J Med Syst (2014) 38:109
DOI 10.1007/s10916-014-0109-y

MOBILE SYSTEMS

# Video summarization based tele-endoscopy: a service to efficiently manage visual data generated during wireless capsule endoscopy procedure

**Irfan Mehmood · Muhammad Sajjad · Sung Wook Baik**

Received: 13 February 2014 / Accepted: 7 July 2014 / Published online: 19 July 2014

**Abstract** Wireless capsule endoscopy (WCE) has great advantages over traditional endoscopy because it is portable and easy to use. More importantly, WCE combined with mobile computing ensures rapid transmission of diagnostic data to hospitals and enables off-site senior gastroenterologists to offer timely decision making support. However, during this WCE process, video data are produced in huge amounts, but only a limited amount of data is actually useful for diagnosis. The sharing and analysis of this video data becomes a challenging task due the constraints such as limited memory, energy, and communication capability. In order to facilitate efficient WCE data collection and browsing tasks, we present a video summarization-based tele-endoscopy service that estimates the semantically relevant video frames from the perspective of gastroenterologists. For this purpose, image moments, curvature, and multi-scale contrast are computed and are fused to obtain the saliency map of each frame. This saliency map is used to select keyframes. The proposed tele-endoscopy service selects keyframes based on their relevance to the disease diagnosis. This ensures the sending of diagnostically relevant frames to the gastroenterologist instead of sending all the data, thus saving transmission costs and bandwidth. The proposed framework also saves storage costs as well as the precious time of doctors in browsing patient's information. The qualitative and quantitative results are encouraging and show that the proposed service provides video keyframes to the gastroenterologists without discarding important information.

This article is part of the Topical Collection on *Mobile Systems*

I. Mehmood · M. Sajjad · S. W. Baik (✉)
College of Electronics and Information Engineering, Sejong University, Seoul, Republic of Korea
e-mail: sbaik@sejong.ac.kr

I. Mehmood
e-mail: irfanmehmood@sju.ac.kr

M. Sajjad
e-mail: sajjad@sju.ac.kr

**Keywords** Wireless capsule endoscopy · Video summarization · Remote monitoring · Visual saliency · Mobile computing

## Introduction

With advancements in remote sensing technology, the quality and accuracy of remote sensing applications have improved significantly. The growing e-health industry uses sensor technologies to monitor patient conditions both locally and remotely. It provides improved patient care through early detection of adverse health conditions and can influence patients' behavior to improve their ongoing health. Power-harvesting technology and improved energy management techniques liberate sensors from bulky power source connections and batteries, allowing them to be used in a wider range of autonomous applications [1, 2]. Disposable low-cost sensor technology is emerging due to miniaturization, embedding and power harvesting. These sensors provide point-of-care monitoring for a broad range of patient conditions [3, 4]. Wireless capsule endoscopy (WCE) [5] is one of the examples of such disposable low-cost sensors. Wireless capsule endoscopy (WCE) is a non-invasive technique that allows doctors to visualize the most inaccessible parts of the human gastrointestinal tract as a means to identify causes of illness. It consists of three components: 1) a disposable wireless capsule, 2) a sensing system with eight antenna arrays, and 3) a battery pack with an image recoding unit (IRU). After a patient swallows the disposable wireless capsule, it travels through the patient gastrointestinal tract by transmitting video frames wirelessly to an IRU worn by the patient. Traditional endoscopy needs hospitalization and specialized medical staff, whereas WCE is free from such compulsions. This property makes it more significant over traditional endoscopy and marks it as a potential candidate for telemonitoring [6–8].

The WCE system allows patients to live an ordinary life during diagnostic procedures. This is a lengthy diagnosis process that lasts approximately 8 h and captures around 50,000 frames.

The underlying value of wireless capsule sensing and monitoring is in the information derived from the data acquisition. The sensing data is important for gastroenterologists to make a correct diagnosis. This collected data is large and contains only a small fraction of informative frames while the rest of the visual data is redundant and non-informative [9]. It is largely agreed that transmission of such huge data is not a feasible solution due to the severe energy and bandwidth constraints of wireless body sensors. This collected data should be processed locally with only relevant data transmitted to gastroenterologists. The processing of such large sensor data on resource constraint WCE becomes a bottleneck. Furthermore, image sensors consume more energy than scalar sensors and streaming of captured WCE videos also requires a large segment of bandwidth that causes a huge amount of power consumption. Addressing the demands of growing WCE data volume and complexity requires a novel framework for efficient managing, and analyzing the data. In this context, video summarization is the most feasible solution. It is the process of reducing video frames to create a summary that retains the visual contents of the original video [10–12]. Ideally, video summarization techniques must utilize high-level semantic details of the WCE video content. However, it is not feasible to extract exact semantically relevant frames (objects such as lesions, bleeding etc.) from WCE videos because of the highly diverse nature of the contents. The majority of the techniques in the literature are therefore either anomalies specific (identification of polyps, bleeding, crohn's disease lesions, etc.) or directly employ low-level features [13–16]. However, the usage of low-level features results in the loss of semantic details, so the summaries generated are not consistent with gastroenterologist perception because of the semantic gap. To cope with these challenges, a visual attention model is used to bridge the gap between the low-level and high-level features. Visual attention is the cognitive process that selectively concentrates on one aspect of the image while ignoring the other contents [17]. In this context, we have developed some state-of-the-art visual attention driven keyframe extraction schemes [10, 11]. However, computational complexity is one of the main drawbacks of these existing summarization techniques. Therefore, these techniques are not feasible for the resource constrained WCE system. To solve this problem, a computationally efficient summarization framework suitable for resource limited devices such as wireless capsule sensors, mobile, etc. is vital.

