the sum of inter-cluster distances. In our approach, we fix it as the ratio of the number of constraints to the total number of points.

In order to optimize (1) with respect to σ_i, we assume that σ_i's are independent from each other and reduce the optimization problem to C independent problems. As the reward and penalty terms do not depend on the scaling parameters σ_i explicitly, setting the derivative of J with respect to σ_i gives the same update equation for σ_i as the LSPL algorithm [14]

$$\sigma_i = \left(\frac{K}{\sum_{j=1}^{N}\sum_{k=1}^{N} u_{ij}^m u_{ik}^m r_{jk}}\right)^{\frac{2}{2+p}} \tag{2}$$

where

$$K = K_1 \frac{2\pi^{2-\frac{p}{2}}|N|}{N} \tag{3}$$

In (3), p is the size of the manifold, and $|N|$ is the cardinality of the neighborhood of j.

In order to optimize (1) with respect to u_{ij}, we rewrite the objective function in (1) as in [14];

$$J = \sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{k=1}^{N} u_{ij}^m u_{ik}^m \hat{D}_{jk}^{i} - \sum_{i=1}^{C}\frac{K_1}{\sigma_i^2} \tag{4}$$

where $\hat{D}_{jk}^{i} = D_{jk}^{i} - w\,SL(\mathrm{j,k}) + w\,SNL(\mathrm{j,k})$, and

$D_{jk}^{i} = \left(1-\exp\left(-\frac{r_{jk}}{\sigma_i}\right)\right)$ is the distance between x_j and x_k using the scaling parameter of cluster i.

$\hat{D}_{jk}^{i}$ can be regarded as the "effective distance" that takes into account the satisfaction and violation of the constraints. For instance, if the pair of points (x_j, x_k) are supposed to be "Should-Link" then $\begin{cases} SL(\mathrm{j,k})=1 \\ SNL(\mathrm{j,k})=0 \end{cases}$.

In this case, the effective distance $\hat{D}_{jk}^{i}$ reduces to $\hat{D}_{jk}^{i} = D_{jk}^{i} - w.$. In other words the actual distance is reduced to help in keeping these points within the same cluster and thus, maintaining the satisfaction of the constraints. Similarly, if a pair of points (x_j, x_k) are supposed to be "Should Not-Link" then $\begin{cases} SL(\mathrm{j,k})=0 \\ SNL(\mathrm{j,k})=1 \end{cases}$.

Thus, the effective distance $\hat{D}_{jk}^{i}$ defined above becomes

$$\hat{D}_{jk}^{i} = D_{jk}^{i} + w \tag{5}$$

That is, the actual distance is increased to help in preventing these points from being assigned to the same cluster.

It can be shown that optimization of J w.r.t u_{ij} yields

$$u_{ij} = \frac{1}{\sum_{t=1}^{C}\left(\frac{d_{ij}^2}{d_{tj}^2}\right)^{\frac{1}{m-1}}} \tag{6}$$

where

$$d_{ik}^2 = \left(\hat{D}^i v_i \right)_k - \frac{v_i^t \hat{D} v_i}{2} \quad (7)$$

and

$$v_i = \frac{\left(u_{i1}^m, \ldots, u_{iN}^m \right)^t}{\sum_{j=1}^{N} u_{ij}^m} \quad (8)$$

SS-LSL algorithm is summarized below.

Algorithm 1: SS-LSL algorithm

Fix number of clusters C and m $\in [1\ \infty)$;
Initialize K_1;
Initialize the fuzzy partition matrix U;
Initialize the scaling parameter σ_i to 1;
Create the sets of pairwise constraints SL and SNL;
Repeat

- Compute D and $\hat{D}^i$ for all clusters.
- Compute v_i **using** (8).
- Compute the distances **using** (7).
- Update the fuzzy memberships **using** (6).
- Update the scaling parameter σ_i **using** (2).

