Video Data Mining: Rhythms in a Movie

Kimiaki Shirahama, Kazuhisa Iwamoto, and Kuniaki Uehera
Graduate School of Science and Technology, Kobe University,
Nada, Kobe, 657-8501, Japan
{*kimi, kazuhisa, uehara*}*@ai.cs.scitec.kobe-u.ac.jp*

Abstract

The task to discover useful editing patterns from a professional video, such as a movie, is one of the main purpose of video data mining. These patterns successfully convey editor's intentions to the viewers. But, data mining on multimedia data like a movie is a challenging task, due to the complicated contents of multimedia data. Particularly, the discovered patterns need to be supported by their semantic features, because these features tell the amateur editors how to use the corresponding patterns during the process of editing a new video.

In this paper, we focus on the rhythm in a movie, consisting of the durations of target character's appearance and disappearance. Based on this rhythm, we divide the movie into topics. Each topic corresponds to one meaningful episode of the character. By investigating such topics, we can discover useful editing patterns of character's rhythm, supported by their semantic features. Also, these rhythms can be used to annotate certain types of topics.

1. Introduction

Video editing is the process of selecting and joining various shots to create a final video sequence. A shot is defined as one exclusive stream of video frames recorded by a single camera. In video editing, there can be numerous ways to make the transition from one shot to another so that the quality of the created video depends on the editor's skill. Professional video editors carry out video editing based on their own editing patterns. So, discovering the useful editing patterns from a professional video, called 'video data mining', is beneficial for amateur editors to produce a new, more attractive video.

In [1], we extracted the syntactical editing patterns from a movie which is indexed by the metadata representing the aspects in every shot. An example of the extracted syntactical patterns is "a long distance shot with a duration of about 6 seconds" followed by "a middle distance shot about 4 sec-

onds". Based on the discovered patterns, our video editing support system [2] helps amateur editors editing videos. But, we could not realize in what kind of context this pattern may be used during the process of editing a new video. That is, it is essential to extract the useful editing patterns supported by their semantic features.

Multimedia streams can be classified into two types, 'discrete' and 'continuous'. A discrete stream (e.g. e-mail stream in Fig. 1 (a)) consists of exclusive elements, where each element makes sense by itself and emerges at a single time point. The semantic information is delivered as a discrete function of time. Kleinberg [3] proposed an efficient method to analyze a discrete stream based on the rhythm of a focused subject. In the e-mail stream in Fig. 1 (a), the rhythm is a sequence of gaps between consecutive message arrivals relevant to a focused subject. Thus, a gap represents a single type of interval where no relevant semantic information is provided.
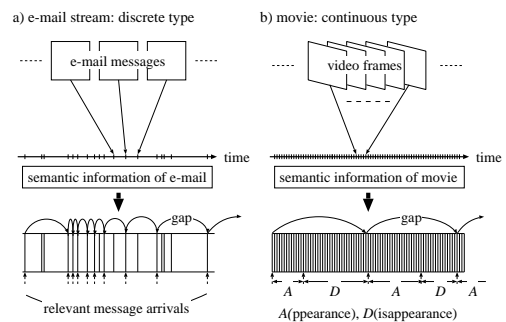

Figure 1. Examples of discrete and continuous streams

On the other hand, a continuous stream (e.g. the movie in Fig. 1 (b)) consists of media quanta. It conveys its semantic information only when media quanta are continuously presented in time. That is, the semantic information is delivered as a continuous function of time. In the movie in Fig. 1 (b), there are two types of intervals: one is an interval

Figure 2. Several shots from Alfred Hitchcock's movie PSYCHO

where he/she appears, the other is an interval where he/she disappears. That is, it is too coarse to analyze a continuous stream based on the rhythm consisting of a single type of intervals, like a discrete stream. We will introduce two types of semantically different intervals (i.e. appearance and disappearance).

In this paper, we propose a more precise rhythm of a focused subject in a continuous stream. It consists of two types of intervals, corresponding to the appearance and disappearance of the focused subject. In particular, we define a rhythm in a movie as a sequence of durations of target character's appearance and disappearance. Based on the target character's rhythm, we divide a movie into 'topics', each one of them presents one meaningful episode of the character. By investigating such topics, we can discover the useful editing patterns of character's rhythm supported by their semantic features, such as conversation, thrilling, and romantic topics.

