Scene-Based Movie Summarization Via Role-Community Networks

Article *in* IEEE Transactions on Circuits and Systems for Video Technology · November 2013

DOI: 10.1109/TCSVT.2013.2269186

4 authors:

Chia-Ming Tsai
National Chung Cheng University
14 PUBLICATIONS **219** CITATIONS
SEE PROFILE

Li-Wei Kang
National Yunlin University of Science and Technology
94 PUBLICATIONS **1,281** CITATIONS
SEE PROFILE

Chia-Wen Lin
National Tsing Hua University
163 PUBLICATIONS **2,828** CITATIONS
SEE PROFILE

Weisi Lin
Nanyang Technological University
411 PUBLICATIONS **8,247** CITATIONS
SEE PROFILE

Scene-Based Movie Summarization Via Role-Community Networks

Chia-Ming Tsai, Li-Wei Kang, *Member, IEEE,* Chia-Wen Lin, *Senior Member, IEEE,* and Weisi Lin, *Senior Member, IEEE*

***Abstract*—Video summarization techniques aim at condensing a full-length video to a significantly shortened version that still preserves the major semantic content of the original video. Movie summarization, being a special class of video summarization, is particularly challenging since a large variety of movie scenarios and film styles complicate the problem. In this paper, we propose a two-stage scene-based movie summarization method based on mining the relationship between role-communities since the role-communities in earlier scenes are usually used to develop the role relationship in later scenes. In the analysis stage, we construct a social network to characterize the interactions between role-communities. As a result, the social power of each role-community is evaluated by the community's centrality value and the role communities are clustered into relevant groups based on the centrality values. In the summarization stage, a set of feasible summary combinations of scenes is identified and an information-rich summary is selected from these candidates based on social power preservation. Our evaluation results show that in at most test cases the proposed method achieves better subjective performance than attention-based and role-based summarization methods in terms of semantic content preservation for a movie summary.**

***Index Terms*—Movie analysis, movie summarization, social network analysis, video adaptation, video summarization.**

I. Introduction

VIDEO summarization aims at condensing a full-length video to a significantly shortened version that still preserves the major content of the original video [1], [2]. Movie summarization is a special class of video summarization. It can be applied for browsing through a movie over a hand-held device or a personal multimedia system in a short period when a viewer does not have enough time to completely watch a movie. Under such a scenario, movie summarization techniques can help a user to trim unimportant or redundant movie content. Generally speaking, the summarization process condensates video content according to motion activities, visual attention, and/or specific event-based criteria. Although manual video summarization by a professional person usually achieves a better viewing experience, different viewing time constraints and user preferences would consume huge amount of manpower for editing video summaries.

The simplest summarization approach is to uniformly downsample video frames with a fixed time interval. However, the major problem of the uniform downsampling approach is to introduce nonuniform information loss. To generate good video summaries, a video summarization method should be able to automatically identify important content in a video. According to the types of content used for video analysis, existing video summarization methods can be classified into cognitive-level approaches [2]–[6] and affective-level approaches [7]–[11].

The cognitive-level approaches extract audio-visual features from a video to identify a set of important key frames/shots to represent the whole video. Several existing cognitive-level methods usually use various low-level features, such as color, texture, audio-visual tempo, and motion features, to identify video highlights [2], [3]. Some cognitive-level methods extract significant video events to represent the original video, which is particularly helpful for domain-specific summarization applications. For example, the method proposed in [4] detects soccer game events, such as goal, referee, and penalty box events, through low-level features, and generates a summary to include these detected events. The method proposed in [5] utilizes concept-expansion trees to construct a relational graph for characterizing the semantic concepts of documentary videos. A graph modeling-based method is proposed in [6] to detect scene changes from the interconnectivity among clusters, followed by a summarization process that chooses significant motion attention content from the scene level to the subshot level.

The affective-level approaches edit a summary by modeling the affective content via user feedbacks. Most existing affective-level methods adopt various kinds of approaches to collect emotion information and map low-level video features into the emotion space. The method proposed in [7] represents and models the affective content of a video as the points of emotion space of the video based on the "dimensional

Manuscript received August 23, 2012; revised January 23, 2013 and March 31, 2013; accepted April 5, 2013. Date of publication June 18, 2013; date of current version November 1, 2013. This work was supported by the National Science Council, Taiwan, under Grant NSC101-2221-E-007-121-MY3 and Grant NSC 100-2218-E-224-017-MY3. This paper was recommended by Associate Editor T. Zhang.

C.-M. Tsai is with the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi 62102, Taiwan.

L.-W. Kang is with the Graduate School of Engineering Science and Technology-Doctoral Program and the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan.

C.-W. Lin is with the Department of Electrical Engineering and the Institute of Communications Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: cwlin@ee.nthu.edu.tw).

W. Lin is with the Division of Computer Communications, School of Computer Engineering, Nanyang Technological University, Singapore 639798.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2013.2269186

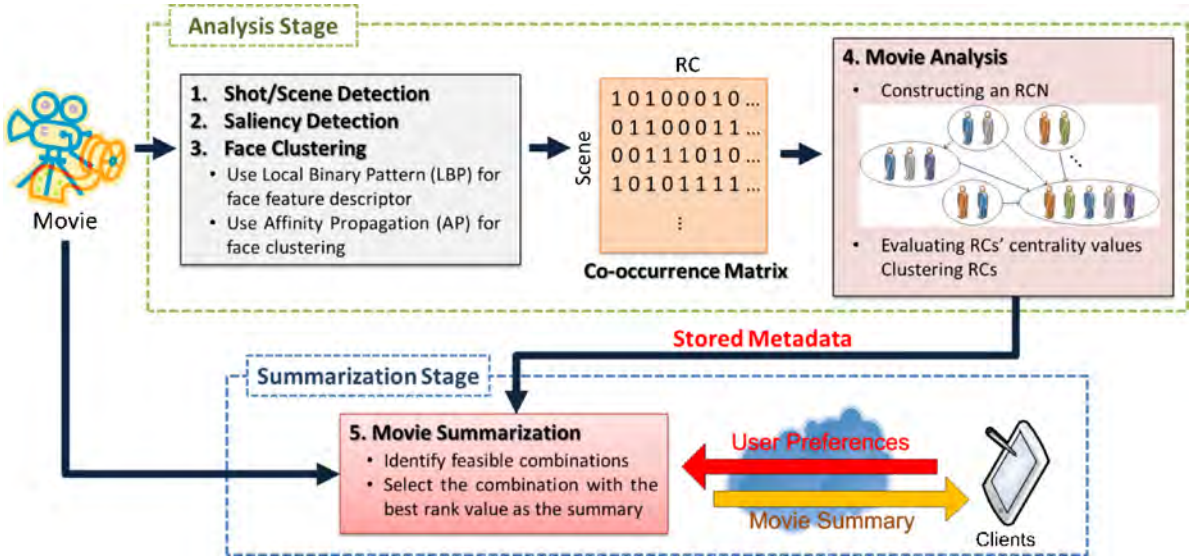


Fig. 1. Overview of the proposed movie summarization framework.

approach to affect" that is known from the field of psychophysiology. In [8], an affective scene classification method was proposed to classify affective audio-visual words based on latent topic driving models. The physiological responses of humans are exploited in [9] to identify the most entertainment segments of a video to produce a summary. The video player system proposed in [10] adaptively controls the playback speed according to the user behavior. Each user is first asked to fast-forward a video and then the player learns the user behavior for adapting playback speed. The method proposed in [11] uses a facial expression recognition method to identify the emotion of a viewer so as to find the highlights of a video.

As a special class of video summarization, movie summarization is particularly challenging as the large variety of movie scenarios and film styles complicate the problem considerably. Many existing movie summarization methods detect movie highlights by bridging the gap between the low-level audio-visual features and the affective content. However, the data used for modeling the affective content or learning the affective model is usually hard to cover all possible movie scenarios and film styles. Moreover, the causal relationships inside a movie are also hard to be directly modeled by using low-level audio-visual features, thus making the condensed version not comprehensive enough to viewers. Recently, a few movie analysis works adopt both the concepts of social network analysis and role recognition to identify roles in a movie [12]–[15], where the role interactions in the movie are treated as a social behavior, and are then modeled via a network structure. The method proposed in [12] identifies leading roles and role communities by constructing a role network. The constructed role network is applied to divide the movie into many independent story segments. The method proposed in [13] further extends the concept of role networks [12] by additionally considering the role appearances between adjacent shots. The method proposed in [14] uses an audio segmentation method and the maximum *a posteriori* probability (MAP) approach to identify the major actors of a movie. The method proposed in [15] uses face clustering and role appearance probability to extract movie highlights. Although the methods in [12]–[15] can be used to characterize the interactions of roles, they still do not take into account the causal relationships between role communities. Besides, the quality of summary is subjective to viewers, and, hence, it is hard to produce a satisfactory summary without considering user preferences.