Until (fuzzy membership do not change)

III. Relational Motion Histogram (RMH) Descriptor

The Relational Motion Histogram (RMH) descriptor is designed to represent the local motion distribution between two sequence of frames $\left(F^i, F^j\right)$. It divides the frame space into 4×4 sub-frames and represents the local motion distribution of each sub-frame by a histogram. In order to generate histograms, motions in all sub-frames are categorized into five types; vertical, horizontal, diagonal, anti-diagonal, and non-defined motion, resulting in a total of 5×16=80 histogram bins.

Let $\left\{\left(F_s^i, F_s^j\right)\right\}_{s\in1\ldots16}$ be the 4×4 sub-sequences resulting from dividing F^i and F^j. For each two sub-sequences $\left(F_s^i, F_s^j\right)$, we compute 5 probabilities, namely the vertical motion probability p_{ijs}^v, the horizontal motion probability p_{ijs}^h, the diagonal motion probability p_{ijs}^d, the anti-diagonal motion probability p_{ijs}^a, and the non-defined motion probability p_{ijs}^n.

p_{ijs}^v is defined as the ratio of the number of pixels of sub-frame F_s^i that moved vertically in its corresponding sub-frame F_s^j over the total number, N_s, of pixels in the sub-frame s. Similarly, p_{ijs}^h, p_{ijs}^d and p_{ijs}^a, are defined as the ratio of the number of pixels of sub-frame F_s^i that moved horizontally, diagonally, anti-diagonally, respectively, in the corresponding sub-frame F_s^j over the total number of pixels in the sub-frame. The non-defined motion probability p_{ijs}^n is the ratio of the number of pixels that are not common to the two sequences of frames $\left(F^i, F^j\right)$ over the total number of pixels in the sub-frame.

In order to determine the number of pixels of sub-frame F_s^i that moved along the 4 directions in its corresponding sub-frame F_s^j, we use the components of the velocity $\left(\mathrm{V}_x, \mathrm{V}_y\right)$ of the pixel at position (x, y) [15]. In fact, the velocity computes the motion between two sequences $\left(F^i, F^j\right)$ at every pixel position (x, y) [16]. It is the solution of (9)

$$\frac{\partial F^i}{\partial x} V_x + \frac{\partial F^i}{\partial y} V_y = F^i(\mathrm{x,y}) - F^j(\mathrm{x,y}) \quad (9)$$

Equation (9) is not sufficient to compute $\left(\mathrm{V}_x, \mathrm{V}_y\right)$, another set of equations is needed, given by some additional constraint. Several existing optical flow methods introduce additional conditions for estimating $\left(\mathrm{V}_x, \mathrm{V}_y\right)$ [15], [16]. In this work, we adopt the method in [15].

Given $\left(\mathrm{V}_x, \mathrm{V}_y\right)$ with respect to each position in the frame (x, y), we compute p_{ijs}^v, p_{ijs}^h, p_{ijs}^d and p_{ijs}^a using the inverse trigonometric function arctan. Fig. 1 displays how the range of usual principal value, in radians, is divided in order to categorize each direction. This categorization is expressed by (10), (11), (12) and (13).

$$\left\{\begin{array}{l} -\frac{\pi}{2} \le \arctan\left(\frac{V_y}{V_x+\varepsilon}\right) < -\frac{3\pi}{8} \\ \quad or\ p_{ijs}^v = p_{ijs}^v + \frac{1}{N_s} \\ \frac{3\pi}{8} \le \arctan\left(\frac{V_y}{V_x+\varepsilon}\right) < \frac{\pi}{2} \end{array}\right. \quad (10)$$

$$-\frac{\pi}{8} \le \arctan\left(\frac{V_y}{V_x+\varepsilon}\right) < \frac{\pi}{8} \qquad p_{ijs}^h = p_{ijs}^h + \frac{1}{N_s} \quad (11)$$

$$\frac{\pi}{8} \le \arctan\left(\frac{V_y}{V_x+\varepsilon}\right) < \frac{3\pi}{8} \qquad p_{ijs}^d = p_{ijs}^d + \frac{1}{N_s} \quad (12)$$

$$-\frac{3\pi}{8} \le \arctan\left(\frac{V_y}{V_x+\varepsilon}\right) < -\frac{\pi}{8} \qquad p_{ijs}^a = p_{ijs}^a + \frac{1}{N_s} \quad (13)$$

where ε is a small real number used to avoid a null denominator.

