

RESEARCH

Open Access



Visual saliency models for summarization of diagnostic hysteroscopy videos in healthcare systems

Khan Muhammad¹, Jamil Ahmad¹, Muhammad Sajjad² and Sung Wook Baik^{1*}

*Correspondence:
sbaik@sejong.ac.kr

¹ Intelligent Media Laboratory, Department of Digital Contents, College of Electronics and Information Engineering, Sejong University, Seoul, Republic of Korea
Full list of author information is available at the end of the article

Abstract

In clinical practice, diagnostic hysteroscopy (DH) videos are recorded in full which are stored in long-term video libraries for later inspection of previous diagnosis, research and training, and as an evidence for patients' complaints. However, a limited number of frames are required for actual diagnosis, which can be extracted using video summarization (VS). Unfortunately, the general-purpose VS methods are not much effective for DH videos due to their significant level of similarity in terms of color and texture, unedited contents, and lack of shot boundaries. Therefore, in this paper, we investigate visual saliency models for effective abstraction of DH videos by extracting the diagnostically important frames. The objective of this study is to analyze the performance of various visual saliency models with consideration of domain knowledge and nominate the best saliency model for DH video summarization in healthcare systems. Our experimental results indicate that a hybrid saliency model, comprising of motion, contrast, texture, and curvature saliency, is the more suitable saliency model for summarization of DH videos in terms of extracted keyframes and accuracy.

Keywords: Video summarization, Image and video processing, Medical image analysis, Visual saliency models, Diagnostic hysteroscopy

Background

The recent advancement in modern technology has shown promising results in human reproductive healthcare. One of the popular method of ensuring reproductive health is diagnostic hysteroscopy (DH), where the sensitive regions of female reproductive system (FRS) are assessed and visualized to diagnose uterine abnormalities (Gavião et al. 2012). An application of the DH is evaluation of glandular openings, concerning with prognosis and reproductive status and other abnormalities mentioned in Gavião and Scharcanski (2007). The DH procedure is performed by a gynecologist using a hysteroscope, which can disseminate the captured sequence of frames to a screen. The DH is performed several times per day, producing a number of DH videos, each of average length 3–4 min. These videos are fully recorded by hospitals in long-term video libraries for later inspection of previous diagnosis, research and training, and as an evidence for patients' complaints in courts (Scharcanski et al. 2006). However, from diagnosis point of view, a small number of frames are required for gynecologists to diagnose the abnormality. To this

end, gynecologists mostly browse the recorded DH videos manually to select the representative frames for supporting DH and as a record in patient history, making this process tedious and time consuming compared to the actual DH examination (Gavião and Scharcanski 2007).

To avoid this time consuming task, video prioritization schemes can be explored to extract keyframes, allowing gynecologists for non-linear browsing of DH contents. Consequently, the extracted keyframes can be used for efficient indexing of DH videos and generation of video summaries, containing relevant DH contents (Ejaz et al. 2013). To evaluate such VS schemes, gynecologists are requested to suggest portions of DH videos, which are diagnostically important and can be represented by a frame. From diagnostic point of view, the video portions with unobstructed view of FRS are important for gynecologists as illustrated in Fig. 1b, c. The DH frames contaminated by lighting and biological effects are discarded and are not of interest to gynecologists. An example of such irrelevant frames is shown in Fig. 1a.

During the DH examination, the specialist spends most of his time in searching for clinically important regions of FRS. Once such areas are found, the hysteroscope is focused on the areas of interest to capture numerous frames (Gavião and Scharcanski 2005). In addition, the surrounding areas near the region of interest are also examined by slowly moving the hysteroscope. Thus, the DH videos contain an enormous amount of redundant frames due to more examination of region of interest and low camera motion. Conversely, the non-important regions are examined quickly with fast movement of hysteroscope (Gavião et al. 2012).

Considering the aforementioned concerns, in this paper, we evaluate the performance of different general-purpose and domain specific VS methods for extraction of keyframes from DH videos. The study covers motion (Mehmood et al. 2015), texture (Ejaz et al. 2013), multi-scale contrast (Mehmood et al. 2014), curvature (Mehmood et al. 2013), and saliency detection using information maximization (SIM) (Bruce and Tsotsos 2005) for summarization. In addition, two general-purpose VS techniques (Ejaz et al. 2012; Ejaz and Baik 2012) are also considered for comparative analysis. The selection of

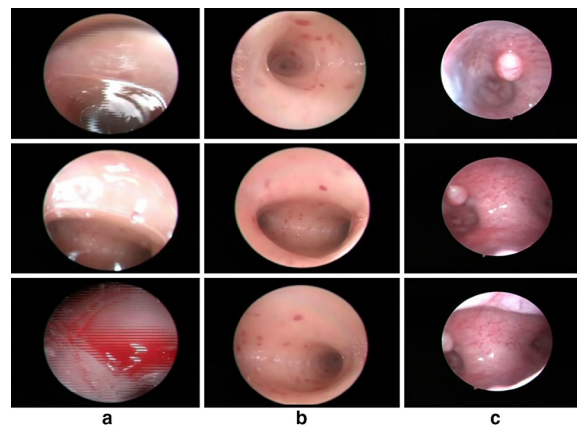


Fig. 1 **a** Non-important frames, indicating irrelevant DH frames contaminated by lighting and biological effects, **b, c** important frames representing diagnostically important DH frames from relevant DH video segments

best saliency detection model for summarization of DH videos is then suggested based on an evaluation criteria, reflecting computational complexity and accuracy.

