UDC 004.275, DOI: 10.2298/CSIS1001085S

Multi-Video Summarization Using Complex Graph Clustering and Mining

Jian Shao[1], Dongming Jiang[1], Mengru Wang[2], Hong Chen[2], and Lu Yao[1]

[1] College of Computer Science, Zhejiang University, China
[2] Zhejiang Radio & TV Group, China
jshao@cs.zju.edu.cn, dmjiang1985@163.com, cs08yl@hotmail.com
{wmr628, ch213}@mail.zrtg.com

Abstract. Multi-video summarization is a great theoretical and technical challenge due to the wider diversity of topics in multi-video than single-video as well as the multi-modality nature of multi-video over multi-document. In this paper, we propose an approach to analyze both visual and textual features across a set of videos and to create a so-called circular storyboard composed of topic-representative keyframes and keywords. We formulate the generation of circular storyboard as a problem of complex graph clustering and mining, in which each separated shot from visual data and each extracted keyword from speech transcripts are first structured into a complex graph and grouped into clusters; hidden topics in the representative keyframes and keywords are then mined from clustered complex graph while at the same time maximizing the coverage of the summary over the original video set. We also design experiments to evaluate the effectiveness of our approach and the proposed approach shows a better performance than two other storyboard baselines.

Keywords: multi-video summarization, complex graph clustering and mining, circular storyboard.

1. Introduction

Multi-video summarization aims to create a summary for a set of topic-related videos having several sub-themes or sub-events around a main topic. Imagine a scenario that a news aggregator gathers various sets of topic-related news videos and provides summaries to represent each video set, such that the users can easily understand the topics after taking a look at the summaries and decide whether to go through watching their interested video set.

Despite that a lot of approaches of automatic video summarization [1-5] and some of multi-document summarization [6] have been proposed in literature, little work has focused on multi-video [7]. In general, multi-video summarization faces two difficult challenges. One is due to the inevitable thematic diversity and overlaps within multi-video than single-video, and hence we need study effective summarization methods to extract the globally main topic information while removing redundancy among different videos as more as possible. The other is due to the multi-modality nature of multi-video over multi-document. Using both text and imagery during summarization is more effective than either modality alone [5, 8], and hence we need effective summarization methods to integrate both textual and visual content in summary creation and visualization.

An intuitive approach for summarization of multi-video with text transcripts is to summarize visual modality and textual modality separately and to visualize as an image-plus-text list storyboard together — employ a shot clustering and mining algorithm to select a list of most representative keyframes from shot clusters to reflect the visual topics, employ multi-document summarization approaches to select a list of most representative keywords to reflect the textual topics, generate a list storyboard by placing a list of keyframes in upper half and a list of keywords in lower half. However, despite that this list storyboard gives brief summaries of both visual content and textual content, it hardly makes use of the relations of visual content and textual content. Actually, given a video set, there usually exist a number of sub-themes, and each sub-theme can be represented as a collection of shots and keywords which have thematic relations with each other. Discovering the relations between shots and keywords will not only offer benefits to find more reasonable topic structure, but also make it possible to further remove redundancy in theme level.

To take advantage of both visual and textural information during summary generation, we propose a novel approach to summarize visual modality and textual modality simultaneously and to visualize as an image-plus-text circular storyboard in this paper. We first build a complex graph consisting of shot nodes and keyword nodes. Shots are linked to each other by visual similarity, while shots and keywords are linked to each other by co-occurrences. In such a way, rather than one-way clustering, i.e., either shot clustering or keyword clustering, we can perform co-clustering of shots and keywords on complex graph to make use of the benefits of shot-keyword relations [9]. Next, we mine representative keyframes and keywords for hidden topics from clustered complex graph with highest importance score while at the same time maximizing the coverage of the summary over the original video set by deeply exploiting the node-level relations and the cluster-level relations. And finally, we design a circular storyboard to present the visual and textual topics and their relations — resize the keyframes and keywords according to their importance and arrange resized keyframes and keywords around or inside a circle according to their relation strength. Experiments carried out on video news demonstrate the effectiveness of our proposed approach.

