Scalable Video Summarization using Skeleton Graph and Random Walk

Rameswar Panda, Sanjay K. Kuanar, Ananda S. Chowdhury
Department of Electronics and Telecommunication Engineering
Jadavpur University
Kolkata 700032, India
Email: {rameswar183, sanjay.kuanar}@gmail.com, aschowdhury@etce.jdvu.ac.in

Abstract—Scalable video summarization has emerged as an important problem in present day multimedia applications. Effective summaries need to be provided to the users for videos of any duration at low computational cost. In this paper, we propose a framework which is scalable during both the analysis and the generation stages of video summarization. The problem of scalable video summarization is modeled as a problem of scalable graph clustering and is solved using skeleton graph and random walks in the analysis stage. A cluster significance factor-based ranking procedure is adopted in the generation stage. Experiments on videos of different genres and durations clearly indicate the supremacy of the proposed method over a recently published work.

Keywords-Scalable video summarization; Skeleton graph; Random Walk; Cluster Significance factor.

I. INTRODUCTION

With recent advances in multimedia technology, there has been a tremendous increase in the number of digital videos in the Internet. For example, as of December, 2013, one of the most primary video sharing websites, YouTube, has reported that more than 1 billion unique users visit their site each month and 100 hours of video are uploaded every minute [1]. It is practically impossible to watch such huge video data with a large number of frames. Video Summarization provides a solution to the above problem by showing the users only some meaningful/key frames from the entire video in form of storyboards or skims [2]- [3]. Although video summarization has been an extensively studied problem in multimedia research, most of the previous methods have focused on short duration videos [4]- [8]. For long duration videos like surveillance feeds, lecture videos or home videos, existing methods may fail to capture proper summary. Moreover, computational complexities of such techniques are found to increase rapidly with increase in the duration of the video, which makes them unsuitable for real-time social multimedia applications. Hence, an efficient and fast technique is necessary for summarizing videos of particularly large duration. This problem of scalable video summarization is addressed in the present work.

Work on scalable video summarization is reported in [6], where the authors have shown the scaling in the generation stage only. However, the concept of scalability can be applied in a more comprehensive manner [9]. In addition to providing different summaries depending on the length constraints during the generation stage, the video summarization framework can also be designed to scale to the size of the input video during the analysis stage. In sharp contrast to [6], we introduce a scalable video summarization framework for both the analysis of the input video as well as the generation of summaries according to the user-specified length constraints. We apply a novel scalable graph clustering technique in the analysis stage using a skeleton graph and a Random walker algorithm (RWA) [10]. A cluster significance factor is used in the ranking procedure in the generation stage.

The rest of the paper is organized as follows: in Section II, we discuss the related work and highlight our contributions. In Section III, we describe the proposed method. Experimental results with detailed analysis are presented in Section IV. Finally, we conclude the paper in Section V with an outline of directions for future research.

II. RELATED WORK

We mention here some of the recently reported works in the field of video summarization. Almeida et al. [4] proposed the key frame extraction technique using the notion of similarity between successive frames. Kuanar et al. [11] proposed a summarization technique based on dynamic Delaunay clustering and information theoretic pre-sampling. Both the above methods can generate summaries of only short duration video clips (maximum duration up to 4 minute). Recently, Zheng et al. [12] introduced a story-driven summarization for egocentric video using a metric of influence. In general, most of the existing video summarization methods follow a single scale approach, i.e., the output is always a single summary. However, a single scale summary is inadequate to provide information with different levels of details, as may be required in a narrative hierarchy. Scalable summarization of video first appears in [6] where the scalability is maintained at the generation stage

using hierarchical clustering with average linkage and a ranking procedure. However, the analysis of input video sequence in [6] is not scaled to the size of the input. As a result, it fails to generate effective summaries for long duration videos due to substantial increase in the computational cost associated with the analysis stage. Recently, Yang et al. [13] proposed a scalable video summarization framework based on sparse dictionary selection. However, this method also provides scalability only in the generation stage and its usefulness is demonstrated solely on the consumer videos. So, a summarization method which is scalable at both the analysis and the generation stages and is applicable to videos from different genres has not been reported to the best of our knowledge.

