Action Movies Segmentation and Summarization Based on Tempo Analysis

Hsuan-Wei Chen
Communication and
Multimedia Laboratory,
Dept. of Computer Science
and Information Engineering,
National Taiwan University
Taipei, Taiwan 106
wei@cmlab.csie.ntu.edu.tw

Jin-Hau Kuo
Communication and
Multimedia Laboratory,
Dept. of Computer Science
and Information Engineering,
National Taiwan University
Taipei, Taiwan 106
david@cmlab.csie.ntu.edu.tw

Wei-Ta Chu
Communication and
Multimedia Laboratory,
Dept. of Computer Science
and Information Engineering,
National Taiwan University
Taipei, Taiwan 106
wtchu@cmlab.csie.ntu.edu.tw

Ja-Ling Wu
Communication and
Multimedia Lab., Dept. of CSIE,
Graduate Institute of
Networking and Multimedia,
National Taiwan University
Taipei, Taiwan 106
wjl@cmlab.csie.ntu.edu.tw

ABSTRACT

With the advances of digital video analysis and storage technologies, also the progress of entertainment industry, movie viewers hope to gain more control over what they see. Therefore, tools that enable movie content analysis are important for accessing, retrieving, and browsing information close to a human perceptive and semantic level. We proposed an action movie segmentation and summarization framework based on movie tempo, which represents as the delivery speed of important segments of a movie. In the tempo-based system, we combine techniques of the film domain related knowledge (film grammar), shot change detection, motion activity analysis, and semantic context detection based on audio features to grasp the concept of tempo for story unit extraction, and then build a system for action movies segmentation and summarization. We conduct some experiments on several different action movie sequences, and demonstrate an analysis and comparison according to the satisfactory experimental results.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – indexing methods.

General Terms

Management.

Keywords

tempo analysis, movie content analysis, movie segmentation, movie summarization, story unit, film grammar.

1. INTRODUCTION

Movies are one of the richest heritages in human civilization. Movies alone constitute a large portion of the entertainment industry. Every year around 4,500 movies are released around the world, spanning over approximately 9,000 hours of digital movie content [1]. With the development of content digitalization and storage technologies, there has been a tremendous increase in the access of digital content on the Internet. In addition, movie viewers expect more control in films and online movie-on-demand systems through movie indexing, retrieval, and browsing. Therefore, it is necessary to automatically analyze the films and provide tools for browsing and retrieving movies.

When we consider the analysis of films, it is important to take into account the unique features of films caused by filmmaking. First, movies are often filmed in an open and dynamic environment with moving cameras and have continuously changing contents, which increase difficulty and uncertainty for movie analysis. Second, directors tend to use different camera techniques and special visual effects that make it difficult to create a suitable content management system. Therefore, the concept of film grammar is adopted, which is the generally used standard and basic guideline for filmmaking. Through film grammar, more semantically matched approaches are expected to be proposed for movie analysis.

In our system, we aim to detect and extract the most semantically important story units and segments of an action movie based on movie tempo analysis. During the process of tempo analysis, there are several problems that need to be solved in advance [2]. First, because a movie is a video sequence with frequent shot changes, it is primary to detect shot boundaries (especially those occurred between two different story units) according to temporal and spatial variations. Second, a director controls a tempo of the movie by different camera and filming techniques, particularly by motion activity. Third, the audio effects influence the viewers' perception and emotional feelings while seeing a movie. Hence we also extract audio features for semantic context detection. We then complete the work of movie segmentation and summarization based on movie tempos, which includes the combination of visual and audio information analyses.

This paper is organized as follows. In Section 2, an overview of film grammar is given. Next, we will give an investigation of previous research work related to movie content analysis in Section 3. In Sections 4 and 5, the system framework and two most important applications of the proposed approach are presented, respectively. The experimental results are illustrated in Section 6. Finally, in Section 7, the conclusions and future work are addressed.

2. FILM GRAMMAR

Film grammar contains rules and conventions that elucidate the relationship among elements that are employed by filmmakers to convey meanings [3][4]. According to the guidelines film

MIR'04, October 15–16, 2004, New York, New York, USA.
Copyright 2004 ACM 1-58113-940-3/04/0010...$5.00.

grammar provides, the analysis of movie content from film theory to computable aesthetics elements can be divided into two parts. First, generally speaking, the filming of a movie is constructed by a general scenario, which is presented by several different story units, and each story unit is composed of a sequence of scenes, and a scene is composed of several shots. In other words, we can tell that the filming of a movie is hierarchical, and in every step the filming follows a basic rule: the editing of a movie emphasizes on psychological guidance that leads to the viewer's impression [5].

