I.D.E.A. — "Intelligent Digital Editing Assistance" for Trailer Production

Isabelle Janda

Diplomarbeit Jahrgang 2011/2012
Masterstudiengang Digitale Medientechnologien, FH St. Pölten

dm101520@fhstp.ac.at

Abstract

"Intelligent Video Editing" is the latest term for a development intended to support editors with their work by applying different and innovative software algorithms. Interactive video browsing, time-synchronized transcriptions, image semantic tracing or 3D video visualization are only a few new features introduced in these small research applications, which show promising results and give editors editing opportunities never explored before. Due to still-existing challenges in interpreting audiovisual content, however, no major commercial application has made serious use of these new developments up to this point. Although especially in the area of trailer generation, where only a small fraction of the often huge and unorganised amount of source footage is eventually used and a creative composition of video and audio material is the desired aim, the application of Intelligent Editing Assistance could simplify and improve the working process immensely.

Therefore, this research focuses on how to create a software tool for Intelligent Digital (Video) Editing Assistance ("IDEA") usable in the professional field of trailer generation. The overall aim is to simplify the working process and to extend editing opportunities for trailer producers. In this research a comprehensive concept has been developed based on the survey of professional trailer producers and evaluated by implementing a basic motion and eye trace detection/visualization approach. This is a first step towards a promising future for Intelligent Video Editing.

1 Introduction

Almost as long as the cinematic film itself, since about 1912, film marketing has existed in the form of so-called "trailers" – short advertising spots, usu-

ally composed of the most representative scenes of a film, aiming to attract viewers. Nowadays, exactly a hundred years later, trailers are still one of the most effective marketing tools in the film and TV industry [11].

Furthermore, trailer producing has become more of an art form over the years and it has developed so many forms of appearance that it has created its own companies specializing in producing movie trailers and has led to big television departments working on trailer marketing. Because it is a creative and demanding process, it is also very time-consuming. Though television broadcast trailers are usually only under a minute long, the production process often takes weeks or even months. Trailer editing is not only about arranging the most exciting scenes of a film, in fact it is so much more. Trailer editors are often responsible for supervising the whole production process and therefore are very often referred to as trailer "producers". They are usually accountable for developing the trailer concept, selecting and cutting music, screening source material, gathering appropriate video clips, creating graphical inserts, including special effects, colour-grading, of course the actual editing process and even writing the voice-over text of a trailer. Therefore, creativity and time are two opposing factors and an effective workflow plays an important role in the outcome of each product. For that reason, this project focused on new and experimental ways of simplifying this process and helping trailer producers to become more effective without abandoning their creativity. With this research project a first serious approach towards intelligent technical assistance for a professional editing field was made and new possibilities were pointed out.

This paper is intended to summarize the work that has been done in [14]. The next section of this paper deals with fundamental principle of trailer production. It is very important to have a thorough knowledge of the area of application when creating a tool which is meant to help people in practice. In Section 3 the possibilities of assisting algorithms, already developed by other research teams, are explored and evaluated. An attempt has been made to create an all-embracing concept by adapting existing approaches to the needs of trailer generation. Eventually, the last and practical part of this paper deals with the implementation of a basic eye trace detection/visualization algorithm and the execution of a comprehensive eye tracking study. By comparing the results of the eye tracking study to the results of the created algorithm findings in this area of research are stated.

2 Trailer Production

This section summarizes fundamental aspects of trailer production found in [14] and introduces the work area IDEA is intended for. For a detailed review please refer to [14] or [15].

The trailer production process is complex and takes longer than one might expect. In comparison to other editing workflows trailer production is different in many ways due to its special characteristics. Trailers can be distinguished, amongst other things, by their appearance medium. It can therefore be differentiated between movie trailers, television trailers and internet trailers. This work revolves around TV trailers which are believed to be best suited for a semi-automatic workflow due to heir variety of appearance forms and their less narrative approach.

Trailer Strategies & Objectives

A trailer has to fulfil several important tasks at once. It is of course supposed to increase the numbers of viewers by acquiring new audiences and activating the established viewership. This is achieved by attracting attention and entertaining people to keep them watching. But a trailer also has to inform the viewers of the current program and convey a brand image. By exaggerating about the quality of the product and by emotionalising its content it draws us in and creates expectations we hope to be met. Although we know that this is not always the case, we still fall for it.

Audio-Visual Design Elements

Trailers have an audio-visual structure built from several different audio-visual elements. They are mainly composed of scenes taken from the, to be advertised, product. But video clips are not the only elements a trailer consists of. Trailers are usually a mix of different audio-visual design elements such as sound bites (e.g. movie quotes), music, sound effects, trailer text (voice over commentary) and motion graphics (titles, 3D animation) in addition to the already mentioned video clips.

Narrative Concepts

This research project has found two narrative concepts TV trailers can be categorized in:

- Trailers with a predominantly narrative approach ("narrative trailers")
- Trailers with a predominantly emotional approach ("image trailers")

Narrative trailers usually promote only one single product – one movie, one episode. Like all narrative art forms they consist of a dramaturgical structure. In contrast image trailer predominantly convey the station's brand image and contain general information about a certain time slot, channel, etc. They normally work with all kinds of source material and only choose so called "highlight" scenes of the used products to create a certain mood.

Narrative trailers have more complex requirements for video retrieval and automation processes. Narrative workflows are very hard to automate due to their complex content structure and dramaturgical build-up. These algorithms do not only consider aesthetical aspects but also require some kind of artificial intelligence to recognize meaning regarding content and its importance.

Image trailers therefore seem to have a big advantage regarding automation and Intelligent Video Editing. Their concept is based on a more aesthetical approach which is based on more easily extractable audiovisual features. It can be stated that trailers with a less narrative approach seem to be better suited and provide more possibilities for automation and Intelligent Video Editing.