This paper presents a cost-effective framework for large sensor data management in the tele-endoscopy system. In the proposed framework, a smartphone collects frame sequences from an IRU and performs video summarization to generate keyframes. In parallel, the smartphone also transmits the generated keyframes to the corresponding medical specialists for analysis. The proposed video summarization algorithm consists of three modules. Initially, the collected frame sequences are converted from RGB color space to color opponent-component (COC) space [18]. Then, the integral image is calculated that makes our summarization method suitable for real time video processing. Finally, image features i.e. moments, curvature, and multi-scale contrast are computed and are fused to obtain the saliency map of each frame. A saliency map is used to select keyframes.

## Background

In an effort to reduce the time required for visual inspection of WCE videos, several methods have been presented in the literature. These methods can be classified into two categories: 1) detection and segmentation of gastrointestinal abnormalities and 2) video summarization. For a detailed review of the existing computer-vision based analysis schemes, the readers may refer to a survey [19].

A lot of attention has been paid to the detection and segmentation techniques that assist physicians by reducing the burden of time-consuming and concentration intensive WCE examination [20, 21]. Concerning the segmentation applied to endoscopy video frames, Tjoa et al. [22] utilized chromatic features. This provides better results for the segmentation of the colon since WCE images contain rich color information. Li and Meng [23] have comparatively studied tumor detection in endoscopy images using texture features. These techniques are domain specific and can only identify and segment certain abnormalities such as bleeding, ulceration, etc. Most of the techniques perform well in detecting only one type of abnormality. However, they generally fail to detect multiple abnormalities. Recently, efforts have been made to generate video summaries instead of detecting specific kinds of abnormalities [24–26, 14]. Video summarization of the WCE aims to select those frames that contain abnormalities. Prominent summarization methods can be classified into two categories i.e. clustering and significant content change. In an effort to reduce the time required for the inspection of WCE videos, Iakovidis et al. and Ioannis et al. [27, 26] presented clustering based WCE video summarization schemes, but these schemes are computationally expensive.

In significant content change based methods, comparison is done between consecutive frames based on the differences of their low-level features. Whereas in clustering based methods, the frames are clustered based on low-level features and a single frame is selected from each cluster. The common features used in these schemes are histogram difference, texture, and shape. There exists a significant semantic gap among summaries generated by physicians and by low-level features that leads to unsatisfactory performance. Researchers have used visual attention theories for video summarization to minimize the semantic gap between human vision and low-level methods [28]. The human visual attention model consists of high-level cognitive processes which lead to saliency maps. The saliency map reduces the complexity of scene analysis and focuses on the most relevant objects. In this context, the visual attention model can be used to efficiently summarize the contents of WCE videos.

## Proposed system

In WCE, a patient swallows a wireless capsule for diagnosing intestine diseases. The micro imager of the wireless capsule captures images and sends it to a data logger (DL) through radio frequency transmitter and the sensor array that are fixed at different locations on the anterior abdominal wall. DL is usually fixed in a belt around the patient's waist. Due to limited resources in terms of battery life and computational power, the implementation of high-level signal processing solutions on wireless WCE's DL is not feasible. For efficient and real time analysis, the captured images are forwarded from DL to smartphone. The image sequences received by the smartphone are summarized with a real time and efficient visual attention model. Finally, the summarized images are transferred to the corresponding gastroenterologists as shown in Fig. 1.

Consider a WCE frame sequence $f_t$; $t$=1, 2, 3…, $N_T$, where $N_T$ denotes the total number of frames. Each frame is in RGB color space (that is, R, G, and B represent the red, green and blue channels respectively). The goal is to extract keyframes from the underlying video and transmit those keyframes to the corresponding gastroenterologists.

### Color space conversion

Color information plays a vital role in the analysis of medical images. Perception of color is usually not accurately represented in RGB space. A better model for improved color perception is COC [29]. This color space is in accordance with the human visual system that helps to efficiently select salient objects in an image. We have incorporated this property in the proposed system by converting frame $f_t$ from RGB color space to COC color space. RGB color channels are converted into four broadly-tuned color channels as $R^{\wedge}$=R-(G + B)/2, $G^{\wedge}$=G-(R + B)/2, $B^{\wedge}$=B-(R + G)/2, and $Y^{\wedge}$=(R + G)/2-|R-G|/2-B [30]. Using these four-color channels, two opponent color pairs red-green and blue-yellow are computed as follows:

$$RG_t = R_t^{\wedge} - G_t^{\wedge} \tag{1}$$

$$BY_t = B_t^{\wedge} - Y_t^{\wedge} \tag{2}$$

The intensity channel is also computed and fused with red-green and blue-yellow channels to get a final aggregated image $F_t$ as given in Eqs. 3 and 4. Then, saliency map is holistically calculated only for this aggregated image.

