Biomedical Signal Processing and Control 33 (2017) 161–168

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Efficient visual attention driven framework for key frames extraction from hysteroscopy videos

Khan Muhammad[a], Muhammad Sajjad[b], Mi Young Lee[a], Sung Wook Baik[a,*]

[a] Intelligent Media Laboratory, Digital Contents Research Institute, College of Electronics and Information Engineering, Sejong University, Seoul, Republic of Korea

[b] Digital Image Processing Laboratory, Department of Computer Science, Islamia College Peshawar, Pakistan

ARTICLE INFO

Article history:
Received 17 June 2016
Received in revised form 28 October 2016
Accepted 12 November 2016
Available online 7 December 2016

Keywords:
Video summarization
Diagnostic hysteroscopy
Visual saliency
Visual attention model
Gynecology
Medical image analysis

ABSTRACT

Recent years have shown enthusiastic research interests in diagnostic hysteroscopy (DH), where various regions of the female reproductive system are visualized for diagnosing uterine disorders. Currently, the hysteroscopy videos produced during various sessions of patients are stored in medical libraries, which are usually browsed by medical specialists *Gynecologists* to visualize previous videos of a patient or to study similar cases. However, the abundant redundancy of frames in DH videos make this searching relatively more difficult for gynecologists, wasting their time browsing such large libraries. In this context, video summarization can be used to reduce this redundancy by extracting key frames, thus making the process of browsing and indexing DH videos more efficient. In this letter, we propose an efficient domain-specific visual attention-driven framework for summarizing DH videos. For key frames extraction, multi-scale contrast, texture, curvature, and motion based saliency features are computed for each frame using integral image, which are then fused by a linear weighted fusion scheme to acquire a final saliency map. Experimental results in comparison with other related state-of-the-art schemes confirm the effectiveness and efficiency of the proposed framework.

1. Introduction

Hysteroscopy is a popular surgical method for assessing and visualizing various regions of the female reproductive system such as the uterine cavity, cervical channel, and tubal ostea [1]. During this procedure, the medical specialist *gynecologist* uses the hysteroscope for diagnosis and treatment of uterine disorders. A hysteroscope is a small lighted fiber-optic technology based telescopic instrument, which can transmit the captured sequence of images to a screen, allowing a gynecologist to focus on guiding the instrument to the regions of interest [2]. Hysteroscopy can be of two types: *diagnostic* and *operative* hysteroscopy. In diagnostic hysteroscopy (DH), the uterus is examined to assess the signs of the abnormality/normalcy while operative hysteroscopy is concerned with treatment of the disorder, when it is diagnosed [1]. Our work is focused on DH.

In practice, several DH sessions are conducted on a daily basis, each having an average time of 3 min. During this examination, a continuous video sequence is produced, which is usually fully recorded by hospitals and clinics for later evaluation and supporting studies of medical research [3]. However, only a limited number of frames from the recorded videos are important for actual diagnosis. In addition, the whole video sequence is linearly browsed for the desired contents whenever the specialists want to review a recorded case or previous videos of a patient. Since there are multiple videos consisting of thousands frames related to a single patient, therefore, browsing for the desired contents can be difficult and significantly more time-consuming than on spot hysteroscopy examination.

To surmount these problems, video summarization [4,5] can be used to prioritize hysteroscopy videos (HV) for extracting key frames, which are diagnostically important for gynecologists. Consequently, they can be used for the indexing of HVs. The current literature of video summarization covers a limited number of articles for summarization of HVs. Literature review dictates that the most recent work on HV summarization has been presented in our previous work [6], utilizing multi-scale contrast, texture, and motion based saliencies for summarization. Our previous work has three limitations: 1) To compute visual features, an RGB color model has been used, where the perception of color is not accurately represented, thus affecting the interests of gynecologists in gener-

* Corresponding author.
E-mail addresses: khanmuhammad@sju.ac.kr, khan.muhammad.icp@gmail.com (K. Muhammad), muhammad.sajjad@icp.edu.pk (M. Sajjad).

