Computational Abstraction of Films for Quantitave Analysis of Cinematography

COMPUTATIONAL ABSTRACTION OF FILMS
for QUANTITATIVE ANALYSIS OF CINEMATOGRAPHY

Gökçe Taşkan

Master's Thesis
Aalto University
School of Art and Design
Department of Media

March 2011

Graphic Design: Ayça Özbank

Aalto University
School of Art and Design

ABSTRACT

Author		*Year of publication*
Gökçe Taşkan		2011
Department	*Degree programme*	
Department of Media	MA in New Media	
Title		
Computational Abstraction of Films for Quantitative Analysis of Cinematography		
Type of work	*Language*	*Number of pages*
Production based	English	70

Abstract

Currently, film viewers' options for getting objective information about films before watching them, are limited. Comparisons are even harder to find and often require extensive film knowledge both by the author and the reader. Such comparisons are inherently subjective, therefore they limit the possibilities for scalable and effective statistical analyses. Apart from trailers, information about films cannot reach viewers audibly or visibly, which seems absurd considering the very nature of film.

The thesis examines repeatable quantification methods for computationally abstracting films in order to extract informative data for visualizations and further statistical analyses. Theoretical background empowered by multidisciplinary approach and design processes are described. Visualizations of analyses are provided and evaluated for their accuracy and efficiency.

Throughout the thesis foundations for the future automated quantification player/plugin, are described aiming to facilitate further developments. Theoretical structures of the website which may act as a gateway that collects and provides data for statistical cinematic research are also discussed.

Materials

http://labs.filmogram.com

Keywords

film quantification, video abstraction, film visualization

Acknowledgements

I would like to thank Media Lab Helsinki for providing me an inspirational and openminded environment to realize this thesis. I would especially like to thank my advisor Markku Reunanen for his dedication, knowledge and valuable feedback.

I am grateful to Niko Knappe, Lauri Kainulainen and Jonathan Cremieux for the fruitful chats and valuable comments that nourished this work. I also want to thank my two opponents Ramyah Gowrishankar, Dimitri Paile for their comments and Nuno Correia, Ahmet Atıf Akın, Enis Özbank for their in depth draft reviews. I appreciate the help from Diane Brown proofreading the thesis.

This work would not have been possible without unconditional support from my family and Ayça Özbank.

Table of Contents

1. Introduction

Invention of Cinématographe in 1895 created a possibility for recording movements. Filmmakers experimented with this new medium by documenting their surroundings. Only after some time did it evolve to be considered an art form. Because the cinema's birth was driven by technology, the two have always been closely related and are in constant interaction with each other. Further inventions and additions helped the cinema industry to grow and create richer experiences; sound, montage, color film, widescreen, special effects, 3D to name a few. One of the things that has not changed since the beginning is the laborious production process. To facilitate post-production, during production phase data about filming—camera lens type, exposure, light conditions, script changes—are written on paper notebooks (Stump 2008). Today, films are mostly made digitally from the first steps until the end of post-production. Modern digital cameras while recording imagery, can save a limited set of data, information about filming from itself and connected peripherals; focus, zoom, f-stop, pan, tilt, boom, dolly data is written on small memory cards.

It is unknown how efficiently this data is used internally by film studios but one thing for sure is none of it is analyzed or made available to the viewer. Maybe analyzing such data is irrelevant or as useless as measuring the ink used by a writer, but if we consider that a film's reach to the viewer has not evolved much, such data might be useful finding new ways to extend this reach. A film offers itself and only itself as a product to be seen from beginning till end. Viewers do no have many

options, to get information before watching. They either look at starred ratings—which can be very subjective and misleading depending on the methods used to collect the ratings—or read reviews, which are, again, very prone to subjectivity, or look at the trailer and at the poster. Reviews can offer in depth accurate analysis of films albeit they don't offer scalable possibilities to compare even different titles, let alone directors, eras or country of origins. A scalable system where films can be analyzed, compared based on their quantitative summaries is lacking.

> *"Now a real film must not be able to be told... The future belongs to the film that cannot be told."* [1] (Dulac cited in Cahill 2008)

As quoted and explained further by Cahill (2008), "A real film requires a cinematography—an expressive system of writing, a language of cinema—whose immediacy evades the constraints of verbal and written language, in order to present (to make present), at the scene of projection, that which can neither be spoken nor written." Dulac's "Visual Film" doctrine and Kubelka's hesitation in regards to talking about his films, in case he might spoil them (Mekas 1967) indicates that films are better not represented by words. Hence, this thesis investigates the use of information visualization for abstraction and representation of films.

In order to abstract and represent films, some form of summarization technique has to be implemented. Lack of standardization of video data across broadcast and film producers obstructs data generation and sharing. Under these circumstances, modernizing Salt's approach and enriching it with various other methods would be a way to provide a heuristic solution for film quantification. Such quantification integrates immutable metadata with mutable extracted and user entered data. Further description can be found in chapter "3. Quantification of Films".

For abstracting films, a multidisciplinary approach is proposed, bringing together some existing analyses and quantification methods from different disciplines, while tweaking them to suit the requirements. Later, adequate data organization, analysis opportunities and clear visualizations will be provided to the film viewers enabling them

1 Or, un vrai film ne doit pas pouvoir se raconter... L'avenir est au film qui ne pourra se raconter... —Germaine Dulac to draw their own conclusions either by looking at ready-made visualizations or by developing their own applications with available data.

This research is positioned, separately from film theory and stylistic film analysis, as one of the aims is to deliver the right tools for researchers and film viewers. Nevertheless, it should be noted that it would be impossible and unreasonable to avoid film theory as some of the film abstraction methods described throughout the thesis are based on this theory. They also contribute to the discussion in chapter "5. Reflections on Visualizations" for justification of proposed abstractions and visualizations.

1.1. Motivation and Goals

The main goal of this thesis is to contribute to the discussion of film quantification by bringing adequate tools to the table. Previous experience with computer vision in interactive environments is taken as a starting point and they are refined to fit this context. Summarization tools will be developed using various programming languages depending on the needs. Resulting analysis software is going to be released as open source, creating possibility for collaboration among the small cinema statistics community and/or experts in analysis methods.

Furthermore, a downloadable video player software, which is based on current open source video player projects will be provided. When a movie source is watched through this player software, it will be analyzed and quantified. Then, software will submit analysis results automatically to the statistics database in the server. On the front-end of the server there will be a website in which film watchers will be able to look for collected movie statistics. The same website will act as a gateway for the community to "play" with data. The film statistics database will be organic and will be constantly growing because it is theoretically and practically impossible to visualize and organize all the data perfectly right from the beginning. To adapt to these changes and to enable the community to make their own little applications or visualizations an

Application Programming Interface (API) will be provided. API's have been very successful in various social networks providing data for the public to play with (Benslimane, Dustdar and Sheth 2008). Use of API in this project is described in "4.3. API" section.

Even though earlier research or artworks touching similar subjects exist, the novel system that is proposed for automated quantification and analysis, enables a wide range of possibilities for comparative film studies and creates the possibility to extend films' reach to the audience. It is hoped that this research will encourage derivative works.

On the personal level, the aim is to provide a foundation for a complex and objective movie recommendation system that is based on analysis data generated in this thesis in correlation with personal film watching history that is collected by a developed movie player or plugin software. In addition to data analysis and statistics, a simple implementation of a user login and watch log function similar to Last.fm's AudioScrobbler [2] system, would provide the necessary tools for achieving this goal.

2 AudioScrobbler is a plugin for logging music listening history for Last.fm users. When installed, it tracks user's songs and sends them to the website, allowing various statistical possibilities.

1.2. Scope

The scope of this thesis is limited to fiction films. Nonfiction films have been left out because of their unpredictability and the director's lessened control on image and camera work due to their purpose, content or circumstances. Nonfiction films include documentaries, sports, instructional, news and home videos.

There are many previous research works focusing on improving media management systems; they implement video identification, genre (Zhou et al. 2010) and mood classification methods (Salway and Graham 2003). Even though some of their methods are used for quantification, implications on a semantic level and other high level feature extractions (e.g. automated movie genre classifications, distinctions between indoor/outdoor, landscape/people shots) are outside of the scope of this thesis.

At this stage of the research, the focus has been given to the visible and audible structures of films. Trivial elements such as references to other films, goofs, hours spent in post-production, etc., are at the moment outside of scope as well because they can be subjective, computationally hard to collect and less relevant to the film itself than the visible and audible properties.