$$I_t = \frac{R_t + G_t + B_t}{3} \tag{3}$$

$$F_t = RG_t + BY_t + I_t \tag{4}$$

### Integral image computation

The integral-image is an attractive concept for those real-time image processing applications which have memory and power constraints, like wireless body sensors and smartphones. Integral-image [31] (or summed area table) is a data structure constructed by replacing each image pixel with a value equal to sum of all pixels above and to the left of the current pixel as follows:

$$\mathcal{F}(x,y) = \sum_{\substack{x' \le x \\ y' \le y}} F_t\left(x', y'\right). \tag{5}$$

Computing an integral-image is a light weight process with time complexity 2WH (where W and H are the width and height of the image, respectively). Once the integral-image is calculated, we can compute the sum of any region in constant time i.e. O(1) rather than O(n^2). This faster computation has been applied to a number of interesting problems [32]. Thus, we can utilize the benefit of the faster computational property of the integral-images to generate video summaries in real-time.

### Visual saliency computation

In this section, a visual saliency is computed to get an efficient video summary. The proposed saliency model is based on three features: image moment, multiscales contrast, and curvature.

#### *Image moment*

A wireless capsule travels through the gastrointestinal tract in natural peristaltic motion. The camera captures mucosal images at different scales and orientations that result in the production of highly redundant data. In order to eliminate this redundant visual data, moments of inertia are used [33]. For this purpose, the four moments of inertia: mean, standard deviation, skewness, and kurtosis are computed from each aggregated image F(t). These statistical moments qualitatively describe the structural shape of a region, its boundaries, texture etc. The four moments: mean, standard deviation, skewness, and kurtosis are computed as:

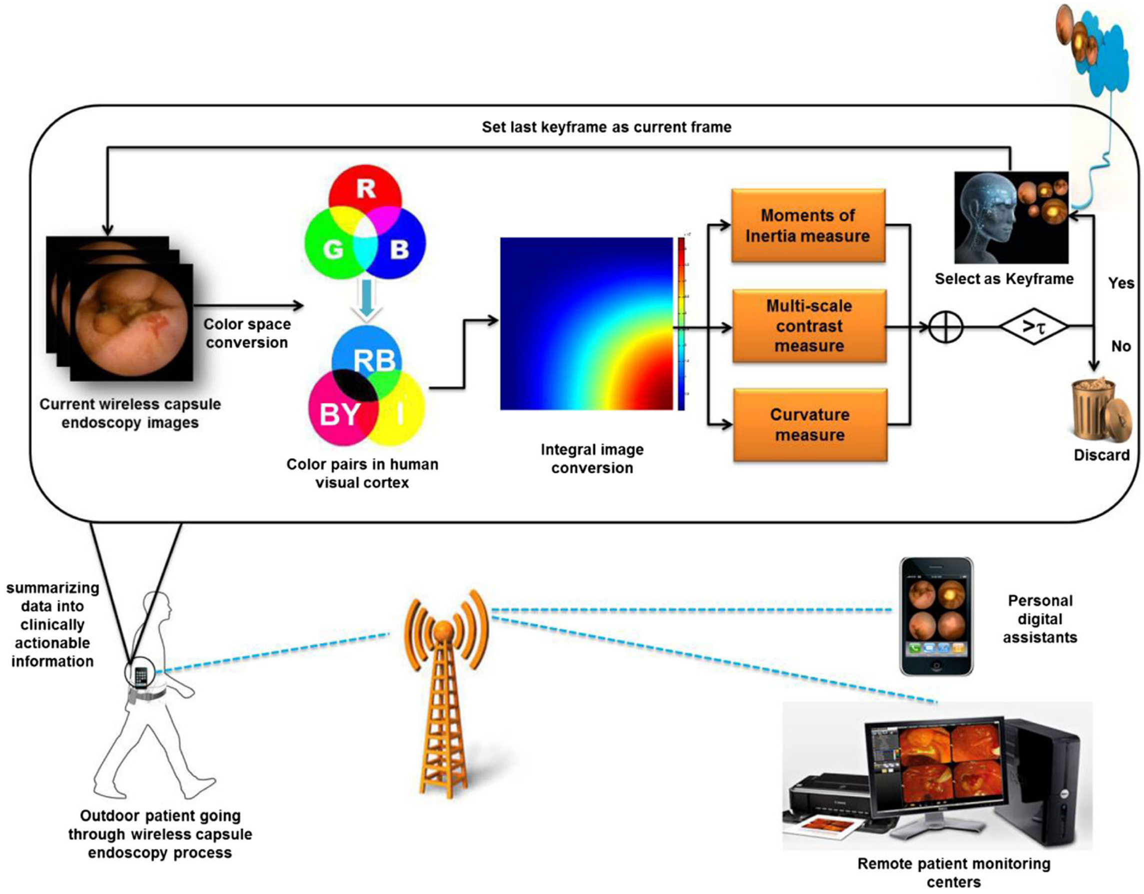
Set last keyframe as current frame
Current wireless capsule endoscopy images
Color space conversion
R
G
B
RB
BY
I
Color pairs in human visual cortex
Integral image conversion
Moments of Inertia measure
Multi-scale contrast measure
Curvature measure
>τ
Select as Keyframe
Yes
No
Discard
summarizing data into clinically actionable information
Outdoor patient going through wireless capsule endoscopy process
Personal digital assistants
Remote patient monitoring centers

