Character-Based Movie Summarization

Jitao Sang[1,2], Changsheng Xu[1,2]
[1]National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China
[2]China-Singapore Institute of Digital Media, Singapore, 119615, Singapore
{jtsang, csxu}@nlpr.ia.ac.cn

ABSTRACT

A decent movie summary is helpful for movie producer to promote the movie as well as audience to capture the theme of the movie before watching the whole movie. Most exiting automatic movie summarization approaches heavily rely on video content only, which may not deliver ideal result due to the semantic gap between computer calculated low-level features and human used high-level understanding. In this paper, we incorporate script into movie analysis and propose a novel character-based movie summarization approach, which is validated by modern film theory that what actually catches audiences' attention is the character [6]. We first segment scenes in the movie by analysis and alignment of script and movie. Then we conduct substory discovery and content attention analysis based on the scene analysis and character interaction features. Given obtained movie structure and content attention value, we calculate movie attraction scores at both shot and scene levels and adopt this as criterion to generate movie summary. The promising experimental results demonstrate that character analysis is effective for movie summarization and movie content understanding.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*abstract method*

General Terms

Algorithm, Performance, Experimentation

Keywords

Movie Summarization, Character Analysis

1. INTRODUCTION

The proliferation of movie content requires efficient and effective techniques for data organization and management. Summarization is one of such key techniques to obtain a brief and to the point representation of the voluminous video data. A decent movie summary is helpful for the producer to promote the movie as well as audience to capture the theme of the movie before watching the whole content. The aim of movie summarization is to select portions that most attract audiences' attention from the original movie. Nevertheless, defining which movie segments is attractive and how to efficiently integrate them into summary is subjective and still remains an open issue and deserves further study. Most recent work on movie summarization heavily relies on video content only. The summary is generated by extracting low-level or mid-level audiovisual features either to identify the key-frame (the salient content) [5, 1] or to explore the movie structure [7, 3]. However, due to the semantic gap between low-level features and high-level understanding, computer calculated audiovisual features cannot characterize attractive movie content at the semantic and affective level.

According to modern film theory, "All films are about nothing - nothing but character" [6], which reveals characters are important for movie summarization. From audiences' perspective, movie is attractive and catches their attention because they want to know about the story of characters. The occurrence and interaction of characters provide meaningful representation of the movie structure and content. Motivated by this, we investigate an alternative way to movie summarization based on character analysis. We first utilize character relation to exploit the movie structure including scene segmentation and substory discovery. Then content attention is evaluated according to three specifically designed character interaction features. Given obtained movie structure and content attention description, we calculate movie attraction scores at both shot and scene levels and use this as criterion to generate movie summary. Compared with low-level or mid-level audiovisual features, character features exhibit high-level meaning and thus can be considered as a more natural representation of movie semantics.

In this paper, our main contribution is to incorporate character analysis into semantic movie summarization. We propose several novel approaches for character-based movie summarization, including 1) A scene segmentation method using analysis and alignment of character co-occurrence in movie and script. 2) A character-based story flow graph to conduct substory discovery. 3) A summarization strategy enabling both informativeness and enjoyability of the generated summary.

The rest paper is organized as follows. In Section 2, we present how to formulate movie structure through character based scene segmentation and substory discovery. Section 3 describes content attention analysis using character interaction features. Movie attraction evaluation and summarization strategy are proposed in Section 4. Experimental results and evaluation are reported in Section 5. We conclude the paper with future work in Section 6.

2. MOVIE STRUCTURE FORMULATION

A typical movie can be structured in the form of substory-scene-

MM'10, October 25–29, 2010, Firenze, Italy.
Copyright 2010 ACM 978-1-60558-933-6/10/10 ...$10.00.

shot. Scene is a set of several shots and substory is built from continuous scenes, which embodies the story flow information [8]. Compared with shot generated during camera shooting process, scene and substory are higher level semantic concepts characterized by composition and interaction of movie characters. Since most movies are related to a story, discovering this hierarchical structure can effectively facilitate movie content understanding and analysis.

2.1 Scene Segmentation

Scene is the elemental unit to constitute a substory in the movie. Accurate scene segmentation not only facilitates movie content understanding, but also affects substory detection. Content-based methods [10] segment movie scenes according to low-level audio/visual features, which lack of necessary semantic guidance and are usually inconsistent with human comprehension. To obtain semantic scenes, we incorporate script into movie segmentation.