2. Basic Concept

A viewer comprehends the semantic information of a movie through the 'action flows' of the characters who appear in this movie. Character's action flow conveys to the viewer what 'action' he/she performs in his/her appearance, and what 'surrounding response' occurs in his/her disappearance. Fig. 2 represents several shots from Alfred Hitchcock's movie "PSYCHO", where there are two characters, *Marion* and *Norman*. We describe *Marion*'s action flow by using her appearance (*A*) and disappearance (*D*) in the bottom row of the table. In *shot 207*, *Marion* drives her car as the action in her appearance. In *shot 208*, looming light is to be seen outside *Marion*'s car as the surrounding response in her disappearance. Note that, a surrounding response means the response by other characters (*Norman* at the end of *shot 254*), or changes in the surrounding setting (*shot 208*), when *Marion* disappears from the screen.

Each action or surrounding response in character's action flow is consistent with the previous one. For example, *Marion*'s action in *shot 209* (she glances outside while driving her car) is relevant to the 'reaction' to the surrounding response in *shot 208*. *Marion*'s action in *shot 251* (she listens to *Norman*) is relevant to the reaction to her action in *shot 250* (she talks to *Norman*). These consistent connections of *Marion*'s reactions convey to a viewer why she glances

outside her car or why she listens to *Norman*. Furthermore, by the connection of character's actions and the surrounding responses over a longer time period, a viewer perceives one meaningful episode, that is, a topic.

Meanwhile, movie editors determine the duration of a character's action (i.e. his/her appearance) based on his/her action and the context of his/her situation. Similarly, the duration of a surrounding response reflects the other characters' actions and the context of their situations, while the target character disappears from the screen. Consequently, we can assume that, as long as the target character performs the similar actions in a particular situation, where the other characters perform the similar actions as the surrounding responses, target character's rhythm stays roughly constant.

3. Dividing a Movie into Topics

For a target character in a movie, we provide the metadata representing the intervals of his/her appearance and disappearance in every shot. Specifically, an interval of character's appearance is defined as the interval where he/she appears on the screen, by contrast, an interval of his/her disappearance is the interval where he/she is completely off screen. Some metadata indexed for *Marion* are shown in Fig. 2. These metadata can be automatically extracted by existing techniques (e.g.[5]), although they may fail to obtain results as good as expected by human users. Since our goal is to discover the useful editing patterns from a movie, we manually annotate the above metadata to extract reliable editing patterns.

Fig. 3 presents *Marion*'s rhythm from shot *194* to *267*, which is obtained by scanning the durations of her appearance and disappearance one by one. Starting from the left, each circle corresponds to *Marion*'s appearance or disappearance in Fig. 2. A positive value indicates the duration of *Marion*'s appearance, and a negative value indicates the duration of her disappearance. Note that, if *Marion*'s disappearance is divided into some durations, we join them into one total duration. The reason is that, *Marion*'s next action is a reaction to the surrounding responses which occurred in the total duration, taken as a whole.

In Fig. 3, the best example reflecting our topic definition described in Section 2, is *topic 2* which is associated with the semantic feature "*Marion* is driving her car at the rainy night". In this topic, *Marion*'s rhythm is almost constant.
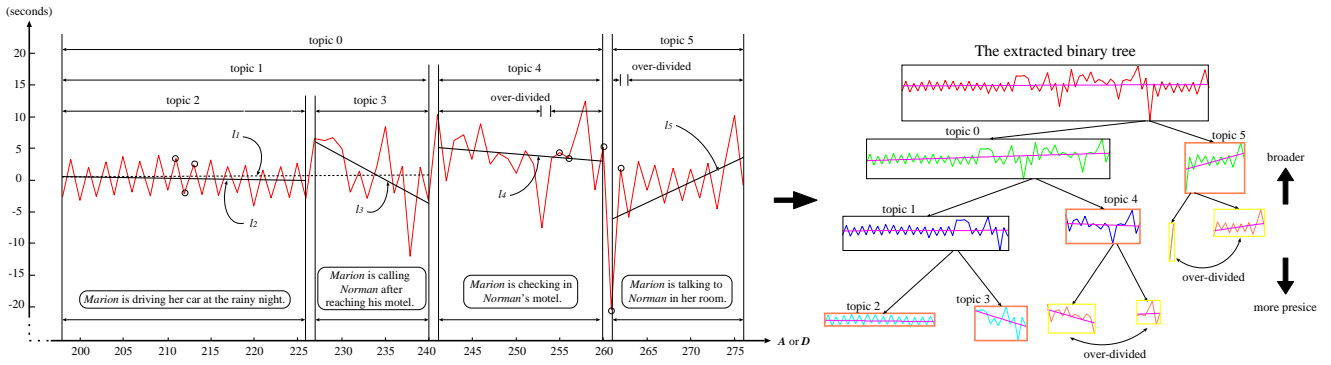
Figure 3. Our rhythmic model of character's actions and surrounding responses