We summarize our main contributions as follows: 1) The novel utilization of complex graph clustering for multi-video summarization; 2) The scheme of mining representative keyframes and keywords for hidden topics by integrating both node-level and cluster-level information in clustered complex graph; 3) the visual presentation of a summary as an image-plus-text circular storyboard.

The remainder of this paper is organized as follows. Section 2 describes our complex graph clustering and mining based multi-video summarization approach. Section 3 describes the data set and experimental results. Finally, we conclude this paper in Section 4.

2. The Multi-Video Summarization Approach by Complex Graph Clustering and Mining

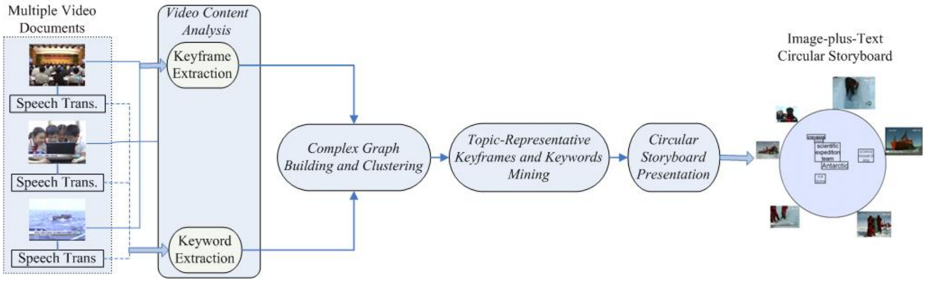
Multiple Video Documents
Speech Trans.
Speech Trans.
Speech Trans
Video Content Analysis
Keyframe Extraction
Keyword Extraction
Complex Graph Building and Clustering
Topic-Representative Keyframes and Keywords Mining
Circular Storyboard Presentation
Image-plus-Text Circular Storyboard

Fig. 1. Proposed multi-video summarization approach

Our proposed multi-video summarization approach is shown in Fig. 1, where the input is multiple video documents with their corresponding speech transcripts, and the output is so-called circular storyboard, an image-plus-text static summary composed of representative keyframes and keywords of hidden topics. There are four stages to generate the summary, of which the first one is to analyze visual content to extract a pool of shots and to analyze textual content of speech transcripts to extract a list of keywords across multiple video documents; the second is to perform complex graph building and clustering on extracted shots and keywords; the third is to mine the representative keyframes and keywords of hidden topics from clustered complex graph; and the last is to use a circle space to present the topic-representative keyframes and keywords as circular storyboard.

2.1. Video Content Analysis

Assume that we have a set of video documents with their speech transcripts, $D = \{d_1, ..., d_m, ..., d_M\}$. To analyze the visual content of given videos, we employ a robust shot boundary detection algorithm in [10] to divide video sequences into a pool of shots denoted as $U = \{u_1, ..., u_i, ..., u_I\}$, where u_i is a shot and I denotes the number of shots in video set. For further visual content processing, we select the middle frame of a shot as the keyframe and represent it as a 423-dimentional vector of 6 features: 256-dim color histogram, 6-dim color moments, 128-dim color coherence, 15-dim texture MSRSAR, 10-dim texture Tamura coarseness, and 8-dim texture Tamura directionality.

In addition, in order to analyze the textual content of given videos, a three-step process is conducted to extract meaningful keywords from speech transcripts. First, a part of speech tagger is used to label out nouns in speech transcript. Next, stop-words are used to further filter unnecessary nouns. And finally, we propose a modified tf-idf formula to score the remaining keywords and select the keywords with highest importance score. Suppose that the selected keywords are denoted as $W = \{w_1, ..., w_j, ..., w_J\}$, the importance score $I(w_j)$ of keyword w_j based on the modified tf-idf formula is given by

$$I(w_j) = \mathrm{mod}\_tf(w_j, D) \times idf(w_j) \quad (1)$$

where $idf(w_j)$ represents the inverse document frequency and can be defined as $idf(w_j) = \log(N / df(w_j))$, where N is the total number of videos in training corpus, and $df(w_j)$ is the number of videos involving the keyword of w_j $\mathrm{mod}\_tf(w_j, D)$ represents the modified term frequency of word w_j in video set D , it is defined by