The rest of this paper is organized as follows: “[Related work](#)” section presents an overview of video summarization and its related schemes. The details of this study are illustrated in “[Methods](#)” section. “[Experimental results and discussion](#)” section presents the experimental results, followed by concluding remarks and future directions in “[Conclusion](#)” section.

Related work

In this section, we present an overview of video summarization along with the previous works related to the process of DH videos abstraction. VS refers to identification of pertinent contents in a video for producing its concise representation known as video abstracts, which can be of two types (Truong and Venkatesh 2007): keyframes extraction and video skims. The former type is concerned with extraction of salient frames from the video. The latter category of VS extracts a condensed form of video clip with short duration, highlighting the main contents of original video (Ejaz et al. 2013). To produce a video abstract, there are two possible ways including manual and automatic summarization. Due to an enormous volume of video data, manual keyframes extraction is difficult and time consuming. Therefore, it is necessary to explore automatic VS for efficient utilization of manpower and other resources.

The current literature indicates that two major categories of features have been used for summarization including low and high-level features. Low-level features based VS methods (Ejaz et al. 2012; De Avila et al. 2011; Almeida et al. 2012; De Avila et al. 2008; Almeida et al. 2013) utilize numerous low-level features such as moments, color, motion, and shape. Due to semantic gap, the low-level features based VS methods do not agree with high-level human perception, decreasing its applicability. Considering this problem, researchers incorporated visual attention models in summarization methods, which extract frames reflecting the human attention. The first visual-attention directed VS scheme is proposed by Ma et al. (2005), utilizing visual, linguistic, and aural features for summary generation. Ejaz et al. (2013) presented a general-purpose keyframes extraction approach utilizing visual attention model. The method is utilizing temporal-gradient directed dynamic visual saliency, which is computationally inexpensive compared to traditional optical flow approaches. In addition, the static visual saliency based on DCT¹ is incorporated in the proposed framework. A non-linear weighted fusion is then used to combine the static and dynamic visual attention measures for generating an attention curve, which is used for producing a video summary.

The previous literature shows that visual attention model based VS schemes are most efficient in finding semantically relevant video summaries in contrast to low-level features based VS methods (Mehmood et al. 2016). Therefore, the focus here is to explore visual attention models based VS methods for extraction of diagnostically important frames from DH videos.

¹ Discrete Cosine Transform.

Scharcanski et al. (2006) presented a VS scheme for extraction of clinically important segments, facilitating quick browsing of DH videos for desired contents. Their scheme can be used to extract keyframes, which are used in record management of patients. Their presented scheme consists of two main steps: (1) firstly, a set of significant video segments are selected using statistical methods and (2) secondly, a post-processing step combines the similar adjacent video segments, avoiding over-segmentation. Gavião and Scharcanski (2007) nominated a VS method for detection of clinically significant segments in DH videos and extracting frames, providing a better visualization of the endometrium details such as glandular openings and vascularization. The approach can generate a video summary containing pertinent frames, enabling quick browsing of video contents. The proposed technique utilizes singular value decomposition characteristics during video abstraction, avoiding parameter adjustment.

Gavião et al. (2012) introduced another method for extraction of clinically important segments for DH videos. The method is capable of associating clinical significance with a DH video clip during the examination session of DH by gynecologist. Using the results of this method, the gynecologists can browse a given DH video non-linearly, saving their analysis time in manually visualizing each frame of the video. Another recent VS method for DH video abstraction is presented by Ejaz et al. (2013), where multi-scale contrast, motion, and texture based saliencies are combined for making a visual attention curve. The keyframes are then extracted using this attention curve, which can be used for analysis and indexing of DH videos.

The above literature designates that numerous proposals have been presented for general-purpose video summarization and DH video abstraction, considering individual factors such as efficiency, computational complexity, and accuracy. The previous VS methods are either too naïve or too complex with significant computational cost. The complex schemes achieve better accuracy in terms of keyframe extraction, however, their extensive computational cost make them less suitable for real-time summarization such as keyframes extraction during wireless capsule endoscopy (Mehmood et al. 2014; Muhammad et al. 2016). The VS methods utilizing simple features are computationally cost-effective, however, their lower accuracy makes them infeasible for sensitive areas of interest such as DH video summarization (Ejaz et al. 2013) and orthoscopic video summarization (Lux et al. 2010). It is therefore important to explore the general-purpose and domain-specific VS methods and exploit a VS framework for keyframes extraction from DH videos, which can maintain a balance between computation cost and accuracy.