**Fig. 1** Overall view of the proposed video summarization based tele-endoscopy service

$$E_t = \frac{\sum_{x=1}^{W} \sum_{y=1}^{H} F_t(x,y)}{W \times H} \tag{6}$$

$$\sigma_t = \sqrt{\frac{\sum_{x=1}^{W} \sum_{y=1}^{H} (F_t(x,y) - E_t)^2}{W \times H}} \tag{7}$$

$$S_t = \frac{\frac{1}{\sigma_t^3} \sum_{x=1}^{W} \sum_{y=1}^{H} (F_t(x,y) - E_t)^3}{W \times H} \tag{8}$$

$$K_t = \frac{\frac{1}{\sigma_t^4} \sum_{x=1}^{W} \sum_{y=1}^{H} (F_t(x,y) - E_t)^4}{W \times H} \tag{9}$$

where $E_t$, $\sigma_t$, $S_t$, and $K_t$ are the mean, standard deviation, skewness, and kurtosis values of frame $F_t$ respectively. The sum of pixels within any rectangular region can be determined with only four array-references using the integral image $\mathbf{F}_t$. This accelerates the computation of moment of inertia features. A function of similarity between the two video frames $F_t$ and $F_{t+1}$ can be defined as:

$$d_{mom}(F_t, \; F_{t+1}) = \sqrt{(E_t - E_{t+1})^2 + (\sigma_t - \sigma_{t+1})^2 + (S_t - S_{t+1})^2 + (K_t - K_{t+1})^2} \tag{10}$$

The resultant distance value $d_{mom}$ is used for removing redundancy in videos.

### *Multi-scale contrast map*

Contrast detection is one of the fundamental properties of the human cognitive process that helps gastroenterologists select the most informative regions in the underlying image. In order to deal with varying size anomalies in WCE frames, multi-scale contrast is used. Multi-scale contrast analyzes WCE frames at different scales that help in efficiently detecting varying size salient objects. For each frame, multi-scale contrast at pixel p(x,y) is computed on aggregated image $F_t$ as:

$$MC^s(x,y) = \|F_t(x,y) - \mu\| \tag{11}$$

$$\mu = \frac{\mathcal{F}(x+N, y+N) + \mathcal{F}(x-N, y-N) - \mathcal{F}(x-N, y) - \mathcal{F}(x, y-N)}{N * N} \tag{12}$$

where s∈ [1 3] is the image' Gaussian pyramid scale and $N$=5 is the neighborhood around pixel p(x,y). μ is computed in constant time that requires operation on only four indexes of integral-image. Finally, multi-scale contrast at a pixel p(x,y) of frame $F_t$ is achieved by adding the contrast at three levels of Gaussian pyramid. This multi-scale contrast produces a gray scale saliency image.

$$MC(x,y) = \sum_{s=1}^{3} MC^s(x,y) \tag{13}$$

*Curvature feature map*

Research in neurosciences suggests that curvature is also an important factor in determining the saliency of an image [34, 35]. Psychophysical studies reveal that curvature influences the gastroenterologist's analysis of the underlying images. Therefore, curvature measurement is used to compute the saliency of each frame. Before computing the image curvature, a Gaussian filter is applied. This efficiently reduces noise by smoothing the overall image.

$$G_t = F_t(x,y) * \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2+y^2 \big/ 2\upsilon^2} \tag{14}$$

Then, first order derivative of $G_t$ is computed that produces a two dimensional column vector as mentioned in equation 15. This vector has an important geometric property that highlights curvature (that is the direction of maximum rate of change) of $G_t$.

$$C = \nabla(G_t) = \begin{bmatrix} G_t^x \\ G_t^y \end{bmatrix} = \begin{bmatrix} \dfrac{\partial G_t}{\partial x} \\ \\ \dfrac{\partial G_t}{\partial y} \end{bmatrix} \tag{15}$$

These partial derivatives are not rotational invariant, however their magnitude is. Thus, to get a rotational invariant curvature map, magnitude of curvature vector C is calculated as:

$$C_{mag} = mag(\nabla(G_t)) = \sqrt{\left(G_t^x\right)^2 + \left(G_t^y\right)^2} \tag{16}$$

The Hu's image moments, multi-scale contrast, and curvature measures are normalized in the range [0 1] and are fused to get a final saliency map. The average of non-zero pixel values is considered as a saliency value. These saliency values are then computed for each frame in the video. The proposed strategy for keyframe extraction is based on the comparison of frames' saliency values. A new keyframe is selected if there is a significant change between the saliency values of the current frame and the previous keyframe. The significant change is a kind of threshold, and users can adjust it according to the level of details they require in summaries.

## Experiments and results

Experiments were conducted to validate the effectiveness of the proposed method. For evaluation, videos were collected from Gastorlab [36] and WCE Video Atlas [37]. In order to accurately evaluate the proposed framework, three evaluation schemes have been used. The subsequent sections provide details of each evaluation scheme.