Before scene segmentation, each movie shot is first represented as a bag-of-characters. We perform character identification [11] to construct character-histogram representation and identify the leading characters. Script is a textual description of movie content that records complete scene structure and related character names, which can be regarded as an objective and accurate external reference. By inspecting the role composition correspondence between script and movie, semantic scene structure can be effectively mapped from script to movie. Specifically, after character identification, the movie is converted into a shot sequence $V = \{v_1, v_2, ..., v_m\}$ (illustrated in Fig. 1(a)) and similarly the script is described as a scene sequence $D = \{d_1, d_2, ..., d_n\}$, where v_i and d_j is character histogram vector for shot i and scene j, respectively.

We formulate the movie/script alignment problem to assigning each shot to a specified script description, which corresponds to finding the optimal assignment sequence $S^* = \{s_1, s_2, \cdots, s_m\}$ that maximizes the a-posteriori probability:

$$S^* = \arg\max_{S \in \mathcal{S}} p(V|S)p(S) \qquad (1)$$

where $\mathcal{S}$ is the set of all possible assignment sequences. We regard the movie shot sequence V as observation sequence and the assignment sequence S as hidden state sequence. Then Equ. 1 is an observation explanation problem and can be solved by Viterbi algorithm under the HMM framework. More technical details can be referred to our previous work [4].

2.2 Substory Discovery

A movie story can be structured narratively via several large narrative elements, i.e. substories. The underlying story flow is important to movie organization and understanding. As presented in [9], movie story development usually accompanies with significant changes of interaction among characters. Therefore, we employ character histogram as the descriptor and exploit the substory structure by scene transition analysis.

We build a story flow graph (SFG) using similar framework proposed in [7] but with several major differences: 1) We describe video segments by character histogram. 2) the structure discovery is performed on the substory level using interscene relationship rather than scene level using intershot relationship. Fig. 1(b) illustrates the proposed substory discovery method. Given movie scenes described by character histograms, scene similarity graph, with scenes as nodes and their similarities as edges, can be constructed. Scene similarity is computed as:

$$Sim(d_i, d_j) = \frac{1}{D(d_i \| d_j)} \qquad (2)$$

where $D(\cdot\|\cdot)$ denotes the symmetric KL divergence between character histograms. We employ Ncut to decompose the scene similarity graph into subgraphs (scene clusters). Temporal relationship is then considered to generate a SFG, with scene clusters as nodes and transition probabilities $P(C_m|C_n)$ between clusters as edges:

$$P(C_m|C_n) = \frac{1}{|C_n|} \sum_{d_i \in C_m} \sum_{d_j \in C_n} \boldsymbol{G}(i - j) \qquad (3)$$

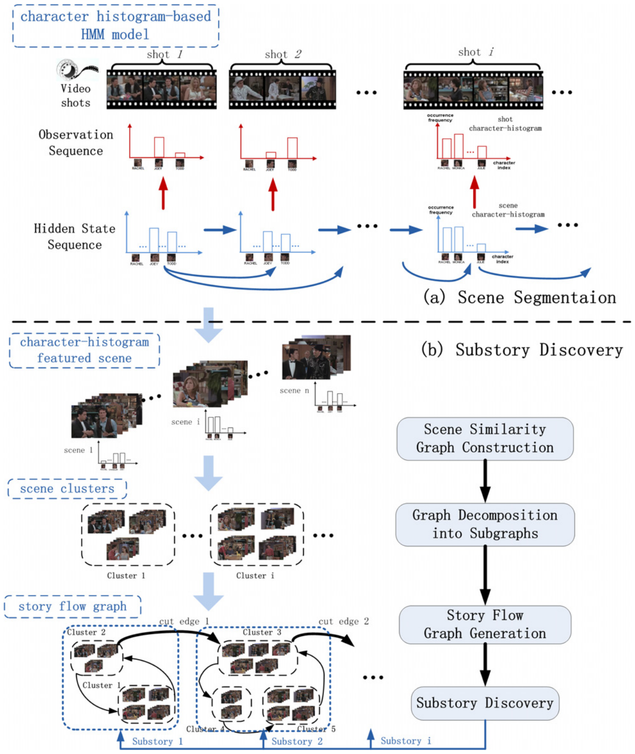

Figure 1: Character-based movie structure formulation: (a) scene segmentation and (b) substory discovery

where $|C_n|$ is number of scenes in cluster C_n, d_i is the i^{th} scene ranked in temporal order, and $\boldsymbol{G}(x) = 1$ if $x = 1$. Substory units are extracted by finding the cut edges of the SFG.

Similar work of [9] also described scene by character histogram and exploited the character interaction for substory discovery. Different from the watershed algorithm in [9], we adopt Ncut for graph decomposition to obtain a global optimum. In addition, in our method the story boundary is directly detected by identifying the cut edge instead of empirically setting a global threshold.