Methods

In this section, we describe the mechanism of all the VS methods, which are considered for evaluation in terms of keyframes extraction and accuracy for DH videos. The methods under consideration include two general-purpose VS schemes, a general-purpose saliency detection model, and numerous domain-specific visual saliency detection models for medical videos. The general-purpose VS methods are our previous works including low-level features based VS (Ejaz et al. 2012) and high-level features based VS (Ejaz et al. 2013). In the former work, three low-level features such as correlation, histogram, and moments of inertia are extracted from the underlying video, which are fused using an aggregation mechanism. An adaptive mechanism is utilized during summarizing

the video by combining the intermediate results, reducing the redundancy. Finally, the keyframes are extracted based on the attention values obtained using the aggregation mechanism.

In the second general-purpose VS method (Ejaz et al. 2013), keyframes are extracted using high-level features of visual attention model. The main bedrocks of this approach is incorporation of temporal-gradient directed dynamic visual saliency and DCT based static visual saliency for summarization, which are computational inexpensive compared to traditional optical-flow schemes. A non-linear weighted fusion is then used to combine the static and dynamic visual attention measures for generating an attention curve, which is used for keyframes extraction. In the coming sub-sections, we describe the various saliency detection models particularly used in summarization of DH videos.

Motion saliency

Motion saliency is one of the prominent saliency detection models used for video summarization in general (Mehmood et al. 2015) and DH video abstraction in particular (Ejaz et al. 2013). In the context of DH videos, motion saliency is effective in finding the inter-frame motion, providing a clue about the importance of a frame. During the DH examination, the gynecologist spends little time in examining the non-important areas by quickly moving the hysteroscope, producing fast inter-frame motion. On the other hand, more time is spend in visualizing the areas of interest by slowly moving the hysteroscope (Ejaz et al. 2013). This produces a significant amount of redundant frames with low inter-frame motion. This gives a clue that the keyframes lie in the sequence of frames having less inter-frame motion. The motion saliency is computed using Eq. 1 as follows:

$$M(DHF_i, P) = \sqrt{M_x^2(P) + M_y^2(P)} \tag{1}$$

Herein, $M_x(P)$ and $M_y(P)$ indicate the x and y components of the motion vector at pixel “P” of the DH frame “DHF_i” relative to the previous frame “DHF_{i-1}”. After computing the motion saliency for each frame, the obtained saliency values are normalized in the range of 0–1.

Texture saliency

In the domain of DH video abstraction, texture saliency can be used to identify the most injurious areas of DH frames. For this purpose, an entropy-directed texture segmentation approach is used. The texture saliency for a DH frame “DHF” can be calculated as follows:

$$E(DHF, P) = - \sum_{k=0}^{\eta-1} Hist_p(k) \log_2 (Hist_p(k)) \tag{2}$$

$$TXI(DHF, P) = \begin{cases} 0 & \text{if } E(DHF, P) < \tau \\ 1 & \text{otherwise} \end{cases} \tag{3}$$

$$TS(DHF, P) = \begin{cases} DHF(P) & \text{if } TXI(DHF, P) = 1 \\ 0 & \text{if } TXI(DHF, P) = 0 \end{cases} \tag{4}$$

Firstly, the entropy “E” of pixel “P” at frame “DHF” is calculated using Eq. 2. A texture segmentation with $\tau = 0.8$ is then applied on “E” as shown in Eq. 3, resulting an injury-free texture image “TXI”. Then the edges of “TXI” are smoothed using closing. Next, the holes in “TXI” are filled, providing the mask image, based on which the injurious parts of the DH frame can be identified (Ejaz et al. 2013). It is worth mentioning that texture saliency “TS” contains only the injurious regions of DH frame. Therefore, a salient frame in this context is the one, whose larger area is injurious. Alternatively, the DH frame with high proportional of injurious regions is assigned a saliency value of 1. The remaining of the frames get their saliency scores relative to the maximum value. To sum up, texture saliency effectively segments the injurious parts of DH frames and assigns them higher saliency scores compared to frames with low-proportional of injurious regions.

Multi-scale contrast map

Contrast map is an effective measurement for finding the uniqueness of a region in a video frame, which has been widely used in computer vision algorithms (Ejaz et al. 2013; Perazzi et al. 2012). In the context of DH video summarization, we explore multi-scale color contrast, which is more effective in salient objects identification of different sizes. The multi-scale color contrast map of a DH frame is calculated in Eqs. 5 and 6 as follows:

$$CCM_c^l(DHF_c, P) = \sum_{q \in N(P)} \left\| DHF_c^l(P) - DHF_c^l(q) \right\|^2 \tag{5}$$

$$MSCCM(DHF, P) = \sum_{l=1}^{\eta} CCM^l(DHF, P) \tag{6}$$

Herein, “DHF_c” indicates one of the three color channels (red, green or blue) for the frame “DHF”. “l” refers to the scale of contrast and “N(p)” shows the neighborhood of the pixel “P”, which is 5×5 . The value of η is set to 3, indicating the levels of Gaussian pyramid (Liu et al. 2011).