In this paper, we propose a novel scene-based movie summarization framework by exploring the causal relationships of role communities and considering user preferences when generating the movie summary. Different from existing role-based methods [12], [13], [15], the proposed method characterizes the role community interactions via role community networks so as to correctly identify the highlight scenes of a movie. As shown in Fig. 1, the proposed method is divided into an analysis stage and a summarization stage. In the analysis stage, we first perform scene/shot boundary detection and use human face clustering to detect and cluster the faces of roles. After that, we construct a role-community network to characterize the relationships between role communities, where a role community contains those roles who appear in a scene. We believe that for movie analysis, using the relationships between role communities in movie scenes is better than using the relationships between individual roles, as described in the book of *The Power of Film*, "Drama is a social art form in which a community speaks to itself about itself" [16]. The concept of role-community networks for movie analysis is important since, following a carefully designed scenario, roles in a movie usually form groups/communities to conduct meaningful social behaviors and interactions to develop the main plots of the movie. Therefore, the proposed method identifies the highlight scenes of a movie based on role-community interactions. In the summarization stage, we formulate the movie summarization as a social network pruning problem subject to a user-specified length constraint so that the summary composed of the extracted highlights

TABLE I
NOTATIONS

Symbols	Meanings
m_k	The k-th scene in a movie, $k = 1, 2, \ldots, N_s$, where N_s is the number of scenes
$\boldsymbol{RC}$	Set of role communities in a movie, where $\boldsymbol{RC} = \{rc_1, rc_2, \ldots, rc_{N_s}\}$
rc_k	Role community in the k-th scene m_k
$\boldsymbol{URC}$	Set of unique role communities in a movie, where $\boldsymbol{URC} = \{urc_1, urc_2, urc_3, \ldots, urc_{N_{\text{URC}}}\}$, where N_{URC} is the total number of unique role communities, $N_{\text{URC}} \leq N_s$
urc_i	The i-th unique role community
SRC	Support role community, which is a subset of $\boldsymbol{URC}$
$src(urc_i)$	Support role community (SRC) of urc_i
G	Directed graph representing a role-community network (RCN), where $G = (V, E)$
$A(G)$	Adjacency matrix of G, where $A(G) = [a_{i,j}]_{\|V\|\times\|V\|}$
$C(urc_k)$	Centrality value of urc_k
$lurc(t)$	Leading node (or leading URC) of the t-th group $gs(t)$ in G
$gs(t)$	The t-th group of URCs in G
$alurc(t)$	Set of associated URCs of $lurc(t)$
$\boldsymbol{URC}^{(t)}$	Set of $\boldsymbol{URC}$ at the t-th iteration in the process of RCN clustering, where $\boldsymbol{URC}^{(t)} = \boldsymbol{URC}$ for $t = 1$
MS	Set of candidate summaries for a movie
ms_k	The k-th candidate summary in MS
GS	Set of N_g URC groups $gs(t)$, $t = 1, 2, \ldots, N_g$, in G, where $GS = \{gs(1), gs(2), \ldots, gs(N_g)\}$
SP	Set of selection priorities for GS, where $SP = \{sp(1), sp(2), \ldots, sp(N_g)\}$
$sp(k)$	The k-th selected URC group from GS
SPL	Set of l URC groups selected from SP, where $SPL = \{sp(1), sp(2), \ldots, sp(l)\}$
M_{SPL}	Set of the q scenes associated with the l URC groups in SPL
m_{spl_i}	The i-th scene in M_{SPL}

can well capture the social behaviors of a movie. The main contributions of this paper are threefold: 1) to construct a role-community social network to characterize the relationships between role communities so as to detect highlights in a movie; 2) to formulate the movie summarization as a social network pruning problem so that the generated video summary can better preserve the social behaviors of a movie; and 3) to optionally incorporate several selections of user preference into the video summarization framework for guiding the selection of movie summary to better fit users' perceptions.

The rest of this paper is organized as follows. In Section II, we define some important terms used in this paper. Section III details the proposed approach of constructing the role-community network for a movie. Our proposed movie summarization method is presented in Section IV. Section V reports and discusses the experimental results. Conclusions are drawn in Section VI.

II. NOMENCLATURE

Before getting into the details of the proposed method, we first define the important terms used in this paper. The main symbols used in this paper are listed in Table I.

A. Scene

In this paper, unlike the traditional definition of a scene, a scene is defined to be composed of several consecutive shots taken in the same place. Suppose in a movie the kth scene is denoted by m_k. Note, the 180° rule is widely used in video and film production [17], where multiple cameras are used to capture different actors in the same place. For example, in a talk show, the first camera takes the speaking host at beginning. Then, the producer switches to the shot that uses the second camera to capture the expression of guests. The screen is then switched between the host and guests in the following video shots that compose the scene. These shots belong to the same scene according to our definition.

B. Role

A role stands for a major actor/actress in a movie. We use a face clustering method to identify the roles involved in each scene.

C. Role Community (RC)

An RC is composed of all roles, who appear in a scene. Suppose there are N_s annotated scenes in a movie. The RCs of scenes m_k, $k = 1, 2, \ldots, N_s$, are denoted as RC = $\{\text{rc}_1, \text{rc}_2, \ldots, \text{rc}_{N_s}\}$, where rc_k denotes the RC in the kth scene m_k. Fig. 2(a) illustrates an example where a movie clip contains 12 scenes, involving six roles labeled as a, b, c, d, e, and f, respectively. In this example, scene m_1 contains two roles, a and b. thereby denoting the RC in m_1 as $\text{rc}_1 = \{1, 1, 0, 0, 0, 0\}$. Similarly, the RC in m_{12} is denoted as $rc_{12} = \{1, 1, 1, 1, 0, 0\}$.

D. Unique Role Community (URC)

In order to identify each unique type of role communities in RC, duplicate RCs need to be removed. As a result, the set of identified URCs are denoted as URC = $\{\text{urc}_1, \text{urc}_2, \text{urc}_3, \ldots, \text{urc}_{N_{\text{URC}}}\}$, where urc_i stands for the ith URC and N_{URC} is the total number of URCs, where $N_{\text{URC}} \leq N_{\text{S}}$. Note that, the RC and URC are indexed in the chronological order, and URC $\subseteq$ RC, as exemplified in Fig. 2(b). For example, in Fig. 2(c), the set of URCs for the movie clip is URC = $\{\text{urc}_1, \text{urc}_2, \text{urc}_3, \text{urc}_4, \text{urc}_5, \text{urc}_6, \text{urc}_7, \text{urc}_8\}$ = $\{\text{rc}_1, \text{rc}_2, \text{rc}_3, \text{rc}_5, \text{rc}_6, \text{rc}_9, \text{rc}_{10}, \text{rc}_{12}\}$.

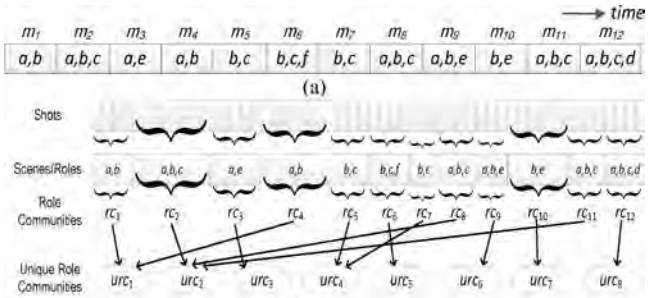
In this example, rc_4 is exactly the same as rc_1, and, hence, rc_1 and rc_4 are both included in URC, where we only use the former one "rc_1" to represent both of them, while the 4th element of URC is set to $\text{urc}_4 = \text{rc}_5$. Note, in this example, including either rc_1 or rc_4 in URC will be exactly equivalent. It is also valid for the settings of the remaining elements.

E. Support Role Community (SRC)

An SRC is a subset of URC. We denote the SRC of urc_i as $\text{src}(\text{urc}_i)$, $i = 1, 2, \ldots N_{\text{URC}}$, where $\text{src}(\text{urc}_i) \subseteq$ URC. More specifically, $\text{src}(\text{urc}_i)$ contains the URCs with the roles also included in urc_i. For example, the set of SRCs of scene m_2 is $\text{src}(\text{urc}_2) = \{\text{urc}_1, \text{urc}_4\}$. Similarly, the SRC in m_{12} is $\text{src}(\text{urc}_8) = \{\text{urc}_1, \text{urc}_2, \text{urc}_4\}$, as illustrated in Fig. 2(d).

III. SCENE ANALYSIS VIA ROLE-COMMUNITY NETWORKS

In this section, we present the analysis stage of the proposed movie summarization framework in detail. This stage includes

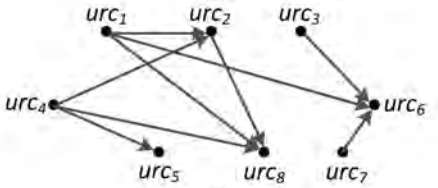


$$RC = \begin{bmatrix} rc_1 \\ rc_2 \\ rc_3 \\ rc_4 \\ rc_5 \\ rc_6 \\ rc_7 \\ rc_8 \\ rc_9 \\ rc_{10} \\ rc_{11} \\ rc_{12} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

(c)

URC	$src(\cdot)$	$C(\cdot)$
urc_1	$\emptyset$	$0-0=0$
urc_2	$\{urc_1, urc_4\}$	$1-1=0$
urc_3	$\emptyset$	$0-0=0$
urc_4	$\emptyset$	$0-0=0$
urc_5	$\{urc_4\}$	$1-0=1$
urc_6	$\{urc_1, urc_3, urc_7\}$	$2-1=1$
urc_7	$\emptyset$	$0-0=0$
urc_8	$\{urc_1, urc_2, urc_4\}$	$3-0=3$

(d)



(e)

$$A(G) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(f)

Fig. 2. Example of a movie consisting of 12 scenes. (a) Roles appearing in the kth scene m_k, $k = 1, 2, \ldots, 12$. (b) Relationship between RC and URC. (c) Role-to-scene co-occurrence matrix for the movie. (d) SRCs and the centrality values of all of the URCs in URC = $\{\text{urc}_1, \text{urc}_1, \ldots, \text{urc}_8\}$, respectively. (e) Directed graph G for the RCN of the movie. (f) Adjacency matrix $A(G)$ for the movie.

four steps including role identification, construction of an Role-community network (RCN), evaluation of the social power for each RCN node, and clustering of RCN nodes.

The basic idea of the proposed scheme is to construct an RCN to characterize the interactions of URCs in a movie. We model a movie as an RCN in which the nodes represent the URCs and the links indicate whether a member of URC is an associated SRC of another URC. As illustrated in Fig. 1, after performing scene detection and face clustering, we construct a role-to-scene co-occurrence matrix for establishing the RCN, as exemplified in Fig. 2(c).

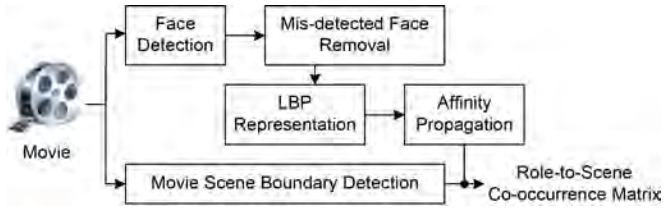


Fig. 3. Process of generating the role-to-scene co-occurrence matrix for a movie.

In order to identify movie highlights based on the constructed RCN, we define the centrality value for each URC and clustered them into relevant groups to classify their relationships. The centrality value of an URC is used to measure the importance of the URC based on its SRC set. Since typically, in a movie, the role communities in earlier scenes are usually used to develop the relationships of roles in later scenes [18], an URC with relatively rich SRC interactions is likely to be an important role community. In our method, an RCN is clustered into different relevant groups, each containing a leading URC and its associated SRCs. Note, in a relevant group, the leading URC indicates the scene containing the most significant URC. The scene with the leading URC in a group should be included in the highlight of a movie, and the scenes with the associated SRCs in the same group should also be highlighted. The operations for RCN construction are elaborated below.