$$\mathrm{mod}\_tf(w_j, D) = \sum_{m=1}^{M} tf(w_j, d_m) \times \left(\frac{tf(w_j, d_m)}{tf(w_j, D)} + \alpha \times \left(1 - \frac{tf(w_j, d_m)}{tf(w_j, D)}\right)\right) \quad (2)$$

$$tf(w_j, D) = \sum_{m=1}^{M} tf(w_j, d_m)$$

where $tf(w_j, d_m)$, $tf(w_j, D)$ represents the frequencies of word w_j in video d_m, in video set D respectively, α is weighting factor. Note that in contrast with $tf(w_j, D)$, $\mathrm{mod}\_tf(w_j, D)$ takes the keyword distribution of video set into account. By setting $\alpha > 1$, the more disperse the keyword distribution of video set is , the larger the $\mathrm{mod}\_tf(w_j, D)$ is, and correspondingly the more important the keyword is.

2.2. Complex Graph Building and Clustering

Considering the task of learning cluster structure from a pool of shots U and a list of keywords W that are extracted from video set D , we can use most one-way existing clustering algorithm to cluster shots and keywords separately [11], or employ co-clustering algorithm [9, 12] to cluster shots and keywords simultaneously. In this paper, we adopt the complex graph clustering algorithm in [9] to simultaneously derive the shot clusters and keyword clusters as well as their relations.

We first organize extracted shots U and keywords W to build a complex graph of shot nodes and keyword nodes. Suppose the instantiated complex graph is denoted as $G=(V^{(1)},V^{(2)},E^{(1,1)},E^{(1,2)})$, where $V^{(1)}=\{v_1^{(1)},...,v_i^{(1)},...,v_I^{(1)}\}$ represents the shot node set, $v_i^{(1)}$ is the ith node in $V^{(1)}$ corresponding to the ith shot in U; $V^{(2)}=\{v_1^{(2)},...,v_j^{(2)},...,v_J^{(2)}\}$ represents the keyword node set, $v_j^{(2)}$ is the jth node in $V^{(2)}$ corresponding to the jth keyword in W; $E^{(1,1)}$ represents the homogeneous edges within shot nodes; $E^{(1,2)}$ represents the heterogeneous edges between shot nodes and keyword nodes. We use affinity matrix $S\in R_+^{I\times I}$ to represent the weights of $E^{(1,1)}$ and use $A\in R_+^{I\times J}$ to represent the weights of $E^{(1,2)}$. Cosine similarity metric is employed to realize matrix S. That is, the edge weight $S_{i,q}$ between node $v_i^{(1)}$ and node $v_q^{(1)}$ is given by

$$S_{i,q}=\frac{fe(u_i).fe(u_q)}{|fe(u_i)|\times|fe(u_q)|} \quad (3)$$

where $fe(u_i)$ denotes the keyframe feature vector of shot u_i. In addition, The co-occurrence between shots and keywords is employed to realize matrix A. That is, the edge weight $A_{i,j}$ between shot node $v_i^{(1)}$ and keyword node $v_j^{(2)}$ is given by

$$A_{i,j}=\sum_{k=1}^{tf(w_j)} sim_t(t(u_i),t(w_j^k)) \quad (4)$$

Note that word w_j may occur several times $tf(w_j)$ in a video which contains shot u_i, so every occurrence is indexed by k. Timing similarity $sim_t(t(u_i),t(w_j^k))$, between $t(u_i)$, the mid-point timing of shot u_i, and $t(w_j^k)$, the mid-point timing of the kth occurrence of word w_j, is defined as follows:

$$sim_t(u_i,w_j^k)=\begin{cases}\exp\{-\dfrac{|t(u_i)^{start}-t(w_j^k)|^2}{2{\sigma_t}^2}\} & (\ t(w_j^k)<t(u_i)^{start}\ )\\ 1 & (\ t(u_i)^{start}\le t(w_j^k)\le t(u_i)^{end}\\ \exp\{-\dfrac{|t(w_j^k)-t(u_i)^{end}|^2}{2{\sigma_t}^2}\} & (\ t(w_j^k)>t(u_i)^{end}\ )\end{cases} \quad (5)$$

This $sim_t(.,.)$ function is a step function having 1 when $t(w_j^k)$ falls in the range between $t(u_i)^{start}$, the start-point timing of shot u_i, and $t(u_i)^{end}$, the end-point timing of shot u_i, but its edges are dispersed using a Gaussian filter with standard deviation σ_t to compensate for the time delay between the shot and keyword occurrence.