Second, film grammar introduces one of the most important components when filming a movie, the *montage*. Montage can be considered as the nerve of a movie. It is an idea of film editing, deriving from the concept that there should be contrast between two different shots that are independent of each other, which is also called the dramatic principle [5]. In 1925, Sergei Mikhailovich Eisenstein proposed the concept of montage for the first time, and presented it in the movie "The Battleship Potemkin". The most famous segment in this movie, the Odessa Steps massacre, which is considered a pioneering and typical filmmaking example with montage [5], is achieved by the principle of montage: the juxtaposition of images of innocence against images of violence; the contrast between long and depersonalizing shots of soldiers and close-ups of the baby carriage; the conflicts between shots of panic citizens and the soldiers walking down from the stairs. Most importantly, montage is claimed to be controlled and presented mainly by "shot length" and "patterns". This claim founds the basis of today's content analysis.

3. PREVIOUS WORK

According to the conception of film grammar, the first step toward movie content management is to segment a movie into semantically meaningful units. In the past research, work has been done by segmenting a movie into units probably according to a temporal basis or low level features. In recent years, research work has shifted a bit from low level features processing to a more semantic high level feature analysis. It aims to extract elements from a movie that are more "expressive" and "representative", which in other words, semantically match the human feelings and perception when seeing a movie.

Jeroen Vendirg et al. [6] proposed an algorithm that extracts the logical story units from a movie and provided a systematic evaluation. The Logical Story Unit (LSU) is defined such that it is applied consistently to a large collection of movies. Since humans perceive LSUs by way of changes in content, and LSU can be defined by its boundaries. The proposed method frames the concept of LSU boundary accommodating the alternation of shots at different locales (also known as the problem of parallel cutting) to create the impression that several events taking place at the same time [6].

Reference [7] proposed approaches to do movie annotation and indexing. Effectively labeling the visual content of movies is essential for annotation. The authors presented the interactive and adaptive i-Notation system, which describes actors' names, automatically processes multimodal information sources, and deals with available sources' varying quality. The approach provides the basis for intelligent interaction and demonstrates significant improvements in annotation efficiency. Reference [8] proposed an algorithm for scene detection of Hollywood movies. The authors proposed a novel two-pass algorithm for scene boundary detection, which utilizes the motion information, shot lengths and color properties of shots as features.

Finally, [2] grasped the concept of film grammar and proposed the term *Tempo* (Pace), which indicates the performance and delivery speed of a movie. Movies ebb and flow, and they seem to happen as the filmmakers respectively hold and overwhelm us with information or activity. Therefore, by taking some cues from psychology, the conception of tempo might be formulated as a function of the information delivery rate that thrusts at the viewer.

4. THE PROPOSED APPROACH

4.1 System Framework

We aim to use film grammar to segment and summarize movie sequences based on tempo analysis. From the previous sections one can see that the following three features control the tempo of movies and the pace of representation at a high level: montage editing, motion activity, and audio effects. In other words, we will analyze movie content from the three corresponding aspects: length of shot, intensity of motion activity, and variation of audio features.

Fig. 1 shows the framework of our system for action movies segmentation and summarization. The framework contains typically two stages. The first stage is to compute features for tempo analysis. The second stage is to apply the segmentation and summarization algorithms to the tempo function to obtain movie segments and summaries.

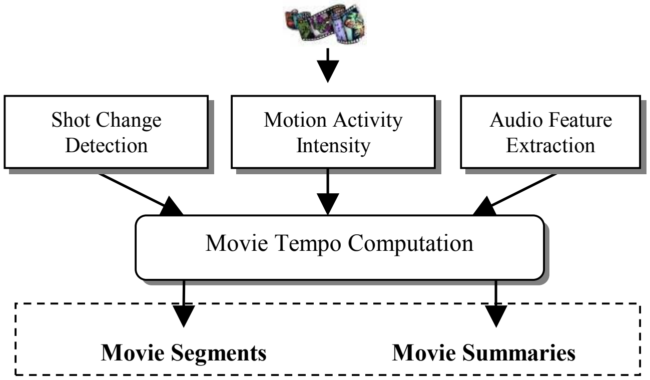

Figure 1. The system framework.

4.2 Shot Change Detection

When dealing with the first step of segmentation of movie sequences, we do not necessarily adopt complicated shot change detection approaches for different and diverse cases. What we require is that the shot boundary between two different stories is clearly detected. Therefore, we use a pixel-based approach for conducting the shot change detection.

Assume X and Y are two frames, and d(X, Y) is the difference between the two frames. $P_X(m, n)$ and $P_Y(m, n)$ represent the values of the (m, n)-th pixels of X and Y, respectively. The following equation shows the approach to do shot change detection:

$$d_{XY}(m,n) = \begin{cases} 1, & if\ \left|P_X(m,n) - P_Y(m,n)\right| > \mathrm{T}_1 \\ 0, & \text{else} \end{cases} \qquad (1)$$

and

$$d(X,Y) = \frac{1}{m*n}\sum_{m}\sum_{n} d_{XY}(m,n)\cdot \qquad (2)$$

If d(X, Y) is larger than a threshold T$_2$, a shot change is detected between frames X and Y.