A. Role Identification

In this step, we perform role identification to identify the individual major roles appearing in each scene. Role identification is an essential step in role-based movie analysis. There exist several research works addressing the role identification problem. The method proposed in [19] uses the detected human faces in the preceding scenes in a movie as training data. After that, each face in the succeeding scenes is associated with one of the roles in the preceding scenes. This approach assumes that, in a movie, all roles in a scene have appeared in the preceding scenes, but it is hard to determine the number of the preceding scenes used for role face training. The affinity propagation clustering algorithm was proposed in [20] to cluster the detected faces into relevant groups. However, the approach may still misclassify some different roles into the same cluster, or misclassify some the same role into different clusters. The method proposed in [21] further combines face clustering and the character name static in the film script to identify roles.

As shown in Fig. 3, before constructing the RCN for a movie, we perform scene boundary detection and human-face clustering to generate the role-to-scene co-occurrence matrix. In our implementation, we first use the method proposed in [22] to detect scene boundaries for a movie. Based on [22], the scene detection of a video is transformed to cluster shots into

scenes by formulating this task as a graph partitioning problem. Scene detection can then be achieved by partitioning the graph constructed for a video, where each node denotes a shot and each edge denotes the similarity between the two nodes (shots). Subsequently, to achieve reliable face detection, we perform both the AdaBoost-based [23] and the neural network-based [24] face detectors to detect faces in each scene. An object is identified as a human face only if both the two face detectors classify it as a face. Moreover, to filter out unreliable faces, if a face does not continuously appear for longer than 20 frames, it will be removed from the list of faces. In the face clustering process, we first project each detected face to the local binary pattern (LBP) feature space [25], then perform the affinity propagation algorithm in [20] for face clustering. Since our method only uses the major roles in a movie to construct the RCN, after the face clustering, the major roles in the clustered face groups are then manually annotated based on the movie information available from the Internet (e.g., the official movie website or IMDb website [32]). Note, facial expressions, lighting change, and face occlusion may lead to incorrect face detection and clustering. According to our experiments, the overall accuracy of face clustering is around 82% after manually merging some oversegmented face groups that belong to the same role. Such an accuracy rate usually can do a good job in the role-community -based summarization as viewers can easily perceptually tolerate the small number of incorrectly selected scenes due to the inaccuracy of face clustering.

B. Construction of RCN

In our method, an RCN is a directed graph G defined as $G = (V, E)$, where $V = URC = \{\text{urc}_1, \text{urc}_2, \ldots, \text{urc}_{N_{\text{URC}}}\}$ denotes the vertex set of G, and $E = \{(v, w) \mid v, w \in V\}$ stands for the edge set of G, which together indicate the relationships between the URCs in URC, as illustrated in Fig. 2(e). The weight of each edge (or path) in G is unity. In Fig. 2(e), the direction of the edge from urc_i to urc_j means urc_i is an element of $\text{src}(\text{urc}_j)$, that is, urc_i is one of the URCs with the roles also included in urc_j.

As mentioned before, the narrative of a movie usually exploits the role relationships in earlier scenes to develop the role relationships in later scenes. Therefore, the RCN for a movie should properly describe the relationships between a URC and the URCs in its SRC set in the movie. Let $(\text{urc}_i, \text{urc}_j)$ denote an edge from urc_i to urc_j in E. The URC relationship between urc_i and urc_j can be expressed by an adjacency matrix $A(G) = [a_{i,j}]_{|V|\times|V|}$, where

$$a_{i,j} = \begin{cases} 1, & \text{if urc}_i \text{ is an SRC of urc}_j \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

as exemplified in Fig. 2(f) for the RCN shown in Fig. 2(e).

C. Evaluation of Social Power of RCN Node

To evaluate the social power of an RCN node (or a URC), we define the centrality value of each node in the RCN (i.e., graph G) for a movie, so that the larger the centrality value is, the higher the social power will be. Since the RCs in earlier scenes are usually used to develop the role relationships in later scenes in a movie, intuitively the centrality value can be defined according to the SRC set of each node. In making a movie, the movie director usually uses several scenes as a prelude to introduce later role interactions and these prelude scenes may be associated with several earlier and later URCs in the SRC set of an URC. To measure the social power of an URC, we define the centrality value of urc_k as the difference of the numbers of the URCs that have appeared and are to appear, respectively, in its SRC set as follows:

$$C(\text{urc}_k) = \sum\nolimits_{jk} a_{j,k} - \sum\nolimits_{j>k} a_{j,k} \tag{2}$$

where $a_{j,k}$ is defined in (1). In (2), $\sum_{j\le k} a_{j,k}$ and $\sum_{j>k} a_{j,k}$ denote the numbers of URCs in $\text{src}(\text{urc}_k)$ for $j \le k$ and $j > k$, respectively, based on the definition shown in (1). For example, recall from the example $\text{src}(\text{urc}_2) = \{\text{urc}_1, \text{urc}_4\}$ described in Fig. 2(d), the centrality value of urc_2 is $C(\text{urc}_2) = 1 - 1 = 0$, that is, when displaying urc_2, urc_1 has already appeared, but urc_4 has not appeared yet. Similarly, Fig. 2(d) shows the centrality values of the other URCs.

TABLE II
PROPOSED RCN CLUSTERING ALGORITHM

Input: an RCN with its URC set (denoted by $\boldsymbol{URC}$), the SRC sets (denoted by $src(urc_k)$, $k = 1, 2, \ldots, N_{\text{URC}}$) of all URCs, and the centrality values (denoted by $C(urc_k)$, $k = 1, 2, \ldots, N_{\text{URC}}$) of all URCs in $\boldsymbol{URC}$, .
Output: the clustering result of the input RCN, denoted by $gs(t)$, $t = 1,2,\ldots, N_g$.

1. Set $t = 1$ and $\boldsymbol{URC}^{(t)} = \boldsymbol{URC}$.
2. Find the leading URC in $\boldsymbol{URC}^{(t)}$ via
$$lurc(t) = arg \max_{urc_k \in \boldsymbol{URC}^{(t)}} C(urc_k).$$
3. Find the set of associated URCs of $lurc(t)$ via
$$alurc(t) = \left\{ urc_i \in src(lurc(t)) \middle| \begin{matrix} urc_i \notin src(urc_j), \\ urc_j \in src(lurc(t)), \text{and } j < i \end{matrix} \right\}.$$
4. Set $gs(t) = \{lurc(t), alurc(t)\}$.
5. Set $t = t + 1$, $\boldsymbol{URC}^{(t)} = \boldsymbol{URC}^{(t-1)} \setminus gs(t-1)$, and go to Step 2 if $\boldsymbol{URC}^{(t)} \neq \emptyset$. Otherwise, go to Step 6.
6. Return the clustering result $gs(t)$, $t = 1,2,\ldots, N_g$.

D. Clustering of RCN Nodes

After constructing the RCN for a movie, the nodes in the RCN will be iteratively clustered into relevant groups according to their centrality values. Each group consists of a leading URC node and its associated URC nodes. As summarized in Table II, the proposed RCN clustering algorithm iteratively performs the following two steps: 1) selection of the leading URC node, and 2) selection of the associated URCs, as elaborated below.

1) *Selection of Leading URC Node:* To select the leading URC node for the t-th group, the node with the largest centrality value in the group is selected as follows:

$$\text{lurc}(t) = \arg\max_{\text{urc}_k} \in \text{urc}^{(t)} C(\text{urc}_k) \tag{3}$$

where $\text{URC}^{(t)}$ represents the set of URC at the t-th iteration in the RCN clustering process, where $\text{URC}^{(t)}$=URC for t =1. For example, the first (t = 1) group of the RCN in Fig. 2(e) is

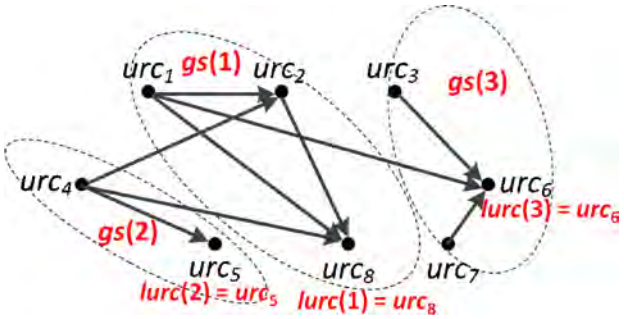


Fig. 4. Clustering result of the RCN shown in Fig. 2(e), where the URCs are clustered into the three groups, gs(1), gs(2), and gs(3).

$\text{URC}^{(1)} = \text{URC} = \{\text{urc}_1, \text{urc}_2, \text{urc}_3, \text{urc}_4, \text{urc}_5, \text{urc}_6, \text{urc}_7, \text{urc}_8\}$ and $\text{lurc}(1) = \text{urc}_8$, that is, "$C(\text{urc}_8) = 3$" is the maximum centrality value among the centrality values of all of the URCs in $\text{URC}^{(1)}$.