1.3. Structure of the thesis

The thesis can be divided theoretically into three main parts. Type conversions occupy the core of the logic behind such division. Films are converted to data, then the data is organized and later analyzed to be presented to the audience as visualizations or through API. (See Figure 1.1)

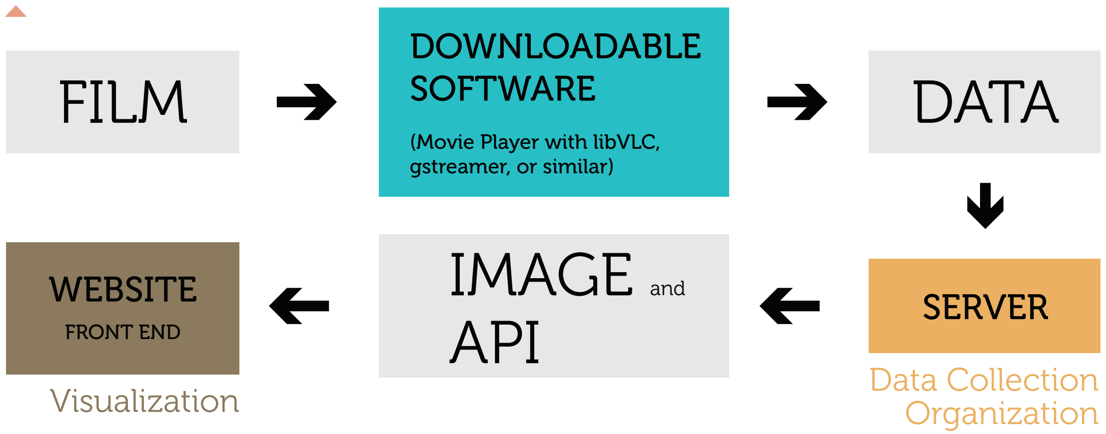


Figure 1.1
Three main parts of the thesis and flow of data

The first part deals with quantification; the first task is to efficiently extract data from films in greater quantity and accuracy. To accomplish this I plan to use implementations from computer vision, audio feature extraction and text analysis. Other tasks include involving viewers to generate data and gathering billing data from various sources. Further discussion about methodology is described in "3. Quantification of Films" chapter.

In the second part, methods from information architecture and statistical analysis adopted to organize and examine the data. This part is not widely covered in this thesis as it is used as a mediatory tool facilitating connection between first and third section. In the mean time, it can be a subject of another thesis by itself and it can serve as a starting point for further research to implement a film recommendation system.

In the last and third part, use of visualization design aims to deliver clear, easy to understand, quickly accessible visual "summaries" of films. The goal with these is to enable visual comparisons between films and use them as an instrument to discover patterns in the data along with mathematical analysis.

As previously mentioned, such divisions are theoretical and in practice quantification and visualization are bound together and represented in the same chapter "3. Quantification of Films". Challenges about data collection and storage are described in chapter "4. Data Collection and Organization". The strengths and shortcomings of visualizations are discussed in chapter "5. Reflections on Visualizations".

2. Context

This chapter familiarizes the reader with various disciplines and key concepts that are referenced throughout the thesis. Ideas from each discipline are interconnected as most of them are used in every step. As described in previous chapter, first of the three theoretical parts of the thesis deals with quantification which makes use of information visualization, statistics, computer vision, audio feature extraction, content summarization as well as statistical cinematic analysis in order to efficiently and accurately abstract films. For the second part, data collection and organization, statistics and open source software development are relevant. Finally, the third part makes more extensive use of information visualization. A brief history of about these disciplines are also provided in their respective sections to explain their relevance while describing examples use cases, necessities, benefits and drawbacks.

2.1. Statistics, Information Visualization

Statistics have been used widely since seventeenth century (Chen et al. 2008), however it was William Playfair who first introduced statistical graph charts in 1786, being aware of their power for effectively communicating data (Playfair et al. 2005). They have been beneficial in many fields of work since their early usages. In medicine, by monitoring and analyzing the health of soldiers Florence Nightingale discovered patterns which were used to campaign for improved sanitary conditions in battlefield treatments (Chen et al. 2008). Charles Joseph Minard communicated mortality of Napoleon's 1812 Russian Campaign with his "Tableaux Graphiques et Cartes Figuratives" (Tufte 1983). Later, it was Tukey who suggested data analysis should be a branch of statistics and researched various data graphs (Chen et al. 2008). Since then, data visualizations have been integrated into communication and decision-making processes in nearly every field. As Larkin and Simon (1987) agree: the famous saying, "a picture is worth 1,000 words" seems to be true.

Since data visualizations deliver a message and, tell a story to the reader, the importance of their efficiency is vital. Tufte (1983) suggests: "What is to be sought is the clear portrayal of complexity. Not the complication of the simple; visual access to the subtle and the difficult, the revelation of the complex." He further states that what is to be aimed is not how one method can be used to explain something, but rather "how can something be explained?" (Tufte 2006). Thus, it can be said that the purpose of statistical graphics' is to provide a clear, simple, understandable graphical summary of a rather complex data so that, not only trained researchers but also untrained viewers can grasp it. Here, the visualization designer's duty is to analyze the data thoroughly, by developing or using existing frameworks and, individual graphics according to data properties. As Gershon and Page state, "Information visualization combines aspects of imaging, graphics, scientific visualization, and human-computer and human-information interactions, as well as information technology" (2001). Another recent article in The Economist (2010) suggests that visualization designers are "melding the skills of computer science, statistics, artistic design and storytelling." Thus, information visualization is multidisciplinary by its very nature.

"Storytelling is an ancient art rooted in our common human culture, as well as in our physiology and psychology"; throughout all human existence, and as early as cave drawings, stories have been used to convey information, events, cultural values and experiences. (Gershon and Page 2001) The development of writing, printing, photography, film and other newer mediums provide increasingly sophisticated means to tell stories. Minard excels in storytelling with his Cartes Figurative (Tufte 1983). As Segel and Heer describe data visualization has much to benefit from storytelling (2010). Harris defines himself more of a storyteller than visualization designer (Segel and Heer 2010; Danzico 2008). His projects Whale Hunt [3], We Feel Fine [4], I Want You To Want Me [5] communicate data with stories. User scenarios which derive from storytelling have been an integral part of Interaction Design. In sum, storytelling is a very powerful tool and it is employed in diverse mediums as well as information visualization.

3 http://www.thewhalehunt.org

4 http://www.wefeelfine.org

5 http://www.iwantyoutowantme.org

2.2. Computer Vision

Computer vision is a subfield of computer science, focusing on image related computational algorithms. Various methodologies and algorithms help construct, meaningful descriptions from digital images. It is an active research field with more than 50 years of history. Some subfields include, image processing, image analysis, machine vision and pattern recognition.

Image processing and analysis is more relevant for this work. Most computer vision applications use these as their foundation. Data that is abstracted from images with various methods are later used to form interpretations about the image content. The most familiar image processing use case would be image editing softwares. Various filters can enable image enhancing for visual properties (photo retouching) as well as computational properties (contrasting various objects in order to highlight them for following operations).

Human perception is a very complex process which involves empirical judgement about surrounding events and objects. More specifically, visual perception is full of low-level capabilities such as light, depth, color, object recognition and high-level capabilities based on knowledge (Ballard and Brown 1982). Imitating this kind of biological system is impossible for machine vision; even achieving simple goals can be very challenging. Therefore, computer vision is a very vast and ever-

growing field. Such complexity requires efficiency measurements, as it is not yet possible to reach perfect analysis results. New and more efficient algorithms are constantly researched for implementations in medicine, manufacturing, military, surveillance, entertainment and many more fields.

For use of computer vision in this thesis, the openCV[6] library is selected as it focuses on real-time image processing. It offers various ready-to-use functions based on related research works. OpenCV is cross-platform and open source while interfaces are already developed for most common programming languages.

For decades, mathematicians and computer scientists have been the only ones who are capable of developing and implementing computer vision. However, the development of relevant new programming environments such as Processing, PureData and openFrameworks, which are user friendly, even to non-programmers, has enabled wider audiences to develop their own applications. This has lead to a prototyping culture where artists, designers and hobbyists started programming for their own needs without much in-depth knowledge about the inner workings of complex algorithms. Of special significance, usage in entertainment increased awareness; Nintendo Wii[7], PlayStation Eye[8] and Kinect[9] offered ready-to-use computer vision solutions to game developers, facilitating design of complex interaction methods for gameplay. Lately, inexpensive small webcams are increasingly embedded with many consumer electronics, including laptops, mobile phones. Camera based do-it-yourself multitouch implementations (NUI Group; Kaltenbrunner and Bencina 2007), augmented reality ad campaigns (Ray-ban 2010) and mobile applications such as Layar[10], Wikitude AR Travel Guide[11], TAT Augmented ID[12] show the wide range of uses, as well as the productivity of utilizing computer vision in diverse areas. Ready to implement software libraries and frameworks such as EyesWeb[13], Touchlib (Wallin 2006), CCV[14], ARToolKit (Kato 1999) and many others empower non-programmers or novice coders with the diverse capabilities of computer vision.