http://dx.doi.org/10.1016/j.bspc.2016.11.011
1746-8094/

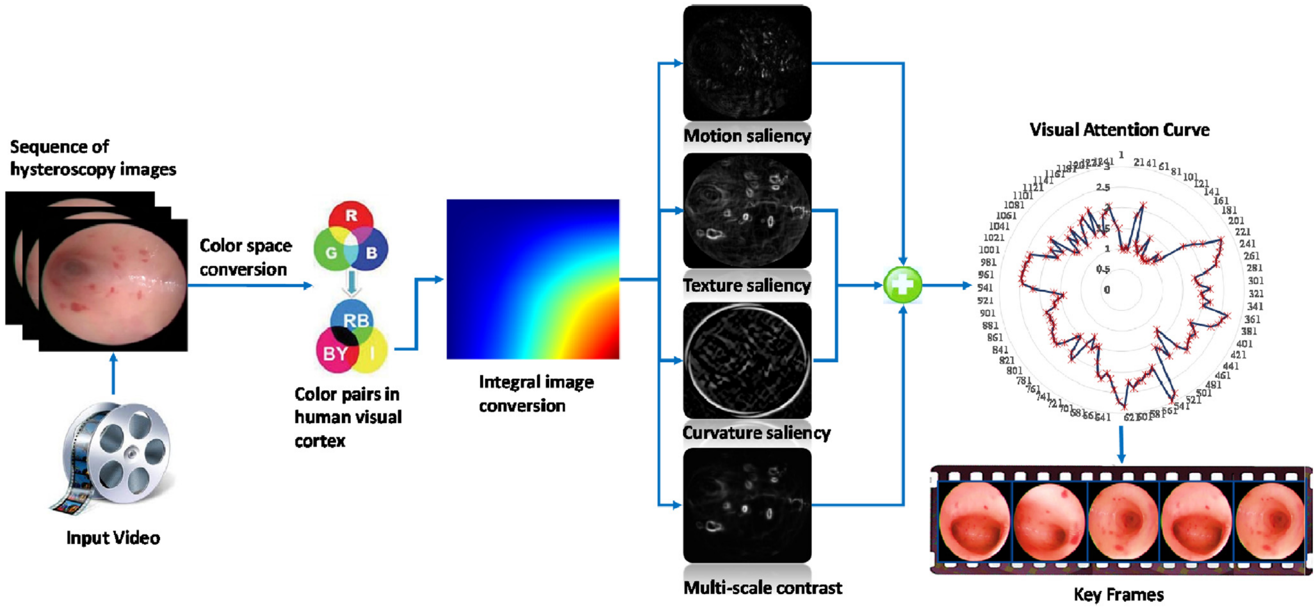


Fig. 1. Framework of the proposed system.

ated summaries, 2) The saliencies are computed on entire frames, which are computationally expensive, making summary generation relatively more time consuming, which in turn degrades its performance, 3) The hystroscope collects images at various scales and orientations, which cannot be captured by features used in previous scheme. According to [6], multi-scale contrast can identify salient objects of various sizes; texture based saliency helps in identification of more injurious regions; and motion based saliency provides indication about frames, having less chances of key frames selection. Thus, our previous approach fails for frames taken from different orientations, requiring orientation invariant features.

To overcome the aforementioned problems, we propose a video summarization framework inspired from visual attention model for HVs. The main contributions of this study are summarized as follows:

1. We propose an efficient video summarization framework, combining the strengths of visual attention model with domain knowledge for key frames selection from HVs.
2. To extract more relevant key frames and generate summaries of gynecologists' interests, our framework uses color opponent color space (COC), which is more in accordance with the human vision system and helps in efficient selection of salient objects in a frame.
3. To reduce the computational complexity, our framework uses integral image/summed-area tables for features computation and generation of summaries.
4. To address the orientation problem, we incorporated curvature feature in our framework. The curvature feature is orientation invariant and is of paramount importance at the intermediate stages of visual signal analysis in the visual cognitive process as suggested by theoretical and psychophysical considerations [7]. Thus, it can detect changes in any direction in a frame, helping in effective extraction of key frames.

The rest of this paper is structured as follows: Section 2 presents the proposed framework. Section 3 explains the experimental results and discussion. Finally, the paper is concluded in Section 4.

2. The proposed framework

In this section, the main embodiments of the proposed framework are described. The proposed scheme consists of three major steps including conversion to the COC color model, integral image based computation of visual features, and key frames extraction. The computed features are fused for getting the saliency map based on which key frames are selected. The pictorial representation of the proposed system is shown in Fig. 1. The detail of the features computation and other intermediate steps are discussed in subsequent sections.

2.1. Color space conversion

In medical image analysis, it is necessary to consider the importance of color information as some of the color spaces such as RGB color space fail to accurately represent the perception of color according to human visual cortex. In this context, COC color model is an optimal choice for improved representation of color perception and efficiency of selecting salient objects due to its accordance with human visual system [8]. We therefore incorporate COC color model in the proposed framework. Consider a hysteroscopy video HV of n_{NF} frames, starting at time t as given in Eq. (1), where "F" indicates a single frame of HV. The goal is to find a set of key frames KF as given in Eq. (2), having n_{KF} frames that are of interest to gynecologists.

$$HV = \left\{ F(t+i) \mid i = 0,\ 1,\ \ldots n_{NF-1} \right\} \tag{1}$$

$$KF = \left\{ F_{KF}(t+1),\ F_{KF}(t+2),\ \ldots\ldots F_{KF}(t+n_{KF}) \mid n_{KF} \le n_{NF} \right\} \tag{2}$$

$$\left\{ \begin{array}{l} R_I = R - (G+B)/2 \\ G_I = G - (G+B)/2 \\ B_I = B - (G+B)/2 \\ Y_I = (R+G)/2 - |R-G|/2 - B \end{array} \right\} \tag{3}$$

$$\left\{ \begin{array}{l} RG = R_I - G_I \\ BY = B_I - Y_I \end{array} \right\} \tag{4}$$

For conversion from RGB to COC, the three channels of the RGB image are transformed to an intermediate representation of four channels, as in Eq. (3). Then the two opponent color pairs RG and BY are calculated as shown in Eq. (4). The final aggregated image F^{AG} is obtained by computing the intensity plane I and combing it with RG and BY using Eq. (5). F^{AG} is then used for integral image calculation based on which various saliencies of the framework are computed [9].

$$\left\{ \begin{array}{l} I = (R + G + B)/3 \\ F^{AG} = RG + BY + I \end{array} \right\} \tag{5}$$

2.2. Visual features computation

In this section, we briefly discuss the process of computing the visual saliencies of the proposed framework. The proposed saliency model uses four features including motion, texture, multi-scale contrast, and curvature. These saliencies are computed using integral image [10], whose computation is relatively less expensive having time complexity 2 MN. Here M and N represent the height and width of the frame, respectively. The value of integral image at a pixel (x, y) is the sum of all pixels above and to the left of (x, y) and is computed as follows:

$$I_t^{\text{int}}(x, y) = \sum_{\substack{xx \le x \\ yy \le y}} F^{AG}(xx, yy) \tag{6}$$

After getting the integral image, one can calculate the sum of any block of the image in constant time with complexity O (1) unlike$O\left(n^2\right)$. This property is the motivational reason of incorporating the integral image in the proposed framework for generating efficient video summaries.