3. CONTENT ATTENTION ANALYSIS

Generally speaking, movie segments having abundant characters involvement (CI), frequent leading character occurrence (LCO), and conflicts between leading characters (LCC) tend to trigger the excitement of the audiences and thus are more attractive. Hence, for content attention analysis, we define three types of character interaction features and compute the content attention value $C\mathcal{A}$ as a linear fusion scheme of these three components:

$$C\mathcal{A} = \lambda_{ci}\mathcal{A}_{ci} + \lambda_{lco}\mathcal{A}_{lco} + \lambda_{lcc}\mathcal{A}_{lcc} \qquad (4)$$

where $\lambda_{ci}, \lambda_{lco}, \lambda_{lcc} \geq 0$ and $\lambda_{ci} + \lambda_{lco} + \lambda_{lcc} = 1$ are normalized combination coefficients, and $\mathcal{A}_{ci}, \mathcal{A}_{lco}, \mathcal{A}_{lcc}$ are number of involved characters, leading character occurrence frequency and number of dialogues between leading characters, respectively. For a video segments Seg_i (either scene or shot), the character related features are defined as

$$\begin{aligned}\mathcal{A}_{ci}(Seg_i) &= \sum_{m=1}^{N_c} I(f_{mi}) \\ \mathcal{A}_{lco}(Seg_i) &= \sum_{m\in\mathcal{L}} f_{mi} \\ \mathcal{A}_{lcc}(Seg_i) &= \sum_{m\in\mathcal{L}}\sum_{n\in\mathcal{L}} Dia(i,m,n)\end{aligned} \tag{5}$$

where N_c is the total number of character involved in the movie, f_{mi} is the occurrence frequency of the m^{th} character in segment i, $I(\cdot) \in \{1,0\}$ is an indication function of a boolean condition, $\mathcal{L}$ is the leading character set, $Dia(i,m,n)$ records the dialogue counts between the m^{th} and n^{th} leading characters in the i^{th} movie segment.

4. MOVIE SUMMARIZATION

Movie summarization can be viewed as a process of selecting video content based on certain criteria. In our method, the criterion is the character-based attraction evaluation for movie segments. The behind philosophy is that attractive video clips are usually essential to understanding the whole movie. We perform the attraction evaluation based on obtained movie structure and content attention description. Besides, the discovered movie structure naturally provides a hierarchical way of shot selection. By utilizing the underlying hierarchical movie structure, we are able to generate readily comprehended summaries and thus improve the audience's visual experience within limited time.

4.1 Movie Attraction Score

Given the identified semantic movie structure and content attention value of each segment, attraction evaluation is performed at scene and shot level, respectively. The structure information and content attention value codetermine the movie attractive degree, i.e. the attraction score. According to the film grammar [8], the beginning and the ending parts of a substory or scene will contain most of its essential information and thus attract more attention. Therefore, we employ the structure information as a weighting term in the attraction score calculation, where the first and last scenes (or shots) in a substory (or scene) deserve higher weights.

At scene level, the attraction score ($\mathcal{AS}$) is defined as

$$\mathcal{AS}(d_i) = \alpha_i \cdot \mathcal{CA}(d_i) \tag{6}$$

where $\mathcal{CA}(d_i)$ is content attention value for scene d_i and computed by Equ.(4) and α_i is structural weight, $\alpha_i = \alpha_s$ ($\alpha_s > 1$ is a parameter emphasizing on the border scenes, we set $\alpha_s = 1.4$ in our experiment) if d_i is the beginning or ending scene of certain substory, otherwise $\alpha_i = 1$.

At shot level, the attraction score for each shot will be further weighted by the belonging scene's attraction score:

$$\mathcal{AS}(v_{ij}) = \mathcal{AS}(d_i) \cdot \beta_{ij} \cdot \mathcal{CA}(v_{ij}) \tag{7}$$

where v_{ij} denotes the j^{th} shot in the i^{th} scene and β_{ij} is also structural weight, $\beta_{ij} = \beta_v$ (we set $\beta_v = 1.3$ in our experiment) if $j = 1$ or $j = |d_i|$ ($|d_i|$ is the number of shots in scene d_i), otherwise $\beta_{ij} = 1$.

4.2 Movie Summarization Strategy

Movie summarization aims to contain as much information as possible in a relatively shorter video clip (i.e. summary) and present to audience in a smooth way [2]. We guarantee the informativeness and smoothness (enjoyability) by additionally considering shot similarity and temporal continuity and then select those refined content as the candidate summary.