6 http://opencv.willowgarage.com/wiki

7 Wii motion controller held by player, featuring an infrared camera, tracks a stationary infrared light emitter station.

8 Eye is a camera itself and used mostly for augmented reality applications for Playstation.

9 Kinect is a Xbox stationary bundle with two cameras and infrared lights. It provides motion tracking without any other peripherals, thus it gives freedom of movement to the player.

10 Layar is a mobile application that allows developers to add their own augmented layers. http://www.layar.com

11 An application that shows Wikipedia information based on users location, for sightseeing during traveling.

12 Augmented ID uses face recognition routines to visualize digital identities of people that are recorded by camera.

13 http://musart.dist.unige.it/EywMain.html

14 Community Core Vision by NUI Group, http://ccv.nuigroup.com

2.3. Audio Feature Extraction

Improved distribution of digital audio files through online stores and peer-to-peer file sharing causes immense growth of individual music libraries. Better ways to manage these libraries has been investigated by audio researchers. How to identify, organize, search, classify and retrieve are a few of the problems of management. In order to manage libraries efficiently, audio files are analyzed with various methods to retrieve information. Overall, audio feature extraction is an extensive research area. There are many dedicated communities and research groups.

Various features can be extracted from audio with signal processing: tempo, pitch, timbre, harmony, structure. Some of these are very specific to music and thus they are not widely used for the film audio. There are also nonstructural properties used to describe sound in film, diegetic and non-diegetic, defining the implied source of the sound. Sounds coming from film world are diegetic while externally added sounds such as narrator's voice and mood music considered as non-diegetic. Rasheed and Shah (2002) focusing on audio-based film analysis, propose a film genre classification method based on audible and visible features; for audio they implement a simple method by computing an energy variable from the peaks in sound levels.

2.4. Statistical Cinematic Analysis

Statistical cinematic analysis is a research area that was first explored by Georg Otto Stindt in 1926 (Thompson 2005) and then formally investigated by Barry Salt. It is about quantifying films to conduct a comparative style analysis. One measurement that has been at the core of previous research is ASL (average shot length); Salt measured shot lengths on printed film rolls using a tape measure (1974). He grounds his analysis and research methods on the basic idea that "the form of films noticeably differ from one to another" (2006). Therefore, according to him it should be possible to characterize and distinguish each film individually.

2.4.1. Film Historian's Point of View

As previously mentioned, Salt notes that variables gathered should be based on the attributes that filmmakers use and have control over (2006). He argues that in this way it would be possible to conduct an accurate style analysis. It seems that he focuses more on the director's style rather than the film itself. Calculations for his metrics involve a very lengthy manual process. Even so that he only grabs a portion of a film—generally the first 40 minutes—to analyze and he assumes that the rest of the film adopts the same style (Salt 1974). Apart from shot lengths, he also measures the scale of each shot and camera movements, by manually assessing and writing them down, making the process even lengthier. In an article, Salt claims that he is spends "only" 3 to 12 hours per film for his analysis while Cutting et al. (2010) spend 15 to 36 hours per film, operator and computation time combined. Even though Cutting et al. (2010) implement a semiautomatic system where an operator double checks the cuts after they are calculated by the software, the amount of time spent on each film solely for distinguishing cuts is considerably significant.

> *"...in enabling comparisons that reveal novel patterns of evidence, statistical findings may defy the historian's expectations, and thus stimulate new research and analysis. In any case, the task here is not to substitute statistical analysis for other film-historical methods..."* (O'Brien 2005)

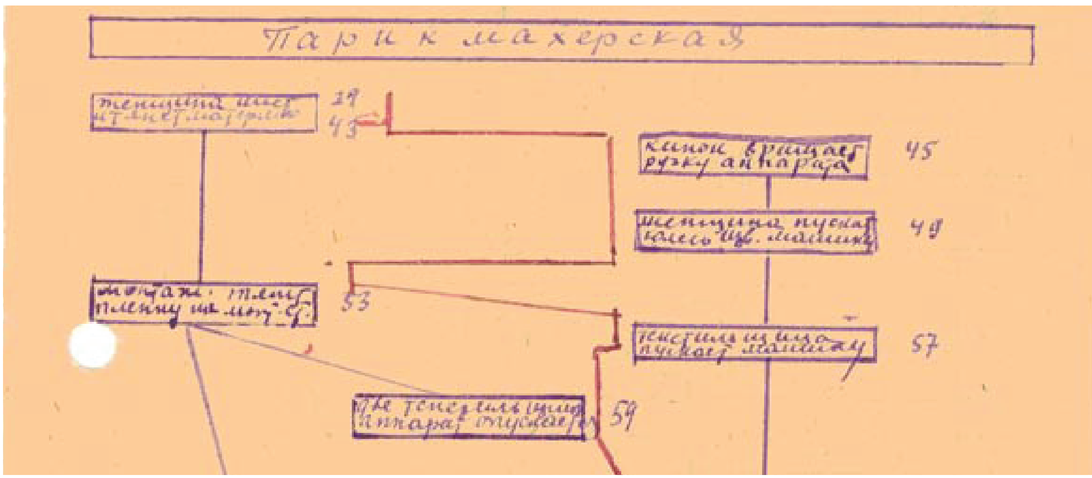
O'Brien suggests that statistical film analysis might create new opportunities for film research but he also warns that it should not be considered as a replacement for other film analysis methods. Bordwell further argues that quantitative analysis tools should be constantly accompanied by qualitative processes such as researcher's film understanding; in his book, he analyzes pace in American films using ASL measurements in correlation with era, director and country of origin data (2006).

2.4.2. Filmmaker's Point of View

From the other side of the looking glass, in the production phase, some directors also use metrics to design their films. Drawings made by Dziga Vertov, a database filmmaker according to Manovich (2002), shows that he has planned and used timing and rhythm while shooting and editing his movies (See Figure 2.1). Kurt Kren's charts show that he also included metrics for planning his films (See Figure 2.2). In Avant-garde cinema, Kubelka makes his films frame by frame, as he states, resulting with very compact and dense films. He argues that cinema has basic rhythm, which derives from film projector's twenty-four impulses. Therefore films have metric rhythm that viewers always feel (Mekas 1967). Kubelka has three metric films: Adebar (1957), Schwechater (1958) and Arnulf Rainer (1960).

Figure 2.1

Excerpt from Vertov's editing interval drawings for Man With a Movie Camera (1929)



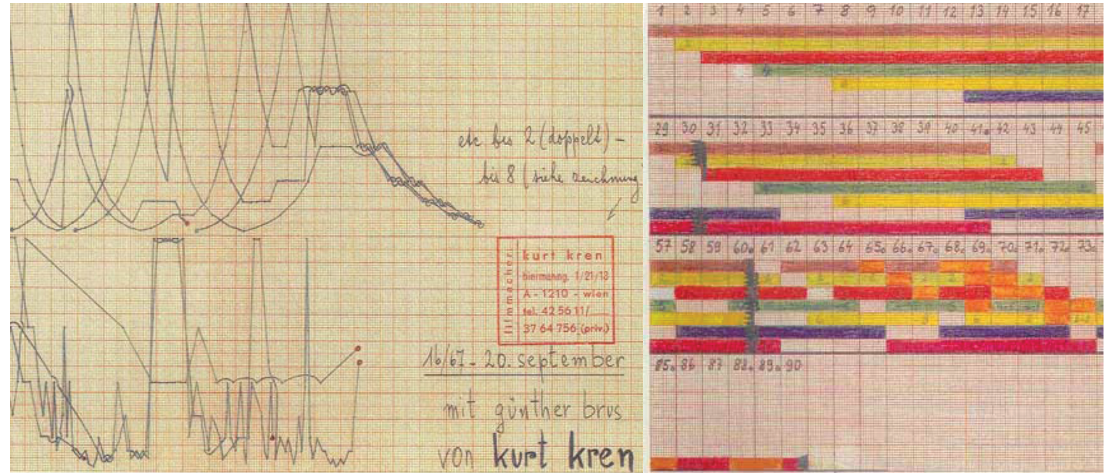


Figure 2.2

Kren's charts for Asylum (1975)

Earlier, Eisenstein defined various montage techniques: tonal, overtonal, intellectual, rhythmic and metric.