Requirements for Quality Trailer Editing

To assure a trailer's quality, several important design principles have to be considered. This subsection mentions a selection of strategies used in the actual trailer production and important for the conception of IDEA.

Apart from regulations given by international TV standards trailers producers have to carefully deal with established design strategies. One of them concerns the audience's attention span. Although there are two different theories about how much attention a trailer should attract – some might consider the subconscious manipulation stronger than attracting the audience's full attention – trailers use some formal design strategies to attract and hold attention. According to [9] they are usually based on certain contrasts (colour, intensity/length, ambiguity/novelty, movement, acoustic signals) or based on specific content (stimuli for current needs, stimuli for basic human needs, storytelling, emotional excitement).

Other strategies concern the area of editing. According to Murch the individual editing process should be based on the "Rule of Six" – a list of editing priorities concluding that all editing decisions should be focused on the audience's perception (Murch's theory is practically reassessed by this project in one of the later sections). Another aspect concerning editing can be defined

by the four dimensions of montage according to [3] defining relations between two shots (graphical, rhythmic, spatial and temporal).

Design principles regarding content usually involve knowledge in the areas of conception and dramaturgy. To honour aesthetic design principles one should deal with visual composition, motion (camera movement, content-based movement), rhythm (physical, emotional, event), the use of edit types and audio design (loudness).

3 The Concept of IDEA

IDEA (Intelligent Digital Editing Assistance) is going to be a practicable software tool that should serve editors in the area of trailer production. The "idea of IDEA" is to include intelligent algorithms which improve the workflow and the scope of action for trailer producers and support them in developing and realizing creative and innovative trailer concepts. IDEA should not only simplify and fasten the trailer production process it should also provide new possibilities of editing approaches. It is intended to be a creative "playground" where trailer producers find different, possibly experimental, ways of applying there editing skills by using new kinds of video visualizations and editing procedures. IDEA is the concept of creating a new working environment for trailer producers that ideally leaves nothing to be desired

3.1 General Structure

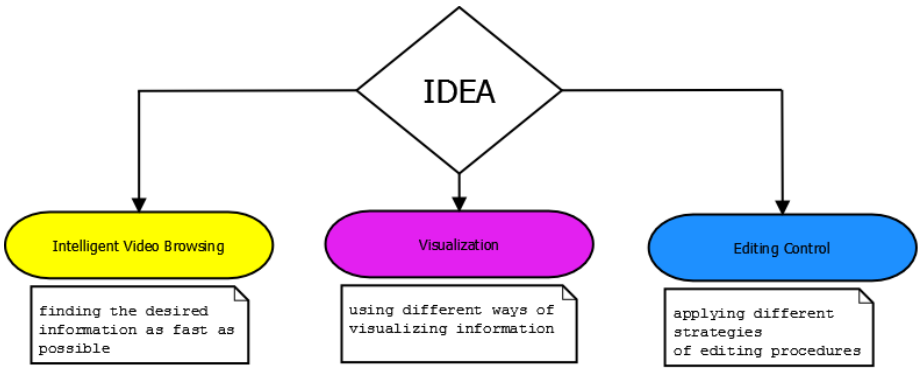

Figure 1 General structure of IDEA

IDEA is an extendible concept that should be improved and updated along with its development. It is based on three fundamental approaches (see Fig-

ure 1). These three basic functions will include new scientific developments that have shown promising results and are going to be evaluated in practice by IDEA. The aim is to first gather the most innovative and recent ideas in these three application areas and to then sort out the best and most valuable ones afterwards. Because of the immense scope of these areas this project can only focus on one of these functions for now. It was decided to concentrate on the part of "Intelligent Video Browsing" as it deals with most of the main issues of trailer editing – time efficiency, large data amount and audio-visual diversity – and can be seen as the launching area of interest for this research.

3.2 Related Work

IDEA is a very unique concept for intelligent video authoring, as it is, according to the author's best knowledge, so far the only existing approach for a professional editing area. Recent projects with similar approaches, like the "Hitchcock" ("Hyper-Hitchcock") system [19] or "Silver" [5] were developed with the intention of simplifying the editing process for the non-professional user and home-video generation but have never been applied to an actual professional editing workflow. While the approach of creating an editing support by intelligent systems is nothing new, first attempts were already made in 1991 by the project of "IMPACT" [23], the need for computer-based help has never been stronger than now. This trend can also be seen in the latest releases of the most common professional video authoring software, such as the introduction of the Avid "PhraseFind" plugin for Media Composer 5.5 [2] as well as Final Cut Pro X's new media organizer using content auto-analysis and smart collections [1], and the improved "speech-to-text" analysis and face detection in Premiere Pro CS 5.5 [7]. These are just a few examples that show how promising the future of intelligent workflows might be.

Furthermore, to the author's best knowledge, IDEA it is the only concept for an editing assistance tool that is customized to the special needs of trailer generation. However, one of the few and groundbreaking projects that should be mentioned here is the project of "SVP" (Semantic Video Patterns) [12], a fully automatic approach for trailer generation, which has laid out down basic principles for the application of intelligent algorithms in the production process of "Hollywood-like" trailers and has, amongst others, inspired this research project.

3.3 Detailed Concept Part I: Intelligent Video Browsing/Video Retrieval

Part one of IDEA's concept deals with the area of intelligent video browsing and semantic retrieval. To enhance trailer producers' workflows it is one of the most effective measures to improve the time-consuming procedure of video and audio search and material screening. The need for an optimised footage management was the trigger for this project. Therefore it was decided that area of Intelligent Video Browsing was going to be dealt with in greater detail than the other conceptual parts (Part II & III are dealt with in [14] but cannot be mentioned here due to the immense scope of this project). It introduces the implementation of a sophisticated footage management system and the use of intelligent algorithms for video browsing. Using the previous theoretical research and a recent survey including the opinion of professional trailer producers the following paragraphs present the most appropriate and suitable concepts for IDEA.