2) *Selection of the Associated URCs:* After selecting the leading node $\text{lurc}(t)$ for the t-th group, its associated URCs will also be selected into this group. To select the associated URCs of $\text{lurc}(t)$, all the URCs in $\text{src}(\text{lurc}(t))$ are identified first. Then, to make the associated URC set as concise as possible, we remove those urc_i in $\text{src}(\text{lurc}(t))$ should urc_i also belong to $\text{src}(\text{urc}_j)$, where $\text{urc}_j \in \text{src}(\text{lurc}(t))$ and $j < i$, that is, the set of associated URCs of $\text{lurc}(t)$ is selected as

$$\text{alurc}(t) = \left\{ \text{urc}_i \in \text{src}(\text{lurc}(t)) \;\middle|\; \begin{array}{c} \text{urc}_i \in \text{src}(\text{urc}_j), \\ \text{urc}_j \in \text{src}(\text{lurc}(t)), \text{and } j < i \end{array} \right\}. \tag{4}$$

Following the above-mentioned example of "$\text{lurc}(1) = \text{urc}_8$," to decide on $\text{alurc}(1)$, i.e., the set of associated URCs of $\text{lurc}(1)$, we first select all the URCs in $\text{src}(\text{urc}_8)$, i.e., $\text{src}(\text{urc}_8) = \{\text{urc}_1, \text{urc}_2, \text{urc}_4\}$. Because urc_4 is already in $\text{src}(\text{urc}_2)$ (as shown in Fig. 2(d), $\text{src}(\text{urc}_2) = \{\text{urc}_1, \text{urc}_4\}$), we remove urc_4 and select $\text{alurc}(1) = \{\text{urc}_1, \text{urc}_2\}$, that is, the constraints used in (4) are mainly designed to prevent the inclusion of those URCs, which already appeared in a previous scene(s) in a movie, into the set of associated URCs of current leading URC. Then, we decide on the t-th group by setting $\text{gs}(t) = \{\text{lurc}(t), \text{alurc}(t)\}$. For example, $\text{gs}(1) = \{\text{urc}_1, \text{urc}_2, \text{urc}_8\}$. After determining $\text{gs}(t)$, we iteratively perform the clustering process by selecting the leading node and its associated URCs for the $(t+1)$-th group as described below.

First, we set $\text{URC}^{(t+1)} = \text{URC}^{(t)} \backslash \text{gs}(t)$. The backslash operator "\" denotes the set difference. That is, we remove the URCs included in $\text{gs}(t)$ from $\text{URC}^{(t)}$ to form $\text{URC}^{(t+1)}$. As a result, $\text{URC}^{(2)} = \{\text{urc}_3, \text{urc}_4, \text{urc}_5, \text{urc}_6, \text{urc}_7\}$. Then, we perform the two processes of leading node selection using (3) and associated URCs selection using (4), respectively. For example, based on (3) and Fig. 2(d), the two URCs with maximum centrality values in $\text{URC}^{(2)}$ are urc_5 and urc_6. In this case, we select the URC in the former scene as the leading node, i.e., $\text{lurc}(2)=\text{urc}_5$. Then, based on (4), we obtain $\text{alurc}(2) = \{\text{urc}_4\}$. Consequently, the second group can be obtained as $\text{gs}(2) = \{\text{lurc}(2), \text{alurc}(2)\} = \{\text{urc}_4, \text{urc}_5\}$. The clustering process is iteratively performed until all nodes (or URCs) in an RCN are classified, as illustrated in Fig. 4.

IV. Proposed RCN-Based Movie Summarization

In this section, we present the proposed movie summarization scheme that formulated the problem of movie summarization as a social network pruning problem, in which a user-specified summarization ratio is used to control the number of scenes containing essential RCs to be preserved, or the number of scenes containing redundant RCs to be removed. In general, a good movie summary should remove as much redundant content as possible while preserving essential semantic information in a movie. However, as the importance and redundancy of movie content is rather subjective, a summarization method should also take into account user preferences. In our method, the movie summarization stage consists of two major steps. First, all feasible combinations of RCs and the associated scenes are identified under a user-specified length constraint. Second, the output summary is selected from the set of identified scene combinations based on the constructed RCN nodes and user preferences. The details of the proposed movie summarization algorithm are elaborated below.

A. Problem Formulation

Suppose a movie contains N_s scenes, denoted as $F = \{m_1, m_2, m_3, \ldots, m_{N_s}\}$,, where m_k denotes the kth scene of the movie. Subject to the user-specified summarization ratio (sr) for a movie, we formulate summarization of the movie as

$$ms = \text{U}_{m_i}, \text{s.t.} \frac{|\text{ms}|}{|F|} \text{sr} \tag{5}$$

where ms is a movie summary consisting of a set of scenes selected from the movie F, $m_i \in F$ denotes one of the scenes selected from F, $|\text{ms}|$ and $|F|$ are the lengths (in terms of presentation time) of ms and F, respectively. The main problem is, given the length constraint (i.e., the sr), how to properly select a set of scenes from the movie to concisely and semantically summarize the movie to satisfy viewers' satisfactions. Obviously, there are multiple feasible combinations of scenes that can meet the length constraint. Assume there are Z feasible combinations in the candidate set represented as $\text{MS} = \{\text{ms}_1, \text{ms}_2, \text{ms}_3, \ldots, \text{ms}_z\}$. The proposed summarization method selects the summary from MS by solving the following optimization problem:

$$\text{ms}^* = \arg \max_{\text{ms}_k \in \text{MS}} R(\text{ms}_k) \tag{6}$$

where ms* denotes the optimal movie summary, and $R(\cdot)$ denotes the retention function defined according to the objective of summarization, which is used to quantify the amount of semantic information preserved in a candidate summary. For example, the cognitive-level methods proposed in [2] and [3] use low-level features to extract important content, where the retention functions are defined as the amount of important low-level features inside the summary. The affective-level methods proposed in [8]–[11] use an affective model to select highlight content into the summary, where the retention function is defined as the degree of affectiveness in the summary.

In our method, the retention function is designed to measure the comprehensiveness of preserved RC relationships. As

aforementioned, the semantic importance of movie content is dependent on the preferences of individual viewers. The preference on the amount of RC relationship and on the number of extracted scenes preserved in a summary for a movie is different from viewer to viewer. However, no matter what the user preference is, in the proposed method, each candidate summary should contain at least the most essential cluster of URCs, i.e., the first group gs(1), determined by the proposed RCN clustering algorithm shown in Table II. Recall from our previous example illustrated in Fig. 4, where the essential URC cluster is gs(1) = $\{\text{urc}_1,\text{urc}_2,\text{urc}_8\}$. The optimal summary is then selected from the candidate list of summaries to satisfy viewers' preference, as described below.

B. Identification of Feasible Summaries

Under a length constraint, there usually exist multiple feasible summaries containing different RC relationships and the associated scenes to concisely describe the movie content. In order to select a good summary, all of the URC groups and the associated scenes should be prioritized according to their importance (e.g., the social powers of scenes) so that an optimal summary with the highest overall importance can be selected from the feasible combinations. To this end, the proposed method first determines the selection priority of each URC group and then determines the selection priority of each movie scene as follows.

1) *Group-Level Selection Priority:* After constructing an RCN for a movie, the proposed method first determines the selection priority of each URC group in the RCN. In order to keep low redundancy in the selected URC groups, we use a greedy selection approach that iteratively selects the next URC group that contains the richest URC interactions in the remaining unselected URC groups. Suppose there are N_g URC groups in the RCN for a movie to be summarized, denoted by GS = $\{\text{gs}(1),\text{gs}(2), \ldots ,\text{gs}(N_g)\}$, obtained in the analysis stage. The corresponding set of selection priority for GS is denoted as SP = $\{\text{sp}(1),\text{sp}(2), \ldots ,\text{sp}(N_g)\}$, where sp($k$) is the kth URC group selected from GS, and the selection priority of sp(k) is higher than that of sp(k+1) and sp(1) = gs(1). The first URC group, i.e., gs(1), always has the highest priority to be selected. Then, the next group is selected by finding from the current nonselected groups, the group with the richest RC interactions, i.e., with the largest number of URCs in it. The selection process is repeated until all of the groups are selected according to their priority from the RCN. For example, for the RCN shown in Fig. 4, after setting sp(1) = gs(1), we select the group with the largest number of URCs from the remaining unselected groups, i.e., sp(2) = gs(3) (with three URCs), and then set sp(3) = gs(2) (with two URCs).

2) *Scene-Level Selection Priority:* After determining the selection priority of each URC group in an RCN, the next step is to prioritize the scenes associated with each URC group. Suppose a total of l URC groups are selected from the set of SP, denoted as SPL = $\{\text{sp}(1), \text{sp}(2), ..., \text{sp}(l)\}$, $1 \le l \le N_g$. Moreover, the q scenes associated with the l URC groups in SPL are denoted as $M_{\text{SPL}} = \left\{m_{\text{spl\_}1}, m_{\text{spl\_}2}, m_{\text{spl\_}3}, \ldots , m_{\text{spl\_}q}\right\}$, which represents the set of candidate scenes which may be selected into the movie summary. In our method, without explicitly determining a specific value of "l," we consider each possible set of SPL for $1 \le l \le N_g$ to generate the corresponding set of associated scenes. For example, when l = 1, we set SPL = $\{\text{sp}(1)\}$ and only consider the combinations of SPLs associated set of scenes. When l = 2, SPL = $\{\text{sp}(1),\text{sp}(2)\}$ and the combinations of its associated set of scenes are considered. We repeat this process until $l = N_g$, where SPL = $\{\text{sp}(1), \text{sp}(2), \ldots , \text{sp}(N_g)\}$. Then, we combine all of the associated scenes for each possible l to form the corresponding set of M_{SPL}, where we totally have $N_{\text{g}}\cdot M_{\text{SPL}}$ sets.

Note that the selected scenes may still have high redundancy since there could exist several scenes associated with the same RCN node (or URC group). The redundancy makes the viewer feel boring and impatient while watching the summary with the same RC repeating several times. Such problem often occurs in the scenes with essential RCs. For example, dialog scenes often appear in a movie, e.g., the main roles have a conversation in a room, and soon or later, they have another conversation on a street. However, these scenes look redundant and should be summarized with only one representative scene, if these dialog scenes contain similar semantics. For another example, in a police chasing scene, at the first moment, a criminal was installing a time bomb and then at the next moment, the main actor/actress is driving a car to stop the accident. The scenes are continuously switched between the criminal and main actor/actress. Although such film style may bring the climax of a story, it is redundant to include too many such scenes into a movie summary, while a viewer just would like to briefly know the evolution of the story. Those repeated scenes should be summarized with fewer representative scenes.