We propose a scalable video summarization method based on scalable graph clustering and apply it to documentary and educational videos. The main contributions of this work are as follows:

i. From the theoretical perspective, we model the scalable video summarization problem as one of scalable graph clustering. A computationally efficient solution is proposed using skeleton graph and random walks.

ii. From the application standpoint, we design a novel architecture that provides scalability to both the analysis and the generation stages of video summarization.

III. PROPOSED FRAMEWORK

The proposed framework consists of a 3-step analysis stage followed by a 1-step generation stage. A block diagram showing all the four steps is presented in Fig. 1. In the first step, we construct a video similarity graph (VSG) from the frames constituting the input video. We then extract a skeleton graph, which is a subgraph of the above VSG. In the second step, a minimum spanning tree (MST) based clustering is used over the skeleton graph to obtain the initial clusters. In the third step, we propagate this initial clustering result into the VSG using a random walker algorithm [10] to obtain the final clusters. In the fourth and final step, the key frames (closest to the centroids of each cluster) are arranged according to the cluster significance factor. We now discuss all the four steps in more details.

A. Extraction of Skeleton Graph

The first step towards video summarization is to split the video stream into a set of meaningful and manageable basic units by the process of temporal video segmentation. Instead of detecting accurate shot changes, which is affected by the presence of variety of transitions (e.g., fade in, fade out, abrupt cut) between successive video frames [2]- [3], we achieve temporal video segmentation by dividing the video stream into a set of frames (still images) [4], [7], [11], [14]. Most of the previous methods consider only a subset of video frames extracted at a predefined sampling rate. However, the choice of the sampling rate greatly influences the content of the video summary [11]. In this work, we consider all the video frames to generate a more effective video summary.
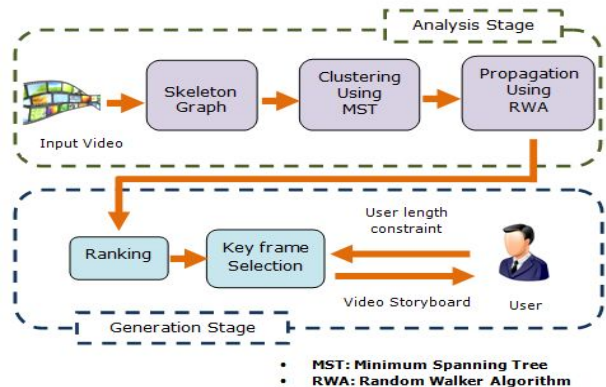

Fig. 1. Overview of the proposed scalable video summarization framework

Color is the most expressive of all the visual features. Hence, we represent each video frame by a 256-dimensional feature vector obtained from a color histogram using HSV color space (16 ranges of H, 4 ranges of S, and 4 ranges of V) [11]. The same color feature is also used in [6]. We then construct the VSG as a weighted complete graph, where each frame is represented as a vertex. The weight W_{ij} of the edge connecting the vertices i and j is given by:

$$W_{ij} = exp\left(-d_{ij}^2/\sigma^2\right) \qquad (1)$$

where d_{ij} is the histogram intersection distance [15] between the frames i and j and σ is a normalization parameter that determines the extent of similarity between any two frames. As suggested in [14], $\sigma = \beta * max(d)$, where $\beta \leq 0.2$ and d is the set of all pair-wise distances. If $d_{ij} \leq \sigma$, then the two frames are deemed to have significant overlap and the edge between them is preserved. Conversely, if $d_{ij} > \sigma$, the frames i and j are treated as dissimilar and the associated edge is removed. In this way, we first reduce the *size* of the VSG. We next reduce the *order* of the size-reduced VSG. High degree vertices play a pivotal role in maintaining the overall structure of a dense graph. Thus, we extract a skeleton of the VSG by choosing vertices whose degrees are higher than a certain threshold [16]. The degree threshold for skeleton graph construction is chosen according to the duration of the input video stream. For example, the initial VSG of the video stream *Health Communication* with duration of 49 minutes 12 seconds contains 85107 vertices whereas the corresponding skeleton graph consists of only 3172 vertices.