> *"He tried various editing patterns, discovering that, for example, film cut metrically to the beat of a typical heart has a profound impact on us precisely because it mirrors our biorhythms."* (Shaw 2004).

However, he also received criticism about formalism towards his metric film October (1928). Although some of the examples were experimental, relying mostly on metrics during film design might not be the best choice outside Avant-Garde cinema. But, evidently usage of metrics is not an unfamiliar concept for filmmakers.

2.4.3. Current Research

Today, statistical cinematic analysis research is still focusing on ASL and relying on lengthy manual measuring process. Rather than using tape measure like Salt did, now researchers use a manual timer soft-

15 A website where a group of scholars publish about their statistical cinematic style analysis research.

ware that is created by cinemetrics.lv[15]. In terms of efficiency it doesn't differ a lot from Salt's methods. It seems that film academics are contributing to their project, however it doesn't attract attention from bigger audience because of the vast workload that it requires.

Their project website also serves as a database for all shot length statistics submitted by volunteers who measured films individually. The database is browsable and each film's ASL graph can be seen separately. Despite having this data in digital format, statistical analysis and visualizations provided are not of great interest. It is not possible to compare different titles.

Reliability of the data gathered in cinemetrics.lv is also questionable since it is not based on a programmed and/or designed repeatable method. It relies on human power, therefore it is highly error-prone. Error rates are likely to increase, especially in films with fast, montaged scenes that have cuts very close to each other. Shower scene in Psycho (2006) and several sequences in Snatch (2000) are some examples that might be hard to mark for a person without stopping and rewinding. At the moment, it seems that digital timer software provided by the website only allow continuous marking.

However, cinemetrics.lv has a big archive, therefore it also seems to have a big volunteer base who manually mark and submit cuts in films. Even though manually measuring cuts is open to human error and the process is time consuming, the resulting cut and shot length data should be more accurate than any existing automated shot boundary detection algorithms. High motion scenes, soft cuts, low video qualities are problematic for automated systems as correct detection rates and false hit rates are not sufficiently good. Nevertheless, it should also be noted that automated systems can extract a very rich palette of data with higher accuracies from films.

2.5. Content Summarization

Summarization has been researched in various fields; MusicBox (Lillie 2008) attempts to create playlists by quantifying music collections and organizing them; Document Cards creates less space consuming visual summaries for text and image based documents (Strobelt et al. 2009). For films, Barry Salt (1974) conducted statistical style analysis quantifying films but only focused on properties that are controlled by the director himself. For this reason, he seems to be interested in quantifying the director, not the film. Later on, other researchers followed his steps by not widening quantitative properties that they use; ASL (average shot length), a term introduced by Salt (1974), occupied the core of their investigations. The methods used to gather ASL has not changed since Salt, even considering the massive increase in computer power and ubiquitous computing[16]. Their process involves watching a film carefully while manually marking each cut in the movie with a digital timer. The data set is also mostly limited to ASL and they don't gather data about other aspects of films.

16 www.cinemetrics.lv author Yuri Tsivian, a professor in the Humanities, as quoted by Bosse and Brisson (2010) in Humanities + Digital Visual Interpretations conference organized by Hyperstudio at MIT in May 2010, is deliberately promoting manual cut marking, arguing that students can enjoy and benefit from the lengthy process. This might be acceptable only to a certain extent as after some level it would be futile to ignore technology and its advantages such as efficiency, rapidity and scalability.

MPEG-7 is an XML based new standard that has been in development for years. MPEG-7 is attempting to merge video with delivered metadata. Another container format for data delivery, MXF (Material Exchange Format) mainly targets production environments for broadcasting (Smith and Schirling 2006). Plenty of other metadata formats are being developed independently for focused purposes by various standards organizations. "However, metadata standardization is proving difficult" as Smith and Schirling claim and they further argue that: "The result is a metadata standards alphabet soup..." (2006). It seems that proper modular standardization needs many more years to be efficient and to be adopted by manufacturers and content creators. Greater public will have to wait to access at production data even if it is intended to be shared.

2.6. Open Source Software Development

Sometimes, software development can be a secretive field where methods and algorithms are kept secret for various reasons like profit, reputation, security, etc. An open source software is a software developed publicly where the code is released without a fee. There are various advantages of open source softwares. Programmers can check the code and submit improvements or bug fixes. Even though not everybody reads the code, it creates a trust between a regular user and the developer, proving that the software is not doing any harm in the background without the user's knowledge. Open source software development initiatives advocate openness, which helps innovation and development.

Mozilla and Linux are some examples of very successful open source projects that are developed as advanced as their closed source commercial equivalents. Linux is especially important as it has reshaped open source software development models; for the first time, the bazaar model as Raymond (2001) describes it, is proving successful in incorporating several important principles. Some of them are:

- *"Release early, release often."*
- *"Users are co-developers."*
- *"Given a large enough beta-tester and co-developer base, almost every problem will be characterized quickly and the fix obvious to someone."*
- *"Any tool should be useful in the expected way, but a truly great tool lends itself to uses you never expected."*

Raymond (2001) further describes that good softwares start from a developer's personal "itch" and good developers know to reuse previously written code, rather than reinventing the wheel. Algorithms, in this thesis are borrowed from various disciplines and re-contextualized thereby conforming to his suggestions.

However, it should also be noted that open source software development is also risky, as bugs in the software would be spotted quickly and if they cause security threats or limit usability, quick fixes would

be needed. If the user base is not big enough, fixes would take longer, causing loss of interest from regular users if. A software that is not maintained regularly won't attract new users either.

Nevertheless, open source nourishes software development and its success is quite obvious, if is evaluated over the past decade. Even big companies, who once were against it, have been initiating their own open source projects alongside their propriety ones. Another good reason to share source code for this project is to establish trust. Google has been successful in this with its mobile platform, Android. It is attracting a growing amount of attention from developers. For this thesis' project, it is crucial that users clearly see from the code that their privacy and their film archive is not compromised.

3. Quantification of Films

Films are artistic, aesthetic expressions. Film making rules were first defined by early directors and are still being constantly redefined. Some directors choose to obey these rules, some break them intentionally, making films with conscious or unconscious choices. Therefore films are structured pieces crafted with careful planning. Since, most of them are not made by random choices—even in that case it won't be random[17]—analyzing a film would result in some sort of data that is specific to that film. Obviously, complexity and correlation of the data depends on analysis methods as well as the variables used. Thus, with enough abstraction, it would be possible to distinguish and compare films with resulting data which can also be referred to as quantitative summaries.

17 Many, believe that every human action is tied to a conscious or unconscious decision-making process. They might seem random only in the absence of information.

Individual methods described in this chapter might not be enough to distinguish each film, however it is the combined usage that will empower fruitful comparisons. It should also be noted that human cognitive process is very complex and filled with diverse low-level and high-level capabilities. As such, it is impossible to quantify what a film might mean for a viewer. Although it should be possible to summarize a film, by considering it as an object, then by dissecting and extracting data from it. In this chapter, proposed methods to dissect a film and proposed visualizations are described. Later, this chapter continues by defining different types of data that can be associated with films; user entered data and information about film's billing. Therefore the three main sections are: Extracted data, User entered data and Billing data.

It should be noted that film is a linear, timed medium. It consists of intervals called frames. A 120 minute, 30 FPS (frame per second) film would have 3600 frames. Quantification methods must take this into account and adopt some optimizations and further abstractions to provide efficient and compact data results.

3.1. Extracted Data

Film is an individual outcome of a production process. It is formed by various elements; sound, image and motion merge seamlessly to create a fictional world.

Various techniques have been invented and used for visual storytelling since early adaptations; montage as Eisenstein first coined the term, lighting, color and its perception, framing and motion. This section covers methods used for quantification purposes. Time based image can be dissected into different properties; shot lengths, color histogram, color energy, motion, sound level and dialogue metrics. Each property require different set to tools and methods to be extracted as data. All the methods are provided with visualizations that are stored in raw numerical data for easier storage and analysis. Visualizations are also kept for quick visual comparisons.