A non-defined motion pixel is a pixel with null velocity components $\left(V_x = 0 \;\; and \;\; V_y = 0\right)$ such that it is not common to the two sequences $\left(F^i(\mathrm{x,y}) \neq F^j(\mathrm{x,y})\right)$. Thus, the the non-defined motion probability p_{ijs}^n can be computed as follows:

$$F^i(\mathrm{x,y}) \neq F^j(\mathrm{x,y}) \;\; and \;\; V_x = V_y = 0 \qquad p_{ijs}^n = p_{ijs}^n + \frac{1}{N_s} \quad (14)$$

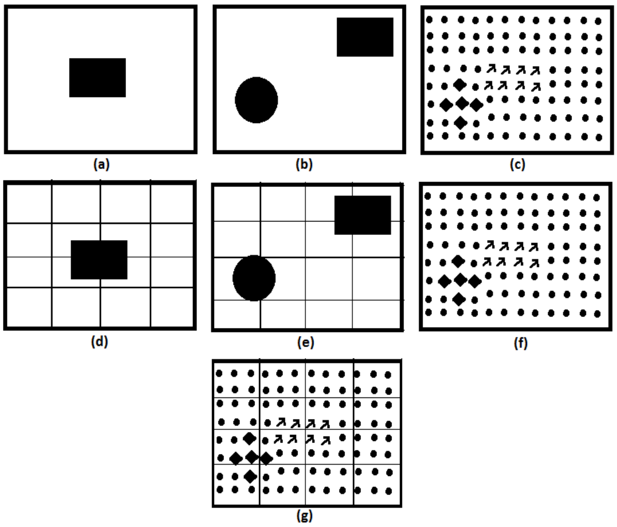

Figure 1. Motion categorization into vertical, horizontal, diagonal, and anti-diagonal directions