Curvature map

During DH examination, the gynecologists move the hysteroscope with a certain orientation to effectively visualize the areas of interests. The previously mentioned saliency detection models are less effective in handling DH frames with such abnormalities. In this context, curvature map is comparatively more effective due to its rotational-invariant property in finding the keyframes with abnormalities, which are captured from different orientations. Furthermore, the neuroscience and psychophysical research also dictates that curvature is an important factor in determining the saliency and improving the decision of gynecologists in selection of keyframes. The curvature map “CM” for a DH frame “DHF” can be calculated using Eqs. 7 and 8 as follows (Mehmood et al. 2013):

$$CM = \left| \nabla^2 g \right| = \sqrt{g_{xy}^2 + g_{xx}^2 + g_{yx}^2 + g_{yy}^2} \tag{7}$$

$$\left\{ \begin{array}{l} g(x, y) = DHF(x, y) \times \Psi \\ \Psi = e^{-\frac{x^2+y^2}{2\sigma^2}} \\ g_{xy} = \frac{\partial^2 g}{\partial x \partial y}, \quad g_{xx} = \frac{\partial^2 g}{\partial x^2}, \quad g_{yx} = \frac{\partial^2 g}{\partial y \partial x}, \quad g_{yy} = \frac{\partial^2 g}{\partial y^2} \end{array} \right\} \quad (8)$$

Fusion scheme and extraction of keyframes

After computing the numerous saliencies for each frame, it is important to combine them to generate a fused saliency map for keyframes extraction. There are several ways to fuse the different saliencies such as linear fusion, linear weighted fusion, max fusion, and non-linear weighted fusion (Ejaz et al. 2013). For ease of understanding, we have used the weighted linear fusion for combining the different saliencies. To this end, the score of each saliency is normalized in the range 0–1. Then the mean of non-zero gray-levels is determined as the saliency score for each feature. The normalized values are then fused to get a final aggregated saliency score for each DH frame. Based on the fused saliency scores, an attention curve is generated, which is then used for keyframes extraction. An illustration of keyframes extraction using attention curve is given in Fig. 2.

After calculating the attention curve, the user/gynecologist is asked to specify the number of keyframes “NKF” for a given DH video. Accordingly, the video is divided into “NKF” number of shots. Within each shot, the frame with highest saliency score is determined as keyframe. By only changing the value of “NKF”, a different set of keyframes can be extracted, enabling gynecologists to analyze the DH video at different summarization levels.

Experimental results and discussion

This sub-section illustrates the performance evaluation of various saliency detection models and general-purpose VS methods for DH videos abstraction. Experiments were performed on a set of DH videos according to Ejaz et al. (2013), each of 2–3 min duration having frame rate of 30 frames/s. MATLAB R2015a was used for conducting the experiments and running the simulation. To obtain the ground truth, gynecologists were asked to select a number of diagnostically important frames from the mentioned DH

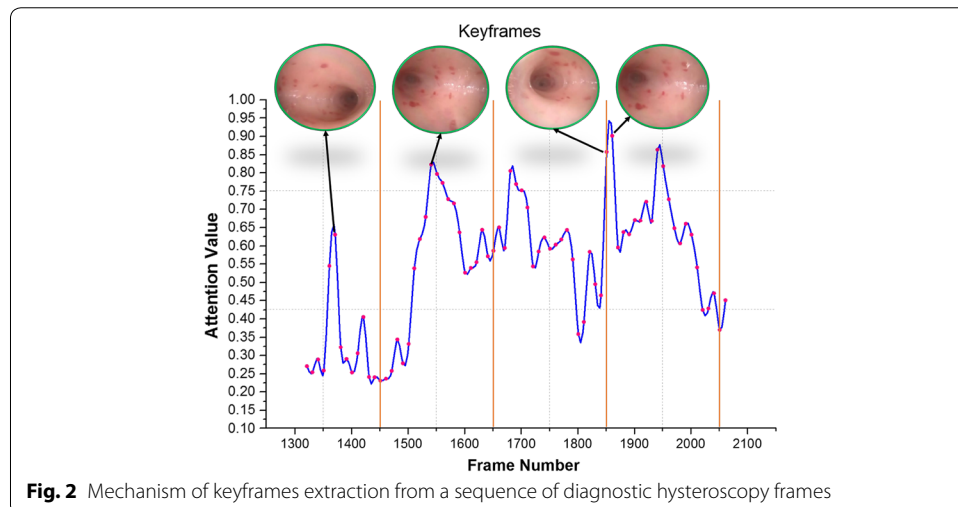


Fig. 2 Mechanism of keyframes extraction from a sequence of diagnostic hysteroscopy frames