B. Clustering of Skeleton Graph via MST

The objective of clustering is to remove visual redundancies between the video frames. Note that the skeleton graph obtained in the previous stage preserves the overall structure of the VSG despite its reduced order and size. So, we cluster the skeleton graph instead of the VSG using a minimum spanning tree (MST) based approach. We adopt the MST-based clustering

because: i) it is capable of detecting clusters with irregular boundaries and ii) it does not require the number of clusters in advance. Using the MST-based clustering, we remove edges that satisfy a pre-defined inconsistency measure in order to obtain separate clusters. Let e denotes an edge in the MST connecting the vertices v_1 and v_2 with weight w. Let N_1 and N_2 be the set of direct neighborhoods of v_1 and v_2. Following the normal distribution of edge weights in MST [17], the inconsistency measure can be formulated as:

$$w > max\left(w_{N_1} + \sigma_{N_1}, w_{N_2} + \sigma_{N_2}\right) \qquad (2)$$

where w_{N_1} and w_{N_2} are the average weights in N_1 and N_2 respectively. Similarly, σ_{N_1} and σ_{N_2} represent the standard deviations of edge weights in N_1 and N_2 respectively. All the edges that satisfy the above inconsistency criterion are removed from the tree because such edges are most likely to be inter-cluster edges. The above procedure results in a set of disjoint subtrees. Each subtree represents a separate cluster.

C. Cluster propagation via Random Walker Algorithm

Vertices in the skeleton graph with known cluster information are now used as seeds for clustering the VSG. Recently, Joyce et al., [16] proposed a cluster propagation algorithm for large-scale social networks using an improved version of the weighted kernel k-means algorithm. However, the work presented in [16] requires several iterative steps to propagate the initial clustering result into the original graph. We obtain the final multi-label clustering in a single interactive step instead using a Random walker algorithm [10], [20].

Let the vertices of a skeleton graph are grouped into k clusters using MST clustering. Assume these vertices as seed/marked vertices (V_M) in the VSG and the remaining vertices as unmarked (V_N). Given a set of weights, the probability that a random walker at node v_i moves to node v_j is given by $p_{ij} = \frac{w_{ij}}{d_i}$, where d_i is the degree of the vertex. The objective of RWA is to calculate the probability that a random walker starting from an unmarked vertex in V_N will first reach each of the k labeled marked vertices in V_M. As suggested in [10], exact solution can be obtained without actually simulating a random walk and translating the problem to a discrete Dirichlet problem. We adopt a method based on anisotropic interpolation on graphs to solve this discrete Dirichlet problem [18]. The discrete Laplacian matrix in this connection is defined as:

$$L_{ij} = \begin{cases} d_i & if \ i = j \\ -w_{ij} & if \ v_i \ and \ v_j \ are \ adjacent \\ 0 & otherwise \end{cases} \qquad (3)$$

Using the marked and unmarked vertices, we can reorder the matrix L as:

$$L = \begin{bmatrix} L_M & B \\ B^T & L_N \end{bmatrix} \qquad (4)$$

In equation (4), L_M and L_N represents the Laplacian matrix between marked vertices and unmarked vertices respectively.

Let the probability at each vertex v_i for each cluster label s, $0 < s \le k$, be denoted by x_i^s. Now, we define a labeled vector with length $|V_M|$ for each cluster label s at a vertex $v_j \in V_M(j = 1, \dots, m)$ as:

$$m_j^s = \begin{cases} 1, & if \ label(v_j) \ = \ s \\ 0, & if \ label(v_j) \ \neq \ s \end{cases} \qquad (5)$$

Subsequently, the potentials for all the cluster labels can be found by solving the system $L_N X = -BM$, where X has columns given by each x^s and M has columns given by each m^s. We have implemented the standard conjugate gradient algorithm to solve the system of equations [19]. Finally, the cluster labels for each node v_i is obtained by the cluster label corresponding to $max(x_i^s)$.