In order to remove the redundant scenes from each set of M_{SPL}, we utilize the latent semantic analysis (LSA) proposed in [26] to measure the similarity between two scenes. The LSA is a widely used technique in information retrieval for analyzing the relationship between words and documents. In our method, a main URC and its associated SRCs in scenes are treated as words, while their associated scenes are treated as documents. In short, we apply LSA based on singular values decomposition (SVD) to analyze the relationships between RCs and scenes. It characterizes the role-to-role relationship and scene-to-scene relationship by projecting the role-to-scene co-occurrence matrix to the RC feature space and the scene feature space. After decomposing the role-to-scene co-occurrence matrix [e.g., Fig. 2(c)], we obtain the scene vectors U, the block diagonal matrix S, and the URC vectors V^{T}. Then, the similarity between the ith and jth scenes in M_{SPL}, i.e., $m_{\text{spl\_}i}$ and $m_{\text{spl\_}j}$, can be measured by the inner product of the ith and jth rows of U.S matrix, i.e., u_iS and u_jS, where u_i and u_j denote the ith and jth rows of U, respectively. Here, we use the following metric to measure the similarity between $m_{\text{spl\_}i}$ and $m_{\text{spl\_}i}$ as

$$\text{Sim}_s(m_{\text{spl\_}i}, m_{\text{spl\_}j}) = \frac{u_i \text{S}^2 u_j^T}{\|u_i\text{S}\| \left\|u_j\text{S}\right\|} \tag{7}$$

where S is a diagonal and square matrix.

Moreover, while watching a movie, we cannot always remember every scene which has appeared, especially when

a scene appeared long ago. Hence, we adopt the temporal proximity weight proposed in [22] to simulate the memory decaying effect to recall a scene in a movie as

$$w_{\text{mem}}(m_{\text{spl}\_i}, m_{\text{spl}\_j}) = \text{e}^{-\frac{1}{d_t} \cdot \left| \frac{p\left(m_{\text{spl}\_i}\right) - p\left(m_{\text{spl}\_j}\right)}{\sigma_t} \right|^2} \tag{8}$$

where $p(m_{\text{spl}\_i})$ and $p(m_{\text{spl}\_j})$ denote the temporal positions of the two scenes, m_{spl_i} and m_{spl_j}, respectively, σ_t is the standard deviation of scene durations in the entire movie, and d_t is used to control the strength of the temporal proximity weighting function, which is set to 20 as suggested in [22].

In (8), the temporal proximity weighting function is an exponentially decreasing function of temporal distance between two arbitrary scenes to characterize the ability of recalling a scene. Then, based on the similarity metric defined in (7) and the temporal proximity weighting function defined in (8), we define the overall weighted similarity (i.e., the redundancy) in the set of candidate scenes M_{SPL} as follows:

$$\begin{aligned} \text{WSim}_M(M_{\text{SPL}}) = &\sum\nolimits_{j=i} \text{Sim}_s(m_{\text{spl}\_i}, m_{\text{spl}\_j}) \\ &\times w_{\text{mem}}(m_{\text{spl}\_i}, m_{\text{spl}\_j}). \end{aligned} \tag{9}$$

Then, the next step is to remove unessential scenes from M_{SPL} to reduce redundancy. To effectively reduce the redundancy of the scenes included in M_{SPL}, we propose to iteratively identify and remove the most redundant scene to minimize the overall redundancy in $M_{\text{SPL}}^{(t)}$, i.e., the set of M_{SPL} at the t-th iteration, so as to generate a compact summary for a movie:

$$m_{\text{SPL}\_r}^{(t)} = \arg\min_i \text{WSim}_M\left(M_{\text{SPL}}^{(t)} \backslash m_{\text{spl}\_i}^{(t)}\right) \tag{10}$$

where $m_{\text{spl}\_r}^{(t)}$ is the scene determined to be removed at the t-th iteration, and $m_{\text{spl}\_i}^{(t)}$ denotes the ith scene in $M_{\text{SPL}}^{(t)}$. We then remove $m_{\text{spl}\_r}^{(t)}$ from $M_{\text{SPL}}^{(t)}$ to obtain $M_{\text{SPL}}^{(t+1)}$. Subsequently, we iteratively apply (9) and (10) to identify and remove redundant scenes until the remaining overall length (or presentation time) of preserved scenes can meet user's requirement, while the retained scenes should still sufficiently preserve the URCs inside SPL.

Note, the above greedy algorithm considers the richness of RC interactions at the group-of-scenes level and the dissimilarity (or diversity) between scenes at the scene level. The movie summaries selected by the greedy algorithm, however, do not take into account user preferences. In our implementation, we relaxed the summarization ratio in (5) to $\text{sr} \pm d_{\text{sr}}\%$ to use the greedy algorithm to select a set of candidate summaries that meet the length range of $\text{sr} \pm d_{\text{sr}}\%$, where d_{sr} is empirically set to 3. We then perform the algorithm described below to select from the set of candidate summaries the final summary based on user preferences.

C. User Preference-Guided Selection of Movie Summary

After performing the algorithm described above, several candidate summaries will be obtained. Let $\text{MS} = \{\text{ms}_1, \text{ms}_2, \text{ms}_3, \ldots \text{ms}_2\}$ denote all feasible combinations that meet the specified length range, where ms_i denotes the ith candidate summary (a set of selected scenes) for a movie. The next step is to identify from these candidates the summary that best satisfy the user preferences. Due to the large variety of user preferences, it is hard to automatically select a combination of URCs and the associated scenes that satisfy a user's preference. To solve the problem, interactions with viewers can be performed by asking the viewers to specify their preferences before the summary selection so as to use the user preferences to guide the summary selection. However, before watching a movie, a viewer might not be able to clearly specify her preference for this movie. Hence, a few works [11], [27] designed user interfaces for viewers to interactively specify their preferred content. Most of the existing works extract low-level features, e.g., color, motion, tempo, from the user specified items to develop the summary. Instead of directly asking a viewer to select her favored low-level features, our method offers three easy-to-select options of user preference to a viewer for movie summary generation, including 1) preference on a summary covering more scenes consisting of major roles, the more the better; 2) preference on a summary with motion activity, the more the better; and 3) preference on a summary with focus on the movie endings.

It is actually difficult to design a set of specific user preferences, especially for viewers who have no idea about a movie to be summarized. Therefore, we intend to design a set of general purpose user preferences for most viewers to easily embed their preferences in a summary. If the endings of story segments or chapters in a movie are known, more elaborate user preferences may be designed. For example, the user preferences could be "Would you like the summary emphasizing on the ending between main actor and main actress?" or "Would you like the summary emphasizing on the roles related to the movie ending?"

For the first user preference selection, i.e., "preference on a summary covering more scenes consisting of major roles, the more the better," we count the total number of nonduplicate RCs associated with each summary candidate We then sort the number of RCs in descending order and assign a rank value to each candidate accordingly, where the smaller the rank value, the higher the priority of the corresponding candidate summary.

For the second user preference selection, i.e., "preference on a summary with motion activity, the more the better," we calculate the motion attention value (MAV) for each summary candidate by summing up the MAV of scenes associated with the summary candidate based on the motion attention model proposed in [6]. Similarly, we also sort the MAVs in descending order and assign a rank value to each candidate accordingly.

For the third user preference selection, i.e., "preference on a summary focusing on the movie endings," we sum up the temporal index of the last video frame in each scene associated with the summary candidate, which measures the degree of closeness between the summary candidate and the movie ending. Similarly, we also sort the index in descending order and assign a rank value to each ms_i accordingly.

Finally, for each summary candidate, we calculate the weighted sum of its rank values corresponding to the

TABLE III
PROPOSED MOVIE SUMMARIZATION ALGORITHM

Input: (i) the RCN clustering result for the input movie, obtained by our RCN clustering algorithm summarized in TABLE II, denoted by $GS = \{gs(1), gs(2), \ldots, gs(N_g)\}$; and (ii) summarization ratio sr for the movie
Output: selected summary ms^* for the movie

1. **Determine the group-level preservation priority** for GS to obtain $SP = \{sp(1), sp(2), \ldots, sp(N_g)\}$, where $sp(k)$ denotes the k-th selected URC group from GS based on the number of URCs in each $gs(t)$.
2. **Determine the scene-level preservation priority** by selecting the first l URC groups from SP, denoted as SPL. Then, iteratively remove the redundant scenes from all of the scenes (denoted as M_{SPL}) associated with SPL by estimating the similarity between two arbitrary scenes in M_{SPL} using (11) to obtain a set of summary candidates, $MS = \{ms_1, ms_2, ms_3, \ldots, ms_z\}$, with summarization ratios closest to the user's preference ratio sr.
3. **Select the final movie summary** ms^* from MS based on the three user preference selections by calculating the three respective measurements for each ms_i, and the corresponding rank of the combination of them, denoted as $\mathrm{Rank}(ms_i)$. Then select $ms^* = \arg\min_{ms_i \in MS} \mathrm{Rank}(ms_i)$.

preference options specified by the viewer, where the weight of a preference is set unity if the preference is specified, and is set zero otherwise. Consequently, the candidate summary with the smallest rank is selected as the final summary. The proposed movie summarization algorithm is summarized in Table III.

V. EXPERIMENTS AND DISCUSSION

To evaluate the performance of the proposed movie summarization method, we conducted subjective tests to evaluate the user satisfaction on movie summaries obtained using the proposed and other summarization methods, including the attention-based method [6] and the role-based method [12] that belong to cognitive-level and social -network-based approaches, respectively. To demonstrate the performance of the analysis stage of the proposed method and make a fair comparison with the method proposed in [12], the role-based method was implemented based on the role analysis technique in [12], followed by our summarization stage with some necessary modifications as follows. For the summarization process for [12], the group-level selection priority is replaced with the co-occurrence frequency of roles and the input of LSA is a role-to-scene co-occurrence matrix to follow the main spirit of "co-occurrence frequency of roles" proposed in [12]. We also evaluated the proposed method without considering user preference selections and with user preference selections.

There is no standard testing procedure for evaluating the performances of video summarization algorithms. As mentioned in [28], the existing evaluation approaches for video summarization can be classified into intrinsic and extrinsic methods. In extrinsic methods, a video summary is usually evaluated with respect to its impact on the performance for a specific information retrieval task. The main problem in extrinsic evaluation is that the applied metric is required to be well correlated with the performance of a task which is the major goal of a video summary to be evaluated. Otherwise, the summary may be scored lower on a task with lower performance regardless of summary quality.