Shots of length smaller than ten frames are merged, as they are deemed to be false positives. We use this straightforward and simple method for shot change detection in order not to over segment the movie sequence. What we want for shot change detection is to detect the clear shot boundary between two different story units, containing various visual special effects within each unit. Fig. 2 shows an example of the result of applying our shot change detection algorithm to movie: “Crouching Tiger, Hidden Dragon”.

Figure 2. An example of shot change detection for a clip taken from the movie: “*Crouching Tiger, Hidden Dragon*”.

4.3 Motion Activity Analysis

We conduct motion activity analysis using the motion activity descriptor defined in MPEG-7 [9]. The descriptor can be extracted directly from compressed videos and is compact, and hence is easy to extract and match. The intensity of motion activity of an action movie shot is part of the directional indication of the movie tempo. When a human is watching a movie sequence, his emotional feelings and impression are often affected by perceiving a sequence as a slow sequence, a fast paced sequence, an action sequence, or a boring sequence. The MPEG-7 motion activity descriptor captures this intuitive notion of “intensity of action” in a film segment. For instance, a “high speed care chase” or a “gunshot in a shocking war”, etc, are perceived as “high action” scenes by most human viewers. On the other hand, the scenes such as a “head and shoulders”, a “pan of scenery”, etc. are considered as slow action scenes.

The shot segmentation is performed first, and then the motion activity descriptor values are extracted based on the motion vector values of a 16 x 16 macroblock. First, the extraction is done for a frame, and in the second stage an entire shot is considered. In the proposed system, the motion vectors are extracted from the MPEG-1 compressed film sequence. For a given P frame, the “spatial activity matrix” C_{mv} is defined as

$$C_{mv}(i,j) = \sqrt{x_{i,j}^2 + y_{i,j}^2}\ , \qquad (3)$$

where $\left(x_{i,j}, y_{i,j}\right)$ is the motion vector associated with the $\left(i,j\right)$ th macroblock. For intra-coded blocks, $C_{mv}(i,j) = 0$.

The average motion vector magnitude per macro-block of the frame C_{mv}^{avg} is given by:

$$C_{mv}^{avg} = \frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N} C_{mv}\left(i,j\right)\ , \qquad (4)$$

where M and N are the width and height of the macroblocks in the frame.

With compressed videos, motion vectors provide the easiest approach to gross motion characteristics of the movie segments. Since motion vector magnitude is an indication of the magnitude of the motion itself, it is natural to use statistical properties of the motion vector magnitude of macroblocks to measure intensity of motion activity. To find the motion vector based parameter for a shot, we first compute the descriptors of all the P frames in the shot. Then, the average motion vector magnitude per macroblock on motion vectors for the entire shot is obtained. Fig. 3 shows how to extract the shot motion activity intensity.

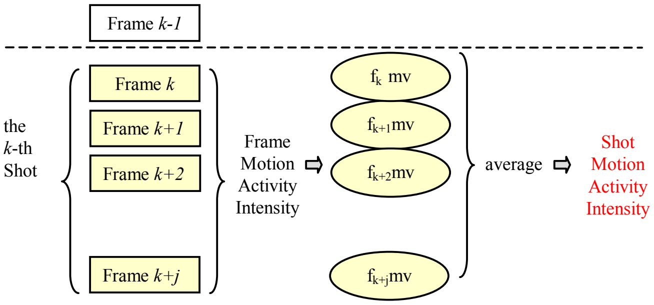

Figure 3. The computation of shot motion activity intensity by using frame motion activity intensity.

4.4 Audio Feature Analysis

In action movies, sound effects play a very important role in performing tempo. Action scenes, particularly, contain mostly a non-repeating shot pattern, non-speech, and large volume of music and audio effects. For typical Hollywood action movies, generally speaking, the larger the volume of audio signal is, the higher the tempo is. We put our emphasis on the scenes that have high tempo values, most of which usually have accompanying large volume of audio sounds. Fig. 4 shows an example of comparison between a low-tempo shot with small volume of audio sounds and a high-tempo shot with large volume of audio sounds.

In our system, the audio energy within every shot is considered. For each shot, the audio energy is computed as the logarithm of the signal energy, and let *N* be the number of audio energy peaks that are greater than a threshold T_a. *N* is further normalized by the number of audio frames within a shot, that is:

$$N' = \frac{N}{\text{Number of Audio Frames Within a Shot}} \qquad (5)$$

Feature *N’* is used to indicate the audio importance score of a shot. Fig. 5 gives a comparison between the average audio energy and the audio energy peaks, *N’*, of an active audio clip. It is shown in the figure that the audio energy peaks amplify the difference between low tempo sound segments and high tempo-high sound segments.

