HERMES, Thorsten; SCHULTZ, Christopher

Automatic Generation of Hollywood-like Movie Trailers

Publiziert auf eCulture Factory:
http://www.eculturefactory.de/eculturetrends/download/hermes_schulz.pdf
Auf netzspannung.org:
http://netzspannung.org/database/378647/de
19. Dezember 2006

Der Beitrag entstand im Rahmen der Veranstaltung »eCulture Trends 06: Zukunft entwickeln – Arbeit erfinden«, die am 20. Oktober 2006 von der Projektgruppe eCulture Factory des Fraunhofer IAIS in Bremen veranstaltet wurde.

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

MARS - EXPLORATORY MEDIA LAB

eCULTURE FACTORY

Automatic Generation of Hollywood-like Movie Trailers

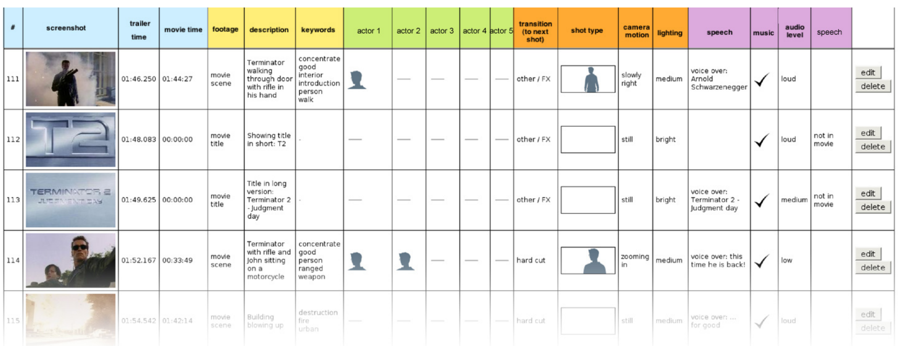
Thorsten Hermes & Christopher Schultz
Technologie-Zentrum Informatik, Universität Bremen
Am Fallturm 1, 28359 Bremen
{hermes, cschultz}@tzi.de

1 Introduction

The movie industry has long perfected advertising its movies. In order to inform audiences about an upcoming movie and to attract it, short previews are presented in cinema or TV – the "movie trailers". A branch of Hollywood's film industry has specialized in trailer production. This process of creating a trailer has always been determined by manual editing of video and audio.

Traditional trailer creation covers the selection of scenes and their arrangement, the composition of a music soundtrack or sound effects, and the generation of additional footage, like a movie title animation. Like this, creation of a unique trailer is a creative process on the one hand. On the other hand, trailers include certain structures or patterns set up by rules of marketing and advertisement, by the expectation of the audience, and by cultural factors. Hence, beside creativity which makes each trailer an artwork, trailer creation follows certain rules which can be defined in algorithms.

There are already several solutions for pure automatic summarization of video footage. *Video abstracting* or *video skimming* analyzes video material and condenses it to important scenes ([CHWC99],[Chr99],[SK98]). Typically the linearity of the input video is preserved in order to give an overview of the complete video so that video abstracting is often in video libraries ([Chr99],[SK98]). *Pictorial summary* as ([KH00],[NT99],[RSC99],[YY97]) aims for a summary of contents in a pictorial way. *Video browsing* is closely related to the pictorial