D. Ranking and Key frame selection

The objective of the ranking stage is to generate suitable summaries for a wide range of potential lengths without increasing the computational complexity. To obtain a scalable representation of video storyboard, we ranked the clusters according to their Cluster Significance Factor (CSF). CSF(s) signifies the content represented by each cluster s and is defined as:

$$CSF(s) = \frac{C_S}{\sum_{j=1}^{k} C_j} \qquad (6)$$

In the above equation C_S denotes the number of frames in the cluster s, C_j denotes the number of frames in the cluster j and k represents the total number of clusters. We rank the clusters in the VSG according to the decreasing CSF values. So, the cluster with the highest CSF value has the rank 1 and so on. For each user request, the ranked list is processed to produce the video summary satisfying the length constraint. Let us assume that the user length request is l and the number of clusters obtained from the analysis stage is k. Then, following cases may occur:

Case 1. $l \le k$: When the user length request is less than the number of clusters, the algorithm captures the overall content of the video by selecting one key frame from each of the top l clusters in the ranked list. For each cluster, the frame closest to the cluster centroid is selected.

Case 2. $l > k$: When the user length request is more than the number of clusters, the algorithm follows an iterative strategy to select l number of frames from k clusters. This strategy is given below:

1. First, choose one frame closest to the centroid from each cluster to obtain a summary consisting of k key frames. This step ensures the proper content coverage of the video storyboard.

2. Next, select the remaining $l - k$ frames from the ranked list. In order to extract more than one frame from any cluster, we pick the frame whose distance from the mean of the already chosen frames from that cluster is largest. The goal of this step

Video Name	Video ID	Length (Hr:Mm: Sec)	Frames	Genre
A New Horizon, Segment 06	1	00:01:05	1944	Documentary
Drift Ice as a Geologic Agent, Segment 03	2	00:01:31	2742	Educational
Space Work 5	3	00:29:49	52,202	Documentary
Health Communication	4	00:49:12	85,107	Educational
Tribute to World War II Nisei Veterans	5	01:34:33	1,64,550	Educational

Measures	Video ID	S1_ [6]	S1_Proposed	S2_ [6]	S2_Proposed	S3_ [6]	S3_Proposed
Fidelity	1	0.56	0.54	0.60	0.63	0.61	0.62
	2	0.42	0.44	0.43	0.58	0.56	0.59
	3	0.47	0.61	0.46	0.58	0.49	0.60
	4	0.50	0.67	0.52	0.72	0.47	0.69
	5	0.32	0.74	0.47	0.68	0.35	0.73
SRD	1	3.67	3.75	4.31	4.41	4.43	4.76
	2	2.98	3.20	3.01	3.86	3.65	3.98
	3	3.20	3.68	3.37	4.31	3.91	5.34
	4	3.21	5.64	3.96	6.48	4.31	7.01
	5	2.90	4.02	3.02	5.78	4.10	6.56

is to capture maximum variation in the summary.

Note that our generation stage is able to produce scalable summaries depending on the user length constraints without incurring additional computational cost from analyzing the video sequence again (*analyze once, generate many*).

Algorithm : SCALABLE VIDEO SUMMARIZATION

INPUT: A video and user length request l.

OUTPUT: l key frames.

Procedure

Extraction of Skeleton Graph

1. Construct the VSG with edge weights given by eq. (1)

2. Reduce the size of the VSG on removing the edges with distance $d_{ij} > \sigma$, where $\sigma = \beta * max(d)$[14].

3. Construct the skeleton graph of the size reduced VSG by choosing vertices whose degrees are higher than a certain threshold.