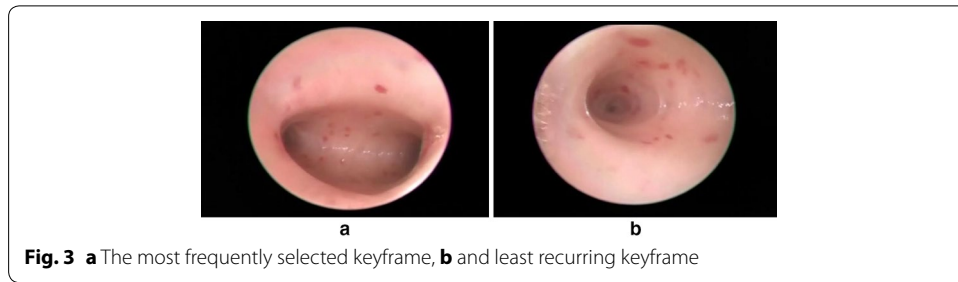
videos. In the current study, a total of five saliency detection models and two general-purpose VS schemes were considered for evaluation, in terms of keyframes extraction, F-measure (Ejaz et al. 2013), and accuracy for DH videos. These models include motion saliency model (Mehmood et al. 2015), multi-scale contrast map (Mehmood et al. 2014), texture saliency (Ejaz et al. 2013), curvature map (Mehmood et al. 2013), and SIM saliency (Bruce and Tsotsos 2005). For keyframes selection, the mean of attention values for each saliency detection scheme was considered as attention curve threshold. The frames with attention values greater than the attention curve threshold are considered as keyframes while the remaining frames are selected as non-keyframes. The frames extracted by these methods were then compared with the ground truth to find the accuracy and F-measure for each VS scheme.

Table 1 illustrates the comparative results of numerous general-purpose and domain-specific saliency detection models for summarization of DH videos. From the results, it can be seen that the performance of SIM and TS is same. Motion saliency reports 30 % accuracy, indicating worse results in this experiment. The best performance of 70 % accuracy is achieved by a hybrid visual saliency model (HSDM), consisting of motion, contrast, texture, and curvature saliencies. The most frequent and least recurring keyframes from Table 1 are shown in Fig. 3. The F-measure based performance evaluation

Table 1 Performance evaluation of numerous saliency detection models for a sample hysteroscopy video

Serial no.	Saliency detection model	Total keyframes	Extracted keyframes	Attention curve threshold	Accuracy (%)	Extracted frame numbers
1	MS	10	3	0.024	30	1584, 2231, 2977
2	MSCM	10	6	0.015	60	1584, 2070, 2289, 2323, 2668, 2977
3	TS	10	5	0.872	50	2070, 2264, 2289, 2323, 2977
4	CM	10	4	0.226	40	2150, 2386, 2668, 2977
5	MS + MSCM + TS	10	6	0.571	60	1584, 2070, 2231, 2323, 2386, 2668
6	HSDM	10	7	0.847	70	1584, 2231, 2264, 2323, 2386, 2668, 2977
7	SIM	10	5	0.984	50	1584, 2070, 2231, 2289, 2977

The score in italic font represents the best accuracy among the given methods



given in Fig. 4 also verifies the fact that HSDM is comparatively more suitable for keyframes extraction from DH videos.

Table 2 presents a comparison of general VS methods, general-purpose saliency detection method, and domain-specific saliency detection schemes. The former category includes two VS methods, utilizing low and high-level features, respectively. The second scheme is a general-purpose saliency detection method which is used here for keyframes

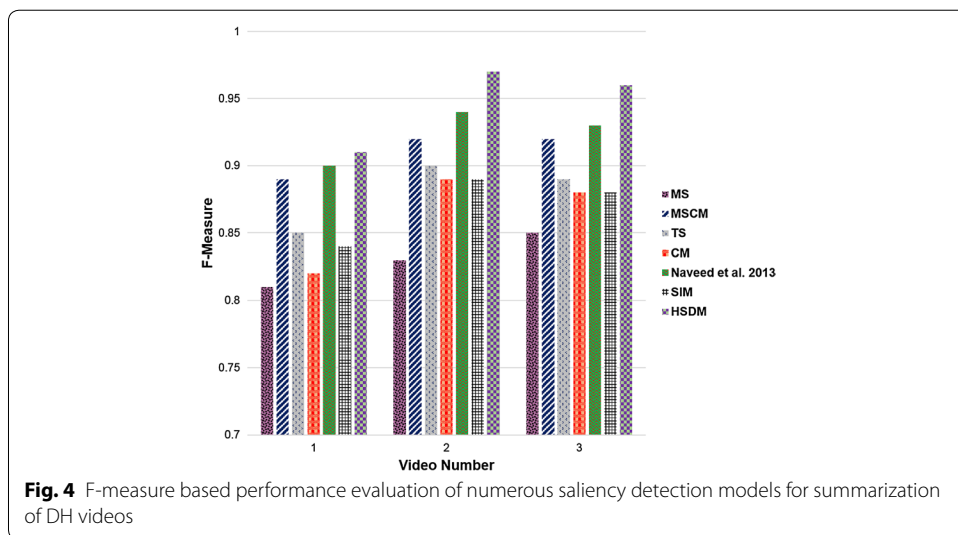
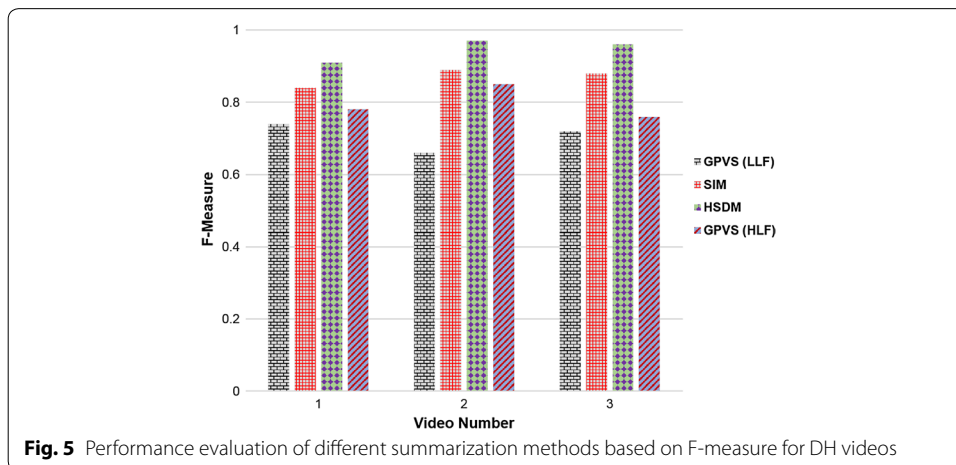