#	screenshot	trailer time	movie time	footage	description	keywords	actor 1	actor 2	actor 3	actor 4	actor 5	transition (to next shot)	shot type	camera motion	lighting	speech	music	audio level	speech	
111		01:46.250	01:44:27	movie scene	Terminator walking through door with rifle in his hand	concentrate good interior introduction person walk		—	—	—	—	other / FX		slowly right	medium	voice over: Arnold Schwarzenegger	✓	loud		edit delete
112		01:48.083	00:00:00	movie title	Showing title in short: T2	-	—	—	—	—	—	other / FX		still	bright		✓	loud	not in movie	edit delete
113		01:49.625	00:00:00	movie title	Title in long version: Terminator 2 - Judgment day	-	—	—	—	—	—	other / FX		still	bright	voice over: Terminator 2 - Judgment day	✓	medium	not in movie	edit delete
114		01:52.167	00:33:49	movie scene	Terminator with rifle and John sitting on a motorcycle	concentrate good person ranged weapon			—	—	—	hard cut		zooming in	medium	voice over: this time he is back!	✓	low		edit delete
115		01:54.542	01:42:14	movie scene	Building blowing up	destruction fire urban	—	—	—	—	—	hard cut		still	medium	voice over: ... for good	✓	loud		edit delete

Figure 1: Extract of the trailer database (Terminator 2)

summarization but furthermore focuses on a hierarchical, not necessarily linear way, of presenting the video content. Most works depend on low level analysis on image and audio ([LPE97],[SK98]) to gather knowledge about the source video. Although the possibility of automatically generating a movie trailer was already explicitly mentioned in ([CKCW04],[LPE97]), no approach yet has specially focused on this field.

Compared to the already existing approaches, automatic trailer generation is a domain dependent process, since each movie genre contains different rules and structures. Presently, the automatic extraction as well as the generation of a narrative, or at least some kind of dramatic arc, seems hardly feasible. Due to the one year time frame, the SVP project focuses on a genre which relies significantly more on visual sensation, speed and effects, than on narrative: Hollywood action movies.

2 Finding Trailer Patterns

Nobody had yet described the structure of trailers accurate enough in order to have a basis for algorithmic processing. But two main research directions can be identified: first, researchers with a background in classic film theory traditionally look at movies from a highly semantic and even interpreting point of view, dealing with the character's abstract roles for the plot, the social relevance or metaphoric value of certain scenes, and the classification of certain types of lighting or camera techniques. And second on the other side, computer scientists mostly concentrate on physical features (like motion estimation) of the movie, related to some practical application, such as database retrieval. As such, much of the research in this field has been done for footage of very formal nature, such as news broadcasts or sports events. The gap between these two directions is tremendous.

Hence, for this work we created our own research data pool. Therefore, 11 action trailers (i.e. "Terminator 2" and "Charlie's Angels") were manually analyzed in detail for aspects such as shot duration, actor appearance, camera position or sound volume. This frame-precise data were stored in a database. This frame-precise data were feed in a database. A visualization of this data is shown in Figure 1.

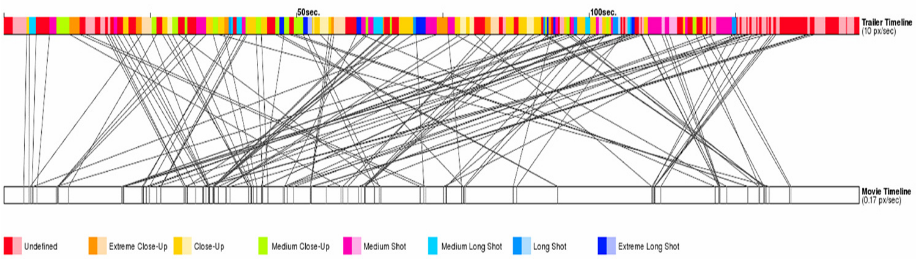


Figure 2: Location of clips in trailers connected to their location in the source movie over time.

Based on the resulting pool of data sets typical trailer patterns are extracted by performing manual and automatic statistical methods. Figure 2 illustrates a self developed data mining tool visualizing the dependency of scene locations within a trailer and its movie. The colored bar at top indicates a timeline of a certain movie trailer. The different colors indicate certain types of a shot (i.e. "close-up" or "long-shot"). The bar at the bottom stands for the timeline of the corresponding movie. So one can see at once which piece of the original movie is used at which position in the corresponding movie trailer.

The extracted knowledge of our research forms the basis for the trailer model, which is used by the system in order to generate a trailer.