3.3.1 General Idea & Problem Specification

The main concept of IDEA is to include a high-end footage browsing and clip retrieval system. The overall aim is to simplify the process of finding and visualizing the most appropriate selection of audio and video material for the individual needs of a certain trailer concept by analysing and extracting low-level audio and video features and interpreting their higher semantic meaning. As pointed out by [4] the video retrieval task, however, raises fundamental questions in computer vision and information retrieval: How to represent video items? What information can directly be extracted from them? And how to explore such information in order to satisfy the user's information need? IDEA therefore faces some major challenges.

The first and main issue concerning video retrieval is one of the greatest human efforts of the last decade: How to make computers understand the context and semantics of audiovisual representations? The problematic differences between low level description of audiovisual media, extracted by content analysis, and concepts that are meaningful to users is generally known as the "semantic gap"[16], and shows the inability of computers to interpret complex correlations ("high-level concepts") of reality correctly. Especially in video retrieval by text the fuzziness of natural language causes major challenges. Unfortunately there is no such thing as a semantic channel between the video's content and the user's need for information and all ef-

forts have to focus on bridging this gap instead of closing it forever.[4] This not only concerns the area of video retrieval but multimedia research in general.

While significant progress has been achieved in the area of image and audio retrieval, there are no satisfactory systems for video retrieval and no true video search systems, capable of analysing the whole spatio-temporal information available in videos so far.[6, 8, 20–22, 25] It is the broad range of applications though that motivates scientific research to constantly come up with new solutions in these research areas [13] and trying to create artificial intelligence to solve human problems. But although the gap is slowly getting narrower it is utopian to believe that a hundred percent perfect and universal solution will ever be achieved. In fact there are, and probably will be, thousands of individual solutions adjusted to each individual multimedia concept [17]. This is all the more reason for scientific approaches like IDEA to focus on their specific areas and to make sure that their corresponding intelligent algorithms are customized to their application area as much as possible to achieve a satisfying success rate.

At last it is worth mentioning that the success of a good intelligent video browsing system depends, to a large extent, on the range and selection of information that is inquired. Again this is a delicate matter as the process of trailer conception is highly individual and cannot be generalized. For this purpose it was decided to introduce a set of high-level search criteria that were evaluated by the personal experience of professional trailer producers.

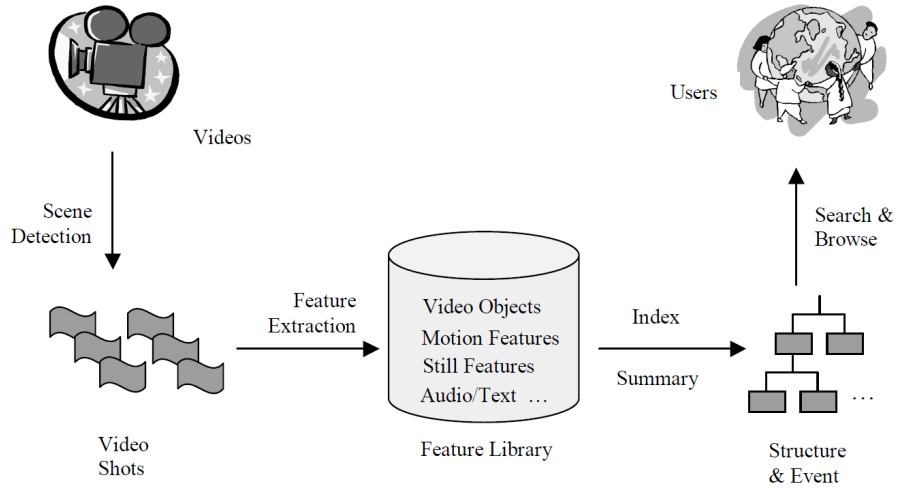

Figure 2 Content based video indexing and retrieval process according to [26]

Figure 2 shows the content-based video indexing and retrieval process according to [26].

3.3.2 Survey

Before intelligent search algorithms can be selected and evaluated for their application in trailer generation it is necessary to define what kind of retrieval criteria should dealt with. This is inasmuch complex as video search always deals with spatio-temporal and therefore multidimensional requests (rich content), a huge amount of raw data and very little prior structure [13]. Therefore requests are on the one hand harder to process and on the other hand more difficult to abstract in an easily understandable structure, compared to image or music retrieval. But it is also a challenging issue as there is no theoretical knowledge about what trailer producers are looking for in their footage. The task of screening footage and choosing potential scenes is one of the most important elements of the conception phase in a trailer producer's workflow though it is relatively unknown how trailer producers (and human beings in general) develop creative concepts. This is on the one hand due to the fact that it is a highly individual process and on the other hand it is understandable that trailer producers are unwilling to share their secrets of success openly. According to marketing theories the basis for this decision process is the requirement of "tonality", which describes how the trailer should be perceived by the audience. The category of tonality therefore defines the creative, verbal and visual style of the concept but does not necessarily determine single design elements [18]. The actual decision of choosing specific scenes over others still lies within the discretion of each individual trailer producer. As there is no relevant survey data in the field to help determine appropriate search criteria and get an overall picture of what could be helpful for trailer producers it was therefore decided to conduct a survey in the course of this project.

20 professional trailer producers from five major television stations in Austria and Germany (PULS 4, ATV, ORF, Red Bull Media House & ProSiebenSat.1 Group) were asked for their opinion on intelligent footage search and specific high-level concept algorithms. The survey was conducted with the help of an online questionnaire and consisted of five different questions (three on intelligent video browsing, one on Murch's editing criteria and one requesting the opinion on eye trace) and an optional field to leave personal commentary.