A Motion Saliency

Motion saliency is an important concept for finding the inter-frame motion of a hysteroscopy video [11]. The gynecologist usually moves the hysteroscope quickly in areas of non-interest. Thus, a low inter-frame motion indicates that the frames are of interest to gynecologists compared to frames with fast inter-frame motion. In the proposed model, the motion saliency is computed as follows [6]:

$$MS\left(F^{AG}, p\right) = \sqrt{MS_x^2(p) + MS_y^2(p)} \tag{7}$$

Where $MS_x(p)$ and $MS_y(p)$ refer to x and y elements of the motion vector respectively, p is the current pixel, and F^{AG} represent the current frame.

B Multi-scale Contrast

Contrast detection is helpful in finding the most informative regions of a hysteroscopy frame. This is due to the fact that the human vision system is comparatively more sensitive to luminance than color [6,12]. The luminance change produces a pattern of contrast, facilitating gynecologists during selection of informative regions in a frame [7]. The human vision system can see numerous resolutions due to large variations in the size of the receptive fields. Therefore, to deal with various sizes of salient objects/anomalies in the hysteroscopy images, multi-scale contrast is used in the proposed scheme. The multi-scale contrast of a frame at pixel p (x, y) is calculated as follows:

$$MSC^{gps}(x, y) = \|F^{AG}(x, y) - \mu\| \tag{8}$$

Where

$$\mu = \frac{I^{\text{int}}(x - n,\ y - n) + I^{\text{int}}(x + n,\ y + n) - I^{\text{int}}(x,\ y - n) - I^{\text{int}}(x - n,\ y)}{n^2} \tag{9}$$

Here, n indicates the neighborhood of the pixel p (x, y) and its value is set to 5. It is experimentally proved in our previous work [7] that n = 5 is the most feasible choice due to its high saliency results and computational feasibility. "gps" represents Gaussian pyramid scale and its value is defined as $gps \in [1,\ 3]$. Finally, a gray-scale saliency image is obtained by summing up the contrast at three levels of Gaussian pyramid as follows:

$$MSC(x, y) = \sum_{gps=1}^{3} MSC^{gps}(x, y) \tag{10}$$

C Texture Saliency

The texture saliency is effective in determining the injurious regions of hysteroscopy frames [6]. To find the texture saliency of a frame, the entropy-based texture-segmentation technique [13] is used in the proposed saliency model. The entropy score of a pixel p at frame F can be computed as follows:

$$ENT(F, p) = -\sum_{j=0}^{gps-1} Hist_p(j)\log_2\left(Hist_p(k)\right) \tag{11}$$

$$TEXI(F, p) = \left\{ \begin{array}{ll} 0 & if\ ENT(F,\ p) < \tau \\ 1 & \quad otherwise \end{array} \right. \tag{12}$$

The resultant image is then segmented for texture with threshold τ=0.8. This eliminates the objects with area less than a threshold, producing a texture image "TEXI" with no injuries. Next, a morphological procedure closing is used to smooth the edges, followed by filling the holes to get an image "MASK". The mask is then used to find the injurious regions of the frame as follows:

$$T(F, p) = \left\{ \begin{array}{ll} F(p) & if\ \ TEXI(F, p) = 1 \\ 0 & if\ \ TEXI(F, p) = 0 \end{array} \right. \tag{13}$$

Finally, a saliency score of 1 is assigned to the frame with largest injurious regions while other frames are allocated values relative to the maximum score.

D Curvature Map

Although the previous features are helpful in identifying various abnormities in hysteroscopy frames, yet they fail when the frames are captured at different orientations. To overcome this limitation, a curvature feature is incorporated in the proposed framework. Attneave [14] reported that the curvature points of a frame are more helpful compared to the edges and straight lines due to difficulty in predicting them using the neighborhood pixels. It is also verified by the current research of neurosciences and psychophysical studies that the rotational invariant curvature map is one of the key factors in saliency computation and can improve the gynecologists' decision making [15]. These reasons necessitate the inclusion of curvature map in the proposed summarization scheme. The curvature map of a hysteroscopy frame "F" can be computed as follows:

$$CM = |\nabla^2 g| = \sqrt{g_{xy}^2 + g_{xx}^2 + g_{yx}^2 + g_{yy}^2} \tag{14}$$