We then perform a hard clustering algorithm described in [9] to derive the shot clusters and keyword clusters as well as their relations between clusters. Suppose that we have grouped the shot nodes $V^{(1)}$ into K clusters $CU = \{cu^{(1)},...,cu^{(k)},...,cu^{(K)}\}$ and grouped the keyword nodes $V^{(2)}$ into L clusters $CW = \{cw_1,...,cw_l,...,cw_L\}$. Let $C^{(1)} \in (0,1)^{I\times K}$ denote the cluster membership matrix for shot nodes $V^{(1)}$ such that $C^{(1)}_{i,k}$ denote the weight that the ith shot node in $V^{(1)}$ is associated with the kth cluster, and $C^{(2)} \in (0,1)^{J\times L}$ denotes the cluster membership matrix for keyword nodes $V^{(2)}$ such that $C^{(2)}_{j,l}$ denote the weight that the jth keyword node in $V^{(2)}$ is associated with the lth cluster. The intra-type cluster relation matrix $D \in R^{K\times K}$ denotes cluster relations within the same type of nodes such that $D_{k,r}$ denotes the link strength between the kth shot cluster cu_k and the rth shot cluster cu_r . The inter-type relation matrix $B \in R^{K\times L}$ denotes the cluster relations between the different types of nodes such that $B_{k,l}$ denotes the link strength between the kth shot cluster cu_k and the lth keyword cluster cw_l .

2.3. Topic-Representative Keyframes and Keywords Mining

This step aims to mine the most representative keyframes and keywords of hidden topics from clustered complex graph. We propose a three-stage mining scheme.

First of all, we define measures to compute the importance of shot nodes and keyword nodes. We employ the modified tf-idf formula described in Sub-Section 2.1 to calculate the importance of keyword nodes. While for computing the importance of shot nodes, we take both the visual and related textual features into consideration. Given a shot u_i in node $v^{(1)}_i$, its importance score $I(u_i)$ is modeled as a combination of its visual informativeness $\inf_{visual}(u_i)$ and related textual informativeness $\inf_{text}(u_i)$ with a weight parameter β specified by users:

$$I(u_i) = \beta \times \inf_{visual}(u_i) + (1-\beta) \times \inf_{text}(u_i) \qquad \textbf{(6)}$$

The computation of textual informativeness of a shot is based on the following intuitions:

– the more important a shot's related keywords are, the more informative it is;
– the more heavily a shot is linked with related words, the more informative it is.

Based on above heuristic intuitions, $\inf_{text}(u_i)$ can be expressed by the following formula:

$$\inf_{text}(u_i) = \sum_{j=1}^{J} I(w_j) \times A_{i,j} \tag{7}$$

On the other hand, the computation of the visual informativeness of a shot is based on the following intuitions:

- the more a shot's similar shots are, the more informative it is;
- the more a shot's similar shots in other videos are, the more informative it is.

Given above heuristic intuitions, $\inf_{visual}(u_i)$ can be expressed by the following formula:

$$\inf_{\text{visual}}(u_i) = \sum_{i \neq p} S_{i,p} \times \delta(u_i, u_p, D)$$
$$\delta(u_i, u_j, D) = \begin{cases} 1 & \text{if } D(u_i) = D(u_j) \\ \gamma & \text{otherwise} \end{cases} \tag{8}$$

where $\delta(u_i, u_p, D)$ is a weight function, and we set $\gamma > 1$ to differentiate the shot edges within one video or across two videos.

Next, we define measures to compute the informativeness of shot clusters and keyword clusters. Different from single-video, multi-video usually contains several sub-themes or sub-events and each sub-theme of sub-event can be viewed by a cluster of theme-related keywords or a cluster of event-related shots. Thus, the computation of the importance of a keyword cluster is based on the following intuitions:

- The more complex a cluster is, the more important it is
- The more important a cluster's contained keywords are, the more important it is.