Clustering of Skeleton Graph via MST

4. Construct a minimum spanning tree of the skeleton graph.

5. Remove the inconsistent edges from the tree using eq. (2).

6. Final disjoint subtrees represent separate clusters (k).

Cluster propagation via Random Walker Algorithm

7. Define the discrete Laplacian matrix (L) as in eq. (4).

8. Find the potentials for all the cluster labels (s) by solving the system $L_N X = -BM$

9. Apply the conjugate gradient algorithm to solve the above system of equations [19].

10. Finally, obtain the cluster label for each node v_i using $max(x_i^s)$.

Ranking and Key frame selection

11. Rank the clusters in decreasing order according to eq. (6).

12. If $l \leq k$, then select one key frame from each of the top l clusters in the ranked list. For each cluster, the frame closest to the cluster centroid is selected.

Else, follow the iterative strategy given below.

First, obtain a summary consisting of k key frames by choosing one frame per cluster.

Next, select the remaining $l - k$ frames from the ranked list by picking a frame whose distance from the mean of the already chosen frames from that cluster is largest.

End Procedure

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present the experimental results. Both qualitative and quantitative performance comparisons are shown.

A. Evaluation dataset and Performance measures

We have experimented with 5 test video segments belonging to different genres (e.g., documentary, educational) from the Open Video (OV) projects [21]. All test videos are in MPEG-1 format with 352×240 pixels and durations range from 1 min 05 sec to 1 hr, 34 min, 33 sec. Detailed information of the test videos are given in Table I. Fidelity, a global descriptor of the summary and Shot Reconstruction Degree (SRD), a local descriptor of the summary [11], [15] are used as the two objective measures. For the subjective study, Informativeness and Visual pleasantness [6] are applied.

B. Comparative performance analysis experiments

In this section we make a comparative performance analysis to evaluate the results of the proposed method. We already mentioned two reported works on scalable video summarization [6], [13]. However, the method in [13] is limited to only consumer videos. For this reason, we have compared our work with [6], which we call the baseline method. We have implemented and tested the baseline method on the five datasets in Table I at three different scales (user length requests) of S1 = 5 frames, S2 = 10 frames and S3 = 15 frames. Table II presents a comparative performance analysis of our proposed method and the baseline method using fidelity and SRD. This table demonstrates that the proposed method outperforms the baseline method in as many as 28 out of 30 cases (5 videos × 3 scales × 2 measures). This improvement is evident at different scales and for videos from different genres with different durations. We now show the summaries for two videos of very different durations using our method and the baseline approach. The key frames in the summaries are arranged according to the CSF. Fig. 2 presents the video summaries produced by these two different approaches for the small duration video *Drift Ice as a Geologic Agent, Segment 03* at a scale of 10. The figure clearly shows the redundancy in the output of the baseline method (inclusion of both first and second frames and similarly both fourth and fifth frames). This redundancy is removed in the video summary obtained from our proposed method. Fig. 3 shows the video summaries produced by both the methods for a long duration video *Health Communication* at a scale of 10. Although the two summaries in Fig. 3 have several overlaps but better summary quality is achieved by our proposed framework which can be confirmed by a visual comparison of the two video summaries as well as from Fidelity and SRD values.

We conduct a user study with 10 volunteers for subjective evaluation. A speed-up version of the original video clips are shown to the users. Then, they are asked to assign a score between 1 and 5 (1 indicates the worst score and 5 indicates the best score) for Informativeness and Visual pleasantness of the summaries. Table III shows the mean results from these 10 volunteers. The results indicate that our summaries are significantly better than [6] for the long duration videos and marginally better for the short duration videos.