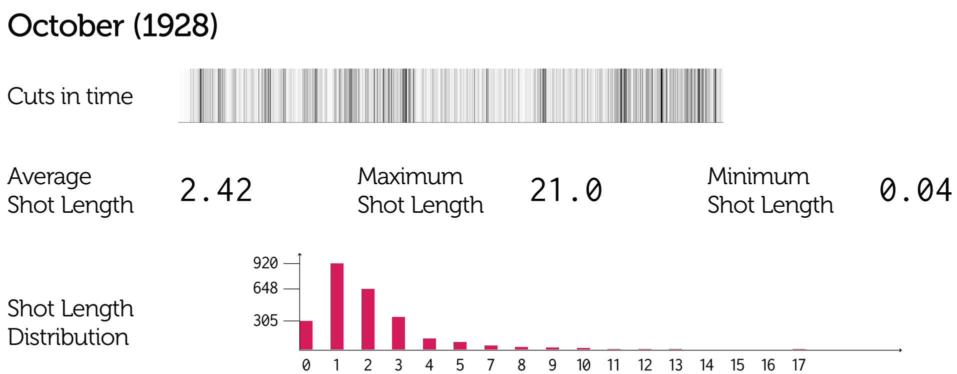
3.1.1. Shot Length

Shots in film are like phrases in literature. They are fundamental units of a film. Shots are put together in a montage to construct stories (scenes) from different, even sometimes unrelated image sequences. As described before, shot length has driven statistical cinematic style analysis research for the past few decades. It is one of the most apparent properties that can be extracted from films. As Salt (1974), Bordwell (2006) and Cutting et al. (2010) investigated in their researches, duration between cuts differ in correlation with era and director. Consequently, shot lengths might provide some kind of signature for films. Although, it's not enough to compare films only with shot lengths, usage along with other metrics would provide useful information.

Shot boundary detection is researched by computer scientists to create automatized scene and shot extraction systems for large media distribution and management systems where scalable automation is important to facilitate organization. Their algorithms are based on the idea that shots consist of frames showing a continuous action taken by a single camera and big changes between consequent frames must mean a cut. These changes are calculated by measuring different variables; color histograms (Zhang et al. 2001), motion activity in the scene and sometimes sound properties (Chen 2004). These algorithms are not flawless and often generate false-positive and false-negatives however they are refined continually by researchers. The method described by Zhang et al. (2001) in combination with optical flow difference measurements is used for shot detection in this thesis project. An example of cut visualization and shot length analysis results are shown in Figure 3.1.

Figure 3.1

Visualization for cuts and shot length analysis results



It is also crucial to investigate and store not only average shot lengths, but also distribution of all shot lengths for a film. Some films might have medium length shots, some might have very long and very short shots; these would provide similar outcomes if only averages were examined.

3.1.2. Colorgram (Color Histogram)

Color in nature and usage of color in general, heuristically communicates emotions, messages and thoughts. Color and color combinations carry different meanings with them depending on the culture and the period. They have become distinctive properties of art and design movements. Color has been a focus of interest for many artists, designers, psychologists, social scientists to name a few.

Filmmakers first used hand coloring and later on film tinting to imply certain narratives before color film was invented. Since then cinematographers have been using color to establish mood. Perception and psychology researchers investigate color's effects on humans and try to establish connections between certain colors and feelings. Apart from the details, such research proves that there is a connection between moods, feelings and color. Color correction in film post-production grounds itself on this; filmmakers might choose to shape colors to establish mood in their films during this final process.

In cinema theaters, lobby cards, prepared in smaller size than a poster, hung on lobbies, present a visual summary of the film to viewers. Some lobby cards feature selected frames from films aiming to quickly communicate this visual feel and look of the film. Proposed summarization technique, named "colorgram", derives from same motivation; it provides a compact color summary of a film. Similar projects have been done by several artists. Salavon (2000), visualizes colors throughout a film in various forms (see Figure 3.2). Dawes, (2004) with his project Cinema Redux, creates "single visual distillation of a film" in his own words (see Figure 3.3). While Salavon is scaling frames to achieve a color reduction, Dawes doesn't modify content of the frames but makes another reduction by selecting a single frame for each second of the film. They both present the films in a vertical, left-to-right orientation; the first frame is at the top left and the last is at the bottom right. A website by an unknown author provides similar visualizations for films that he selects[18].

18 http://moviebarcode.tumblr.com

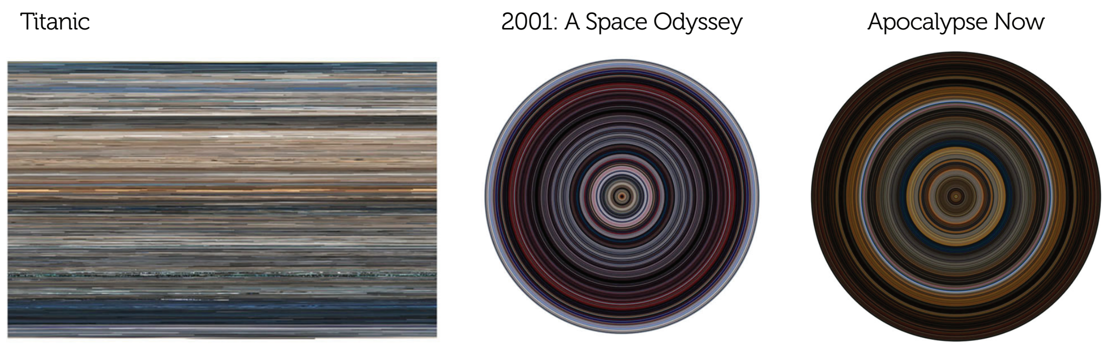


Figure 3.2
Color abstractions for films made with different methods, by Salavon

Another important aspect is that neither of these works are targeting digital; they are realized for print medium. In addition they are singular works realized for specific purposes with varying levels of intuition, therefore they are artistic rather then systematic.

Figure 3.3
Cinema Redux by Dawes

For this thesis, I propose using color histogram visualization "colorgram" for films. It is an outcome of an automated process which is used for analysis and comparisons later on. Colorgram is separated into columns, each representing a slice of the film. Columns are generated by taking individual frames and compressing them horizontally. Later, they are added horizontally one after another, hence x axis (width of

the colorgram) represents the length of the film. Final iteration of colorgrams I made conclusively for five different films can be seen in Figure 3.4.

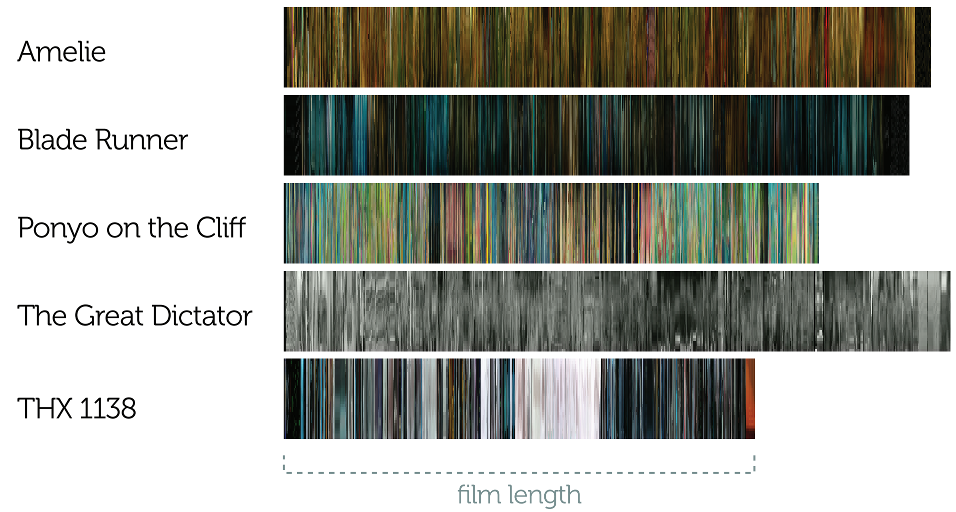


Figure 3.4
Colorgram examples for five films

More detail about design decisions behind colorgrams and comparisons of colorgram from various films are explained in chapter 5. Reflections on Visualizations.

3.1.3. Color Energy

Color in HSV or HSB color representation has three variables: Hue, saturation and value (See Figure 3.5). They have diverse effects on the perception of color; red hues are more energetic and therefore have higher energy. Blue hues are relaxing and thus have lower energy. Saturation is also an important factor, as more saturation means more energy. Hue value is represented by angle, between 0–360. Red is zero; blue is 240; green is 120. Saturation and value is represented by percentage, between 0–100.