The first question of the survey concerning intelligent video browsing was designed as an open question, thus giving trailer producers a chance to formulate their own ideas on which audiovisual aspects seem most important when gathering shots and scenes from footage. Statements were collected and used to design the search concept in the following subsection.

The second question category was designed to learn more about trailer producer's opinions on special high-level search concepts. The following algorithms were proposed and participants were requested to comment their persuasion on whether these concepts would be helpful in trailer editing or not. Results clearly showed trailer producers' favourites and which algorithms are not considered to be valuable (see Table 1).

Table 1: Result evaluation for possible video retrieval algorithms

Category	**Percentage**	**Votes (out of 20)**
Object Retrieval	25%	5
Character Retrieval / People Search	90%	18
Location-Identification	35%	7
Text Search	25%	5
Classification of Shots	40%	8
Motion Detection	30%	6
Visual Highlight Search	65%	13
Emotion & Atmosphere Detection	60%	12
Audio Event Detection	70%	14
Rhythmical Retrieval	15%	3
Sound Bite Retrieval	85%	17
Speech Search	50%	10
Human Activity Recognition	45%	9
Other	10%	2

After giving participants an overview of the scope of intelligent video retrieval they were asked to evaluate its overall meaningfulness for the area of trailer editing. Results clearly show that opinions are divided on this issue. Figure 3 shows that there is no tendency towards a generally positive or negative opinion but that a high number of answers can be found in midrange. These results could be interpreted as uncertainty and cautiousness about the perspectives of this product (combined with a certain positive or negative expectation) and could also indicate the lack of detailed knowledge about the

concept that was not further explained. As IDEA is a totally new approach results like these are to be expected and should not be overrated as they only show user's expectations and are not based on experience. Ultimately the results can however be appraised positively as 35% of the participants expect intelligent video retrieval for trailer editing to be somewhat useful and another 30% are neutral or uncertain about its realization.

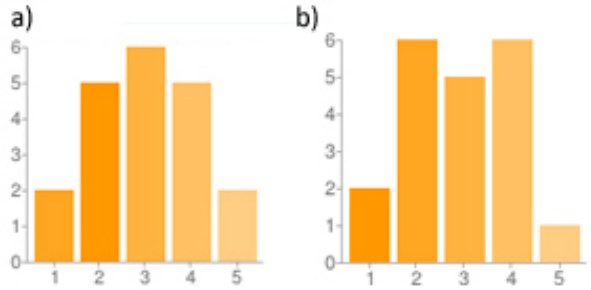

Figure 3 Overall opinion on a) intelligent video retrieval for trailer editing and b) eye trace and motion visualization (1 = very useful, 5 = not useful)

The survey also showed that that the overall priority ranking by trailer producers concerning their editing criteria corresponds with Murch's personal ranking. Emotion and Story achieved the highest priority for editing decisions and aspects strongly related to movement (Eye-Trace, Rhythm) are ranked third and fourth place before Editing Grammar. It has to be mentioned though, that the range of answers was very diverse which indicates that producers' opinions can diverge from each other strongly and that editing decisions are highly individual. Looking at the average grades given by the participants it can be seen that to some extent all of these categories seem to be important in the editing process (see Table 2).

Table 2. Survey results on Murch's "Rule of Six"

Ranking	Category	Priority Percentage
1	Emotion	23,80%
2	Story	21,93%
3	Rhythm	21,12%
4	Eye Trace	18,45%
5	Editing Grammar	14,71%

Although only 30% of all surveyed trailer producers found that the implementation of motion detection is useful for trailer generation, the survey also showed that 40% liked the combined visualization of motion and eye trace and that at least 25% have an open mind about its implementation for trailer generation. 35% of all participants did not believe that eye trace algorithms are helpful (see Figure 3). However, because of the controversy of its applicability for trailer generation it was our intent to further research this special area and to investigate its possibilities for IDEA in the following practical part.

3.3.3 IDEA's Intelligent Search Concept

IDEA defines six semantic levels on which video browsing can be applied. As we deal with trailers IDEA proposes that the highest semantic structure should be a highlight summarization, followed by the video itself, then further structured by scenes, groups of shots (video groups), shot and key frames. As scenes, shots and frames are self-explanatory due to their natural video origin and highlight summarization is further explained in the chapter on "visualization/abstraction", this only leaves the level of groups unexplored. The group level in IDEA's context can also be described as the "exploratory browsing level" and is intended to group video shots after certain high-level concepts. The "exploratory browsing level", though, is not limited to groups only. It can also be browsed using threads or clouds instead.

Consequently, IDEA's overall browsing structure (see Figure 4) guarantees the right size of scale for each individual video clip by letting each semantic level be browsed and scrubbed individually. This concept also meets the needs of the two trailer producers mentioned in the survey who suggested "structure detection algorithms" as useful for trailer generation. The second search concept focuses on the area of video retrieval and therefore on the inquiry of special trailer needs. In contrast to the index of a book, video content for trailer generation cannot be structured that easily as explained in the problem definition (see semantic gap, etc.). However, the first aspect that can be defined here is the observation that trailer producers usually look for certain events in trailer footage. These events cannot be determined by the logical hierarchical structure of the video material itself as their duration varies and can overlap segments such as shots and scenes. For example, a movie quote usually lasts longer than a shot but does not fill a whole scene. Although events can of course be roughly displayed by the collection of shots or the overall scene it is found in it is probably more efficient for trailer pro-

ducers if their search inquiry exactly results in the selection of frames that contain this certain event. Therefore IDEA's concept suggests treating event search separately in execution and visualization.