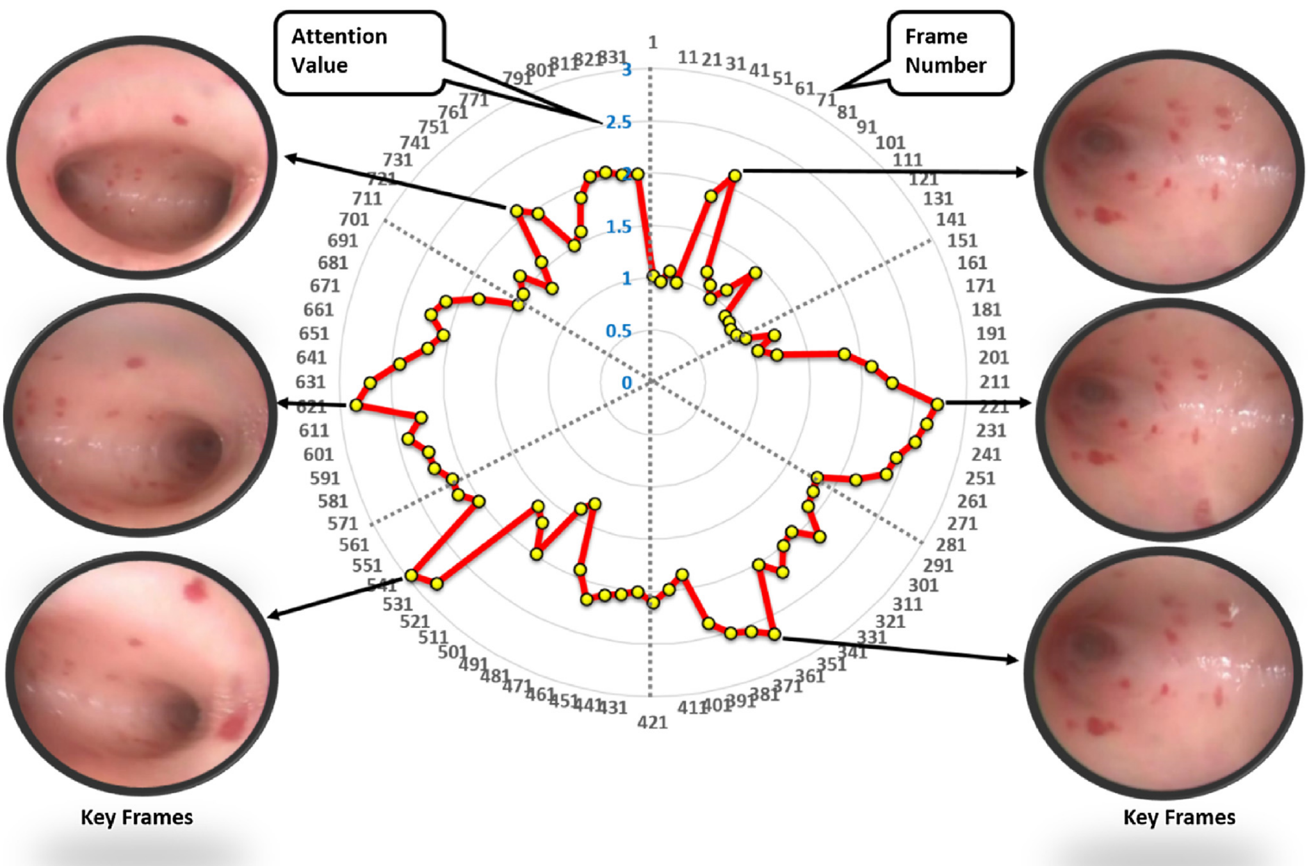


Fig. 2. An illustration of key frames selection process.

Where

$$\left\{\begin{array}{l} g\,(x,y)=F\,(x,y)\times\Psi \\ \Psi=e^{-\dfrac{x^2+y^2}{2\sigma^2}} \\ g_{xy}=\dfrac{\partial^2 g}{\partial_x\partial_y},\quad g_{xx}=\dfrac{\partial^2 g}{\partial_{x^2}},\quad g_{yx}=\dfrac{\partial^2 g}{\partial_y\partial_x},\quad g_{yy}=\dfrac{\partial^2 g}{\partial_{y^2}} \end{array}\right\} \tag{15}$$

Here Ψ indicates a Gaussian operator with σ = 3, which smoothens the frame at spatial and transform domains, hence reducing the amount of noise.

2.3. Aggregation mechanism and key frames extraction

The diagnostically significant frames are extracted based on the final aggregated attention values of the HV frames. To get the final attention values, an aggregation mechanism is used in the proposed framework. Firstly, the values of four features are normalized in the range of [0,1] known as "suppression". This step globally suppresses the saliency scores of individual features, helping in smoothing and attaining robustness. Next, the average of non-zero pixels is computed and is nominated as the individual saliency value of each feature for a frame. These saliency scores of four features are then fused to get a single attention value for each frame as follows:

$$A\,(F)=(1-MS\,(F))+MSC\,(F)+T\,(F)+CM\,(F) \tag{16}$$

It is worth mentioning that the values of MS are subtracted from 1, assigning more weight to the frames with slow inter-frame motion. The reason for this is that gynecologists are not interested in certain frames of the video. Thus, they move the hysteroscope quickly, resulting in high speed motion. Finally, an attention curve is generated based on the attention values of an entire video. To visualize the frames of interest, the gynecologist is requested to specify the number of desired key frames n_{KF}. Based on n_{KF}, the video is divided into "S" segments where S is the ratio of total number of frames n_{NF} to desired number of key frames n_{KF}. A frame having the highest attention score within a video segment is then selected as a key frame. An overview of the proposed key frame selection process is depicted in Fig. 2. This facilitates the gynecologists to visualize the video at several summarization levels by changing n_{KF} only. Thus, the proposed scheme also avoids the problem of re-computing the attention values for visualization of DH videos at different summarization levels.

3. Experimental results

In this section, we evaluate the performance of the proposed video summarization method compared to other state-of-the-art schemes. The experimental results were collected based on a dataset of 10 real hysteroscopy videos. The duration of each video varies from 2 to 3 min with a frame rate of 30 f/sec. For evaluation, we requested two gynecologists to select ground truth from the selected ten videos. The ground truth is selected in terms of small video segments as mentioned in [2]. However, to compare all the methods in a straightforward way for easy interpretation of the results by readers, a single representative frame is selected as a ground truth for each video segment. This mechanism is already adopted in a recent work of VS for DH videos in [6]. The key frames selected by our summarization scheme are then compared with

Table 1
Quantitative evaluation based on recall (R), precision (P), and F-measure (F) scores obtained by the proposed method and other state-of-the-art schemes.

Serial#	Avila et al., Method [17]			Gaviao et al. Method [2]			Naveed et al. Method [6]			Proposed Method		
	R	P	F	R	P	F	R	P	F	R	P	F
1	0.48	0.61	0.53	0.85	0.98	**0.91**	0.86	0.95	0.90	0.89	0.94	**0.91**
2	0.66	0.71	0.68	0.93	1	0.96	0.94	0.98	0.95	1	0.95	**0.97**
3	0.61	0.58	0.59	0.86	0.92	0.88	0.92	0.99	**0.95**	1	0.85	0.92
4	0.45	0.43	0.43	0.83	0.9	0.86	0.95	0.97	**0.95**	1	0.91	**0.95**
5	0.51	0.56	0.53	0.81	0.87	0.83	0.89	0.88	0.88	0.91	0.88	**0.89**
6	0.52	0.5	0.50	0.82	0.88	0.84	0.95	0.98	**0.96**	0.95	0.97	0.95
7	0.43	0.58	0.49	0.93	0.97	**0.94**	0.86	0.96	0.90	0.95	0.91	0.92
8	0.45	0.49	0.46	0.8	0.85	0.82	0.82	0.9	0.85	0.85	0.89	**0.86**
9	0.61	0.71	0.65	0.75	0.88	0.80	0.94	0.93	0.93	0.94	0.94	**0.94**
10	0.54	0.63	0.58	0.83	0.79	0.80	0.96	0.98	**0.96**	0.92	0.92	0.92
Average	0.526	0.580	0.550	0.841	0.904	0.870	0.909	0.952	0.929	0.944	0.920	**0.931**

the ground truth for computing the scores of recall, precision, and F-measure. These metrics are calculated as follows [11]:

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

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

$$F - Measure = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall}\right) \tag{19}$$

Herein, a frame is said to be true positive (TP) if it is selected by the summarization technique and the gynecologist as key frame. A false negative (FN) represents a frame that is chosen by the specialist and not by the summarization scheme. Finally, a false positive (FP) shows a frame that is nominated as key frame by the summarization method but not by the medical doctor [6]. Recall computes the probability of selecting a relevant frame as key frame. Precision determines the relevancy of chosen key frames. Recall and precision cannot be used alone for performance evaluation as they are complementary to each other [16]. Therefore we use F-measure, which reflects the average of both precision and recall. The proposed scheme is compared with three methods whose comparative results are shown in Table 1. The results indicate that the general-purpose key frame extraction methods such as scheme in [17] are relatively less effective in summarizing HVs. Furthermore, the proposed scheme outperforms other domain-specific HV summarization methods due to incorporation of COC color space and curvature feature. The proposed framework is also computationally in-expensive compared to our recent scheme [6] as shown in Fig. 3. This is due to utilization of integral image in the proposed framework.

In Fig. 4, the key frames chosen by the gynecologist and the schemes under consideration are shown. It can be noted that the existent methods fail to extract the key frames, which are captured by hysteroscope from different orientations. The reason for this is that the current schemes do not utilize any rotation-invariant feature in their frameworks. The proposed framework resolves this limitation and is capable to extract those key frames, which are captured at different orientation. Overall, the proposed framework is efficient requiring less computation and generates summary that is of more interest to gynecologists compared to other schemes.

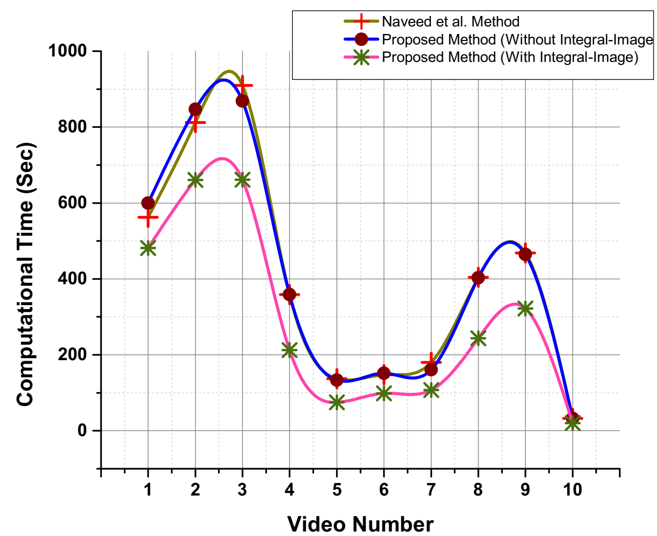


Fig. 3. Time analysis of our recent technique [6] and the proposed scheme on different HVs.

4. Discussion

In this sub-section, we discuss about the visual characteristics of diagnostically unimportant frames, the reasons why a generic VS method would not be feasible for summarization of DH videos, and how our work is addressing these problems. The diagnostically unimportant frames are mostly corrupted due to lighting and biological conditions. Such frames are discarded as they are not important for analysis by gynecologists. Fig. 5 shows an example of irrelevant frames along with diagnostically important frames.

It is not guaranteed for other generic and domain specific medical video summarization techniques [9,17] to perform well for keyframes extraction from DH videos. The reasons are obvious as mentioned in our previous work [6] and the current research work. For instance, the summarization scheme of WCE [9] uses image moments, curvature feature, and multi-scale contrast, which cannot fulfill all the requirements needed by hysteroscopy video summarization. The framework of [9] lacks two important features: motion and texture saliency. Motion saliency is the most important feature for summarization of DH videos as it provides indication about a frame's importance by computing the inter-frame motion. This is evident from the fact that gynecologists moves the hysteroscope quickly in areas of non-interest, resulting in faster inter-frame motion. Conversely, areas of gynecologist's interest are checked thoroughly by slow movement of hysteroscope, resulting in redundant images with lower inter-frame motion [6]. This information facilitates our framework to focus on frames having lower inter-frame motion.

Another important feature missing in the framework of [9] is texture saliency, which is particularly important for DH video abstraction to identify the frames with most injurious regions. Considering these limitations, the work of [9] is not suitable for extraction of representative frames from DH videos. Our framework is more suitable for extraction of representative frames from