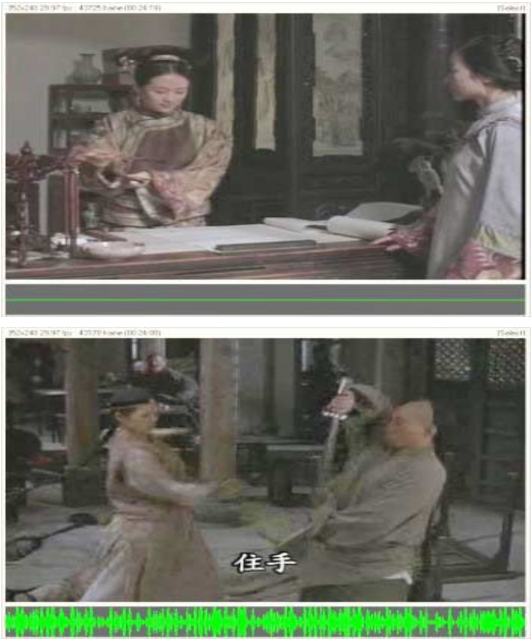

Figure 4. A comparison between a small volume audio scene and a high volume audio scene. The corresponding audio sound segments are also shown in the bottom of the scenes.

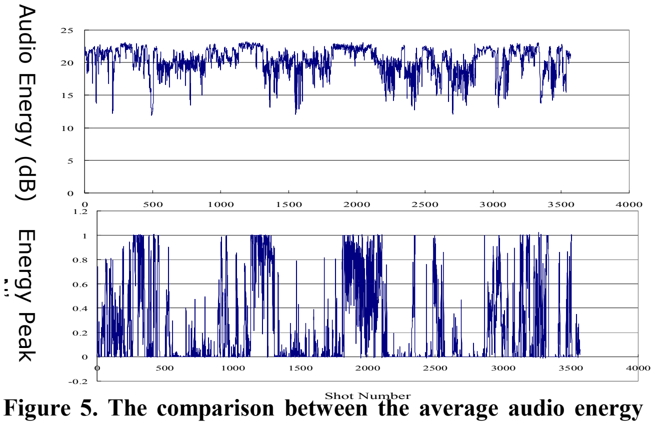

Figure 5. The comparison between the average audio energy and the audio energy peaks, *N'*.

4.5 Tempo Computation

The concept and computation of movie tempo were proposed and analyzed in [2]. We further modify and improve the tempo function by taking into account the audio effect, which plays an important role in action movies analysis. Movie tempo is considered as the rate of performance or delivery of a movie and is constructed mainly according to the guidelines film grammar defines. Because tempo is a continuous measure to capture the ebb and flow of a film, the tempo function depends on shot length, motion activity intensity, and audio effects and is defined as follows:

$$T(n) = -\frac{\alpha(s(n)-\mu_s)}{\sigma_s} + \frac{\beta(m(n)-\mu_m)}{\sigma_m} + \frac{\gamma(a(n)-\mu_a)}{\sigma_a}, \quad (6)$$

where s denotes shot length in frames,

m represents motion activity intensity,

a stands for audio importance score,

σ_s, σ_m, *and* σ_a are standard deviations of shot length, motion intensity, and audio importance score, and

μ_s, μ_m, *and* μ_a are averages of shot length, motion intensity, and audio importance score, respectively.

Here, the weighting parameters α, β, and γ are initially assigned as α = β = γ = 0.5, which assumes that shot length, motion activity intensity, and audio effect contribute equally to the tempo of a given film. Other weighting schemes are investigated in Section 6. In Fig. 6, an example of a movie tempo plot and its corresponding two story units are illustrated.

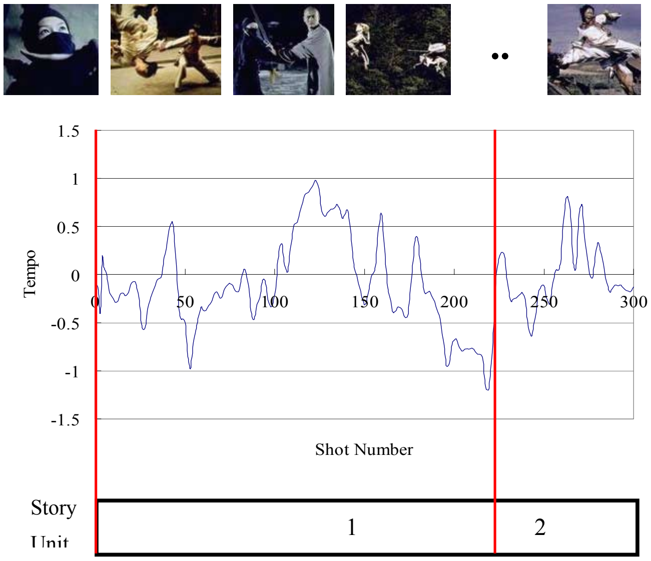

Figure 6. An example movie tempo plot of a 20-min clip taken from the Movie: *Crouching Tiger, Hidden Dragon* and its corresponding two story units.