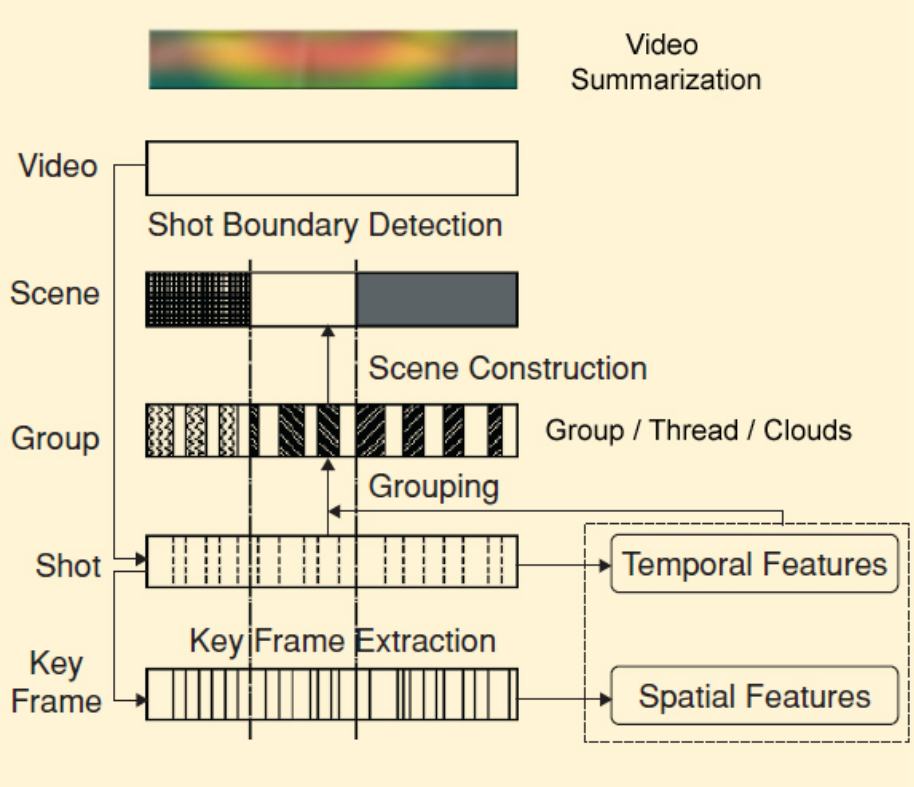

Figure 4 IDEA's browsing structure based on the structure of "hierarchical video presentation for scripted content" by [24]

As displayed in Figure 5 a first general structure of search criteria was developed according to suggestions and results that were determined by the conducted survey. It can be categorized into the following levels of inquiry:

- The Frame Level (deals with visual aspects in a single frame – no time correlation)
- The Shot Level (analyses content within the duration of a shot)
- The Scene Level (analyses content within the duration of a scene)
- The Event Level (deals with trailer-characteristic audiovisual events)

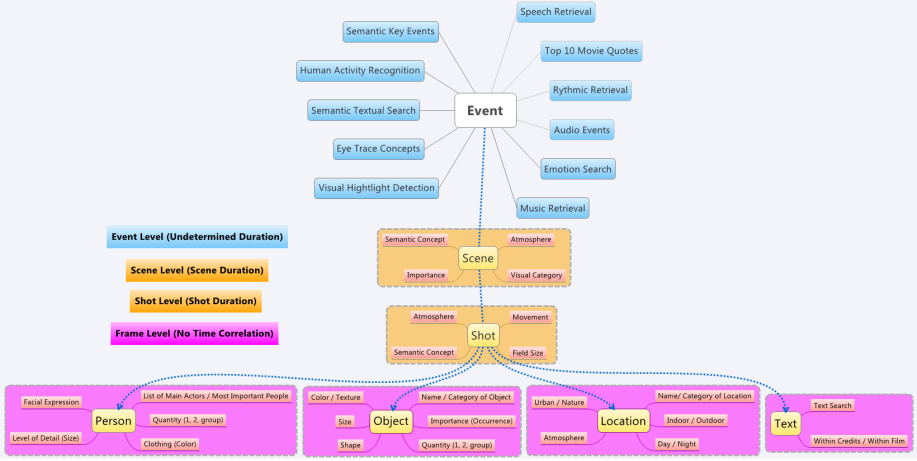

Figure 5 Search Criteria for IDEA

All these Levels and subcategories of inquiry can be combined or individually applied to form the final request. Depending on the level of search there are four possibilities of how results can be displayed:

- Level 1 (Frame Level): In key frames or series of adjacent frames
- Level 2 (Shot Level): In shots
- Level 3 (Scene Level): In scenes
- Level 4 (Event Level): in smart selections of audiovisual parts (see project "Silver")

3.3.4 IDEA's Technical Realization

Although many suggestions about IDEA's technical realization were made in the course of [14] this paper cannot provide a detailed review due to its immense scope. However the following areas of research were treated using methods for audio-visual segmentation and analysis:

- Segmentation:
  - Shot Boundary Detection
  - Key Frame Extraction
  - Scene Change Detection
- Analysis:
  - Person Detection
  - Object Detection / Recurring Motifs
  - Location Identification
  - Text Search
  - Atmosphere, Mood & Genre Detection

- Movie Shot Type Classification
- Motion Detection / Estimation
- Semantic Concept Detection / Semantic Textual Search
- Human Activity Recognition
- Semantic Key Event / Narration Detection
- Emotion Stimuli Search
- Speech Retrieval / Transcription
- Audio Event Detection / Music Information Retrieval
- Rhythmic Retrieval

4 Concept Realization

Measuring the possible value of IDEA without actually having applied it is almost impossible. Therefore this chapter is intended to give a first insight into functions of IDEA and their possible results. As this paper is not meant to actually deal with the implementation of algorithms, this first practical attempt was conducted with the help of Katrin Lasinger in cooperation with Vienna University of Technology. With the help of Katrin Lasinger results were evaluated and interpreted by using the synergetic effects of our two areas of expertise.