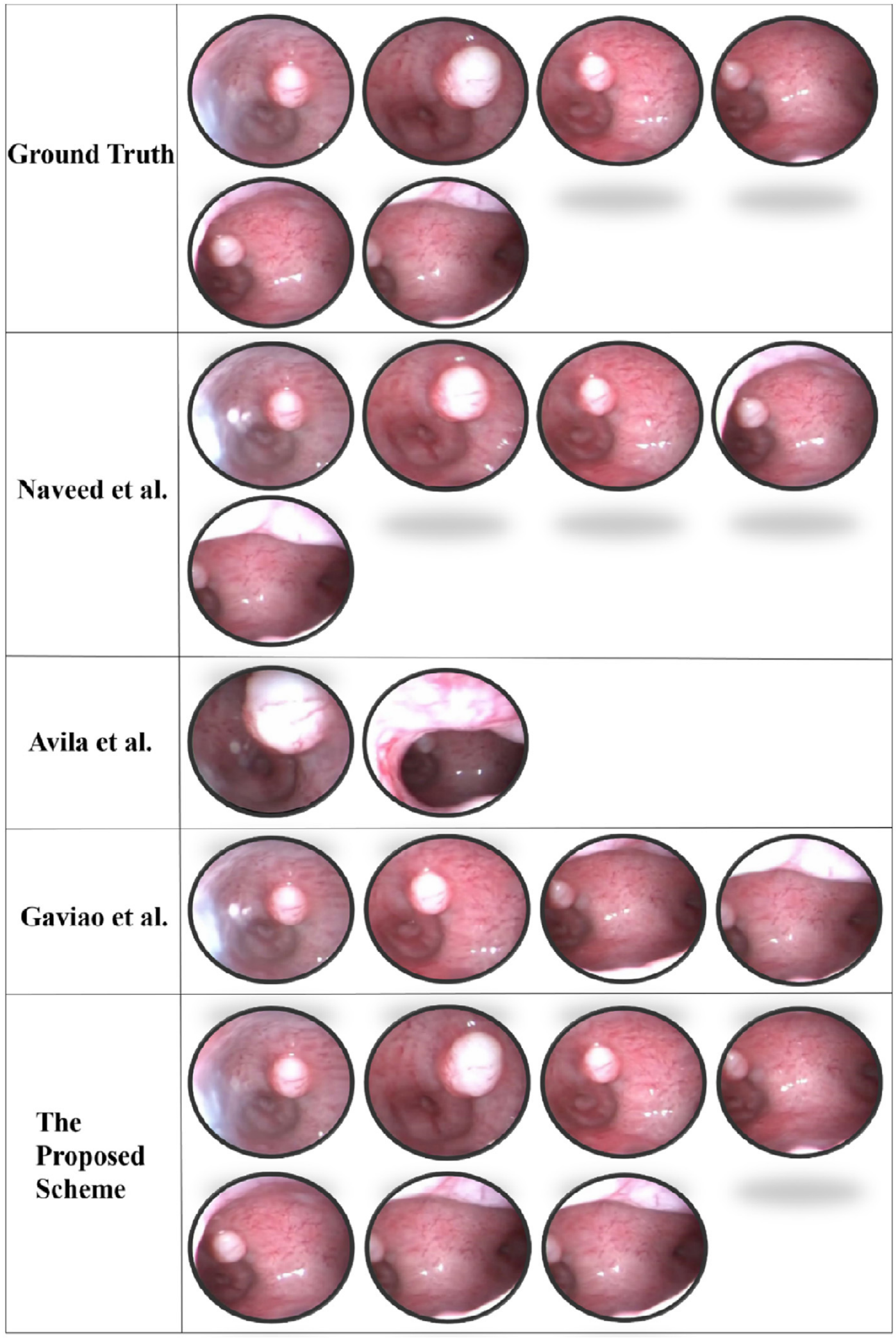


Fig. 4. Visual results of key frames generated by the existing summarization methods and the proposed scheme on a sample HV.

DH videos as it considers all these features along with other embodiments as mentioned in the introduction section.

5. Conclusion

In this study, a cost-effective domain-specific visual attention-driven framework is proposed for extraction of key frames from diagnostic hysteroscopy videos. Novel integral image based features are used to efficiently compute the attention values, considering multi-scale contrast, curvature, motion, and textural information. Interestingly, the proposed visual attention-driven framework was found to be effective enough to highlight the diagnostically important regions of hysteroscopy frames. Experimental results indicate that incorporating the COC color space in the proposed framework generate more effective summarizes of HVs. Furthermore, the proposed scheme is capable of highlighting the abnormalities in frames captured from different orientations due to usage of the curvature map. Overall, the proposed framework is cost-effective and comparatively more effective in extraction of diagnostically important key frames compared to general-purpose and domain-specific video summarization schemes.

Although, the current work improved the results up to some extent but still more research is required to fully extract all the keyframes which lies in the video segments captured from different orientations. In future, we have intention to further improve the performance of the proposed framework by incorporating adaptive differential evolution algorithms such as [18] in frames segmentation stage and to utilize it for other medical video abstraction