C. Comparative computational complexity experiments

In this section, we make a comparative analysis to validate the scalability in the analysis stage. All the experiments are performed on a desktop PC with Intel(R)core(TM) i5-2400 processor and 8 GB of DDR2-memory. Table IV shows the performance comparison with [6] on the total processing time required to produce the summaries that includes time required for both analysis and generation stages. From Table IV, it can be observed that the processing time required by [6] increases rapidly with an increase in the duration of the video. However, our proposed framework is highly efficient in producing video summaries irrespective of their duration. For example, a speed up factor (ratio of processing times) of 6 is achieved for the video stream *Space Work 5* at a scale of 5 and a speed up factor of 5 is achieved for the video stream *Tribute to World War II Nisei Veterans* at a scale of 10.This fast processing time is a result of the underlying scalable graph clustering adopted in the proposed framework. This scalability in the analysis stage makes our framework highly suitable for real-time social multimedia applications where huge amount of video data need to be processed.

V. CONCLUSION

In this paper, we proposed a novel framework for scalable video summarization using scalable graph clustering and a ranking procedure. Our framework provides scalability during both the analysis and the generation stages of video summarization, thereby making it highly amenable for real-time social multimedia applications. Experimental results clearly demonstrate the superiority of our method over a recently published work [6]. In future, we will focus on the integration of more extensive set of video features to further improve the summarization result. Another direction of future research will be to examine the scalability during the analysis stage of videos from different genres with same length.

REFERENCES

[1] http://www.youtube.com/t/press_statistics.

[2] A.G. Money, H.W. Agius, Video summarization: a conceptual framework and survey of the state of the art, Journal of Visual Communication and Image Representation, 19 (2), pp. 121143, 2008.

[3] B.T. Truong, S. Venkatesh, Video abstraction: a systematic review and classification, ACM Transactions on Multimedia Computing, Communications, and Applications, 3 (1) pp. 137, 2007.

[4] J. Almeida, N.J. Leite and R.S. Torres, VISON: Video Summarization for Online applications, Pattern Recognition Letters, vol. 33, pp. 397-409, 2012.

[5] S.E.F. Avila, A.P.B. Lopes, A. Jr. Luz and A.A. Araujo. VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method. Pattern Recognition Letters, 32 (1), pp. 5668, 2011.

[6] L. Herranz, J. M. Martinez, A Framework for Scalable Summarization of Video, IEEE Transaction on Circuit System for Video Technology, 20(9), pp. 1265-1270, 2010.

[7] A.S. Chowdhury, S. K. Kuanar, R. Panda and M.N. Das, Video Storyboard Design using Delaunay Graphs, In: Twenty First International Conference on Pattern Recognition, pp. 3108-3111, 2012.

[8] N. Ejaz, T. B. Tariq and S. W. Baik, Adaptive key frame extraction for video summarization using an aggregation mechanism, Journal of Visual Communication and Image Representation, vol. 23, pp. 1031-1040, 2012.

[9] L. Herranz and J. M. Martnez, An integrated approach to summarization and adaptation using H.264/MPEG-4 SVC, Signal Process.: Image Communication., vol. 24, no. 6, pp. 499509, 2009.

[10] N. Paragios, Y. Chen and O. Faugeras, Handbook of Mathematical Models in Computer Vision, Springer, 2006.

[11] S. K. Kuanar, R. Panda, A.S. Chowdhury, Video Key frame Extraction through Dynamic Delaunay Clustering with a Structural Constraint, Journal of Visual Communication and Image Representation, 24 (7), pp. 12121227, 2013.

[12] Lu Zheng, K. Grauman, Story-Driven Summarization for Egocentric Video, In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2714 - 2721, 2013.

TABLE III

PERFORMANCE COMPARISON USING SUBJECTIVE MEASURES.

Measures	Video ID	S1_ [6]	S1_Proposed	S2_ [6]	S2_Proposed	S3_ [6]	S3_Proposed
Informativeness	1	3.7	3.8	4.2	4.2	4.4	4.6
	2	2.6	2.8	2.8	3.9	4.0	4.3
	3	2.8	3.9	3.1	4.3	3.2	4.6
	4	2.1	3.2	2.8	3.8	3.1	4.5
	5	2.4	3.7	2.7	3.9	3.0	4.3
Visual Pleasantness	1	4.1	4.1	4.3	4.4	4.1	4.4
	2	3.9	4.0	3.8	4.5	4.2	4.5
	3	2.3	3.9	2.3	4.0	2.5	4.2
	4	3.1	4.3	3.5	4.3	3.5	4.5
	5	2.2	3.6	2.0	3.9	3.1	4.2