But, character's rhythm is generally inconstant even in a topic, because the topic may include some actions and their surrounding responses. So, we need to extract the topics, where character's rhythm is relatively constant with some degree of approximation error.

For a character's rhythm, we figure out topics by employing time series segmentation method [4]. Our method recursively divides a segment of character's rhythm into two sub-segments in a top-down approach, beginning with the whole range of his/her rhythm. In each recurrence, our method considers every possible division of a segment into two sub-segments, while approximating each sub-segment by the line computed by least-square method. Then, our method finds the optimal pair of sub-segments within the current segment, which minimizes the sum of the approximation errors in the two sub-segments. That is, in each sub-segment of the optimal pair, the correlation between character's rhythm and the approximation line is as strong as possible within the segment.

Subsequently, based on the reduction of the approximation error by the division of a segment into the optimal pair ($error\_reduction$), our method decides whether this segment should be further divided into sub-segments of the optimal pair, according to the following policy: Dividing the segment into the sub-segments having different correlations, that is, character's rhythm is not constant in the segment, causes a $error\_reduction$ larger than the threshold values. By contrast, dividing the segment into sub-segments having common correlations, that is, character's rhythm is constant, causes a $error\_reduction$ smaller than the threshold.

We present the above procedures with *Marion*'s rhythm in Fig. 3. In *topic 1*, (*topic 2*, *topic 3*) is firstly selected as the optimal pair of sub-segments. Then, the large $error\_reduction$ reveals that *topic 2* and *topic 3* have the different correlations of *Marion*'s rhythm. So, *Marion*'s rhythm in *topic 1* is not regarded as constant, and divided

into *topic 2* and *topic 3*. In *topic 2*, *Marion*'s rhythm is almost constant. In such *topic 2*, the sub-segments of the optimal pair inevitably have common correlations, which is indicated by the small $error\_reduction$. So, our method doesn't divide *topic 2* any more. In this way, two *Marion*'s topics, *topic 2* and *topic 3*, are detected.

Finally, a part of PSYCHO is organized into the binary tree consisting of several topics where *Marion*'s rhythm is constant at multiple abstraction levels. And this tree shows *topic 1* is a large topic which includes two sub-topics, *topic 2* and *topic 3*. Using this kind of inclusions, we can define the structure of a movie from the perspective of a certain character.

4. Comparing Topics with Semantic Features

We analyze the movies based on the leading characters' rhythms, *Marion* and *Roger* in Alfred Hitchcock's movies "PSYCHO" and "North by Northwest", and *Qui-Gon* in George Lucas's movie "Star Wars Episode I", respectively. By investigating the binary tree extracted for each target character, we effectively find the meaningful topics associated with their semantic features. The text descriptions in Fig. 3 are annotated by human interpretation. Note that, a meaningful topic may be the parent of over-divided topics (e.g. *topic 4* and *topic 5* in Fig. 3), where the character's rhythm is relatively constant from a broader perspective. Based on the detected topics, we compare the rhythms with their semantic features. As a result, we can discover the following editing patterns of character's rhythm, supported by their semantic features.

[Hitchcock's movies:] Character's movement, including other character's movement as a surrounding response, is associated with its duration. For instance, if a character actively moves, the action is presented in a long duration, on the other hand, a short duration is yielded if the character

hardly moves. This means that, an uneventful action in a long duration bores viewers, in addition, the editors use as long duration as needed to thoroughly present character's eventful action. Based on this association, we discover the following editing patterns, specific to Hitchcock's movies.