5. MOVIE SEGMENTATION AND SUMMARIZATION

In our system, we aim to provide a movie analysis based on the tempo function. We describe the algorithm for movie segmentation and summarization in this section.

5.1 Movie Segmentation

It is more likely and popular that a user would prefer to browse through the movie to search for some interesting scenes or particular information. A likely application is movie browsing, letting viewers jump to their favorite stories. Therefore, we present the shots in a way that facilitates users in browsing and searching.

We adopt the hierarchical clustering algorithm proposed by Zhang et al. in [10], but tailor it to our application: the system for action movies segmentation and summarization. We derive the algorithm based on the tempo curve. In addition, we also deliberate the temporal relation between shots, since shots in each story unit have to be sequentially continuous. Therefore, a tempo-based story unit clustering algorithm is constructed. The key consideration is that almost all movie sequences are structured in the way that they can be segmented into several different story units, each of which contains and delivers a sub-topic. Furthermore, more segmentation can be done for each story unit to discover the important events at a finer level. In our system, we provide two different granularities regarding an action movie: story units and important events. It is a basic brief that almost all types of movies can be viewed in such a structured way, and the benefit is that movie viewers could focus on certain important stories or events in their browsing.

From the tempo curve of a movie, we can see that the most impressive or dramatic segments are more often accompanied by a higher tempo value. Each peak in the tempo curve should correspond to the shot in the segment that depicts the most interesting action.

The proposed segmentation and summarization algorithm follows the following four steps:

i. Preparing the candidates for high-tempo shots: The *N* top tempo shots over all shots ($S_1 \sim S_N$) are selected as the candidates. The ideal value of *N* should be an appropriate number of story units in a particular movie.

ii. Finding the peak high-tempo shots: Since the high-tempo values of neighboring shots could be very close, one significant story hill may have several high-tempo shots. Therefore, we modify the algorithm to determine the "real" story peaks in the tempo curve by the following steps:

a. Calculate the average number of frames, $f_{average}$, between every 2 consecutive shots.

b. For each pair of consecutive shots, if their distance is smaller than a heuristic $\gamma^* f_{average}$, then the shot with a smaller tempo value is dismissed from the set of candidate shots, where γ is an empirically determined scalar.

After this step, each of the remaining shots will be the peak of one story unit.

iii. Cutting story boundary: After identifying the high-tempo peak shots, we need to decide the boundary between two neighboring story hills. This is done by taking the shot with the lowest tempo value between two neighboring peaks of story hills. Such a "boring" shot indicates the end of the previous story and/or the start of a new story.

iv. Recursion: The algorithm is applied one more time within every story unit with the same γ value, to detect the so-called "important events" of that story unit. In other words, in our current implementation, a 2-level hierarchy structure is adopted: story units and events.

5.2 Movie Summarization

In our proposed approach, we provide two different types of summarizations: movie trailer and movie preview. In this section, we will present the details of theses two applications.

i. Movie Trailer: Movies trailers are usually purposed to draw audiences' attention by presenting the most attractive and compact scenes in a movie. In order to achieve the goal, we simply consider the shots with high tempo values, where tempo is deemed to represent the most exciting plots. We generate the movie trailer by the following steps:

(1) Pick up shots with top *k* tempo values.
(2) Concatenate the selected shots to be the whole summary for the film, while preserving the temporal order of visual and audio content.

ii. Movie Preview: Movie previews are usually used to provide a brief and general picture of the whole movie. Instead of presenting the most attractive scenes in a movie, movie previews tend to give a terse but thorough description of the movie, letting movie viewers understand the movie better and get a more concrete concept about the movie. We generate the movie preview by the following steps:

(1) For each story unit S(i), pick up shot(n) with the highest tempo value in this unit.
(2) To find the summary in every story unit, extract two more shots before and after shot(n), until the accumulated frame number of these shots $\geq$ k,

> k= x • |S(i)|, where x is the compression ratio of the summary decided by humans, and |S(i)| is the total frame number of S(i).

(3) Concatenate the summaries of every story unit to be the whole summary for the film.
Fig. 7 shows the concept of movie preview summarization.

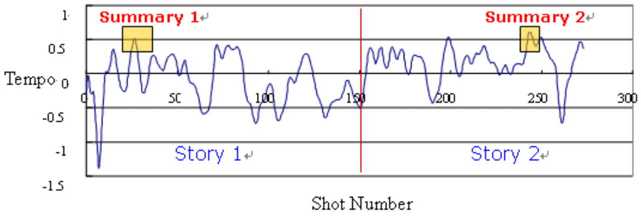

Figure 7. The movie preview summarization based on story units.