4.1 First Practical Approach – Motion Detection and Eye Trace Visualization

Our project was divided into two parts: implementation of an eye trace/gaze prediction algorithm and the execution of an eye-tracking viewer study. Extracted information from this study was used to evaluate the quality and performance of our algorithm and to describe further findings concerning the question of applicability of intelligent algorithms for trailer production.

4.1.1 Eye Trace Algorithm

IDEA's eye trace algorithm focuses on the extraction of simple features instead of machine learning approaches. The advantage of our approach is that no training data and less computing time is required. This is especially beneficial for IDEA as its execution does not depend on any databases and can already be used at this early stage of realization. Nevertheless, the project tried to achieve universally valid results about IDEA's applicability based on

the assumption that previous research indicates that people in generally tend to look at the same points of interest within a radius of 12% of the movie scene.[10]. Based on these investigations and the results of our eye-trace study the algorithm focuses on the detection of motion and human faces. A single focus point is detected, representing the predicted eye-trace.

4.1.2 Eye-tracking Viewer Study

The eye-tracking viewer study was carried out at St. Pölten University of Applied Science using SMI's "BeGaze" eye tracking analysis software[1]. 18 participants (six men, twelve women) were invited to view seven different short video clips including award-winning movie and TV trailers as well as two excerpts from Hollywood-movies. Figures 6 and 7 give image examples of the video clips enhanced with gaze information and clearly show the focus on people and movement.

Figure 6 Heat Map from "Luther"

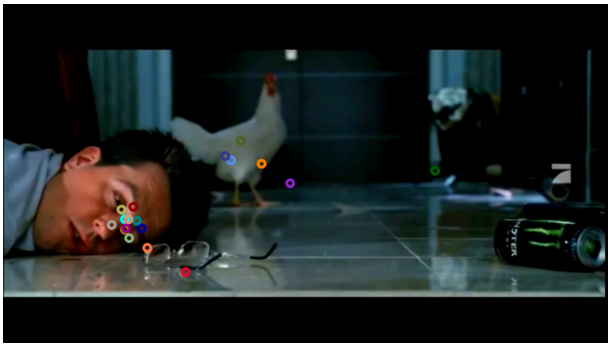

Figure 7 Bee Swarm from ProSieben Megablockbuster Autumn 2011

1 http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/begaze-analysis-software.html

4.1.3 Results

The first question dealt with was whether eye trace is applicable for trailer production at all, assuming that the audience can only focus on the screen centre due to it's the fast pace. Considering the resulting gridded AOI's data views of our study this assumption might seem true at first glance (see Figure 8) as the results clearly show that the screen centre, even in slow paced movie excerpts, received the most dwell time by viewers. However, this finding is not surprising as human perception is naturally centred (and balanced), and no audiovisual media would frame its content dominantly unbalanced over time. Hence, for this question of research it was more relevant to take a look at the position range of views rather than their quantity of dwell time. Therefore one can argue that if the theory of stationary user perception was true, views would exclusively focus on the screen centre and gridded AOI's data views would show no dwell time on the edges of the screen. Reassessing this theory one can also take a look at the enhanced video clips. Though this may only be true for short periods of time bee swarms, heat maps and focus maps clearly show deviations of attention from the screen centre (see Figure 7).

Another very useful discovery made during the eye-tracking study was the fact that if there is a larger shift of attention from one point on the screen to another (mostly caused by cuts) the inertia of human perception causes viewers to dwell on the focus point of the previous frame for a certain time (about 11–14 frames depending on distance) before focusing on the next point of interest. In editing those jumps of attention should be avoided, if possible, as they often disrupt the flow of movement and rhythm. Nevertheless, especially in trailer editing such cuts are often inevitable and sometimes acceptable if the viewer is given enough time to adjust. It is therefore even more important for trailer editing and IDEA to provide a tool that visualizes these jumps and advises trailer producers on the shot time appropriate for viewers.

The extracted fixation points from the eye-tracking viewer study were used as ground truth for the evaluation of our implemented eye trace algorithm. For each frame the extracted points were clustered in a centre focus point which was then compared to the predicted focus point of our algorithm. The Euclidean distance between these two points was calculated. Since there is no exact focus position and since viewers usually do not look at the exactly same position we introduced an acceptance radius for the focus position. If the estimated focus point was within the acceptance radius the position was detected correctly. As previously mentioned, Goldstein et al. claim that peo-

ple tend to look at the same points of interest within a radius of 12% of the movie scene [10]. Based on this claim we defined our acceptance radius as 12% of the movie scene width. All error rates of experiments conducted within the course of this thesis can be seen in the Table 3.