On the other hand, in intrinsic evaluation methods, the quality of a generated video summary is judged directly based on summary analysis, where the criteria can be user judgment of fluency of the video summary, coverage of key scenes of the source material, or similarity to an ideal video summary edited by an expert. More specifically, intrinsic evaluation methods usually assess the subjective qualities of video summaries by adopting a questionnaire methodology to evaluate subjects' experiences about the summaries. However, film styles and genres are usually diverse, and, hence, in intrinsic evaluation, it is usually hard to derive ideal video summaries for evaluation, even professionals may not agree on which parts of the scenes to be included in a summary. Moreover, a video summary may receive different scores under different measures, or when compared to different reference summaries created by different experts. Therefore, in our subjective experiments, we conducted an intrinsic evaluation and designed a questionnaire to ask subjects to rank the qualities among different video summaries generated by our method and the two compared methods [6], [12]. The main reason why we adopted an intrinsic evaluation in our experiments is that the main goal of the proposed approach is to generate movie summaries to satisfy viewers' user preferences while preserving as much information as possible, rather than designing for a specific information retrieval task, as mentioned in [28].

A. Subjective Quality Evaluation

In order to evaluate the subjective quality of various movie genres, we selected 12 Hollywood movies as listed in Table IV. We set the summary length to be about 15–18 min for the compared methods and the lengths of the generated summaries are listed in Table IV. In our experiments, we invited 36 subjects to participate in the subjective tests. None of the subjects had knowledge about the algorithm implementations. In [5] and [10], subjects were asked to participate in all subjective tests. However, since there were 12 test movies in our experiments, asking subjects to evaluate all test movies would give them too heavy workload to keep their concentration on the evaluation. To avoid the problem, each subject was asked to participate in the evaluation of four movies, which were randomly chosen from the 12 test movies. Each test movie was thus evaluated by 12 subjects—an enough number of subjects to ensure the results not be biased by a few subjects. Each subject specified her own preferences, and, hence, our system generated 12 different summaries for each movie for the 12 different subjects, respectively. In our experiments, the preferences were specified before watching the movies based on the fact that viewer usually watch a summary of a movie before watching the complete movie to judge whether I am worth to watch or not. A possible extended application of user preference selection is that a viewer may specify her preferences after watching parts of a movie if she has no enough time to watch the whole movie. This viewer can then specify the user preferences based on her preferences and the knowledge learned from watching the partial content, so

TABLE IV
DETAILED INFORMATION ABOUT THE TEST MOVIES

ID	Movie Title (Release Year)	Genres [32]	Length	Summary Length		
				Attention-based method [6]	Role-based method [12]	The proposed method With all 3 preferences
1	Up (2009)	Animation/Adventure/Comedy	96 m	17m 16s	16m 18s	16m 54s
2	Evan Almighty (2007)	Comedy/Family /Fantasy	96 m	15m 35s	16m 38s	15m 30s
3	The Hurt Locker (2008)	Drama/Thriller/War	131 m	15m 50s	16m 15s	16m 14s
4	The Devil Wears Prada (2006)	Drama/Romance	109 m	15m 52s	13m 17s	16m 31s
5	Ratatouille (2007)	Animation/Comedy/Family	111 m	15m 58s	17m 15s	24m 49s
6	Salt (2010)	Action/Crime/Mystery	100 m	15m 21s	16m 40s	16m 13s
7	Quantum of Solace (2008)	Action/Adventure/Thriller	106 m	16m 08s	16m 07s	17m 35s
8	Lost in Translation (2003)	Drama	104 m	16m 34s	17m 18s	17m 32s
9	The Rebound (2009)	Comedy/Romance	95 m	15m 26s	15m 47s	15m 07s
10	Notting Hill (1999)	Comedy/Romance	124 m	16m 06s	17m 45s	17m 34s
11	The Pursuit of Happiness (2006)	Biography/Drama	117 m	17m 48s	17m 44s	17m 51s
12	The Da Vinci Code (2006)	Mystery/Thriller	149 m	15m 54s	17m 33s	17m 40s

TABLE V
QUESTIONNAIRE USED IN THE SUBJECTIVE TEST (IN Q1 - Q6, SUBJECT RANKS EACH METHOD FROM THE BEST TO THE WORST)

No.	Questions
Q1	How about the enjoyability of each summary?
Q2	How about the informativeness of each summary?
Q3	In order to understand the major plots of the movie, please rank the suitability of each summary.
Q4	In order to understand the major role relationship, please rank the suitability of each summary.
Q5	How does the summary capture the development of the movie plots?
Q6	Please rank the overall quality of each summary.
Q7	How do you make your decision in Q6?

that the proposed method may generate a summary that better meets the viewer's preferences. The subjects were not asked to do the test in a controlled environment. Instead, they were allowed to watch the assigned movies and the corresponding summaries as many times as they want in their own chosen places.

In our experiments, we designed the questionnaire based on the set of questions used in [29] and [30]. More discussions about questionnaire-based subjective ranking for video summarization can be found in [31]. Table V lists the seven questions used for subjective evaluation, denoted by Q1–Q7, respectively. In questions Q1–Q6, subjects were asked to rank the movie summaries from best to worst. In Q7, each subject was asked to give the reasons for the decisions in Q6. Q7 was used to understand the feeling of a subject as well as to screen out unattended answers. Fig. 5(a)–(f) depicts the evaluation results of Q1–Q6 for the 12 test movies. The rank assigned to an evaluated summarization method is denoted by a score of 4, 3, 2, or 1 to indicate that it is ranked as the best, the second, the third, or the worst, respectively. Consequently, the score of each summary for a question is the average of the individual scores received from all the subjects. Fig. 6 illustrates the weights of the proposed three user preference selections for each test movie.

In question Q1, subjects were asked to rank the enjoyability of a movie summary. The enjoyability is to measure whether the movie summary can extract semantically complete and coherent scenes, for which the proposed methods significantly outperform the other two compared methods for most test movies. For movie #1 and #9, a few subjects felt that the movie summaries obtained by the proposed methods miss some exciting fighting scenes or important prelude scenes, thereby making the average scores of the proposed methods close to the other two compared methods. In Q2, subjects were asked to rank the summaries in terms of informativeness [6], which is used to measure the degree of richness of content contained in a movie summary. The proposed methods still significantly outperform the other two methods in terms of informativeness. For movie #11, the two summaries obtained by the proposed methods miss some critical content of the main actress so that the summaries were ranked to be close to the role-based method [12]. In Q3, subjects were asked to rank the abilities of movie summaries in preserving the major plots of the movie, whereas Q4 was used to evaluate the abilities of the summaries in preserving the major role relationships. The difference between the major plots and the informativeness in Q2 is that a movie summary receiving a high score in informativeness may still contain unessential scenes, whereas the major plots of a movie should not contain unessential scenes. The results of Q3 and Q4 show that our methods achieve the best performances for most test movies. With the proposed RCNs, the extracted movie summaries not only preserve important role relationships, but also keep the major plots of a movie. In Q5, subjects are asked to answer how the summary captures the movie plot. However, for some movies (e.g., movies #11 and #12), a few subjects felt the proposed methods miss some critical subplots, thereby may not outperform the compared methods in this aspect. The reason is that the proposed method may miss some subplots of a movie if the movie consists of multiple main (near-)independent subplots, where the role communities among them are less correlated. In case of such situation, the summary obtained

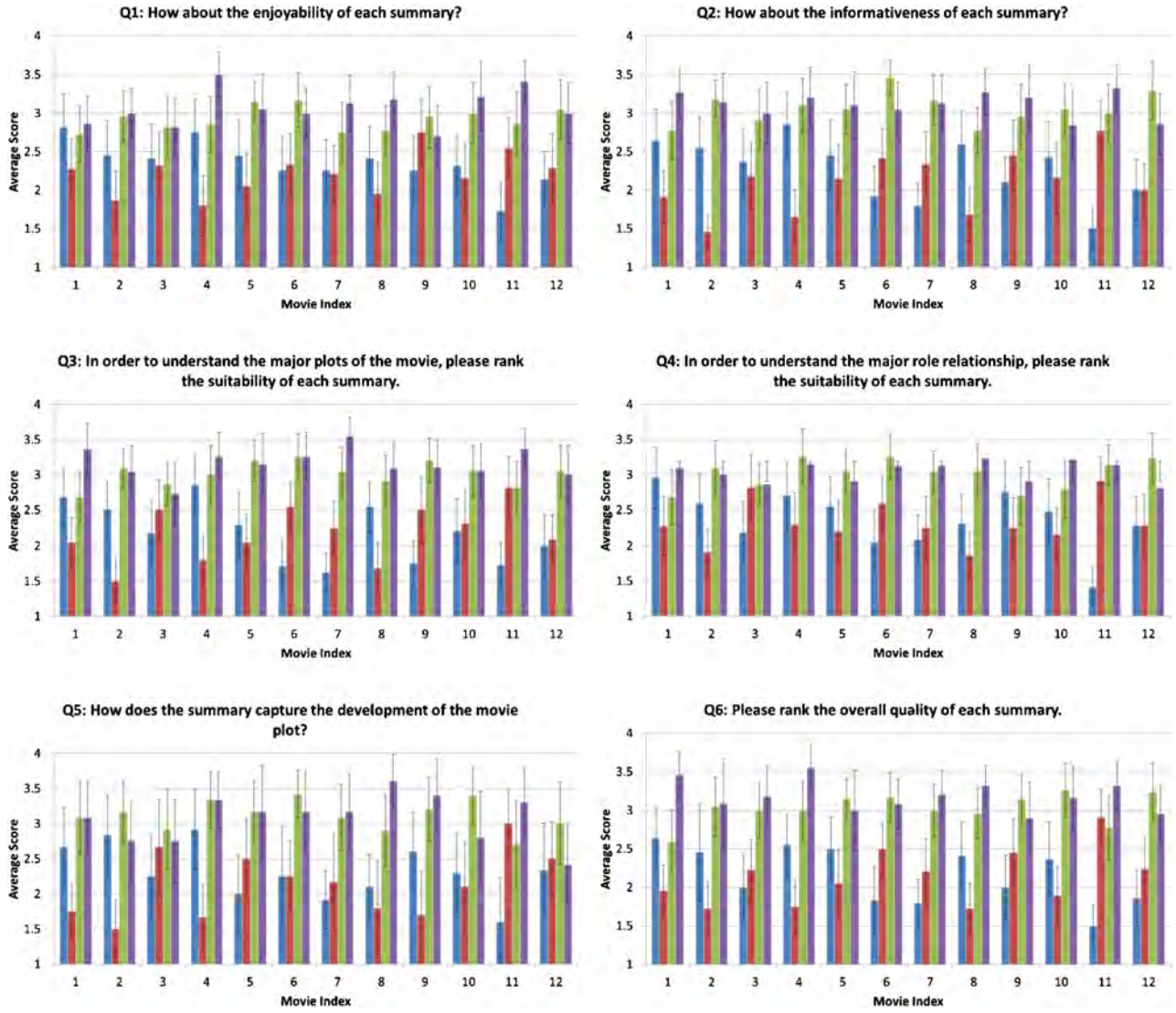


Fig. 5. Subjective evaluation results of the Q1 - Q6 for the 12 test movies: the blue, red, green, and purple lines indicate the attention-based method, the role-based method, the proposed method without user preference selections, and the proposed method with user preference selections, respectively.

by the proposed method may concentrate on the scenes corresponding to role communities with higher social power, which usually belongs to the same subplot(s). In Q6, subjects were asked to rank the overall qualities of the summaries, where the summaries obtained by the proposed methods (with and without user preference selections) were ranked as the first and second for almost all test movies (except movies #1 and #11). In Q7, subjects were asked to identify their main reasons about the decisions for Q6. According to the feedbacks for Q7, the subjects' evaluations were mainly based on whether a summary can well capture the plot development and major role interactions in a movie. Most subjects commented that the proposed methods can bring better user experiences in terms of the completeness of relationships among major roles, and the preservation of the main plots in a movie.