6. EXPERIMENTAL RESULTS

Based on the proposed approach, we perform experiments on three action movies, including *Crouching Tiger, Hidden Dragon* (2000), *Minority Report* (2002), and *Charlie's Angels II* (2003). The following table shows the details of each movie sequence.

Table 1. Details of three experimented movies.

Movie Title	Crouching Tiger, Hidden Dragon	Minority Report	Charlie's Angels II
Runtime	120 min	145 min	106 min
Movie Genre	Action + Drama	Action + Sci-fi	Action
File Format	MPEG-1		
Resolution	320 x 240		
Audio Format	16 bits/sample, mono, 1.6kHz		
Delivery	30 frames/second		

First, results from the three movies are generated with α= 0.4, β= 0.4, and γ= 0.2, empirically, providing a detailed analysis and presentation for movie segmentation and summarization. Fig. 8, 9, and 10 respectively show the tempo plot of each movie, while detected story units are reported in Tables 2, 3, and 4, accordingly. Examining the alignments of Figure 8 v.s. Table 2, Figure 9 v.s. Table 3 and Figure 10 v.s. Table 4, we found that every report in the tables matches the automatically generated story units.

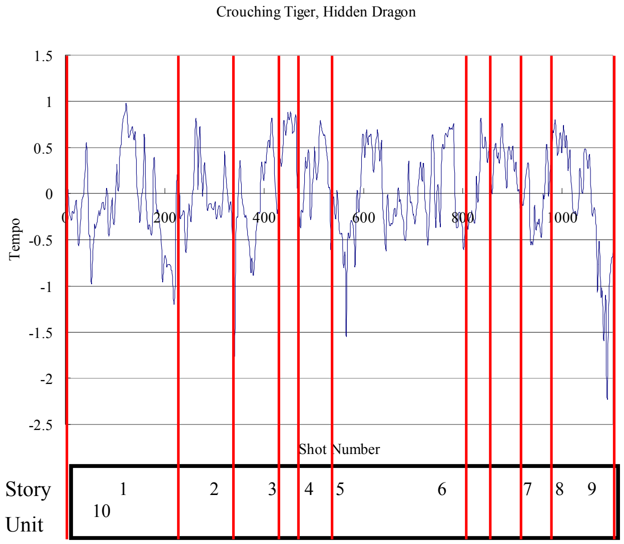

Figure 8. The tempo plot of the movie *Crouching Tiger, Hidden Dragon*.

Table 2. Story units of the movie *Crouching Tiger, Hidden Dragon*.

Story Unit Detected	
1	• The beginning appearance of Li Mu Bai, Yu Shu Lien and Yu Jiou Long. • The master sword was stolen.
2	• The fight among Jade Fox, Li Mu Bai, and Police Inspector Tsai. • Focus on not only the stolen sword but the resentment between Jade Fox and Li Mu Bai as well.
3	• The first met between Yu Jiou Long and Li Mu Bai. • The initial recall of the first time Dark Cloud met Yu Jiou Long.
4	The former half of the fight between Dark Cloud and Yu Jiou Long.
5	The latter half of the fight between Dark Cloud and Yu Jiou Long.
6	The romance between Dark Cloud and Yu Jiou Long, and Yu's escape from the wedding.
7	The beginning of the fight between Yu Jiou Long and Yu Shu Lien.
8	The fight between Yu Jiou Long and Yu Shu Lien.
9	The classic fight between Li Mu Bai and Yu Jiou Long in the bamboo forest.
10	The high tide ending of the movie.

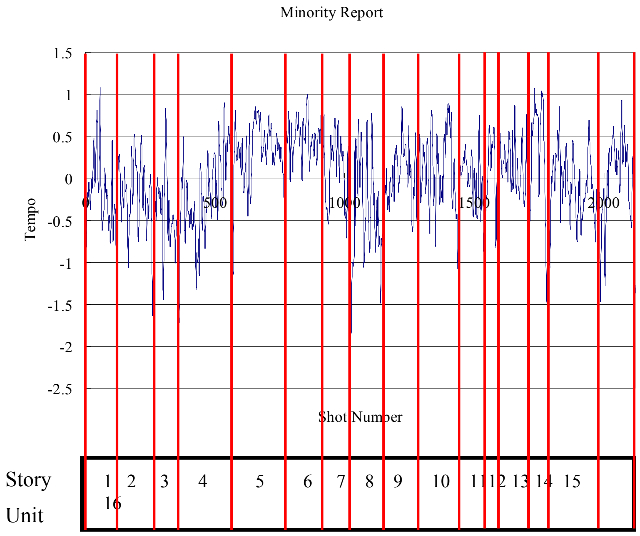

Figure 9. The tempo plot of the movie *Minority Report*.