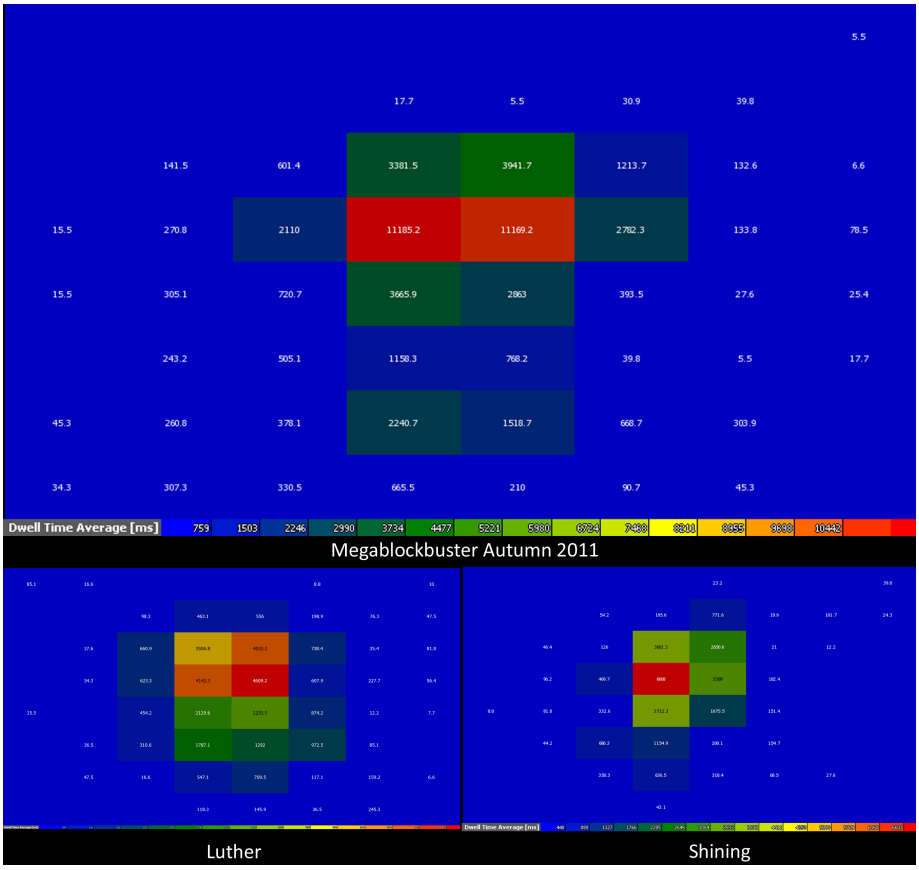

Figure 8 gridded AOI's data views from eye-tracking study
(colour range from red to blue and numbers indicating dwell time in ms)

Table 3: Error Rates of IDEA's eye trace algorithm

Video Clips	Error Rate
Transformers 3	44.5%
Luther	45.7%
Megablockbuster Autumn 2011	45.9%
Tree of Life	40.3%
Pirate of the Carribbean	48.5%
Shining	29.7%

Figures 9 and 10 show two exemplary frames of the movie clip "Pirates of the Caribbean: On Stranger Tides" and "Transformer 3".

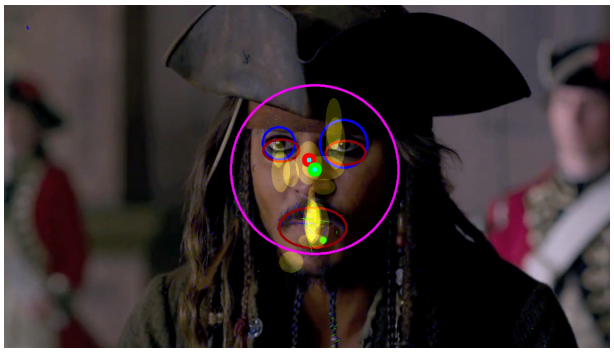

Figure 9 Frame from "Pirates of the Caribbean: On Stranger Tides" with correct focus point prediction

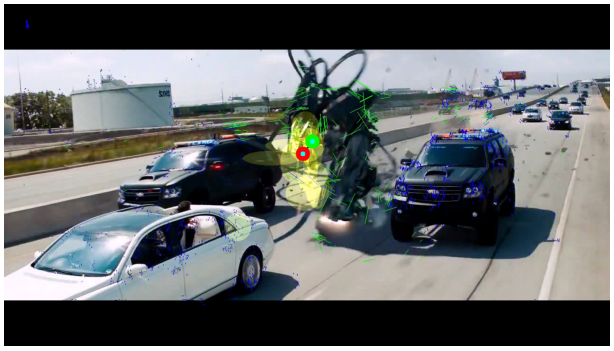

Figure 10 Frame from "Transformer 3" with correct focus point prediction

5 Conclusion

Generating an intelligent tool for editing assistance is indeed a challenging task. On the one hand this is due to the nature of editing as trailer producers and other editing specialists have to draw on a complex combination of experience, know-how and intuition during the creation phase. It is a sensitive and difficult task to define computer-based solutions that can be fitted into this creative and individual working process. On the other hand certain limitations of technology make it difficult to interpret complex correlations of real life (and therefore video). These might never be resolved completely and constrain the applicability of intelligent algorithms. Despite all this we be-

lieve that this paper clearly shows the distinctive advantages of Intelligent Digital Editing Assistance and its promising future.

The main objective of this paper was to clarify whether intelligent algorithms can be applied meaningfully to the process of trailer production or not and to create a basic concept for an applicable tool (IDEA) by proposing possible approaches for realization. The survey results showed that trailer editing is focused on the trailer's impact on the viewers (emotion, story, rhythm, eye trace) rather than editing grammar, and that opinions are deeply divided on the usefulness of computer-based help in this production process. The survey, however, also showed that most trailer producers wish for an intelligent video retrieval system in the areas of people identification (90%), sound bite detection (85%), or audio event detection (70%). Taking the survey's results and the unique structure of scripted media into consideration a comprehensive concept for search criteria, based on different levels of inquiry, was developed. The large potential of certain approaches was shown but it was also pointed out that these are fuzzy areas of investigation.

In the last chapter, the practical part of this thesis, it was tried to further clarify the issue of application with the example of eye trace visualization. The aim was to develop an algorithm that helps trailer producers to determine their editing flow of attention and to avoid overloading viewers with busy trailers they cannot process. The algorithm was then evaluated by a comprehensive eye-tracking study (18 participants). Despite the theory that the viewers' attention in trailers is stationary the eye tracking study clearly showed that shifts of attention go hand in hand with a certain dwell time, thus making it impossible for the viewers to instantly process the content of a new scene. The algorithm, which was developed in cooperation with Vienna University of Technology by Katrin Lasinger, included face detection and motion detection approaches but failed to achieve an appropriate success rate for trailer editing. Nevertheless it succeeded in showing the potential and usefulness of such algorithms for IDEA.

Both practical approaches clearly demonstrated that appropriate eye trace visualization could be very useful in the trailer editing process and enabled us to make important discoveries concerning human perception of video.