Based on above intuitions, $I(cw_l)$, the importance of l_{th} keyword cluster cw_l is given by

$$I(cw_l) = \sum_{j=1}^{J} C_{j,l}^{(2)} \times I(w_j) \tag{9}$$

While for computing the importance score of a shot cluster, we take both the shot nodes inside cluster and the relations with other clusters into consideration. Given a shot cluster cu_k, its importance score $I(cu_k)$ is modeled as a combination of its node-level informativeness $\inf_{node}(cu_k)$ and cluster-level informativeness $\inf_{clust}(cu_k)$ with a weighting factor ξ:

$$I(cu_k) = \xi . \inf_{node}(cu_k) + (1 - \xi) . \inf_{clust}(cu_k)$$
$$\inf_{node}(cu_k) = \sum_{i=1}^{I} C_{i,k}^{(1)} \times I(u_i) \tag{10}$$

$$\inf_{clust}(cu_k) = \sum_{q \neq k} D_{k,q} + \psi . \sum_{l=1}^{L} B_{k,l}$$

where the node-level informativeness $\inf_{node}(cu_k)$ is directly calculated by summing up the importance of the shots that belong to the cluster; the cluster-level informativeness $\inf_{clust}(cu_k)$ is defined as a combination of its intra-type relations with other shot clusters and its inter-type relations with keyword clusters with a weighting factor ψ .

Finally, we select representative keyframes and keywords of hidden topics. Each cluster indicates a hidden sub-theme or sub-event. Therefore, we propose a two-step selection procedure to maximize the coverage of the topics while at the same time removing redundancy as more as possible.

- Select a number of shot clusters with highest importance score, and only choose the keyframe of the most important shot in each cluster as its representation.
- Select a number of keyword clusters with highest importance score, and only choose the most important keyword in each cluster as its representation.

2.4. Circular Storyboard Presentation

After mining topic-representative keyframes and keywords, we use a circle space to present the topic-representative keyframes and keywords as circular storyboard. The representative keyframes of visual topics are averagely placed at regular intervals on the boundary of the circle. The representative keywords of textual topics are displayed inside the circle at a position which is determined to their relevance with the shot cluster.

$$po\,\mathrm{int}(cw_l) = \sum_{k=1}^{K} po\,\mathrm{int}(cu_k) \times B_{k,l} \tag{11}$$

In order to guild user's attention to important topics, keyframes and keywords are resized according to their importance score, so that higher importance shot clusters are represented with larger keyframes, higher importance keyword cluster are represented with larger font size.

$$\begin{gathered} size(cw_l) = I(cw_l)/I(cw_l)_{\max} \\ size(cu_k) = 0.5 + 0.5 \times I(cu_k)/I(cu_k)_{\max} \end{gathered} \tag{12}$$

Table 1. Test video set

No.	Category	Video num.	Shot num.	Total length
I-1	Sci-Tech	3	27	2:33
I-2	Sci-Tech	9	105	7:05
II-1	Business	10	117	9:02
II-2	Business	15	226	15:22

3. Experiments

3.1. Data set

We have collected a corpus of CCTV Broadcast News during January 1, 2008 and May 31, 2008 from CCTV News Broadcasting Program website (http://news.cctv.com/program/xwlb). Speech transcript for each video was also collected using a large vocabulary continuous speech recognition engine [13]. The corpus contains a total of 2080 video clips with wide diversity in topics. Following the news classification scheme of Baidu News site（http://news.baidu.com/）, we use transcript text to classify videos into 13 categories: world, China, sports, entertainment, society, business, internet, science & technology (Sci-Tech), house, auto, culture, education, health and game. And then we employ an affinity propagation clustering algorithm [14] to group videos in each category into clusters. Each cluster contains a set of videos having a main topic. Four clusters from Sci-Tech and Business categories are chosen as our test set. Table 2 shows the cluster information including: the category, the number of videos, the total number of shots and the total length of each cluster.