Table 2 Comparison of general video summarization methods, general-purpose and domain-specific saliency detection based summarization schemes for keyframes extraction from a sample hysteroscopy video

Serial. no	Method name	Category of video summarization	Total keyframes	Feature extraction model	Total number of frames	Accuracy (%)
1	Ejaz et al. (2012)	General-purpose VS	10	Low-level features	3529	20
2	Ejaz and Baik (2012)	General-purpose VS	10	High-level features	3529	40
3	SIM (Bruce and Tsotsos 2005)	General-purpose VS	10	General-purpose saliency detection	3529	50
4	Hybrid saliency detection model	Domain-specific VS	10	Domain-specific saliency detection	3529	70

The score in italic font represents the best accuracy among the given methods

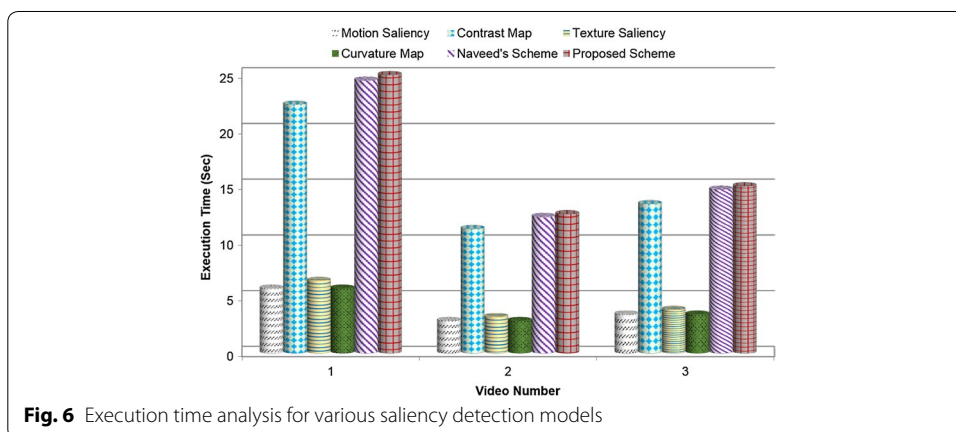


extraction from DH videos. The latter category illustrates a hybrid saliency detection framework, specific for DH videos. From the experiments, it can be noted that the suggested HSDM produces promising results by giving an accuracy of 70 %, hence dominating other related VS approaches. The same fact is also verified using F-measure based performance evaluation as given in Fig. 5.

Figure 6 highlights the computational complexity of numerous saliency detection models in terms of execution time for keyframes extraction based on a set of DH videos. The graph indicates that the running time of motion, texture, and curvature saliency is almost same. Multi-scale contrast map is computationally expensive compared to the former saliencies. The running time of the suggested HSDM is slightly greater than Ejaz et al.'s scheme (2013) but it provides higher accuracy and F-measure compared to other general-purpose and domain specific VS methods.

Conclusion

During the process of diagnostic hysteroscopy, several hysteroscopic sessions are conducted for a single patient per day. Due to the large number of patients and their multiple hysteroscopic sessions, an enormous amount of hysteroscopic videos are collected. However, a limited number of frames are required for actual diagnosis process, whose



manual extraction by gynecologists is comparatively difficult and time consuming due to large-sized hysteroscopic videos. To facilitate the gynecologists in browsing for desired diagnostically important contents, video summarization schemes are used. In this work, we have conducted a comprehensive study of numerous generic and domain-specific video summarization schemes for hysteroscopic videos. Further, we have investigated the performance of various visual attention models combined with domain knowledge for summarization of DH videos. Our findings based on numerous experiments are reported as follows:

1. The general-purpose video summarization schemes are less suitable for hysteroscopic videos due to their significant similarity in color and texture, and absence of shot boundaries.
2. Among the evaluated visual saliency models, a hybrid saliency detection model comprising of motion, texture, multi-scale contrast, and curvature is found as the best combination of visual saliencies for hysteroscopic video abstraction, considering its accuracy and extracted keyframes.