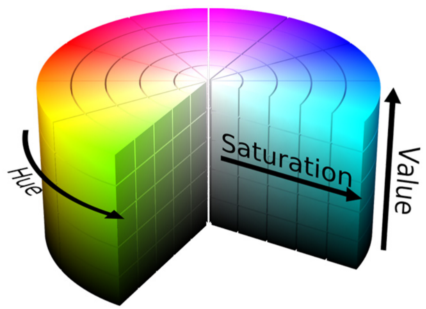


Figure 3.5
HSV Cylindrical-coordinate representation of color

Wang and Cheong (2006) suggest a color energy calculation based on psychological studies. They attempt to map saturation, brightness and hue's affective relationships with arousal and valence. They calculate raw color energy with the following formula:

$$E_{color} = \sum_{k}^{M} F(h_k) s_k v_k$$

M is total number of pixels in a frame, where k iterates. s_k, v_k are saturation and lightness values. Hue energy function $F(h_k)$, assigned a range between 1.25–0.75, depending on the angular distance to blue and red, is calculated as:

$$F(h_k) = \frac{1.25 \times dist_{blue} + 0.75 \times dist_{red}}{dist_{blue} + dist_{red}}$$

While calculating $dist$, shortest distance—which can also wrap around from 360 to zero—is taken. Visualization examples for color energy levels throughout films are illustrated in Figure 3.6.

Quantification of Films

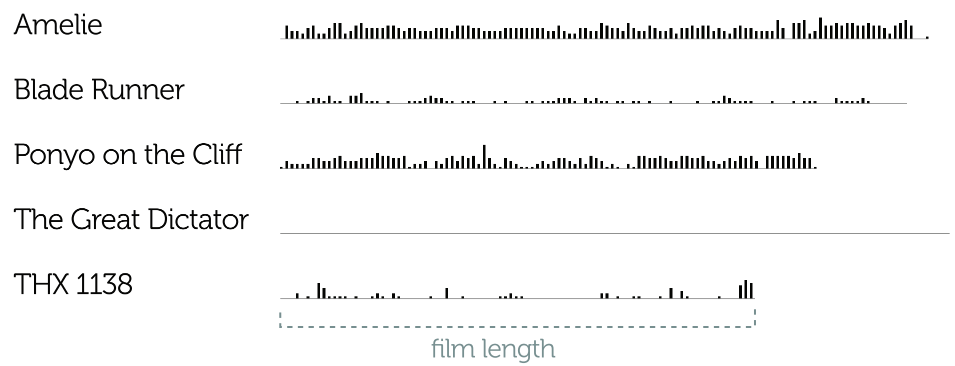


Figure 3.6
Color Energy Examples

Color energy calculation for The Great Director, which is a black and white film, outputs zero for every frame of the film. It is caused by absence of hue, therefore "saturation" is always zero. This is discussed more in chapter "5. Reflections on Visualizations".

3.1.4. Movement Energy

An object that moves in our angle of sight, immediately attracts attention. Film and other derivative mediums benefit from this and they have a very unique position in art and design as they have movement at their core and they wouldn't exist without movement.

Movement in film can be separated into two types: Camera movement and subject movement. It might be computationally expensive and error-prone to distinguish between these two types at the moment, therefore it is omitted for this thesis and it might be regarded as future development. For now, all the movements in a given moment in film are added together to calculate momentary motion.

Movement can be calculated by simplifying it to patterns of motion, defined as optical flow. Typically, from the digital point of view, displacement of pixels between consecutive frames represent motion.

There are various optical flow measuring techniques named after their creators: the most common is Lucas–Kanade. Other dense tracking techniques are Horn–Schunck, Gunnar Farneback. OpenCV has implementations of them as well as many helper functions. Dense tracking techniques are computationally very heavy and in most cases unnecessary. Simpler sparse optical flow techniques calculate motion by simplifying source images and relying on some basic assumptions. (Bradski 2008, pp. 322–337) Even though sparse techniques are faster, in this context Horn-Schunuck is empirically found to produce more accurate results.

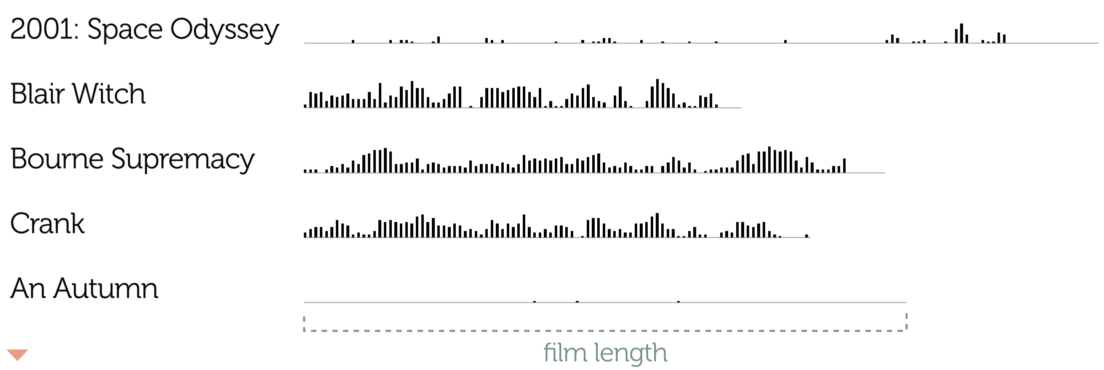


Figure 3.7
Movement Energy Examples

For calculating movement energy in films, the detail of the each frame is reduced by smoothing with the “bilateral smoothing” algorithm which also preserves the edges. Then, optical flow is measured with the Horn-Schunuck technique. After measuring the total motion difference between each set of frames, a per minute average for the included frames is calculated in order to be efficiently visualized. Resulting visualizations are similar to the previously described analysis results in this chapter. Some examples are shown in Figure 3.7.

3.1.5. Audio Energy

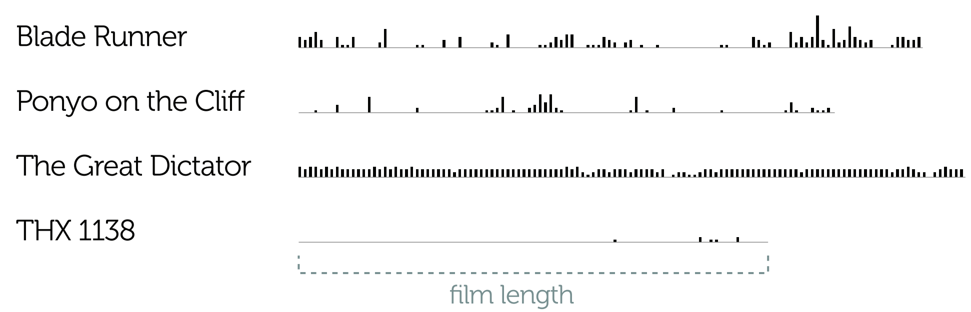
In film, audio is used to enhance cinematic experience through dialogues, soundtracks and effects. It is also used to increase tension, transfer emotions and establish moods. Even though audio and music semantics can vary among people, various properties can be used at least to define boundaries. It can be said that in film volume increases the viewer’s excitement and involvement. Therefore, peak detection through waveform analysis is investigated. At the moment, for simplicity’s sake, other audio feature extractions are considered as future research and development.

Audio energy is measured by analyzing peaks in the audio’s waveform. Calculating it throughout a film would produce interesting results in correlation with other analysis results. Audio energy is calculated as:

$$E_{audio} = \sum_{i \in interval} (A_i)^2$$

In which, A_i is the audio sample indexed by time i. (Rasheed and Shah 2002) Full audio is analyzed with intervals to reduce complexity and possible clutter. Interval length is 50 milliseconds. An example of audio energy visualization is illustrated in Figure 3.8.

Figure 3.8
Audio Energy Examples



3.1.6. Dialogue Metrics

Keyword extraction is used in literature for characterization of texts. It serves various purposes from style analysis to author identification. In the context of film, if the same methods are applied to dialogues, word selections and the frequency of their occurrence might create new opportunities for analysis and comparisons. Two examples of visualizations for outcome of keyword extraction results are shown in Figure 3.9. Most common 10 words are extracted; later, stop words[19] are filtered out; then the results are represented by circles. Circle size is the relative occurrence of a word in the film's dialogues.