3 Trailer Generation System

The automatic creation process of a trailer from a movie is realized in two main phases - this analysis and the generation. These two phases reflect in two main software components – the "Analyzer" and the "Generator".

First, the source movie is extensively analyzed by the Analyzer for a multitude of audio and video features. Based on these features the logical units (i.e. shots or clips) are categorized into different categories like "explosion", "scream", "fast action" and so on.

And second, this knowledge is used by the Generator to calculate and render a trailer based on the patterns and formulated models. The whole process can be run fully automatically. In the following, the two main software components "Analyzer" and "Generator" are described in more detail.

3.1 Knowledge About Trailers

The analyzing operations of the system gather knowledge about features such as cuts, actor appearance, speech, text, sound events and many more. Therefore, analyzing software was customized or self written.

The resulting low level analysis data are combined to higher level knowledge about the movie. For example, the combined low level analysis results of the speak detection and the sound volume analysis determine the higher level knowledge about possible shouts occurring in the movie. Like this, we obtain a large collection of knowledge about the source movie.

Visual Features:

- Face detection: Faces even of moving persons or not frontal perspective (see figure 3) are detected in the movie. Additional the system tries to distinguish for example the main actor by clustering all found faces.
- Cut detection: Film shot boundaries are detected based on the visible changes between two different scenes.

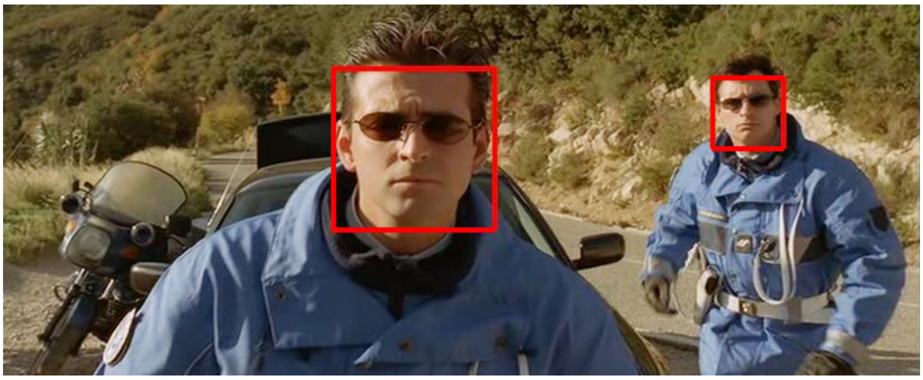


Figure 3: Results of the face detection ("The Transporter")

- Motion analysis: The image motion and cluster ranges of homogeneous optical flow are calculated by a so called correlation approach. Scenes with high image motion indicate action.
- Text detection: Scenes with overlaid text are found automatically. Movie title or the credits at the end of a movie are not suitable for a trailer because they are disturbing while the rendering process.
- Background information: Additional movie data such as title, director and actor names, famous quotes and won awards are automatically extracted from the Internet Movie Database (IMDb[1]).

Audio Features:

- Sound volume detection: The movie is automatically segmented into ranges of homogeneous sound volume. The volume calculation was adapted to the human perception of sound.
- Sound event detection: This certain piece of software was trained by audio examples in order to find and differentiate between certain audio events. The used classifier (support vector machine) is able to identify explosions, gunshots, and screams as well as other dramatic sound events.
- Speech detection: The software analyzes the audio track for speaking people. Scenes with speech are a good indicator for appearing actors. Furthermore, fast action scenes do not include speech.
- Quote detection: The software can detect important quotes within the movie. Movie speech is matched with quotes registered in the IMDb. So most famous statements or dialogues in movies - like "Hasta la Vista, Baby" form Terminator - can be found for the trailer.
- Music detection: Based on a beat detection the system can distinguish soundtrack music. Scenes without music are more suitable for a trailer than scenes with background music (figure 4).

