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

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/

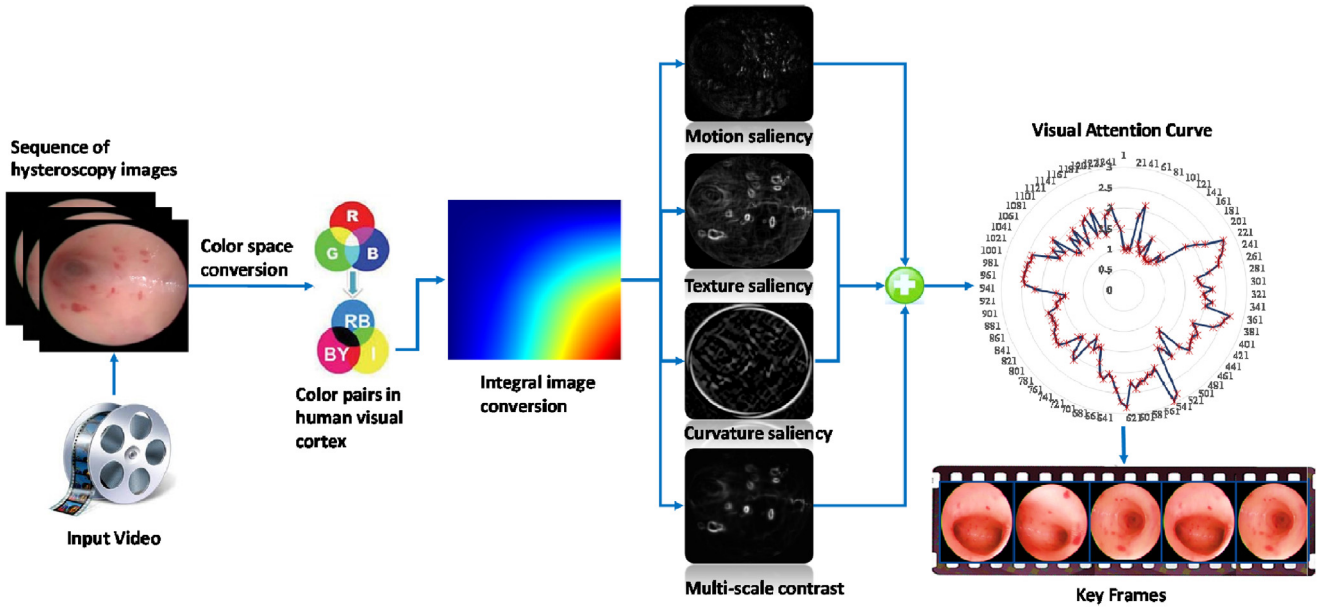

Fig. 1. Framework of the proposed system.

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.

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

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

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

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