In future, we have intension to focus on minimizing the computational complexity of the system by extracting light-weight features from DH videos. Another possible future direction is to combine data hiding [watermarking (Liu et al. 2015; Liu et al. 2016), image and video steganography (Mstafa and Elleithy 2015; Muhammad et al. 2015; Lin et al. 2015)] with the video summarization frameworks by embedding the patient and gynecologists data in DH videos/keyframes, resulting in secure and privacy-preserving VS framework as presented in (Muhammad et al. 2015) for secure visual contents retrieval from personalized repositories and other mobile healthcare applications (Lv et al. 2016). Furthermore, we are also planning to explore deep learning and incorporate GPUs based processing (Mei and Tian 2016; Mei 2014) for efficient keyframes extraction, their indexing and retrieval (Rho et al. 2008; Rho et al. 2011; Rho and Hwang 2006).

Authors' contributions

KM and SWB proposed the basic idea of this work. KM carried out the experiments. KM, JA, and MS analyzed the experiments. KM wrote the paper and other authors helped in revising the paper. All authors read and approved the final version of the manuscript.

Author details

¹ Intelligent Media Laboratory, Department of Digital Contents, College of Electronics and Information Engineering, Sejong University, Seoul, Republic of Korea. ² Digital Image Processing Laboratory, Department of Computer Science, Islamia College Peshawar, Khyber Pakhtunkhwa, Pakistan.

Acknowledgements

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

Authors' Information

Khan Muhammad received his BCS degree in Computer Science from Islamia College, Peshawar, Pakistan in 2014 with research in information security. Currently, he is pursuing MS leading to Ph.D. degree in digital contents from Sejong University, Seoul, Republic of Korea. He is working as a researcher at Intelligent Media Laboratory (IM Lab) since 2015. His research interests include image and video processing, data hiding, image and video steganography, video summarization, diagnostic hysteroscopy, and wireless capsule endoscopy. Jamil Ahmad received his BCS degree in Computer Science from the University of Peshawar, Pakistan in 2008. He received his Master's degree in 2014 from Islamia College, Peshawar, Pakistan. He is also a faculty member in the Department of Computer Science, Islamia College, Peshawar. Currently, he is pursuing Ph.D. degree in Sejong University, Seoul, Korea. His research interests include image analysis, content based multimedia retrieval and computer vision. 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 Ph.D. degree in Digital Contents from Sejong University, Seoul, Republic of Korea. He is now working as a research associate at Islamia College Peshawar, Pakistan. He is also the head of Digital Image Processing Laboratory (DIP

Lab) at Islamia College Peshawar, Pakistan. His research interests include digital image super-resolution and reconstruction, sparse coding, video summarization and prioritization, image/video quality assessment, and image/video retrieval. Sung Wook Baik is a Professor in the Department of Digital Contents at Sejong University. He 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 School of Electronics and Information Engineering, Sejong University, Seoul, Korea, where he is currently a Full Professor and Dean of Digital Contents. His research interests include computer vision, multimedia, pattern recognition, machine learning, data mining, virtual reality, and computer games.

Competing interests

The authors declare that they have no competing interests.