### Significance of the proposed framework

In this section, the results of the proposed method have been presented on a single shot video downloaded from SYNMED,[1] ,[2] In this video shot, wireless capsule captures phlebectasia while recording the small bowl of a patient. Phlebectasia is an abnormality caused due to unnatural dilation of veins in human body. This abnormity appears as a dark spot in this video shot. The attention curves of the Hu's image moments, multi-scale contrast, curvature and fused saliency curves are shown in Fig. 2. The phlebectasia captured in this video is a salient object and the multi-scale contrast and curvature saliency efficiently highlight this abnormality. Also, it is evident from the fused attention curve that the frames containing phlebectasia have high attention values and are selected as a keyframe.

### Subjective evaluation

In this section, the comparison of the proposed summarization framework has been performed with state-of-the-art video summarization techniques. This comparison is based on a subjective rating because the video summarization quality assessment is inherently subjective. For this purpose, two gastroenterologists were requested to select significant frames in each video. The frames selected by gastroenterologists thus acted as a ground truth. The frames selected by a particular summarization scheme were then compared with the ground truth to find Recall, Precision, and F-measure. Recall is the ratio of the number of

[1] http://www.synmed.co.uk/index.htm

[2] http://www.youtube.com/watch?v=zBYbFQzldtU

relevant frames chosen as keyframes to the total number of relevant frames in the ground truth database. Precision is the ratio of the number of relevant frames chosen as keyframes to the total number of relevant and irrelevant frames selected as keyframes. Recall and Precision are complementary with each other. Thus, to get a better insight of the comparative analysis, F-measure was also calculated. A value of F-measure close to 1 indicates high values of both Recall and Precision and vice versa. These three matrices are computed as:

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{17}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{18}$$

$$\mathrm{Recall} = 2\left(\frac{\mathrm{Recall} * \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}}\right) \tag{19}$$

where true positive (TP) is a frame selected by both the gastroenterologists and the technique. False negative (FN) is a frame chosen as keyframe by the gastroenterologist but not by the technique. False positive (FP) is a frame selected as keyframe by the technique but not by the gastroenterologist.

The proposed method was compared with our general purpose summarization technique [10] and a domain specific endoscopy video summarization scheme presented by Pan et al. [14]. It can be seen from Table 1 that the proposed technique outperformed other mentioned techniques by yielding higher values of Recall, Precision, and F-measure. This extraordinary performance is achieved primarily because of the saliency model. Image moments effectively removed redundant frames. Multi-scale contrast and curvature measures had precisely identified the important (abnormal) areas in the frame as shown in Fig. 3. The visual results show that the proposed saliency model was very effective in detecting various gastrointestinal abnormalities. This saliency map had a key-contribution in the overall summarization performance of the proposed scheme.