[1]The IMDb can be accessed under the URL: `http://www.imdb.com`

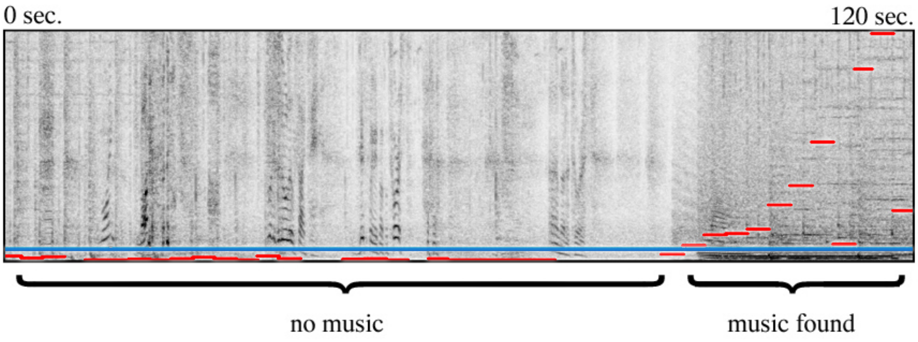


Figure 4: Results of the music detection; an audio sequence is segmented into music and music-free section. The short horizontal lines in the right over a threshold (the blue line) indicate (disturbing) background music.

3.2 Composing Trailers

Based on extracted movie information our generation software creates a fully fledged movie trailer. In order to create a new trailer first the footage of the source movie using 19 labels is categorized: the systems intersection algorithm calculates the probability and labels sequences belonging to a certain category. As a result, a list of suitable clips for each category is created, like "explosion clips", "scream clips", or "slow action clips". From this pool of clips a new trailer can be composed.

As the "brain" the system uses an ontology. The ontology stores all information needed to create a trailer, for example abstract models of trailer patterns, or rules about the composition of a new trailer (like this it is defined for example that no scene should be shown twice). When creating a trailer, this system dynamically calculates a trailer structure according to the stored facts and rules, and according to the amount of categorized clips. This sequence is enhanced with all kinds of effects in order to create "real trailer feeling".

- Automatic Editor: The new trailer is automatically cut out of the source movie. Therefore, open source software is used which can be controlled by scripts.
- Music and Sound Effects: The system combines and mixes pre-produced music and sound effect media of our self created audio library in order to produce a unique trailer soundtrack during each generation.
- 3D Animations: The system creates complex 3D animations automatically at runtime. In order to control modeling and rendering of the external 3D software Python scripts are used. Like this, it is possible to create 3D text animations for the trailer, presenting the movie title or actor names (see figure 5).
- Slow motion, special transitions, volume, location: The software manages several footage properties for fine-tuned trailer results according to trailer semantics.

Figure 5: Dynamically created 3D animations, (a) movie title, (b) punchline, (c) movie credits.

4 Comparing Trailers

In addition to our own evaluation, we performed a test viewing with a group audience of 59 people. Since producing movie trailers is a creative process, a purely statistical analysis, like a shot-by-shot comparison of generated and official trailers, would not sufficient. Therefore, we presented professional trailers, random clip sequences and trailers produced by our software. Test people were asked to rate the same six aspects for each video.

As expected, the overall rating of the random trailers is significantly lower than any of the others, while professional performed best. Our automatically generated trailers received high ratings for good composition and “cuts & effects”, and lower ratings for “narrative aspects”. This shows that our system succeeds in timing the cuts and adding animation screens with emphasizing audio effects to enhance the genre-typical powerful appearance. The aspect “character introduction” and “plot introduction”, which depend highly on sophisticated high-level analysis, received the lowest scores.