TABLE IV

PERFORMANCE COMPARISON USING PROCESSING TIME (IN SECONDS)

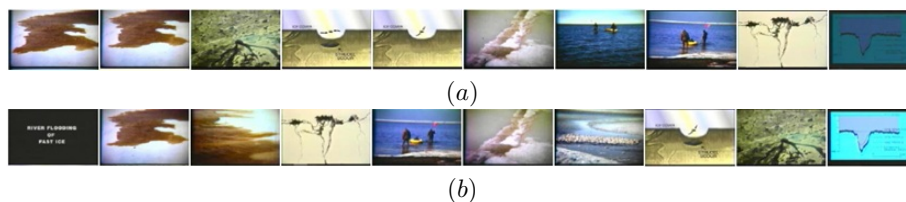| Video ID	S1_ [6]	S1_Proposed	S2_ [6]	S2_Proposed	S3_ [6]	S3_Proposed
1	2.74	2.61	2.80	2.77	2.91	2.92
2	3.21	2.04	3.37	2.34	3.56	2.61
3	600.5	100.3	695.6	123.3	787.9	148.6
4	802.3	164.1	867.4	187.7	989.0	196.9
5	1172.6	245.7	1197.3	261.8	1253.5	282.4

(a)

(b)

Fig. 2. Summary produced by both Baseline method [6] and our proposed framework for the video Drift Ice as a Geologic Agent, segment 02.
(a) Baseline [6]: Fidelity = 0.432, Shot Reconstruction Degree = 3.012, Informativeness = 2.8, Visual pleasantness = 3.8
(b) Proposed Method: Fidelity = 0.587, Shot Reconstruction Degree = 3.864, Informativeness = 3.9, Visual pleasantness = 4.5

(a)

(b)

Fig. 3. Summary produced by both Baseline method [6] and our proposed framework for the video Health Communication.
(a) Baseline [6]: Fidelity = 0.527, Shot Reconstruction Degree = 3.967, Informativeness = 2.8, Visual pleasantness = 3.5
(b) Proposed Method: Fidelity = 0.721, Shot Reconstruction Degree = 6.482, Informativeness = 3.8, Visual pleasantness = 4.3

[13] Y. Cong, J.Yuan and J. Luo, Towards Scalable Summarization of Consumer Videos Via Sparse Dictionary Selection, IEEE Transactions on Multimedia, vol.14, pp. 66-75, 2012.
[14] J. Shi and J. Malik, Normalized cuts and image segmentation, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 22, pp. 888905, 2000.
[15] G.Ciocca and R.Schettini, A innovative algorithm for key frame extraction in video summarization, Journal of Real-time Image Processing, vol. 1, pp. 69-88, 2006.
[16] J.J. Whang, Su Xin, I.S Dhillon,. Scalable and Memory-Efficient Clustering of Large-Scale Social Networks, In: IEEE International Conference on Data Mining, pp. 705-714, 2012.
[17] C. Zahn. Graph-theoretical methods for detecting and describing gestalt clusters. IEEE Transactions on Computers, C-20:6886, 1971.
[18] L. Grady and E. Sschwartz. Anisotropic Interpolation on Graphs: The Combinatorial Dirichlet Problem, Technical Report CAS/CNS-TR-03-014, Department of Cognitive and Neural Systems, Boston University, Boston, MA, July 2003.
[19] R. Barrett et. al. Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, Number 43 in Miscellaneous Titles in Applied Mathematics Series. SIAM, November 1993.
[20] Hanghang Tong, Christos Faloutsos, Jia-Yu Pan, Fast Random Walk with Restart and Its Applications:, In: IEEE International Conference on Data Mining, pp. 613-622, 2006.
[21] The Open Video Project: http://www.open-video.org.