Summarizing all these previous findings it can be stated that although IDEA might still face some serious challenges some advantages and approaches introduced in this paper could improve a trailer producer's working process immensely. Although the survey showed that many trailer producers fear to be under even more pressure if computer-based help is available this

is an effect that is certainly *never* been intended by this project as repetitively stated in the theoretical discussion. To stress this once again, the concept of *IDEA aims to be supplemental to the individual and highly valuable work of a trailer producer. It is meant to shorten tedious work processes to give trailer producers more time for the creative part of their work.* This aspect should always be taken into consideration and any misuse of this concept should be prevented to guarantee the production of high-quality trailers.

References

[1] Apple – Final Cut Pro X – Top Features: www.apple.com/finalcutpro/top-features/#powerful. Accessed: 2012-06-09.

[2] Avid | PhraseFind: http://www.avid.com/US/products/phrasefind. Accessed: 2012-06-09.

[3] Bordwell, D. & Thompson, K. (2010). *Film art: an introduction*. McGraw-Hill.

[4] Calistru, C., Ribeiro, C. & David, G. (2012). High-Dimensional Indexing for Video Retrieval, in: I. Karydis (Ed.): *Multimedia – A Multidisciplinary Approach to Complex Issues*. Available from: www.intechopen.com/books/multimedia-a-multidisciplinary-approach-to-complex-issues/high-dimensional-indexing-for-video-retrieval. Accessed: 2012-10-28.

[5] Casares, J. P. (2001). Silver: an intelligent video editor, in: *CHI'01 extended abstracts on Human factors in computing systems* (2001), 425–426.

[6] Celma, Ò. (2006). Foafing the music: Bridging the semantic gap in music recommendation, in: *Proceedings of the 5th International Semantic Web Conference* (2006), 927–934.

[7] Digital video editing software | Adobe Premiere Pro CS6 – Features: www.adobe.com/products/premiere/features.html. Accessed: 2012-06-09.

[8] Ekin, A., Tekalp, A. M. & Mehrotra, R. (2004). Integrated semantic-syntactic video modeling for search and browsing, in: *IEEE Transactions on Multimedia*. 6, 6 (Dec. 2004), 839–851.

[9] Felser, G. (2001). *Werbe- und Konsumentenpsychologie*. Spektrum-Akademischer Verlag.

[10] Goldstein, R. B., Woods, R. L. & Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place?, in: *Comput. Biol. Med.* 37, 7 (Jul. 2007), 957–964.

[11] Hediger, V. (2001). *Verführung zum Film: der amerikanische Kinotrailer seit 1912*. Schüren.

[12] Hermes, T. & Schultz, C. (2006). Automatic Generation of Hollywood-like Movie Trailers. Available from: http://www.eculturefactory.de/eculturetrends/download/hermes_schulz.pdf. Accessed: 2012-10-28.

[13] Hu, W., Xie, N., Li, L., Zeng, X. and Maybank, S. (2011). A Survey on Visual Content-Based Video Indexing and Retrieval, in: *Systems Man and Cybernetics Part C Applications and Reviews IEEE Transactions on*. 41, 6 (2011), 797–819.

[14] Janda, I. (2010). *Lineare vs. Interaktive Trailer*. Bachelor Thesis. FH St. Pölten.

[15] Janda, I. (2012). *I.D.E.A. – "Intelligent Digital Editing Assistance" for Trailer Production*. Master Thesis. FH St. Pölten.

[16] Ksibi, A., Elleuch, N., Ben Ammar, A. & Alimi, A. M. (2011). Semi-automatic soft collaborative annotation for semantic video indexing, in: *EUROCON – International Conference on Computer as a Tool (EUROCON), 2011 IEEE* (Apr. 2011), 1–6.

[17] Minding the Gap: 2012. http://www.computer.org/csdl/mags/mu/2012/01/mmu2012010002.html#bibmmu20120100024. Accessed: 2012-04-14.

[18] Schmidt, S. J. (2004). *Handbuch Werbung*. LIT Verlag Münster.

[19] Shipman, F., Girgensohn, A. & Wilcox, L. (2008). Authoring, viewing, and generating hypervideo: An overview of Hyper-Hitchcock, in: *ACM Trans. Multimedia Comput. Commun. Appl.* 5, 2 (Nov. 2008), 15:1–15:19.

[20] Smeulders, A. W. M., Member, S., Worring, M., Santini, S., Gupta, A. & Jain, R. (2000). Content-based image retrieval at the end of the early years, in: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 22, (2000), 1349–1380.

[21] Smith, J. R. (2007). The real problem of bridging the "semantic gap", in: *Proceedings of the 2007 international conference on Multimedia content analysis and mining* (Berlin, Heidelberg, 2007), 16–17.

[22] Snoek, C. G. M. & Worring, M. (2005). Multimedia event-based video indexing using time intervals, in: *Multimedia, IEEE Transactions on*. 7, 4 (Aug. 2005), 638–647.

[23] Ueda, H., Miyatake, T. & Yoshizawa, S. (1991). IMPACT: an interactive natural-motion-picture dedicated multimedia authoring system, in: *CHI '91 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM New York, NY, 343–350.

[24] Xiong, Z., Zhou, X. S., Tian, Q., Yong, R. & Huang, T. S. (2006). Semantic retrieval of video – review of research on video retrieval in meetings, movies and broadcast news, and sports, in: *Signal Processing Magazine, IEEE*. 23, 2 (Mar. 2006), 18 –27.

[25] Yan, R. & Hauptmann, A. (2007). A review of text and image retrieval approaches for broadcast news video, in: *Information Retrieval* 10, 4 (2007), 445–484.

[26] Zhong, D. (2001). *Segmentation, Index and Summarization of Digital Video Content*. Doctoral Thesis. Graduate School of Arts and Science, Columbia University.