Table 3. Story units of the movie *Minority Report*.

Story Unit Detected	
1	The beginning of the murder, predicated by the crime prophets.
2	The murder, which was going to be done by Howard Marks, leaded to the introduction to Pre-Crime Cop.
3	• Anderton recalled the miserable past happening to his son and his family. • The introduction to the three crime prophets, especially the most powerful one, Agatha.
4	The sudden but unclear crime image from Agatha, leading to the set-up murder done by Anderton.
5	Anderton's first escape.
6	Anderton's second escape from Detective Witwer.
7	Anderton visited Hineman, the system inventor, and discovered the existence of the minority report.
8	Anderton underwent the operation for changing the eyeballs.
9	Anderton was temporarily blind and investigated by the detective spiders.
10	Anderton took out Agatha from the Pre-Crime Cop.
11	Andeton's escape with Agatha.
12	Anderton was meeting the time at which the murder would happen.
13	Anderton was going to find the truth.
14	The murder happened.
15	Anderton visited her ex-wife, Lara, with Agatha.
16	• The discovery of the truth. • The end.

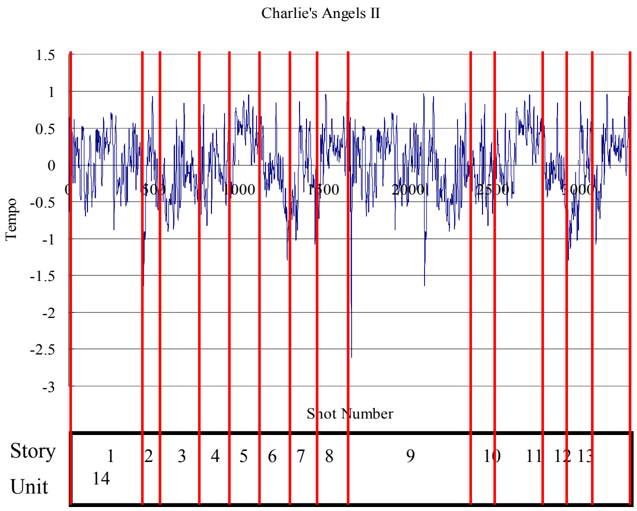

Figure 10. The tempo plot of the movie *Charlie's Angels II.*

Table 4. Story units of the movie *Charlie's Angels II.*

Story Unit Detected	
1	The introductory beginning of the movie: the case, and the three angels.
2	• The private life of three angels. • Charlie assigned the mission.
3	The first careful investigation for the mission, leading to the monitor action at the beach.
4	The appearance of Madison, played by Demi Moore.
5	The motorcycle extreme race.
6	• The absurd past of Dylan with O'Grady. • The past of Thin Man.
7	The monitor action at San Pedro Harbor.
8	The dance and stealing at the bar.
9	• The dangerous fight at the harbor warehouse. • The leaving of Dylan.
10	The fight between three angels and Madison.
11	The movie premiere of Jason.
12	The gathering of the enemies.
13	The final fight among three angels, Madison, and O'Grady.
14	The happy ending of the movie.

From Figs. 8, 9, and 10, the shot numbers normalized by the whole movie runtimes, the shot change ratios per minute, are 9.20, 14.63, and 31.24 for "*Crouching Tiger, Hidden Dragon*", "*Minority Report*", and "*Charlie's Angels II*", respectively. The difference in shot change ratio per minute indicates the "intensity" of an action movie. That is, although the three movies are all action movies, they have somewhat different primary characteristics: "*Crouching Tiger, Hidden Dragon*" can be considered an action movie with more drama and plot design, "*Minority Report*" can be regarded as a combination of an action movie and a sci-fi movie, and "*Charlie's Angels II*" is generally a typical action movie.

In our experiment, a five-level subjective human evaluation is performed, where 5 is the best and 1 is the worst. Two different metrics are used in the subjective human evaluation. First, metric "relevance" calibrates the perceptive similarity between the subjective human-generated movie segments and the automatic computer-generated movie segments. Second, metric "representative" indicates how expressive and readable a movie summary is. In performing the subjective human evaluation, the "relevance" and "representative" scores are provided by the members of Communication and Multimedia Laboratory at Dept. of CSIE, National Taiwan University. (Details of the personnel can be found in http://www.cmlab.csie.ntu.edu.tw/~wei/experiment.html/.) Finally, the experimental results of the 5-level subjective human evaluation are listed in the following tables.

Table 5. The 5-level subjective evaluation for the movie "*Crouching Tiger, Hidden Dragon*".