Let r be the required skim ratio. We present the movie summarization strategy below. The output of this algorithm is the selection decision of boolean variable t_{ij} for each shot, where $t_{ij} = 1$ denotes shot v_{ij} is selected and $t_{ij} = 0$ discarded. The input includes skim ratio r and attraction score $\mathcal{AS}$ for each shot v_{ij} and scene d_i. According to the hierarchical movie structure, the top-down selection strategy is conducted on three levels.

Substory Level: To guarantee a complete story line instead of discarding low-attention sections as previous work [7], we maintain every substory unit and perform the top-down selection strategy beginning from the substory level. We set the compression ratio proportional to the normalized sum of included scenes' attraction score. The compression ratio of shots is computed as:

$$r_{ss_k} = r \cdot \frac{\sum_{d_i \in ss_k} \mathcal{AS}(d_i)}{\sum_{d_i} \mathcal{AS}(d_i)} \tag{8}$$

Scene Level: For each scene, the ratio of shots remained in a scene is proportional to the normalized scene attraction score. The number of shots contained in generated summary for each scene $d_i \in$ substory ss_k is:

$$|d'_i| = \left\lfloor r_{ss_k} \cdot \frac{\mathcal{AS}(d_i)}{\sum_{d_i \in ss_k} \mathcal{AS}(d_i)} |d_i| \right\rfloor \tag{9}$$

where $\lfloor \cdot \rfloor$ is floor function.

Shot Level: Informativeness and smoothness are major criteria for movie summarization. According to information theory, more dissimilar shots contained in generated movie summary, more informative the summary holds. Smoothness is guaranteed by considering temporal consistence of selected shots. Therefore, at shot level, given computed number of remained shots in each scene, we select the given number of shots which minimize temporal continuity Q_t as well as maximizing appealing score Q_{as} and shot difference Q_d. The decision variable t_{ij} for each shot $v_{ij} \in d_i$ is determined by:

$$t_{ij} = \arg \max_{\sum_{v_{ij}\in d_i} t_{ij} = |d'_i|} \frac{\gamma_{as} Q_{as} + \gamma_d Q_d}{\gamma_t Q_t} \tag{10}$$

where

$$\begin{aligned} Q_{as} &= \sum_{t_{ij}} t_{ij} \cdot \mathcal{AS}(v_{ij}) \\ Q_d &= \sum_j \sum_p t_{ij} t_{ip} \cdot D(v_{ij} \| v_{ip}) \\ Q_t &= \sum_j \sum_p t_{ij} t_{ip} |p - j| \end{aligned}$$

and $\gamma_{as}, \gamma_d, \gamma_t$ are weighting parameters for the shot selection.

Finally, movie summary is obtained by combining those shots v_{ij} whose selection variable $t_{ij} = 1$ in temporal order.

5. EXPERIMENTS

We conducted experiments on one Hollywood movie: You've got mail (YGM) and three TV episodes from the second season of sitcom "Friends"(F202, F205 and F208), respectively. The 4 test movies are segmented into $82, 14, 7$ and 13 scenes. The number of identified leading characters is $2, 4, 3$ and 4. We compared our approach with [5], which proposed a user stimulus-driven attention model built on visual, aural and linguistic attention to estimate movie attraction.

5.1 Attraction Evaluation

To evaluate effectiveness of the proposed character based attraction evaluation, we perform comparison of automatically vs. manually derived movie attraction score curve for 30 minutes of the

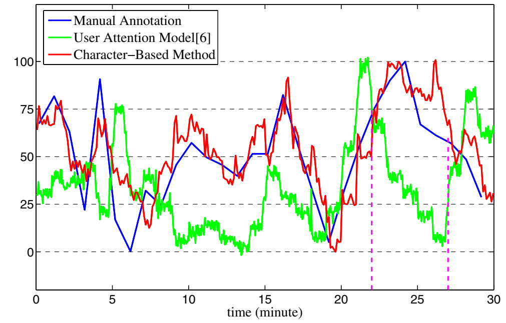

Figure 2: Manual vs. automatic attraction score

Table 1: Performance evaluation of movie summarization

		Informativeness			Enjoyability		
		10%	20%	30%	10%	20%	30%
YGM	Ma [5]	50.2	**67.6**	**78.2**	47.6	65	77.8
	Ours	**55.4**	65	73.5	**57.2**	**68**	**79.2**
F202	Ma [5]	53.8	69	83.6	59.2	70.8	85
	Ours	**65**	**74**	**85.2**	**68.6**	**79.4**	**89**
F205	Ma [5]	60	73.8	**85**	57.8	69	82.7
	Ours	**64.4**	**75**	84.2	**66.4**	**77**	**86.4**
F208	Ma [5]	54.4	69	79.4	52.2	65.8	80.2
	Ours	**59**	**71.2**	**82.4**	**63**	**73.4**	**86.6**

movie YGM in Fig. 2. Ground truth of the attraction score is the average of two graduate students' manual annotation according to the attention the movie attracts (one score per minute).