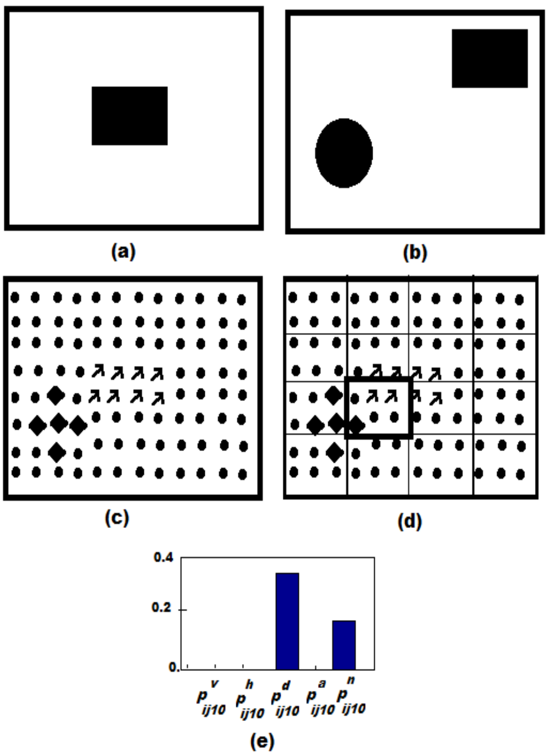

Figure 2. Local motion histogram generation

Fig. 1 illustrates the generation of local motion histogram. Fig. 1a and Fig. 1b are two sequence frames F^i and F^j, respectively. We notice that the rectangular object of frame F^i moved diagonally in F^j, and the circular object is not common to F^i and F^j. Fig. 2c shows the direction of the motions obtained by computing the velocity and representing its direction at each location by an arrow. The circular points represents a null velocity direction, and the diamond points represent the non-defined motion points where $\left(V_x = 0 \;\; and \;\; V_y = 0\right)$ and $\left(F^i(\mathrm{x,y}) \neq F^j(\mathrm{x,y})\right)$. In Fig. 1d we divide the frame space into 4x4 sub-frames. For each sub-frame, we compute a local motion histogram using (10), (11), (12), (13) and (14). Fig. 2d highlights sub-frame 10. We notice that this sub-frame includes 2 diagonal arrows and one diamond over a total of 6 pixels in the sub-frame. This can be expressed in terms of probabilities as $p_{ijs}^d = \frac{2}{6}, p_{ijs}^n = \frac{1}{6}, p_{ij1}^v = p_{ij1}^h = p_{ij1}^a = 0$. Fig. 2e shows the local motion histogram of sub-frame 10.

The resulting RMH of the two contiguous frames $\left(F^i, F^j\right)$ is the concatenation of the five probabilities $\left(p_{ijs}^v, p_{ijs}^h, p_{ijs}^d, p_{ijs}^a \; and \; p_{ijs}^n\right)$ with respect to the 16 blocks. It is defined as follows:

$$RMH\left(F^i, F^j\right) = \left[p_{ij1}^v, p_{ij1}^h, p_{ij1}^d, p_{ij1}^a, p_{ij1}^n, \ldots, p_{ij16}^v, p_{ij16}^h, p_{ij16}^d, p_{ij16}^a, p_{ij16}^n\right] \quad (15)$$

The proposed CE video summarization approach relies on: *(i)* RMH Descriptor extraction; *(ii)* Constraint formulation; and *(iii)* SS-LSL based clustering. RMH descriptors are used to represent the motion captured between two contiguous frames. Then, SS-LSL is used to cluster capsule endoscopy frames into two categories. The frames that belong to the first category represent the transition frames between two different scenes. The second category groups the "smooth" frames where no noticeable motion has been captured. Since this problem involves clustering sparse and high dimensional data. The supervision information consists of pairs of frames that should not be included in the same cluster. These constraints are deduced the RMH descriptors. In fact we consider $RMH\left(F^i, F^j\right)$ does not exceed a given threshold then the frames F^i and F^j should not belong to the same cluster.

Unlike the naive solution that relies on estimating the flow between successive frames, and discarding them if their score is less than a given threshold, the proposed CE video summarization framework along with the proposed RMH feature, and the semi-supervised learning algorithm is able to group visually similar frames in the same cluster even if there are dissimilar frames appear between them. Moreover, the proposed approach solves the problem of finding the appropriate threshold to discard irrelevant frames.

IV. EXPERIEMNTS

In this section, we evaluate the proposed video summarization system on real CE videos. A range of experiments were performed to assess the strengths and weaknesses of the proposed approach. We use 4 CE videos collected on four different patients. Three of them (patient 1, 2 and 4) are showing ulcer and bleeding symptoms. This collection of color CE videos generated 96,963 frames. Table I summarizes the considered CE frames data.

TABLE I. SUMMARY OF THE CAPSULE ENDOSCOPY VIDEO COLLECTION

	Duration	Nb of frames	Nb of ulcer frames	Nbre of bleeding frames
Patient 1	15 min	15377	73	207
Patient 2	16 min	16902	89	232
Patient 3	19 min	23193	0	0
Patient 4	46 min	41491	141	379

Fig. 3 displays representative frames from the obtained clusters on patient 1 data.

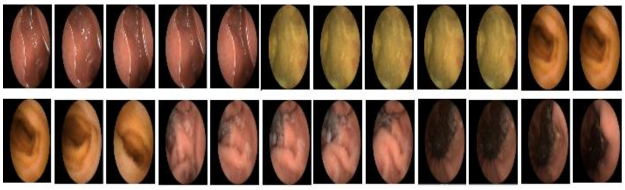

Figure 3. Sample frames from the obtained summary of the CE video captured on patient 1.