Story Unit	Relevance Score (1-5)
1	4.2
2	3.7
3	4.0
4	3.8
5	4.6
6	4.3
7	3.9
8	4.5
9	4.8
10	4.2

Table 6. The 5-level subjective evaluation for the movie "*Minority Report*".

Story Unit	Relevance Score (1-5)
1	4.3
2	4.7
3	3.8
4	4.1
5	4.0
6	4.1
7	4.8
8	4.3
9	4.3
10	4.0
11	4.5
12	3.6
13	3.7
14	4.2
15	4.0
16	3.9

Table 7. The 5-level subjective evaluation for the movie "*Charlie's Angels II*".

Story Unit	Relevance Score (1-5)
1	4.2
2	3.9
3	4.3
4	3.7
5	4.8
6	4.2
7	4.1
8	4.1
9	4.2
10	3.9
11	3.6
12	3.7
13	4.0
14	4.2

The following table shows the subjective evaluation results of two types of movie summaries, the movie trailer and the movie preview.

Table 8. The 5-level subjective evaluation of movie summaries.

Movie Title	Summary Type	Representative Score (1-5)
Crouching Tiger, Hidden Dragon	Trailer	4.0
	Preview	4.2
Minority Report	Trailer	3.9
	Preview	4.5
Charlie's Angels II	Trailer	3.5
	Preview	3.7

In addition, the ratio of each movie summary to the entire sequence is given in Table 9.

7. CONCLUSIONS AND FUTURE WORK

In this paper, a movie tempo analysis based action movie segmentation and summarization framework is proposed. In this work, first, we take into account the film grammar, on which tempo design of a movie is dominated by directors. Three techniques, shot change detection, motion activity analysis, and audio energy extraction, are investigated to compute the tempo of action movies. Based on tempo, the movie segmentation and summarization, in turn, can be performed, successfully. For movie segmentation, a hierarchical clustering algorithm is provided and for each movie, story units and important events are further segmented and extracted. For movie summarization, two different types of summaries, movie preview and movie trailer, are presented for movie viewers to efficiently browse or skim the movies.

Table 9. The summarization ratio of each movie summary.

Movie Title	Movie Summary Runtime		Summary Ratio
Crouching Tiger, Hidden Dragon	Trailer	1 min 17 sec	1 %
	Preview	1 min 50 sec	1.5 %
Minority Report	Trailer	1 min 22 sec	0.9 %
	Preview	3 min 54 sec	2.7 %
Charlie's Angels II	Trailer	2 min 02 sec	2 %
	Preview	3 min 15 sec	3 %

We adopt the conception from film grammar that the editing technique- the montage, and human perceptual elements- motion and audio, together contribute the formation of movie tempo. However, a lot more sophisticated filmmaking techniques are applied to modern movies, and more guidelines in film grammar should be further investigated for semantic film analysis. In addition, there are a lot of different categories of movies in addition to action movies, such as music movies, horror movies, etc. How to generalize the proposed system for analyzing different kinds of movie genres is the main concern of our future work.

8. ACKNOWLEDGMENTS

This work was partially supported by the National Science Council and the Ministry of Education of Taiwan, ROC under the contract No. NSC92-2622-E-002-002, NSC92-2213-E-002-023 and 89E-FA06-2-4-8.

9. REFERENCES

[1] Howhard D. Wactlar. The Challenges of Continuous Capture: Contemporaneous Analysis and Customized Summarization of Video Content. CMU, USA.

[2] Brett Adams, Chitra Dorai, and Svetha Venkatesh. Toward Automatic Extraction of Expressive Elements From Motion Pictures: Tempo. *IEEE Trans. on Multimedia*, Vol. 4, No. 4, December 2002.

[3] Arijon, D.. Grammar of the Film Language. *Silman-James Press*, 1976.

[4] Chitra Dorai and Svetha Venkatesh. Media Computing: Computational Media Aesthetics. *Kluwer Academic Publishers*, 2002.

[5] Leo Braudy and Marshall Cohen. Film Theory and Criticism: Introductory Readings. *Oxford University Press*, 1999.

[6] Jeroen Vendirg and Marcel Worring. Systematic Evaluation of Logical Story Unit Segmentation. *IEEE Trans. on Multimedia*, Vol. 4, No. 4, December 2002, pp. 492-499.

[7] Jeroen Vendirg and Marcel Worring. Interactive Adaptive Movie Annotation, *IEEE Multimedia*, Vol. 10, Issue 3, 2003.

[8] Zeeshan Rasheed and Mubarak Shah. Scene Detection in Hollywood Movies and TV Shows. In Proc. of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[9] Extraction and Use of MPEG-7 Descriptions, ISO/IEC TR 15938-8:2002(E).

[10] Bin Yu, Wei-Ying Ma, Klara Nahrstedt, and Hong-Jiang Zhang. Video Summarization Based on User Log Enhanced Link Analysis. *ACM MM '03*, pp. 382-391.