19 They are the most common function words found in a language.

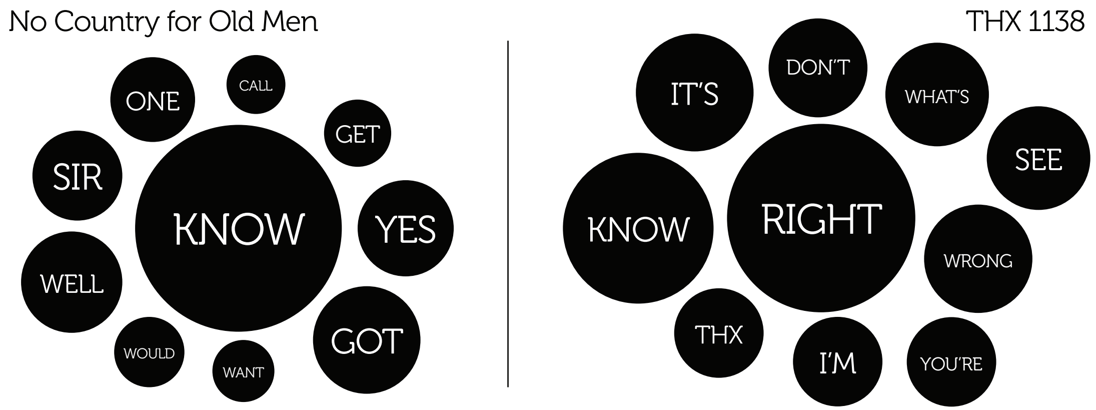


Figure 3.9

Most used words in No Country for Old Men (2007) and THX 1138 (1971)

Dialogues for films are gathered in an unconventional way; subtitles for films are examined to extract the dialogues. It is possible to find the subtitles online from various sources and they are often created from digital discs. It is likely that these contain small errors, like spelling or timing but it is acceptable for the sake of convenience, since it would be almost impossible to get full dialogues for a film using any other method. Errors can also be reduced by examining multiple subtitle files from different sources for the same film.

3.2. User Entered Data

Online video sharing services have long been benefiting from user contribution. YouTube[20] offers its users various opportunities for interacting with uploaded video content: Tags, comments, annotations, video responses, rating. This section deals with user-generated data's role in film quantification.

20 http://www.youtube.com

Interaction between video based content and users can occur in two ways: Synchronous and asynchronous. Synchronous microblogging[21] throughout live media events as researched by Shamma et al. (2009) brings attention to the fact that viewers liked to communicate via Twitter[22] concurrently during broadcasts. Many live TV shows, including Current TV's[23] "Hack the Debate" campaign, tried to integrate Twitter into their own broadcasts to create a conversation with the viewer. Nathan et. al (2008) demonstrate that users can also be enthusiastic about engaging in TV show related social communication through asynchronous interaction in video-on-demand services. It seems that viewers are eager to interact with content and with each other while watching. This practice is becoming more common and easier as technology becomes more ubiquitous. A recent similar example is mobile and tablet based statistics applications' usage while watching football during 2010 FIFA World Cup in South Africa.

21 "The posting of very short entries or updates on a blog or social networking site" (Oxford Dictionaries. April 2010)

22 A short microblogging service where users can share up to 140 characters at a time.

23 A pioneering online and cable television network that offers prizes for viewer submitted content. http://www.current.com

The software and website system proposed in this thesis is intended to be used with computers[24]. Therefore an asynchronous communication solution where interactions are saved is necessary. Saving interactions would help to create community generated data where some amount of qualitative variables can be collected and shared.

24 Requiring user to watch films either by inserting DVD's or by opening digitally encoded files.

3.2.1. Temporal Commenting

For a time-oriented medium like films, user input should also be temporal like YouTube's annotations and SoundCloud's[25] comments. Several efforts like Social TV, AmigoTV, CollaboraTV have also been made for TV. Temporal Commenting can be especially important for film, in two ways. First, it enables content creators to distribute an official, text-based commentary independently from the video content. Second, users are offered a platform where they can share their thoughts as tags or comments, their inner film associations or outside film associations as links with other viewers.

25 An audio sharing site for artists, http://www.soundcloud.com

As shown in Figure 3.10, SoundCloud collects time based comments from users and visualizes them on top of the audio player. In SoundCloud if an audio has too many comments, a clutter forms and it causes interaction problems because selecting and viewing individual comments gets harder. Therefore, an additional feature of adding context to the comments (e.g. official, trivia, associational) would be beneficial for temporal film commenting. It would offer more scalability than a simple on/off switch as the viewer would be able to filter which category of comments are shown while watching.

Figure 3.10

SoundCloud's player and temporal comments placed on top

3.2.2. Temporal Tags

Tagging systems are important factor for improved organization, search and discovery. They usually consist of one or two words. While being shorter than comments, they offer more flexibility than predefined categories. In this context users would be able to temporally tag throughout the films. Some examples of temporal tags can be, marking goofs, scenes, characters.

Users will be able to share temporal tags via provided links. Linking and referencing dynamic content on the web can be tricky, especially if they are long and include many variables. Users may not notice if a part of a long link is missing, especially if it is not semantic. URL shortening systems[26] have been recently become popular due to their effectiveness in this manner. Many internet services implemented their own, internally developed url shortening systems[27] after seeing the popularity among their users. In this project, an internal url shortening system is going to provide easy-to-remember links for temporal tags. Sharing would enable viewers, researchers, filmmakers and producers to reference moments in films, further allowing different sets of data to be quantified, if enough of it is gathered.

26 Some examples are: http://bit.ly, http://tinyurl.com, http://goog.gl

27 http://t.co by Twitter, http://fb.me by Facebook, http://lnkd.in by LinkedIn

3.3. Billing Metadata

Billing metadata is a film's immutable production data which is provided by filmmakers. It is fact-based and it does not change. Release year, director, actors, length, language, country of origin are different examples of metadata (see Figure 3.11). They are obtainable from various online sources: IMDb[28], MetaCritic[29], tagChimp[30] themoviedb.org. IMDb is the biggest and most up to date online cinema information source. However it is propriety and it doesn't provide free real-time access to its database. To its credit static, but frequently updated plain text data files are provided on its ftp sites. A considerable advantage is

28 http://www.imdb.com (Internet Movie Database)

29 http://www.metacritic.com

30 http://www.tagchimp.com

that the film producers provide the data themselves and it has become an industry standard at least for major film industries. MetaCritic uses data from IMDb. TagChimp provides community-generated data. According to their claim it has 142,445 movies, TV shows and music videos combined, which is very low compared to IMDb's feature film only count: 259,193 (IMDb 2011). Themoviedb also has a community-driven model. Even though they don't disclose the actual number of films listed, it seems to be in good condition; as they claim, it is used by many popular media player softwares such as Moovida, XBMC, Plex, MythTV and MediaPortal.

Directed by	D. W. Griffith	**Starring**	Mae Marsh	**Cinematography**	Billt Bitzer
Produced by	D. W. Griffith		Robert Harron	**Editing by**	D W. Griffith
Written by	D. W. Griffith		Constance Talmadge		James Smith
	Hettie Grey Baker		Lillian Gish		Rose Smith
	Tod Browning		Gino Corrado	**Distributed by**	Triangle Distributing
	Anita Loss		Douglas Fairbanks		Corporation
	Mary H. O'Connor		Madame Sul-Te-Wan	**Relsease Year**	1916
	Wlat Whitman		King Vidor	**Country**	United States
	Frank E. Woods	**Music by**	Joseph Carl Breil	**Language**	Silent film
			Carl Davis		English intertitles

Figure 3.11
A billing metadata example

Billing metadata can be used in correlation with analysis data to discover patterns. For example, 1960–1970 French films can be compared with 1980–1990 Japanese ones. Such comparisons would result in interesting discoveries and would enable in-depth stylistic qualitative analysis for cinema researchers. Statistical cinematic analysis researchers have already been investigating films made by the same director to conduct a style analysis.

4. Data Collection and Organization

Data collection and organization is an important step for this project. It is impossible to have access to all the films made until today, thus it is necessary to provide adequate tools (software) for people to contribute to the project by quantifying their own film archives. Personal archives might consist of various formats, such as film rolls, Laserdiscs, DVDs, High Definition discs, high quality and low quality digital files. Varying sources creates a fragmentation and such fragmentation requires standardized tools. For the sake of simplicity and controllability, formats that cannot be played through a computer software are omitted as the quantification is done by the developed software.

It is necessary to establish a standard and effective communication between client side software and server side which collects all the raw data and visualizes it. This chapter defines methods used to increase reliability and efficiency of the process. Technical and social problems that will arise and ways to tackle them are also described.

4.1. Volunteer Computing / Crowd Sourcing

CPU power and especially the number of computers in use increase almost everyday. Supercomputer centers are no longer the sole centers of calculation. A large amount of CPU power is now distributed around the world. This distribution creates possibility of new computing methods. Volunteer computing is a distributed computing arrangement

where people voluntarily dedicate CPU time (computation power) and hard disk space of their own computers for heavy scientific calculations. Many projects have benefited and continue to benefit from this practice: GIMPS[31], distributed.net, SETI@home[32], Folding@home[33] just to name a few. Platforms like Berkeley Open Infrastructure for Network Computing (BOINC), World Community Grid, Grid.org offer the possibility for various academic projects to use a readymade infrastructure and formed community.