SS-LSL clustering based on RMH descriptor achieves reasonable image region clustering. By analyzing and comparing the content of the different clusters generated by the proposed clustering approach, we observed that SS-LSL without supervision cannot find homogeneous clusters. On the other hand, with 5% of supervision constraints, the two clusters reflect the true structure of the frame collection.

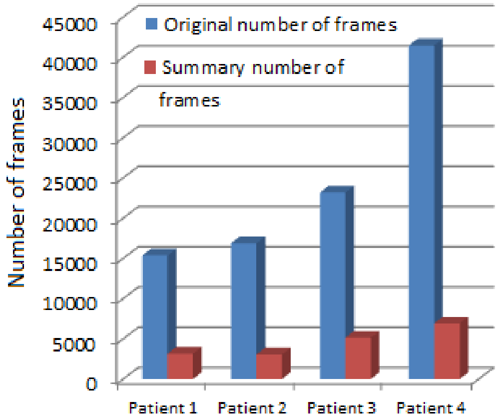

Figure 4. Original and final number of frames of 4 patient CE videos.

The original number of frames extracted from each video, and the final number of summary frames are shown in Fig. 4. As one can see, the final number of frames is 9 times smaller than the original one on average. These results have been obtained with RMH similarity threshold set to 0.5.

In Table II, we report the comparison result of the proposed CE video summarization process with three different approaches. Namely we compare our method with; *(i)* the proposed method when we do not provide supervision information to guide the clustering algorithm, *(ii)* the summarization approach based on estimating RMH feature between successive frames, and discarding frames corresponding to scores below the threshold (0.5), *(iii)* and the method in [17]. As it can be seen the resulting number of frames differs from one approach to the other. Further investigations showed that the obtained clusters obtained using the proposed method match better the expert analysis of the CE video. More specifically, despite the fact that the methods *(i)*, *(ii)* and *(iii)* yield larger number of frames, their content is not more informative than the summary obtained using the proposed method. This can be attributed to the fact the supervision information used to guide the clustering algorithm yield visually homogeneous clusters, and detects more efficiently noise frames.

TABLE II. COMPARISON OF THE RESULTING NUMBER OF FRAMES OBTAINED USING (A) THE PROPOSED METHOD, (B) THE PROPOSED METHOD WITH NO SUPERVISION INFORMATION, (C) THE ESTIMATION OF THE FLOW BETWEEN SUCCESSIVE FRAMES, AND (D) THE METHOD IN [17]

	Original number of frames	Nbre of frames using the proposed method	Nbre of frames using the proposed method (no supervision)	Nbre of frames using the method in [17]
Patient 1	15377	4000	4648	4789
Patient 2	16902	5214	5965	5747
Patient 3	23193	6843	7491	7311
Patient 4	41491	8016	8727	8391

V. CONCLUSION

In this paper, we have proposed a novel summarization system for capsule endoscopy video. The system is based on a Semi-supervised Clustering and Local Scale Learning (SS-LSL) algorithm, and a novel relational motion histogram descriptor (RMH) that is designed to represent the local motion distribution between two contiguous frames. SS-LSL is used to group image regions into prototypical region clusters that summarize the capsule endoscopy frames. The constraints consist of pairs of frames that should not be included in the same cluster. These constraints are deduced from the training frame collection to help in guiding the clustering process. On the other hand, RMH descriptor is intended to identify “highlight” frames which contain typical variations within the frame collection. These variations could be due to different pathologies, small tumors and other subtle abnormalities of the small intestine, etc. The proposed video summarization system has been trained, field-tested, evaluated, and compared using a large-scale cross-validation experiment that uses four endoscopy videos acquired from four patients at different geographic locations.

REFERENCES

[1] J. D. Zeitoun, “A. C. Portal hypertensive duodenal polyp: A case report,” *World J. Gastroenterology*, vol. 13, no 9, pp. 1451-1452, 2007.