The subjective evaluation results also show that the attention-based method [6] outperforms the role-based method [12] for about half of the test movies. The main reason is that the role-based method tends to select the scenes with roles that have appeared more frequently in a movie, which are usually in the rear part of the movie, thereby missing many important scenes in the front part of the movie. Moreover, the main reason why the attention-based method performs the worst for several movies is that the attention-based method cannot discover the importance of each role and the relationships among roles, thereby leading to the missing of some essential scenes for story development in the generated video summaries.

Note, the genres of the test movies include action, romance, and biography. Intuitively, the summary of an action movie should contain more action scenes. However, when the summarization ratio is too low, the preserved contents selected by the attention-based method mainly contain fighting scenes, making subjects hard to understand the development of role interactions. Similarly, the romance and biography movies typically use many static dialog scenes to describe the interactions among roles. The attention-based method usually fails to select these important dialog scenes into the summary.

Besides the above subjective user study, we also conducted two quantitative evaluations. In the evaluations, we first asked a professional with film editing expertise to manually edit

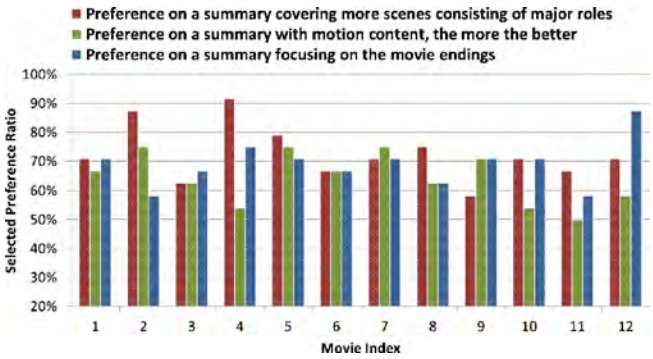


Fig. 6. Preference weights of the three user preference selections for the 12 test movies.

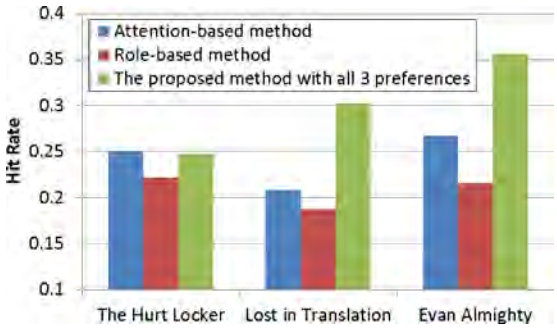


Fig. 7. Comparison of the hit-rate performances of selected scenes of three summarization schemes for three test movies.

summaries for three movies, "The Hurt Locker," "Lost in Translation," and "Evan Almighty," that the professional was familiar with. The manually edited movie summaries are then used as the references for calculating two quality metrics. The first is the hit-rate of selected scenes, which is defined as the ratio of the total length of the scenes that are selected both in an automatically generated summary $\mathrm{MS}_{\mathrm{auto}}$ and in its corresponding reference summary $\mathrm{MS}_{\mathrm{ref}}$ edited manually to the total length of the reference summary itself

$$R_{\mathrm{hit}} = \frac{|\mathrm{MS}_{\mathrm{auto}} \cap \mathrm{MS}_{\mathrm{ref}}|}{|\mathrm{MS}_{\mathrm{ref}}|} \tag{11}$$

where $|\cdot|$ represents the time length.

Note, the higher the hit-rate, the more consistent the scenes selected in a movie summary to its reference summary. The comparison result is shown in Fig. 7. Although a manually edited summary is subject to the editing person's own preference, it is still a good reference for qualitatively evaluating a movie summary.

Besides the hit-rate, based on the reference summaries, we further compare the similarity between the distribution of RCs selected in an automatically edited summary and that of the corresponding reference summary. Assume the RC distribution of the role communities selected in an automatically edited summary is $p_{\mathrm{MS}} = \{p_{\mathrm{RC\_1}}, p_{\mathrm{RC\_2}}, \ldots, p_{\mathrm{RC\_}N}\}$, and that in the reference summary is $p_{\mathrm{ref}} = \{p^{\mathrm{ref}}_{\mathrm{RC\_1}}, p^{\mathrm{ref}}_{\mathrm{RC\_2}}, \ldots, p^{\mathrm{ref}}_{\mathrm{RC\_}N}\}$, the RC distribution similarity is defined as the histogram intersection of the two distributions [34] as

$$\mathrm{RC}_{\mathrm{sim}}(p_{\mathrm{MS}}, p_{\mathrm{ref}}) = \sum_{k=1}^{N} \min(p_{\mathrm{RC\_}k}, p^{\mathrm{ref}}_{\mathrm{RC\_}k}). \tag{12}$$

In (12), the higher the similarity metric, the more consistent the selected RCs between the automatically generated summary and the reference. Table VI compares the proposed method with the role-based method [12] in terms of the RC distribution similarity, evidently showing that the proposed method generates significantly more consistent role community selections with that of the reference summaries, compared to the role-based scheme.

TABLE VI
COMPARISON OF THE RC DISTRIBUTION SIMILARITY PERFORMANCES OF THE PROPOSED METHOD AND THE ROLE-BASED METHOD [12]

Movie / Method	*The Hurt Locker*	*Lost in Translation*	*Evan Almighty*
The proposed method	37.73%	47.09%	33.95%
Role-based method	26.16%	21.15%	2.77%

Our method is implemented in MATLAB 2010a and is executed on a personal computer with Intel Core i5-2430 M 2.4 GHz CPU and 6 GB memory. When excluding the time for role identification and scene boundary detection, the complexity of the analysis stage mainly depends on the number of groups in the RCN and the number of redundant scenes in each group. In contrast, the complexity of the summarization stage depends on the number of feasible combinations. On average, the analysis stage takes about 14 min to construct an RCN, obtain the centrality values of RCs and cluster RCs, whereas the summarization stage only consumes less than 0.3 s to generate a movie summary in all tests.

B. Limitations

The proposed method also has its limitations. Although our method is ranked the best for most test movies, there are still few subjects who disliked the summaries obtained by the proposed method. They argued that the summaries obtained by our method ignore some exciting scenes, e.g., martial arts clips, dialog scenes of main actors/actresses, or inner feeling of main actors/actresses. Essentially, it is a tradeoff between including more exciting scenes and including more role interactions into a summary. One possible solution is to provide an additional option in the user preference selection to allow replacing some less important role-interaction scenes in a summary with exciting scenes. Besides, as our method uses scene as the basic unit for movie summarization, should be the selected important scenes contain long-duration scenes, a few important scenes may be crowded out from a summary due to its length constraint. For some rare extreme cases, the scene-based summary selection of our method could lead to inaccurate length control, while selecting most essential URCs into the summary (e.g., in Table IV, for movie "Ratatouille" the length of the summary generated by our method is longer than the others). This problem can be solved by refining the granularity of summarization from scenes to shots.

Besides, as the granularity of scene-level summarization is not fine enough, a summary of a too short length may hurt the perceptual comprehension of the summary to a viewer, thereby limiting the length of the generated summary. Nevertheless, the problem can be addressed by combining our method with a shot-level and/or a keyframe-level summarization scheme to generate a very short movie summary (e.g., a 1-min to 2-min summary). For example, one can use a multilevel scheme which first applies our scene-level method to remove unimportant or redundant scenes to generate 15-min to 30-min scene-level summary for a movie, followed by a shot selection and/or keyframe selection method [33] to further shorten the summary to the desired length. As a result, with the multilevel approach, a semantically important scene can be represented with much more shots or frames, compared to shot/keyframe-based approaches. This not only preserve main plots of a movie in a compact way, but also avoid the representations of scenes with too few shots/frames caused by only applying keyframe-based or skimming-based approaches.

VI. Conclusion

In this paper, we proposed a novel two-stage scene-based movie summarization framework based on role-community networks to eliminate semantically redundant scenes in a movie. In the analysis stage, the proposed method constructed a social network to characterize the relationships between the role-communities in a movie. Based on the role-community network, the centrality values of role communities were evaluated and then used to cluster role communities into relevant groups. In the summarization stage, our method formulated the movie summarization as a social network pruning problem, where a set of feasible summary combinations is identified and an information-rich summary is finally selected from these summary candidates. In the selection of movie summary, three types of user preference were offered as an optional feature to further improve the subjective quality of the generated movie summary. Our subjective evaluation results demonstrated the proposed method outperforms the attention-based and role-based methods in most test cases. The feedback comments of subjects also indicated that the proposed method not only preserve the highlights of a movie, but also keeps the evolution of role relationship in movies.

References

[1] A. G. Money and H. H. Agius, "Video summarization: A conceptual framework and survey of the state of the art," *J. Vision Commun. Image Represent.*, vol. 19, no. 2, pp. 121–143, Apr. 2008.

[2] Y. Li, S.-H. Lee, C.-H. Yeh, and C.-C. Jay Kuo, "Techniques for movie content analysis and skimming," *IEEE Signal Process. Mag.*, vol. 23, no. 2, pp. 79–89, Mar. 2006.