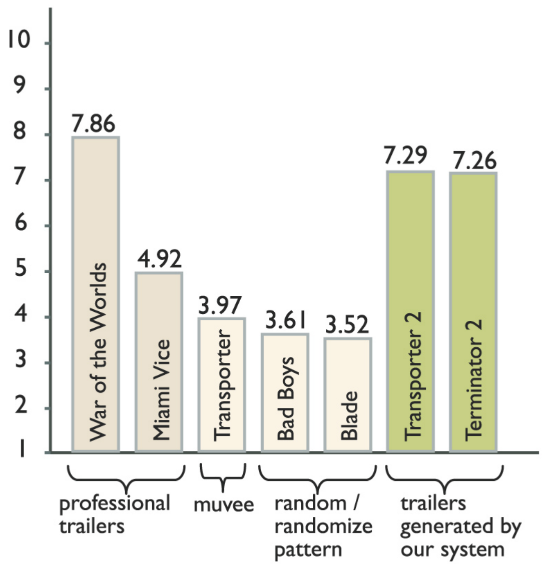


Figure 6: Evaluation results

The results show that our attempts of automatic categorization and composition appear to be generally successful (compare figure 6). They suggest that our automatic trailers are a clear improvement over random shot selection methods. Furthermore, it seems that wrongly chosen shots, resulting from inaccuracies of the analysis modules, typically do not disturb the flow of the trailer.

5 Conclusion

In this paper we presented our approach of an automatic trailer generation system. Based on the automatic analysis of the audio and video sources of certain movies the introduced system is able the produce a trailer with the help of knowledge (stored in an ontology) of the corresponding domain. In this project we concentrate on action movies since this genre offers a lot of opportunities for automatic analysis. The test results are very promising since the asked probands judged the trailers very similar to the professional made trailers. Again, based on the test results our future work will cover the story telling phase during the trailer. There we have the weakest points in our work.

6 Acknoledgements

This work could never be happen without the tremendous contribution by all the other project members. These are in alphabetical order: C. Brachmann, H. I. Chunpir, S. Gennies, B. Haller, O. Herzog, A. Jacobs, P. Kehl, A. P. Mochtarram, D. Möhlmann, C. Schrumpf, B. Stolper, and B. Walther-Franks. Thank you a lot!

Sample trailers and further material about the SVP project can be found under the URL `www.tzi.de/svp`.

References

[Chr99] M. Christel. Visual digest for news video libraries. *Proceedings ACM Multimedia Conference, Orlando, FL*, 1999.

[CHWC99] Michael G. Christel, Alexander G. Hauptmann, Adrienne Warmack, and Scott A. Crosby. Adjustable filmstrips and skims as abstractions for a digital video library. In *ADL*, pages 98–104, 1999.

[CKCW04] Hsuan-Wei Chen, Jin-Hau Kuo, Wei-Ta Chu, and Ja-Ling Wu. Action movies segmentation and summarization based on tempo analysis. In *MIR '04: Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*, pages 251–258, New York, NY, USA, 2004. ACM Press.

[KH00] C Kim and J Hwang. An integrated scheme for object-based video abstraction. *Proceedings of ACM Conference on Multimedia Los Angeles, CA*, page 303311, 2000.

[LPE97] Rainer Lienhart, Silvia Pfeiffer, and Wolfgang Effelsberg. Video abstracting. *Communications of the ACM*, 40(12):54–62, 1997.

[NT99] J Nam and A Tewfik. Dynamic video summarization and visualization. *Proceedings of ACM International Conference on Multimedia, Orlando, FL*, 1999.

[RSC99] K Ratakonda, M Sezan, and R Crinon. Hierarchical video summarization. *IS&T/SPIE Conference on Visual Communications and Image Processing99, San Jose, CA*, 3653:1531–1541, 1999.

[SK98] M.A. Smith and T Kanade. Video skimming and characterization through the combination of image and language understanding. *IEEE International Workshop on Content-Based Access of Image and Video Database*, pages 61–70, 1998.

[YY97] M Yeung and B Yeo. Video visualization for compact presentation and fast browsing of pictorial content. *IEEE Trans CSVT*, 7:771–785, August 1997.