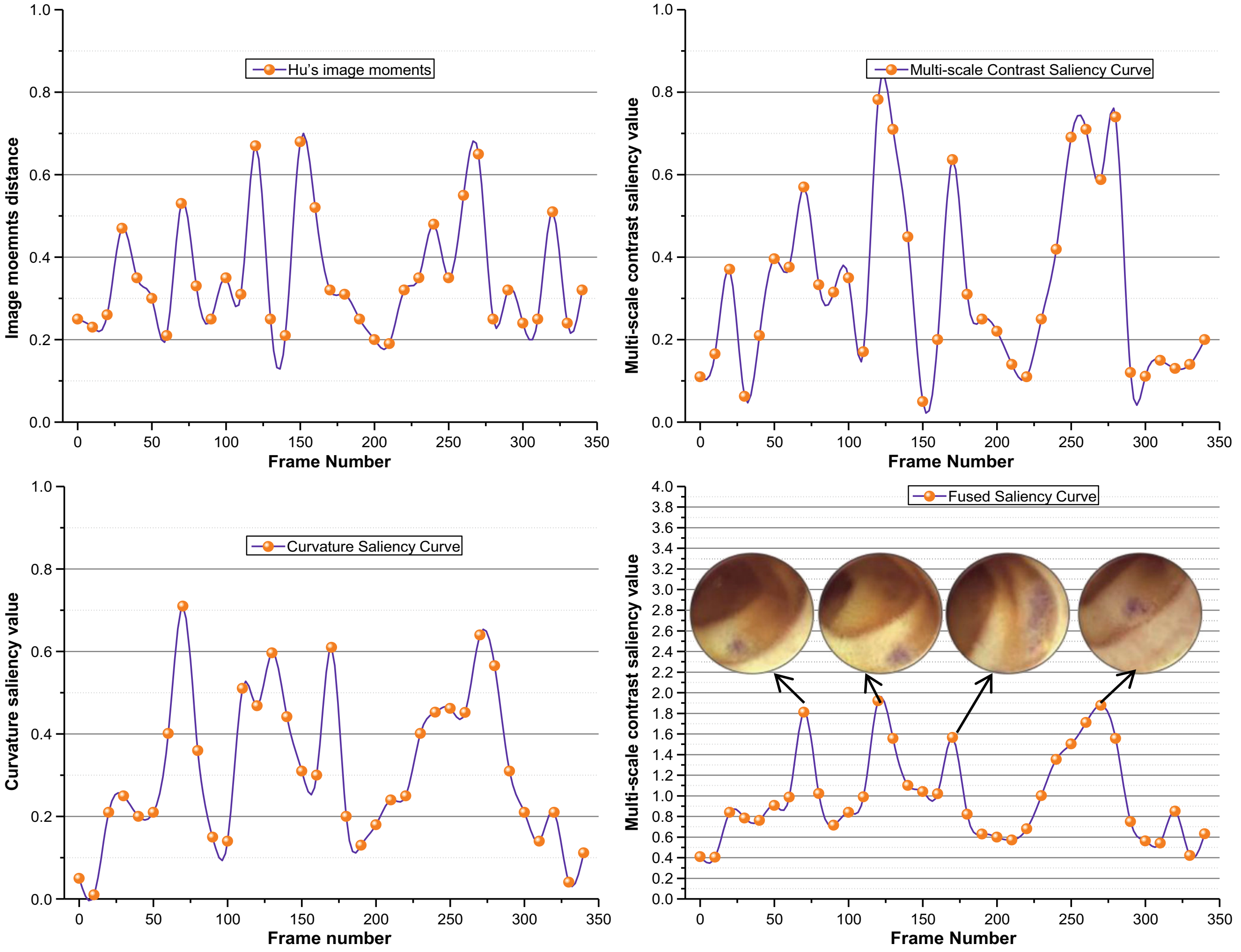

**Fig. 2** The visual attention curves for the sample WCE video shot

**Table 1** Recall (R), Precision (P), and F-measure (F) achieved by different summarization techniques

| | Pan et al. [14] | | | Ejaz et al. [10] | | | Proposed | | |
|---|---|---|---|---|---|---|---|---|---|
| Video No. | R | P | F | R | P | F | R | P | F |
| 1 | 0.63 | 0.71 | 0.67 | 0.71 | 0.74 | 0.72 | 0.75 | 0.8 | 0.77 |
| 2 | 0.67 | 0.66 | 0.66 | 0.75 | 0.80 | 0.77 | 0.83 | 0.91 | 0.87 |
| 3 | 0.72 | 0.75 | 0.73 | 0.83 | 0.85 | 0.84 | 0.91 | 0.88 | 0.89 |
| 4 | 0.75 | 0.73 | 0.74 | 0.80 | 0.74 | 0.77 | 0.8 | 0.82 | 0.81 |
| 5 | 0.69 | 0.72 | 0.70 | 0.77 | 0.80 | 0.78 | 0.85 | 0.83 | 0.84 |
| 6 | 0.66 | 0.70 | 0.68 | 0.75 | 0.81 | 0.78 | 0.84 | 0.87 | 0.85 |
| 7 | 0.71 | 0.68 | 0.69 | 0.85 | 0.82 | 0.83 | 0.89 | 0.92 | 0.90 |
| 8 | 0.72 | 0.74 | 0.73 | 0.77 | 0.80 | 0.78 | 0.89 | 0.91 | 0.90 |
| 9 | 0.70 | 0.67 | 0.68 | 0.79 | 0.82 | 0.80 | 0.86 | 0.81 | 0.83 |
| 10 | 0.61 | 0.65 | 0.63 | 0.78 | 0.75 | 0.76 | 0.84 | 0.82 | 0.83 |
| Average | 0.68 | 0.70 | 0.69 | 0.78 | 0.79 | 0.785 | .84 | .85 | .85 |

## Application

In remote diagnosis applications like WCE, a huge amount of collected data becomes a bottleneck in healthcare services. Difficulties include storage, search, sharing, analysis, and visualization. The proposed method is implemented for smartphones because it automatically extracts diagnostically important visual data and sends it to the medical centers for analysis. The proposed telemedicine service is a resource-aware application that provides real-time analysis in order to improve overall patient healthcare. The major contributions and future possibilities of the proposed work are:

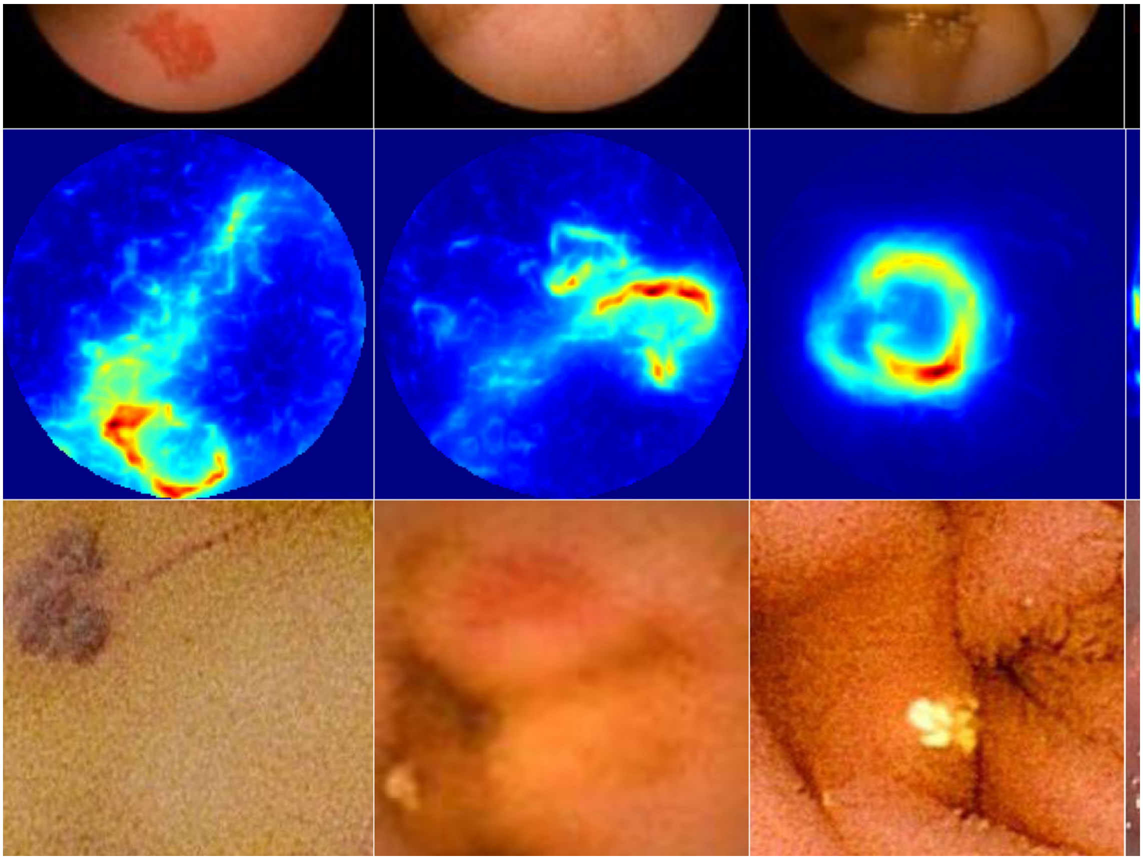

**Fig. 3** Saliency maps. From top to bottom: input image and its corresponding saliency map

- Provides medical specialists fast and easy access to vital information anytime/anywhere during the WCE procedure.
- Reduces the storage and transmission cost.
- Increases access to healthcare opportunities.
- Once video keyframes are extracted, there are further possibilities of organizing them for browsing and navigation purposes, rather than the strict sequential display of WCE videos.
- This framework can be extended to other remote diagnostic procedures to monitor and timely tackle abnormal findings.

## Conclusion

With the proliferation of medical imaging data in remote health monitoring applications such as WCE, remote biosensing data processing becomes extremely challenging. This large sensor data also leads to an increased demand on storage and bandwidth transmission capabilities. Thus, an efficient framework is required to process capsule sensor data and disseminate actionable intelligence. To deal with such data-intensive remote health care systems, we have presented an efficient video summarization framework. The proposed framework is based on a mobile assisted video summarization mechanism that extracts semantically relevant frames from video sequence captured by a wireless capsule sensor. Smartphones provide a convenient platform for local processing as they have not only processing capability but are able to connect medical centers through reliable communication networks. In the proposed video summarization, data processing capabilities have been streamlined and automated through the use of an efficient visual attention model that allows near real-time extraction of significant video frames. The extracted information is transferred to specialists using the infrastructure of the Internet and wireless capsule sensor-to-mobile communications. This empowers gastroenterologists to make correct decisions in real time. The experimental evaluation represents significant advances over traditional tele-endoscopy systems with difficult access. The high performance of the proposed system is further demonstrated by the integral-image that facilitates real-time measurement of the visual saliency model. Quantitative and qualitative evaluation validates the effectiveness of the proposed framework.

Future work will be carried out to investigate the integration of cloud services with mobile-computing to further improve performance. We have the intention to include cloud services for optimized processing keeping in view the trade-off between energy consumption and computational complexity. The availability of such a high performance remote data management system will enable a cost-effective solution for mining large wireless capsule sensing data.