3.2. Results and Evaluations

In our proposed approach, we formulated multi-video summarization as a complex graph clustering and mining problem and represented the selected representative keyframes and keywords as an image-plus-text circular storyboard. To compare the performance with our proposed method, we implement two storyboard approaches as the baselines. As for one of them, baseline-I, we simplify the complex graph clustering and mining algorithm into an extremely special case by setting all elements of affinity matrix A described in Sub-Section 3.2 as zero to ignore the relations of shots and keywords. The representative keywords with highest importance score are directly selected; the representative keyframes with most visual informativeness are selected from clusters grouped from a simplified homogeneous graph clustering. And the finally selected keyframes and keywords are visualized as a list storyboard by placing keyframes in upper half and keywords in lower half. The goal of baseline-I is to evaluate the effectiveness of our proposed complex graph clustering and mining algorithm. As for another storyboard system, baseline-II, we just replace circular storyboard presentation in our proposed approach as a list storyboard presentation while keeping the others not changed. The goal of baseline-II is to evaluate the presentation quality of circular storyboard.

In order to quantitatively evaluate the effectiveness of our proposed approach, we evaluate informativeness between the two baselines and our proposed one. The criterion informativeness measures whether the summary can bring users as much as the original video set do or not. We invited 12 graduate students, including 8 males and 4 females, to give subjective scores ranging from 1 (worst) to 100 (best) to the criterion. After browsing the three types of video summaries, they can watch the original video and write down their decisions.

Table 2 lists the results. Compared to baseline-I with an average score of 69.5, baseline-II achieves an average score of 74.1 with a relative improvement of 6.6%. These results demonstrate the effectiveness of proposed complex-graph clustering and mining algorithm. In addition, compared to baseline-II with an average score of 74.1, our proposed approach obtain an average score of 77.6 with a relative improvement of 4.7% which indicates that the presentation of summary as circular storyboard can offer more information to users.

Table 2. Evaluation results

Video set	Baseline-I	Baseline- II	Proposed
I-1	82.3	84.8	86.8
I-2	66.0	73.1	77.4
II-1	71.6	78.7	83.3
II-2	58.1	59.8	62.7
Avg.	69.5	74.1	77.6

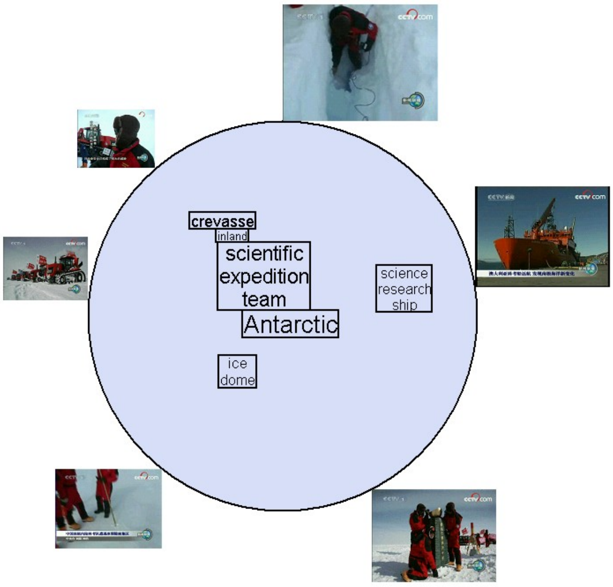

Fig. 2. Circular storyboard based on complex graph clustering and mining

Figure 2 illustrate the example of the summarized results using our approach, in video set 1-1 which has a main topic “science expedition in

Antarctic" (Note that the original Chinese keywords have been mapped into English keywords for illustration and explanation).

4. Conclusions

Multi-video summarization faces technical challenges due to the wider diversity of topics in multi-video than single-video as well as the multi-modality nature of multi-video over multi-document. We propose techniques to analyze both visual and textual content across a set of videos and to create a circular storyboard composed of topic-representative keyframes and keywords. We formulate the generation of circular storyboard as a complex graph clustering and mining problem, in which the divided shots from visual data and the extracted keywords from speech transcripts are first structured into a complex graph and grouped into clusters; the representative keyframes and keywords of hidden topics are then mined from clustered complex graph to maximize the coverage in topics while at the same time removing redundancy as more as possible. Experiments carried out on four video clusters of CCTV Broadcast News videos show the effectiveness of our proposed approach.