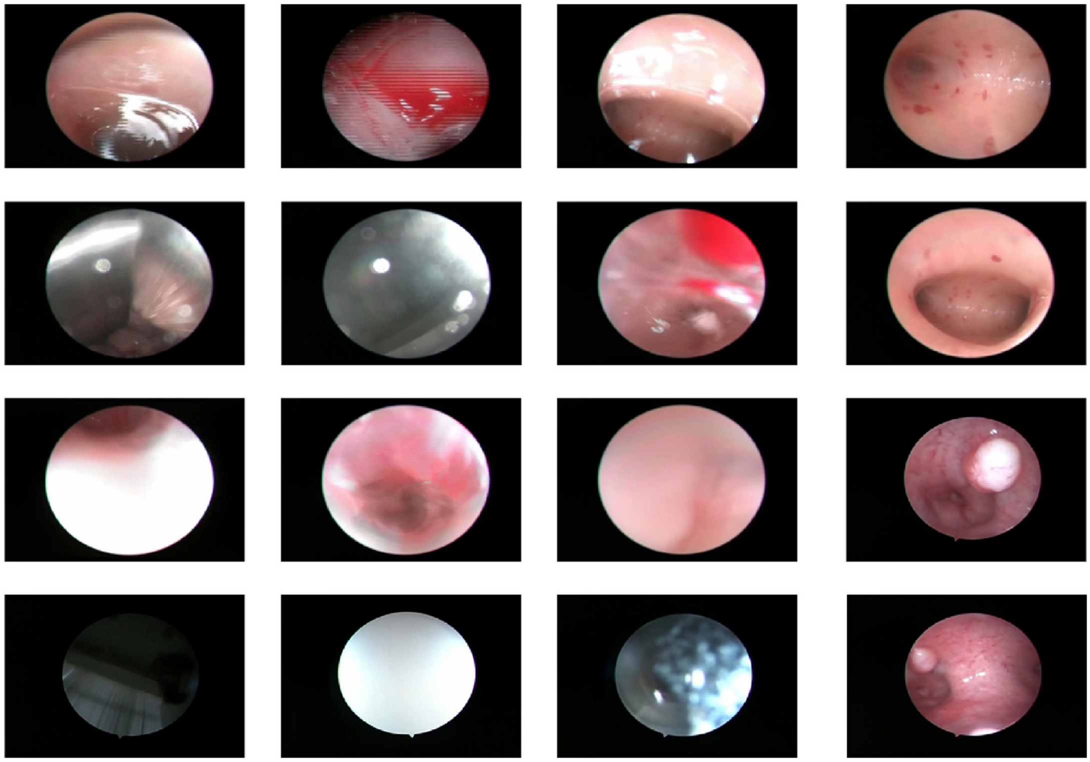


Fig. 5. Visual representation of important and non-important frames. The first three columns from left to right are non-important frames taken from different hysteroscopy videos. The last column shows diagnostically important frames.

domains such as wireless capsule endoscopy [19,20]. In addition, we tend to utilize more discriminative feature extraction methods [21] for developing an efficient indexing and retrieval system [22,23], improving the indoor and mobile healthcare facilities [24,25]. Finally, the data hiding techniques such as steganography [26–29] and watermarking [30,31] in combination with image encryption [32] can be used to embed the sensitive medical data in diagnostic hysteroscopy videos, preserving the patient's privacy and reduces the chances of modification by attackers, thus results in improved diagnosis.

Acknowledgement

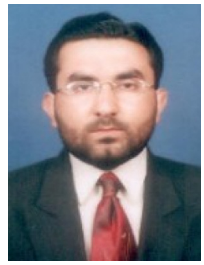
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.2016R1A2B4011712).

References

[1] W. Gavião, J. Scharcanski, Evaluating the mid-secretory endometrium appearance using hysteroscopic digital video summarization, Image Vision Comput. 25 (2007) 70–77.
[2] W. Gavião, J. Scharcanski, J.-M. Frahm, M. Pollefeys, Hysteroscopy video summarization and browsing by estimating the physician's attention on video segments, Med. Image Anal. 16 (2012) 160–176.
[3] J. Scharcanski, W.G. Neto, J.S. Cunha-Filho, Diagnostic hysteroscopy video summarization and browsing, 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference (2006) 5680–5683.
[4] S.E. De Avila, A. da Luz, A. de Araujo, M. Cord, VSUMM: an approach for automatic video summarization and quantitative evaluation, Computer Graphics and Image Processing, 2008. SIBGRAPI'08. XXI Brazilian Symposium on (2008) 103–110.
[5] K. Muhammad, J. Ahmad, M. Sajjad, S.W. Baik, Visual saliency models for summarization of diagnostic hysteroscopy videos in healthcare systems, SpringerPlus 5 (2016) 1495.
[6] N. Ejaz, I. Mehmood, S.W. Baik, MRT letter: visual attention driven framework for hysteroscopy video abstraction, Microsc. Res. Tech. 76 (2013) 559–563.
[7] I. Mehmood, N. Ejaz, M. Sajjad, S.W. Baik, Prioritization of brain MRI volumes using medical image perception model and tumor region segmentation, Comput. Biol. Med. 43 (2013) 1471–1483.
[8] S. Engel, X. Zhang, B. Wandell, Colour tuning in human visual cortex measured with functional magnetic resonance imaging, Nature 388 (1997) 68–71.
[9] I. Mehmood, M. Sajjad, S.W. Baik, Video summarization based tele-endoscopy: a service to efficiently manage visual data generated during wireless capsule endoscopy procedure, J. Med. Syst. 38 (2014) 1–9.
[10] F.C. Crow, Summed-area tables for texture mapping, ACM SIGGRAPH computer graphics 18 (1984) 207–212.
[11] I. Mehmood, M. Sajjad, W. Ejaz, S.W. Baik, Saliency-directed prioritization of visual data in wireless surveillance networks, Inf. Fusion 24 (2015) 16–30.
[12] F. Perazzi, P. Krähenbühl, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (2012) 733–740.
[13] S.J. Wernecke, L.R. D'Addario, Maximum entropy image reconstruction, IEEE Trans. Comput. 100 (1977) 351–364.
[14] F. Attneave, Some informational aspects of visual perception, Psychol. Rev. 61 (1954) 183.
[15] D.D. Hoffman, M. Singh, Salience of visual parts, Cognition 63 (1997) 29–78.
[16] I. Mehmood, M. Sajjad, S.W. Baik, Mobile-cloud assisted video summarization framework for efficient management of remote sensing data generated by wireless capsule sensors, Sensors 14 (2014) 17112–17145.
[17] S.E.F. de Avila, A.P.B. Lopes, A. da Luz, A. de Albuquerque Araújo, VSUMM: a mechanism designed to produce static video summaries and a novel evaluation method, Pattern Recognit. Lett. 32 (2011) 56–68.
[18] A. Khan, M.A. Jaffar, L. Shao, A modified adaptive differential evolution algorithm for color image segmentation, Knowl. Inf. Syst. 43 (2015) 583–597.
[19] K. Muhammad, M. Sajjad, S.W. Baik, Dual-level security based cyclic18 steganographic method and its application for secure transmission of keyframes during wireless capsule endoscopy, J. Med. Syst. 40 (2016) 1–16.
[20] Muhammad Khan, Jamil Ahmad, Muhammad Sajjad, Seungmin Rho, Sung Wook Baik, 2016. Evaluating the Suitability of Color Spaces for Image Steganography and Its Application in Wireless Capsule Endoscopy. In 2016 International Conference on Platform Technology and Service (PlatCon), IEEE, pp. 1–3.
[21] A. Irtaza, M.A. Jaffar, E. Aleisa, T.-S. Choi, Embedding neural networks for semantic association in content based image retrieval, Multimedia Tools Appl. 72 (2014) 1911–1931.
[22] J. Ahmad, M. Sajjad, I. Mehmood, S. Rho, S.W. Baik, Saliency-weighted graphs for efficient visual content description and their applications in real-time image retrieval systems, J. Real-Time Image Process (2015) 1–17, http://dx.doi.org/10.1007/s11554-015-0536-0.