Received: 26 May 2016 Accepted: 30 August 2016

Published online: 06 September 2016

References

- Almeida J, Leite NJ, Torres RDS (2012) Vison: video summarization for online applications. *Pattern Recogn Lett* 33:397–409
- Almeida J, Leite NJ, Torres RDS (2013) Online video summarization on compressed domain. *J Vis Commun Image Represent* 24:729–738
- Bruce N, Tsotsos J (2005) Saliency based on information maximization. In: *Advances in neural information processing systems*, pp 155–162
- De Avila SE, da Luz A, de Araujo A, Cord M (2008) VSUMM: an approach for automatic video summarization and quantitative evaluation. In: *XXI Brazilian symposium on computer graphics and image processing, 2008. SIBGRAPI'08*, pp 103–110
- De Avila SEF, Lopes APB, da Luz A, de Albuquerque Araújo A (2011) VSUMM: a mechanism designed to produce static video summaries and a novel evaluation method. *Pattern Recogn Lett* 32:56–68
- Ejaz N, Baik SW (2012) Video summarization using a network of radial basis functions. *Multimed Syst* 18:483–497
- Ejaz N, Tariq TB, Baik SW (2012) Adaptive key frame extraction for video summarization using an aggregation mechanism. *J Vis Commun Image Represent* 23:1031–1040
- Ejaz N, Mehmood I, Baik SW (2013a) MRT letter: visual attention driven framework for hysteroscopy video abstraction. *Microsc Res Tech* 76:559–563
- Ejaz N, Mehmood I, Baik SW (2013b) Efficient visual attention based framework for extracting key frames from videos. *Sig Process Image Commun* 28:34–44
- Gavião W, Scharcanski J (2005) Content-based diagnostic hysteroscopy summaries for video browsing. In: *18th Brazilian symposium on computer graphics and image processing, 2005. SIBGRAPI 2005*, pp 21–28
- Gavião W, Scharcanski J (2007) Evaluating the mid-secretory endometrium appearance using hysteroscopic digital video summarization. *Image Vis Comput* 25:70–77
- Gavião W, Scharcanski J, Frahm J-M, Pollefeys M (2012) Hysteroscopy video summarization and browsing by estimating the physician's attention on video segments. *Med Image Anal* 16:160–176
- Lin C-C, Liu X-L, Tai W-L, Yuan S-M (2015) A novel reversible data hiding scheme based on AMBTC compression technique. *Multimed Tools Appl* 74:3823–3842
- Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X et al (2011) Learning to detect a salient object. *IEEE Trans Pattern Anal Mach Intell* 33:353–367
- Liu Z, Zhang F, Wang J, Wang H, Huang J (2015) Authentication and recovery algorithm for speech signal based on digital watermarking. *Sig Process* 123:157–166
- Liu Z, Huang J, Sun X, Qi C (2016) A security watermark scheme used for digital speech forensics. *Multimed Tools Appl*. doi:10.1007/s11042-016-3533-9
- Lux M, Marques O, Schöffmann K, Böszörményi L, Lajtai G (2010) A novel tool for summarization of arthroscopic videos. *Multimed Tools Appl* 46:521–544
- Lv Z, Chirivella J, Gagliardo P (2016) Bigdata oriented multimedia mobile health applications. *J Med Syst* 40:1–10
- Ma Y-F, Hua X-S, Lu L, Zhan H-J (2005) A generic framework of user attention model and its application in video summarization. *IEEE Trans Multimed* 7:907–919
- Mehmood I, Ejaz N, Sajjad M, Baik SW (2013) Prioritization of brain MRI volumes using medical image perception model and tumor region segmentation. *Comput Biol Med* 43:1471–1483
- Mehmood I, Sajjad M, Baik SW (2014) Video summarization based tele-endoscopy: a service to efficiently manage visual data generated during wireless capsule endoscopy procedure. *J Med Syst* 38:1–9
- Mehmood I, Sajjad M, Ejaz W, Baik SW (2015) Saliency-directed prioritization of visual data in wireless surveillance networks. *Inf Fusion* 24:16–30
- Mehmood I, Sajjad M, Rho S, Baik SW (2016) Divide-and-conquer based summarization framework for extracting affective video content. *Neurocomputing* 174:393–403
- Mei G (2014) Evaluating the power of GPU acceleration for IDW interpolation algorithm. *Sci World J* 2014:1–8
- Mei G, Tian H (2016) Impact of data layouts on the efficiency of GPU-accelerated IDW interpolation. *SpringerPlus* 5:1
- Mstafa RJ, Elleithy KM (2015) A video steganography algorithm based on Kanade-Lucas-Tomasi tracking algorithm and error correcting codes. *Multimed Tools Appl*. doi:10.1007/s11042-015-3060-0
- Muhammad K, Sajjad M, Mehmood I, Rho S, Baik SW (2015) A novel magic LSB substitution method (M-LSB-SM) using multi-level encryption and achromatic component of an image. *Multimed Tools Appl*. doi:10.1007/s11042-015-2671-9

- Muhammad K, Mehmood I, Lee MY, Ji SM, Baik SW (2015b) Ontology-based secure retrieval of semantically significant visual contents. *J Korean Inst Next Gener Comput* 11:87–96
- Muhammad K, Sajjad M, Baik SW (2016) Dual-level security based cyclic 18 steganographic method and its application for secure transmission of keyframes during wireless capsule endoscopy. *J Med Syst* 40:1–16
- Perazzi F, Krähenbühl P, Pritch Y, Hornung A (2012) Saliency filters: contrast based filtering for salient region detection. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, 2012, pp 733–740
- Rho S, Hwang E (2006) FMF: query adaptive melody retrieval system. *J Syst Softw* 79:43–56
- Rho S, Han B-J, Hwang E, Kim M (2008) MUSEMBLE: a novel music retrieval system with automatic voice query transcription and reformulation. *J Syst Softw* 81:1065–1080
- Rho S, Hwang E, Park JH (2011) M-MUSICS: an intelligent mobile music retrieval system. *Multimed Syst* 17:313–326
- Scharcanski J, Neto WG, Cunha-Filho JS (2006) Diagnostic hysteroscopy video summarization and browsing. In: *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp 5680–5683
- Truong BT, Venkatesh S (2007) Video abstraction: a systematic review and classification. In: *ACM transactions on multimedia computing, communications, and applications (TOMM)*, vol 3, p 3

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