1. **In an uneventful topic,** character's rhythm is fast and constant. For example, in a conversation topic, *Marion* is talking to someone while standing still. If *Marion*'s utterance in her appearance is long, in order to keep her rhythm fast and constant, the presentation of her action is split by inserting the listener's action.

2. **In a thrilling topic,** despite the active movements of a character (e.g. *Roger*) and the surrounding elements (e.g. the spray plane from which *Roger* is running away), such active actions are presented in short durations, which makes viewers more tensed.

3. **In a romantic topic,** character's rhythm is very distorted. The primary cause is that, despite the inactive movements of a character and his/her partner (they are mainly hugging and kissing), these actions are presented in very long durations to create a romantic mood.

[Star Wars Episode I:] In this movie, we can not observe the association between character's movement and its duration. The reason is that, the editor omits some minor details of character's eventful action, in order to present what is happening to the several leading characters in Star Wars Episode I.

1. **In a battle topic,** when *Qui-Gon* is fighting, the durations of his fighting actions and the surrounding responses are less than 2 seconds.

2. **In a viewer's topic,** the short durations of *Qui-Gon*'s actions and the long durations of the surrounding responses alternate. That is, the editor mainly presents what the other characters are doing, and sometimes presents *Qui-Gon*'s action in short durations.

5 Conclusions and Future Works

In this paper, we concentrate on discovering useful editing patterns supported by their semantic features, by utilizing character's rhythm in a movie. By looking through the binary tree extracted for a target character, we can effectively find the meaningful topics of this target character. That is, our method enables a user to effectively browse a movie from each character's perspective. But, the binary tree is too complicated for a user to grasp the story's outline of the movie. The reason is that, the binary tree only represents that consecutive topics differ from each other in character's rhythm, it does not represent the relations between consecutive topics. Also, there are several over-divided topics in the binary tree. So, we must carefully identify the meaningful topics in the binary tree regarding the adjacent topics. For this purpose, we will use an automaton, where each state is associated with one rhythm determined by classifying the topics by their least-square approximation lines.

Furthermore, we treat character's appearance and disappearance in one dimension, but these are semantically different informations. So, character's appearance and disappearance should be treated independently by adopting the multiple linear regression method [6]. This extension allows us to analyze a movie based on character's rhythm consisting of several types of intervals, such as the durations of character's active-movement, inactive-movement, and disappearance. Furthermore, this rhythm can be extended to multiple target characters. Then we can explicitly confirm the discovered patterns described in Section 4. For example, a thrilling topic will be characterized by the rhythm consisting of the short durations of one character's active-movement and the short durations of another character's active-movement. Finally, the discovered patterns not only can be applied to our video editing support system [2], but also can be implemented as additional tool for browsing the whole story at a glance.

References

[1] Y. Matsuo, K. Shirahama and K. Uehara, "Video Data Mining: Extracting Cinematic Rules from Movie", Proc. of 4th International Workshop on Multimedia Data Mining (MDM/KDD), pp. 18-27 (2003).

[2] M. Kumano, Y. Ariki, M. Amano, K. Uehara, K.Shunto and K. Tukada, "Video Editing Support System based on Video Grammar and Content Analysis", Proc. of 16th International Conference on Pattern Recognition (ICPR), pp. 1031-1036 (2002).

[3] J. Kleinberg, "Bursty and Hierarchical Structure in Streams", Proc. of 8th ACM SIGMOD International Conference on Knowledge Discovery and Data Mining, pp. 91-101 (2002).

[4] V. Guralnik and J. Srivastava, "Event Dection from Time Series Data", Proc. of 5th ACM SIGMOD International Conference on Knowledge Discovery and Data Mining, pp. 33-42 (1999).

[5] J. Fan and Y. Ji, "Automatic Moving Object Extraction toward Content-Based Video Representation and Indexing", Journal of Visual Communications and Image Representation, Vol. 12, No. 3, pp. 217-239 (2001).

[6] G. Ridgeway, "Strategies and Methods for Prediction", Book of THE HANDBOOK OF DATA MINING, N. Ye (eds), Lawrence Erlbaum Associates, pp. 159-191 (2003).