[23] S. Rho, B.-j. Han, E. Hwang, M. Kim, MUSEMBLE: a novel music retrieval system with automatic voice query transcription and reformulation, J. Syst. Softw. 81 (2008) 1065–1080.
[24] Z. Lv, J. Chirivella, P. Gagliardo, Bigdata oriented multimedia mobile health applications, J. Med. Syst. 40 (2016) 1–10.
[25] R. Patsakis, P. Venanzio, A. Solanas, M. Bouroche, Personalized medical services using smart cities' infrastructures, Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on (2014) 1–5.
[26] R.J. Mstafa, K.M. Elleithy, A video steganography algorithm based on Kanade-Lucas-Tomasi tracking algorithm and error correcting codes, Multimedia Tools Appl. (2015) 1–23.
[27] M. Sajjad, K. Muhammad, S.W. Baik, S. Rho, Z. Jan, S.-S.et al. Yeo, Mobile-cloud assisted framework for selective encryption of medical images with steganography for resource-constrained devices, Multimedia Tools Appl. (2016) 1–18, http://dx.doi.org/10.1007/s11042-016-3811-6.
[28] K. Muhammad, I. Mehmood, M.Y. Lee, S.M. Ji, S.W. Baik, Ontology-based secure retrieval of semantically significant visual contents, J. Kor. Inst. Next Gener. Comput. 11 (3) (2015) 87–96.
[29] K. Muhammad, M. Sajjad, I. Mehmood, S. Rho, S.W. Baik, A novel magic LSB substitution method (M-LSB-SM) using multi-level encryption and achromatic component of an image, Multimedia Tools Appl. (2015) 1–27, http://dx.doi.org/10.1007/s11042-015-2671-9.
[30] Z. Liu, F. Zhang, J. Wang, H. Wang, J. Huang, Authentication and recovery algorithm for speech signal based on digital watermarking, Signal Process. 123 (2015) 157–166.
[31] H. Zhao, H. Wang, M.K. Khan, Statistical analysis of several reversible data hiding algorithms, Multimedia Tools Appl. 52 (2011) 277–290.
[32] R. Hamza, F. Titouna, A novel sensitive image encryption algorithm based on the Zaslavsky chaotic map, Inf. Secur. J.: Glob. Perspect. (2016) 1–18 http://dx.doi.org/10.1080/19393555.2016.1212954.

Khan Muhammad received his BS degree in Computer Science from Islamia College Peshawar, Pakistan with research in information security. Currently, he is pursuing MS leading to PhD degree in digital contents from Sejong University, Seoul, Republic of Korea. He is working as a researcher at Intelligent Media Laboratory (IM Lab). His research interests include image and video processing, data hiding, image and video steganography, video summarization, diagnostic hysteroscopy, wireless capsule endoscopy, deep learning, and CCTV video analysis. He has published 15+ papers in peer-reviewed international journals and conferences such as Future Generation Computer Systems, Journal of Medical Systems, Biomedical Signal Processing and Control, Multimedia Tools and Applications, SpringerPlus, KSII Transactions on Internet and Information Systems, Journal of Korean Institute of Next Generation Computing, NED University Journal of Research, Technical Journal, Sindh University Research Journal, Middle-East Journal of Scientific Research, MITA 2015, PlatCon 2016, and FIT 2016.



Muhammad Sajjad received his Master degree from Department of Computer Science, College of Signals, National University of Sciences and Technology, Rawalpindi, Pakistan. He received his PhD degree in Digital Contents from Sejong University, Seoul, Republic of Korea. He is now working as an assistant professor at Department of Computer Science, Islamia College Peshawar, Pakistan. He is also head of "Digital Image Processing Laboratory (DIP Lab)" at Islamia College Peshawar, Pakistan. His research interests include digital image super-resolution and reconstruction, medical image analysis, video summarization and prioritization, image/video quality assessment, and image/video retrieval.

Mi Young Lee is a research professor at Sejong University, Republic of Korea. She received her MS and PhD degrees from Department of Image and Information Engineering at Pusan National University. Her research interests include interactive contents, UI, UX and developing digital contents.

Sung Wook Baik received the B.S degree in computer science from Seoul National University, Seoul, Korea, in 1987, the M.S. degree in computer science from Northern Illinois University, Dekalb, in 1992, and the Ph.D. degree in information technology engineering from George Mason University, Fairfax, VA, in 1999. He worked at Datamat Systems Research Inc. as a senior scientist of the Intelligent Systems Group from 1997 to 2002. In 2002, he joined the faculty of the College of Electronics and Information Engineering, Sejong University, Seoul, Korea, where he is currently a Full Professor and Dean of Digital Contents. He is also the head of Intelligent Media Laboratory (IM Lab) at Sejong University. His research interests include computer vision, multimedia, pattern recognition, machine learning, data mining, virtual reality, and computer games.