We present the attraction curve based on the user attention model [5] for comparison. All attraction scores are normalized to $0 - 100$. As shown in Fig. 2, our character-based attraction curve is more consistent with the manual annotation, especially in 22 - 27 minutes when the hero and heroine meet at the first time. We can see that audiovisual feature-based attention model fails to capture the attention triggered by the character conflicts and story development.

5.2 Movie Summarization

Since currently there is no objective ground truth to evaluate summarization result, we conduct subjective evaluation. Five graduate students are invited to participate in user study. They are familiar with all 4 test videos. Generated summaries with skim ratio $c = 10\%, 20\%, 30\%$ are presented and the subjects evaluated the informativeness and enjoyability in a scale from $0 - 100$ where 100 denotes the quality of original video.

Table 1 compares our movie summarization method with [5] in 4 test movies. The result is the average evaluation value of five subjects. Note that shot is the basic selection unit in our summarization framework, hence the skim ratio in our method is actually the shot compression ratio. As most shots are basically the same length in our dataset, the summary duration generated by [5] and our method under the same skim ratio are almost the same. In Table. 1, as skim ratio decreases, our character-based movie summarization remains acceptable performance because the audiences' interested character conflict are preserved. We achieved better enjoyability than [5] for two reasons: 1) the character-based movie structure analysis is incorporated in the top-down selection strategy and the story completeness is guaranteed, 2) we considered temporal continuity in summarization strategy to improve smoothness and the minimum unit for selection is shot instead of key-frame. Besides, we notice that the informativeness in complex character-configuration movies (movie YGM in the first line of Table 1) is limited by the performance of character identification, which reveals that effective character representation serves as the bases for our method.

6. CONCLUSION

We have presented a novel movie summarization framework by analyzing role composition and interaction. The main module include movie structure formulation, content attention analysis and a character-based summarization strategy. Subjective evaluation has demonstrated that character features embodies more informative cues for movie content understanding and character analysis provides an alternative way for semantic movie summarization. In the future, we will be working towards extracting more external event information from movie scripts (e.g. action descriptors) as well as investigating optimal feature configuration for different movie genres.

7. ACKOWNLEDGEMENT

This work was partially supported by the National Natural Science Foundation of China (60833006, 90920303), and 973 Program(2010CB327905)

8. REFERENCES

[1] G. Evangelopoulos, K. Rapantzikos, A. Potamianos, P. Maragos, A. Zlatintsi, and Y. Avrithis. Movie summarization based on audiovisual saliency detection. In *ICIP 2008*, pages 2528–2531, October 2008.

[2] Y. Li and C.-C. J. Kuo. *Video Content Analysis Using Multimodal in Formation: For Movie Content Extraction, Indexing and Representation*. Kluwer, Norwell, 2003.

[3] Y. Li, S.-H. Lee, C.-H. Yeh, and C.-C. J. Kuo. Techniques for movie content analysis and skimming. *IEEE Signal Processing Magazine*, 23:79–89, March 2006.

[4] C. Liang, Y. Zhang, J. Cheng, C. Xu, and H. Lu. A novel role-based movie scene segmentation method. In *PCM 2009*, pages 917–922, March 2009.

[5] Y.-F. Ma, X.-S. Hua, L. Lu, and H.-J. Zhang. A generic framework of user attention model and its application in video sumarization. *IEEE Trans. Multimedia*, 7(5):907–919, October 2005.

[6] J. Monaco. *How to Read a Film: The Art, Technology, Language, History and Theory of Film and Media*. Oxford Univ. Press, New York, 1982.

[7] C.-W. Ngo, Y.-F. Ma, and H.-J. Zhang. Video summarization and scene detection by graph modeling. *IEEE Trans. Circuits Syst. Video Technol*, 15(2):296ÍC305, February 2005.

[8] S. Sharff. *The Elements of Cinema: Towards a Theory of Cinesthetic Impact*. Columbia University Press, 1982.

[9] C. Weng, W. Chu, and J. Wu. Rolenet: Movie analysis from the perspective of social networks. *IEEE Transactions on Multimedia*, 11(2):256ÍC271, February 2009.

[10] Y. Zhai and M. Shah. Video scene segmentation using markov chain monte carlo. *IEEE Trans. on Multimedia*, 8(4):686 – 697, July 2006.

[11] 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*, 11(7):1276–1288, November 2009.