**Acknowledgments** This research is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2012904).

**Author disclosure statement** No competing financial interests exist.

## References

1. Jovanov, E., and Milenkovic, A., Body area networks for ubiquitous healthcare applications: opportunities and challenges. *J. Med. Syst.* 35(5):1245–1254, 2011.
2. Baig, M. M., and Gholamhosseini, H., Smart health monitoring systems: an overview of design and modeling. *J. Med. Syst.* 37(2): 1–14, 2013.
3. Ahn, C. H., Choi, J.-W., Beaucage, G., Nevin, J. H., Lee, J.-B., Puntambekar, A., and Lee, J. Y., Disposable smart lab on a chip for point-of-care clinical diagnostics. *Proc. IEEE* 92(1):154–173, 2004.
4. Yuce, M. R., Ng, S. W., Myo, N. L., Khan, J. Y., and Liu, W., Wireless body sensor network using medical implant band. *J. Med. Syst.* 31(6):467–474, 2007.
5. Karargyris, A., and Bourbakis, N., Wireless capsule endoscopy and endoscopic imaging: A survey on various methodologies presented. *Eng Med Biol Mag IEEE* 29(1):72–83, 2010.
6. Nesbitt, T. S., Cole, S. L., Pellegrino, L., and Keast, P., Rural outreach in home telehealth: assessing challenges and reviewing successes. *Telemed J E-Health* 12(2):107–113, 2006.
7. Chakraborty, C., Gupta, B., and Ghosh, S. K., A Review on Telemedicine-Based WBAN Framework for Patient Monitoring. *Telemed e-Health* 19(8):619–626, 2013.
8. Touati, F., and Tabish, R., U-healthcare system: State-of-the-art review and challenges. *J. Med. Syst.* 37(3):1–20, 2013.
9. Lee, H.-G., Choi, M.-K., Shin, B.-S., and Lee, S.-C., Reducing redundancy in wireless capsule endoscopy videos. *Comput. Biol. Med.* 43(6):670–82, 2013.
10. Ejaz, N., Mehmood, I., and Wook Baik, S., *Efficient visual attention based framework for extracting key frames from videos*. Image Communication, Signal Processing, 2012.
11. Mehmood, I., Ejaz, N., Sajjad, M., and Baik, S. W., Prioritization of brain MRI volumes using medical image perception model and tumor region segmentation. *Comput. Biol. Med.* 43(10):1471–1483, 2013.
12. Sajjad, M., Mehmood, I., and Baik, S. W., Sparse Representations-Based Super-Resolution of Key-Frames Extracted from Frames-Sequences Generated by a Visual Sensor Network. *Sensors* 14(2): 3652–3674, 2014.
13. Li, B., and Meng, M.-H., Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection. *Inform Technol Biomed IEEE Trans* 16(3):323–329, 2012.
14. Pan, G., Yan, G., Qiu, X., and Cui, J., Bleeding detection in wireless capsule endoscopy based on probabilistic neural network. *J. Med. Syst.* 35(6):1477–1484, 2011.
15. Chen D, Meng M-H, Wang H, Hu C, Liu Z A novel strategy to label abnormalities for wireless capsule endoscopy frames sequence. In: Information and Automation (ICIA), 2011 I.E. International Conference on, 2011. IEEE, pp 379–383
16. Sainju, S., Bui, F. M., and Wahid, K. A., Automated bleeding detection in capsule endoscopy videos using statistical features and region growing. *J. Med. Syst.* 38(4):1–11, 2014.
17. Kundel, H. L., History of research in medical image perception. *J. Am. Coll. Radiol.* 3(6):402–408, 2006.
18. Shapley, R., and Hawken, M. J., Color in the cortex: single-and double-opponent cells. *Vis. Res.* 51(7):701–717, 2011.

19. Chen, Y., Lee, J., (2012) A Review of Machine-Vision-Based Analysis of Wireless Capsule Endoscopy Video. Diagnostic and therapeutic endoscopy 2012
20. Kumar, R., Zhao, Q., Seshamani, S., Mullin, G., Hager, G., and Dassopoulos, T., Assessment of crohn's disease lesions in wireless capsule endoscopy images. *Biomed Eng IEEE Trans* 59(2):355–362, 2012.
21. Li, B., and Meng, M. Q.-H., Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. *Comput. Biol. Med.* 39(2):141–147, 2009.
22. Tjoa, M., Krishnan, S., Doraiswami, R., Automated diagnosis for segmentation of colonoscopic images using chromatic features. In: Electrical and Computer Engineering, 2002. IEEE CCECE 2002. Canadian Conference on, 2002. IEEE, pp 1177–1180
23. Li, B.-P., and Meng, M. Q.-H., Comparison of Several Texture Features for Tumor Detection in CE Images. *J. Med. Syst.* 36(4): 2463–2469, 2012.
24. Li, B., Meng, M.-H., Zhao, Q., Wireless capsule endoscopy video summary. In: Robotics and Biomimetics (ROBIO), 2010 I.E. International Conference on, 2010. IEEE, pp 454–459
25. Bashar, M. K., Kitasaka, T., Suenaga, Y., Mekada, Y., and Mori, K., Automatic detection of informative frames from wireless capsule endoscopy images. *Med. Image Anal.* 14(3):449–470, 2010.
26. Ioannis, K., Tsevas, S., Maglogiannis, I., Iakovidis DK Enabling distributed summarization of wireless capsule endoscopy video. In: Imaging Systems and Techniques (IST), 2010 I.E. International Conference on, 2010. IEEE, pp 17–21
27. Iakovidis, D. K., Tsevas, S., and Polydorou, A., Reduction of capsule endoscopy reading times by unsupervised image mining. *Comput. Med. Imaging Graph.* 34(6):471–478, 2010.
28. Ma, Y.-F., Hua, X.-S., Lu, L., and Zhang, H.-J., A generic framework of user attention model and its application in video summarization. *Multimedia IEEE Trans* 7(5):907–919, 2005.
29. Engel, S., Zhang, X., and Wandell, B., Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature* 388(6637):68–71, 1997.
30. Itti, L., Koch, C., and Niebur, E., A model of saliency-based visual attention for rapid scene analysis. *Pattern Anal Mach Intell IEEE Trans* 20(11):1254–1259, 1998.
31. Crow, F. C., Summed-area tables for texture mapping. In: ACM SIGGRAPH Computer Graphics, 1984. vol 3. ACM, pp 207–212
32. Viola, P., Jones, M., Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 I.E. Computer Society Conference on, 2001. IEEE, pp I-511–I-518 vol. 511
33. Hu, M.-K., Visual pattern recognition by moment invariants. *Inform Theory IRE Trans* 8(2):179–187, 1962.
34. Murphy, T., Matlin, M., Finkel, L.H., Curvature covariation as a factor in perceptual salience. In: Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on, 2003. IEEE, pp 16–19
35. Hoffman, D. D., and Singh, M., Salience of visual parts. *Cognition* 63(1):29–78, 1997.
36. GastroLab http://www.gastrolab.net/.
37. WCEVideoAtlas http://www.wceatlas.org/index.php.