[2] M. Mackiewicz, "Capsule endoscopy - state of the technology and computer vision tools after the first decade," in *Book New Techniques in Gastrointestinal Endoscopy*, InTechOpen, 2011.
[3] B. Li, Q. H. Meng, and Q. Zhao, "Wireless capsule endoscopy video summary," presented at the IEEE International Conference on Robotics and Biomimetics, China, 2010.
[4] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," presented at 5th Berkeley Symposium on Mathematical Statistics and Probability, USA, 1967, pp. 281-297.
[5] D. K. Iakovidis, E. Spyrou, and D. Diamantis, "Efficient homography-based video visualization for wireless capsule endoscopy," presented at the IEEE 13th International Conference on Bioinformatics and Bioengineering (BIBE), Greece, 2013.
[6] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.
[7] M. C. Cunha, "MPEG-7 visual descriptors: Contributions for automated feature extraction in capsule endoscopy," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 16, no. 5, pp. 628-637, 2006.
[8] O. A. H. Okun, "Unsupervised data reduction," *Signal Processing*, vol. 87, no. 9, pp. 2260-2267, 2007.
[9] S. Zhe, *et al.*, "Removal of non-informative frames for wireless capsule endoscopy video segmentation," presented at the IEEE International Conference on Automation and Logistics (ICAL), China, 2012.
[10] R. Zhou, *et al.*, "A novel method for capsule endoscopy video automatic segmentation," presented at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Japan, 2013.
[11] S. Yao, P. Guturu, and B. P. Buckles, "Wireless capsule endoscopy video segmentation using an unsupervised learning approach based on probabilistic latent semantic analysis with scale invariant features," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 1, pp. 98-105, 2012.
[12] M. Mackiewicz, "Capsule endoscopy - state of the technology and computer vision tools after the first decade," in *Book New Techniques in Gastrointestinal Endoscopy*, InTechOpen, 2011.
[13] Z. Qian and Q. H. M. Max, "An abnormality based WCE video segmentation strategy," presented at IEEE International Conference on Automation and Logistics - ICAL, Hong Kong, 2010.
[14] O. Bchir and H. Frigui, "Fuzzy relational kernel clustering with local scaling parameter learning," presented at IEEE MLSP, Kittila, Finland, 2010.
[15] R. Battiti, E. Amaldi, and C. Koch, "Computing optical flow across multiple scales: An adaptive coarse-to-fine strategy," *Int. J. Comput. Vis*., vol. 6, no. 2, pp. 133-145, 1991.
[16] H. Liu, T. H. Hong, M. Herman, and R. Chellappa, "Accuracy vs. efficiency trade-offs in optical flow algorithms," presented at 4th Eur. Conf. Computer Vision, Cambridge, U.K., 1996, pp. 174-183.
[17] S. Tsevas, S. Iakovidis, D. E. Maroulis, and E. Pavlakis, "Automatic frame reduction of wireless capsule endoscopy video," presented at 8th IEEE International Conference on BioInformatics and BioEngineering (BIBE), Greece, 2008.

Dr. Mohamed Maher Ben Ismail is assistant professor at the computer science department of the College of Computer and Information Sciences at King Saud University. He received his PhD. degree in Computer Science from the University of Louisville in 2011. His research interests include Pattern Recognition, Machine Learning, Data Mining and Image Processing.

Dr. Ouiem Bchir is assistant professor at the computer science department, College of Computer and Information Sciences (CCIS), King Saud University, Riyadh, Saudi Arabia. She got her PhD from the University of Louisville, KY, USA. Her research interests are Spectral and kernel clustering, pattern recognition, hyperspectral image analysis, local distance measure learning, and Unsupervised and Semi-supervised machine learning techniques. She received the University of Louisville Dean's Citation, the University of Louisville CSE Doctoral Award, and the Tunisian presidential award for the electrical engineering diploma.