5. Acknowledgement

This work is supported by National Natural Science Foundation of China (No. 60833006), National Key Technology R&D Program (2007BAH11B01), Science and Technology Project of Zhejiang Province (2008C13G1410001，2008C13075，2008C13081，2008C13G2010039)

6. References

1. Christel M. G. and Hauptmann A. G. and Lin W. H. and Chen M. Y. and Yang J. and Maher B. and Baron R. V.: Exploring the utility of fast-forward surrogates for BBC rushes. Proc. TRECVID Summarization Workshop, Vancouver, British Columbia, Canada. (2008)
2. Christel M. G.: Evaluation and user studies with respect to video summarization and browsing, IS&T/SPIE Symposium on Electronics Imaging, San Jose, CA. (2006).
3. Yeung, M.M. and Yeo, B.L.: Video visualization for compact presentation and fast browsing of pictorial content. IEEE Trans. Circuits and System for Video Technology, vol. 7, no. 5, 771-785. (1997)9.
4. Chen, B. W. and Wang, J. C. and Wang, J. F.: A novel video summarization based on mining the story-structure and semantic relations among concept entities. IEEE Tran. Multimedia, vol. 11, no. 2. (2009)
5. Ding, W.: Multimodal surrogates for video browsing. Proc. ACM Digital Lib, Berkeley, CA, 85-93. (1999)

6. Wan, X.J. and Yang, J.W.: Improved affinity graph based multi-document summarization. Proc. Human Language Technology Conf. of the NAACL, 181-184. (2006)
7. Wang F. and Merialdo B.: Multi-document video summarization, Proc. ICME'09. (2009)
8. Liu Y. N., Wu F., Zhuang Y. T., Xiao J.: Active post-refined multi-modality video semantic concept detection with tensor representation, ACM Multimedia 2008，91-100. (2008)
9. Long, B. and Zhang, Z. F. and Yu P. S.: Clustering on Complex Graph, Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence. (2008)
10. Ye, Z.Y. and Wu, F., A robust fusion algorithm for shot boundary detection. Journal of Computer Aided Design and Computer Graphics (In Chinese with English Abstract), vol.15, No.11, 950-955. (2003)23.
11. Odobez, J.M. and Gatica-Perez, D.: Video shot clustering using spectral methods. Proc. 3rd Int. Workshop on Content-Based Multimedia Indexing, Rennes, France, 94-102. (2003)
12. Deodhar M. and Ghosh J. and Gupta G. and Cho H. and Dhillon I.: A scalable framework for discovering coherent co-clusters in noisy data, ICML’09. (2009)17.
13. Young S. and Everman G and et al.: The HTK book (for HTK version 3.4), http://htk.eng.cam.ac.uk. (2007)
14. Xia D.Y. and Wu F. and Zhang X.Q and Zhuang Y.T.: Local and global approaches of affinity propagation clustering for large scale data. Journal of Zhejiang University SCIENCE A, vol. 9 no. 10 pp. 1373~1381. (2008)

Jian Shao has received his B.S. degree in electrical science and engineering from Nanjing University in 2003, Ph.D. degree from Institute of Acoustics, Chinese Academy of Sciences in 2008. He is currently a post-doctor of College of Computer Science, Zhejiang University. His research is focused on multimedia analysis and retrieval.

Dongming Jiang has received his B.S. degree in computer science and technology from Zhejiang University (ZJU), China, in 2007. He is currently pursuing the M.S. degree in the computer application technology at ZJU. His research interests include video analysis and information retrieval.

Mengru Wang has received his B.S. degree in electrical science and engineering from Zhejiang University in 1994. Now he is a senior engineer and team leader of Zhejiang radio & TV Group. His research is focused on video processing and transmission

Hong Chen has received his M.S. degree in electrical science and engineering from Zhejiang University in 2005. Now he is an engineer of Zhejiang radio & TV Group. His research is focused on video processing and transmission.

Yao Lu has received his B.S. degree in software engineering from Zhejiang University (ZJU), China, in 2008. He is currently pursuing the M.S. degree in computer science at ZJU. His research interests include video processing and personalized recommendation.

Received: May 10, 2009; Accepted: September 21, 2009.