[3] L. Herranz and J. M. Martińez, "A framework for scalable summarization of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 9, pp. 1265–1270, Sep. 2010.

[4] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Trans. Image Process.*, vol. 12, no. 7, pp. 796–807, Jul. 2003.

[5] B.-W. Chen, J.-C. Wang, and J.-F. Wang, "A novel video summarization based on mining the story-structure and semantic relations among concept entities," *IEEE Trans. Multimedia*, vol. 11, no. 2, pp. 295–312, Feb. 2009.

[6] C.-W. Ngo, Y.-F. Ma, and H.-J. Zhang, "Video summarization and scene detection by graph modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 2, pp. 296–305, Feb. 2005.

[7] A. Hanjalic and L.-Q. Xu "Affective video content representation and modeling," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 143–154, Feb. 2005.

[8] G. Irie, T. Satou, A. Kojima, T. Yamasaki, and K. Aizawa, "Affective audio-visual words and latent topic driving model for realizing movie affective scene classification," *IEEE Trans. Multimedia*, vol. 12, no. 6, pp. 523–535, Oct. 2010.

[9] G. M. Arthur and A. Harry, "ELVIS: Entertainment-led video summaries," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 6, no. 3, pp. 17:1–17:30, Aug. 2010.

[10] K.-Y. Cheng, S.-J. Luo, B.-Y. Chen, and H.-H. Chu, "SmartPlayer: User-centric video fast-forwarding," in *Proc. ACM SIGCHI Conf. Human Factors Comput. Syst.*, Boston, MA, USA, Apr. 2009, pp. 789–798.

[11] W.-T. Peng, W.-T. Chu, C.-H. Chang, C.-N. Chou, W.-J. Huang, W.-Y. Chang, and Y.-P. Hung, "Editing by viewing: Automatic home video summarization by viewing behavior analysis," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 539–550, Jun. 2011.

[12] C.-Y. Weng, W.-T. Chu, and J.-L. Wu, "RoleNet: Movie analysis from the perspective of social networks," *IEEE Trans. Multimedia*, vol. 11, no. 2, pp. 256–271, Feb. 2009.

[13] M.-C. Yeh, M.-C. Tseng, and W.-P. Wu, "Automatic social network construction from movies using film-editing cues," in *Proc. Int. Workshop Social Multimedia Comput. (in Conjunction With IEEE ICME)*, Jul. 2012, pp. 242–247.

[14] H. Salamin, S. Favre, and A. Vinciarelli, "Automatic role recognition in multiparty recordings: Using social affiliation networks for feature extraction," *IEEE Trans. Multimedia*, vol. 11, no. 7, pp. 1373–1380, Nov. 2009.

[15] J.-T. Sang and C.-S. Xu, "Character-based movie summarization," in *Proc. ACM Multimedia*, Firenze, Italy, Oct. 2010, pp. 855–858.

[16] H. Suber, *The Power of Film*. Studio City, CA, USA: Michael Wiese Productions, 2006.

[17] S. Sharff, *The Elements of Cinema: Toward a Theory of Cinesthetic Impact*. New York, NY, USA: Columbia Univ. Press, 1982.

[18] D. Bordwell and K. Thompson, *Film Art: An Introduction*. New York, NY, USA: McGraw-Hill, 1993.

[19] O. Arandjelovic and A. Zisserman, "Automatic face recognition for film character retrieval in feature-length films," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit.*, San Diego, CA, USA, Jan. 2005, pp. 860–867.

[20] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.

[21] Y.-F. Zhang, C. Xu, H. Lu, and Y.-M. Huang, "Character identification in feature-length films using global face-name matching," *IEEE Trans. Multimedia*, vol. 11, no. 7, pp. 1276–1288, Nov. 2009.

[22] Z. Rasheed and M. Shah, "Detection and representation of scenes in videos," *IEEE Trans. Multimedia*, vol. 7, no. 6, pp. 1097–1105, Dec. 2005.

[23] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comput. Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[24] H. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 23–38, Jan. 1998.

[25] T. Ojala, M. Pietikä inen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[26] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. Am. Soc. Inform. Sci.*, vol. 41, no. 6, pp. 391–407, Sep. 1990.

[27] M. Ellouze, N. Boujemaa, and A. M. Alimi, "IM(S)^2: Interactive movie summarization system," *J. Vision Commun. Image Represent.*, vol. 21, no. 4, pp. 283–294, Jan. 2010.

[28] C. M. Taskiran, "Evaluation of automatic video summarization systems," in *Proc. SPIE Conf. Multimedia Content Anal. Manage. Retrieval*, vol. 6073, Jan. 2006, pp. 178–187.

[29] N. Shroff, P. Turaga, and R. Chellappa, "Video précis: Highlighting diverse aspects of videos," *IEEE Trans. Multimedia*, vol. 12, no. 8, pp. 853–868, Dec. 2010.

[30] Y. Fu, Y. Guo, Y. Zhu, F. Liu, C. Song, and Z.-H. Zhou, "Multi-view video summarization," *IEEE Trans. Multimedia*, vol. 12, no. 7, pp. 717–729, Nov. 2010.

[31] L.-X. Tang, T. Mei, and X.-S. Hua, "Near-lossless video summarization," in *Proc. ACM Multimedia*, Beijing, China, Oct. 2009, pp. 351–360.

[32] The Internet Movie Database (IMDb) [Online]. Available: http://www.imdb.com
[33] Y.-H. Ho, W.-R. Chen, and C.-W. Lin, "A rate-constrained key-frame extraction scheme for channel-aware video streaming," in *Proc. IEEE Int. Conf. Image Process*., Singapore, Oct. 2004, pp. 613–616.
[34] M. J. Swain and D. M. Ballard, "Colour indexing," *Int. J. Comput. Vision*, vol. 7, no. 1, pp. 11–32, Nov. 1991.

Chia-Ming Tsai (S'09) received the B.S. degree from the Feng Chia University, Taichung, Taiwan, in 2003, and the M.S. degree from the National Chung Cheng University, Chiayi, Taiwan, in 2005, both in computer science and information engineering. Since 2005, he is pursuing the Ph.D. degree in the Department of Computer Science and Information Engineering, National Chung Cheng University.

From August 2012 to March 2013, he was a Software Engineer with the CyberLink Inc., Taipei, Taiwan. His research interests include video coding and video content adaptation.

Li-Wei Kang (S'05–M'06) received the B.S., M.S., and Ph.D. degrees in computer science from National Chung Cheng University, Chiayi, Taiwan, in 1997, 1999, and 2005, respectively.

From 2005 to 2010, he was a Post-Doctoral Research Fellow, and from 2010 to 2013, he was an Assistant Research Scholar with the Institute of Information Science, Academia Sinica, Taipei, Taiwan. Since February 2013, he has been an Assistant Professor with the Graduate School of Engineering Science and Technology—Doctoral Program, and the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan. His research interests include multimedia content analysis and multimedia communications.

He served as an Editorial Advisory Board member for the book *Visual Information Processing in Wireless Sensor Networks: Technology, Trends and Applications*, IGI Global, 2011, a Guest Editor of a special issue on Advance in Multimedia, *Journal of Computers*, Taiwan, a special session Co-Chair of APSIPA ASC 2012, a registration Co-Chair of APSIPA ASC 2013, and a Co-Organizer of special sessions of VCIP 2011–2012, and APSIPA ASC 2011–2013. He won four paper awards presented at Computer Vision, Graphics, and Image Processing Conferences, Image Processing and Pattern Recognition Society, Taiwan in 2006–2007 and 2012, respectively.

Chia-Wen Lin (S'94–M'00–SM'04) received the Ph.D. degree in electrical engineering from the National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2000.

Prior to joining academia, he was with Information and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu, from 1992–2000. He was with the Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan, from 2000–2007. He is currently an Associate Professor with the Department of Electrical Engineering and the Institute of Communications Engineering, NTHU, Hsinchu, Taiwan. His research interests include video content analysis and video networking.

He is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE Multimedia, and the *Journal of Visual Communication and Image Representation*. He is also an Area Editor of *EURASIP Signal Processing: Image Communication*. Since September 2013, he has been a Chair of the Multimedia Systems and Applications Technical Committee. He served as Technical Program Co-Chair of the IEEE International Conference On Multimedia & Expo (ICME) in 2010 and Special Session Co-Chair of the IEEE ICME in 2009. He was a recipient of the 2001 Ph.D. Thesis Awards presented by the Ministry of Education, Taiwan. His paper won the Young Investigator Award presented by VCIP 2005. He received the Young Faculty Awards presented by CCU in 2005 and the Young Investigator Awards presented by National Science Council, Taiwan, in 2006.

Weisi Lin (M'92–SM'98) received the B.Sc. degree in electronics and the M.Sc. degree in digital signal processing from Zhongshan University, Guangzhou, China, and the Ph.D. degree in computer vision from King's College, London University, London, U.K.

He has been the Project Leader of over ten major successfully delivered projects in digital multimedia technology development. He also served as the Laboratory Head, Visual Processing, and the Acting Department Manager, Media Processing, for the Institute for Infocomm Research. Currently, he is an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, perceptual modeling, video compression, multimedia communication, and computer vision. He has published over 190 refereed papers in international journals and conferences.

Dr. Lin is a Chartered Engineer, U.K., a fellow of the Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He organized special sessions in ICME 2006, ICME 2012, the IEEE International Workshop on Multimedia Analysis and Processing in 2007, IEEE ISCAS 2010, PCM 2009, VCIP 2010, APSIPA ASC 2011, and MobiMedia 2011. He gave invited, keynote, and panelist talks in International Workshop on Video Processing and Quality Metrics in 2006, IEEE ICCCN 2007, VCIP 2010, and the IEEE Multimedia Communication Technical Committee (MMTC) Interest Group of Quality of Experience for Multimedia Communications in 2011, and tutorials in PCM 2007, PCM 2009, IEEE ISCAS 2008, IEEE ICME 2009, APSIPA ASC 2010, and IEEE ICIP 2010. He is currently on the editorial boards of the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, and the *Journal of Visual Communication and Image Representation*, and four IEEE Technical Committees. He is the Co-Chair of the IEEE MMTC Special Interest Group on Quality of Experience. He has been elected as a Distinguished Lecturer of APSIPA in 2012.