Volunteer computing is especially important for this project as it would be quite impossible to have access to a large number of films and quantify them in one server. If one attempted to quantify all the films registered to IMDb, and if we assume that the quantification takes place in real time, and that average film length is 90 minutes; the whole process would take:

259 805 × 90 = 23 382 450 minutes, which is more than **44 years**.

However, with help from volunteers who are willing to dedicate their CPU power and their film archives, it would be possible to quantify a huge number of films in much less time. If the project attracts attention from as few as 100 volunteers, quantification could be finished in around **5 and a half months**. Of course, this last estimate is based on an ideal situation where those 100 people have all the films; they coordinate so they don't quantify the same films twice and they quantify 24 hours a day, 7 days a week. Although perhaps unrealistic, this gives an idea of how this project can immensely benefit from the volunteer computing.

It should also be noted that distributed computing in this project differs from traditional volunteer computing projects. First, it requires data to be input by the volunteer whereas most of the others provide data, formulas and only require computation resources from the volunteer. In this project, input data is taken from individual films from the

31 Great Internet Mersenne Prime Search, http://www.mersenne.org/

32 Search for Extra-Terrestrial Intelligence (University of California), http://setiathome.berkeley.edu/

33 For computing simulations of protein foldings (Stanford University)

volunteer's archive. Consciously choosing which films to quantify is already an active participation. Even more active involvement is also possible; the volunteer is able to add temporal comments, tags throughout the film. Therefore the project has some aspects of the crowd sourcing model where "the crowd" helps in solving problems or generating data. "Crowd" wisdom is claimed to be greater than the individual member's intellect (Brabham 2008). In this context, each volunteer's film knowledge and archive is different, so tags and comments submitted by them will differ. For this reason, it might be better to describe user involvement in this project as a combination of volunteer computing and crowd sourcing, wherein users can choose the amount and type of their participation.

4.1.1. Reputation System

Online content can be reached instantly from all over the world, thus in online communities it is very normal to have an immense activity level, causing content management problems. Different models can be used to tackle these problems; moderated, semi-moderated, community-driven, anarchic. Most community driven sites use a combination of these models in different levels, but most employ a reputation system to encourage and reward the users. Community driven models work better than conventional moderated models in terms of management, especially when the content and user interaction grows beyond the usual limits. Reputation systems are used by many online merchandising sites such as Amazon, eBay as well as other community sites as Stack Overflow, Digg and Yelp. Each buyer and seller in the auction site eBay gives reputation points to each other; in Stack Overflow—a question and answer site for programmers—users simply vote on each others input and gain more reputation by doing so. Some of these reputation systems are enriched by badges, achievements and peer testimonials which provide orientation and achievable small rewards for community members while distinguishing valuable content from the valueless.

In this project, it is especially important to have a reputation system, because user submitted quantification data has to be filtered depending on the quality. The reasons why filtering is necessary and filtering procedures are described in detail in the next section "4.1.2 Data Pollution."

Each user has their cordial goals, therefore a reputation system addressing multiple aspects and rewarding actions without discrimination is necessary. Various actions to be rewarded can be listed as following in a decreasing order of importance:

- Submitting quantification data for a new film
- Submitting quantification data for a previously quantified film
- Writing temporal comments
- Watching films from the software without quantifying

For every submission of film quantification data, the quality of the source is calculated and the resulting quality variable is used to recalculate a new reputation for the user. Consequently, it is better for the user to quantify higher quality sources that are not previously quantified.

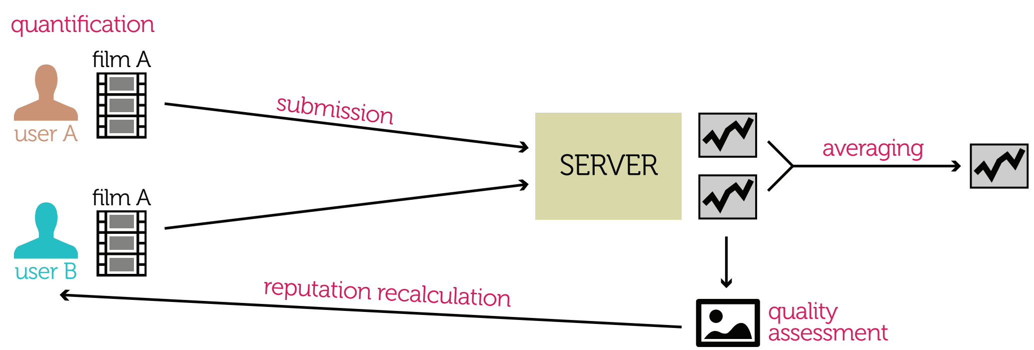
Stack Overflow, a highly community driven question/answer site, is granting certain administrator privileges such as to the community members, even at low reputation levels. As the user's reputation increases, the level and amount of privileges increases with it. This creates a self-controlled community where the authority is distributed among the members. A similar model can also be used for a film quantification project where temporal comments are filtered based on the community feedback. Similar systems in simpler forms are also used by YouTube; flagging a comment or a video is brought to the attention of the moderators.

4.1.2. Data Pollution

Distributed computing for film quantification brings with it another challenge with itself. The same film can be watched, quantified and submitted by multiple users. Each quantification might produce slightly different results, depending on the source—DVD, Blu-Ray, encoded high-quality file or low quality file. Therefore, a ranking and averaging algorithm that is able to produce a single final quantification is necessary. A user reputation system and a quality assessment are the key points for this. When a user submits a quantification data, several variables about the source (e.g. dimensions, encoding format) are also sent to the server to make a quality assessment of the source. This calculation later affects data averaging that takes place in case there are several quantifications for a single film as well as it affects the submitter's reputation. This reputation is used mainly for easing quantification data assessment and it does not have to be shared with users. Perhaps it is better if it is not shared, to prevent frustrations that might arise from having low reputations or fraud attempts made to gain more. Nevertheless it is beneficial for internal calculations. Overview of this system is illustrated in Figure 4.1.

Figure 4.1

Overview of reputation system and averaging for data pollution



All the submitted data coming from users are also kept in the server for future referencing and for re-ranking and re-averaging calculations. If there is a new quantification data coming either from same user or another user for the same film, data averages and the user's reputation are recalculated. In this way, the system can be defined as dynamic and changing according to the quantification submissions.

4.1.3. Legal Issues

Several legal issues related to volunteer computing and film copyrights might have to be addressed. Data protection acts state that softwares and websites should not gather and store any personal information of its users. In this project, no personal information is collected by quantifying software and this is going to be ensured by releasing the source code. Film copyright issues can be tricky to address. No film is copied or made available to the public, therefore the project should not have copyright problems as it would be considered derivative work. Even thought the films are abstracted to different forms, new representations would not cause viewers to fully experience the film and consequently won't keep them from paying to see the actual film. Therefore these abstractions fall into fair usage terms. In fact, the situation is favorable for film producers because this system is offering alternative ways for film research and personal film discovery.

4.2. SMIL as Data Format

In this context, standardizations are needed for data transfers between user's computer (quantification software), server, database and visualization software (website). Even though various formats and markup languages have been suggested for temporal data, at the moment there is not a single widely accepted standard. One of the standardization formats, SMIL, is selected to be used in this project because of its lightweight and flexibility. All quantification data submitted by users is transferred to the server by a single SMIL file (see Figure 4.2 for overview). The data is then written to the database and analyzed for visualizations. SMIL allows integration of metadata, timed captions, links and other various textual information into a single file. When the user adds a temporal comment or tag, the transferred SMIL file only contains new data and the film id, hence the amount of transferred data decreases while increasing the communication speeds. An example of SMIL format is shown in Appendix.

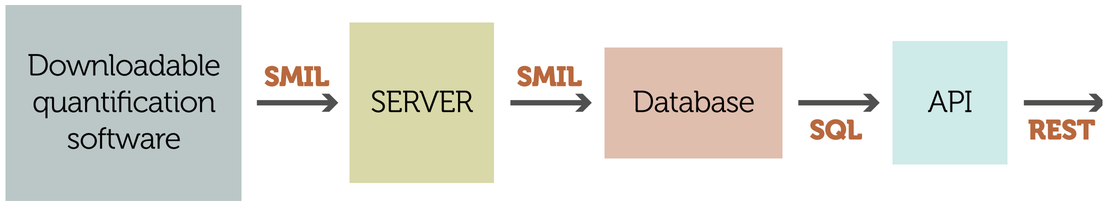


Figure 4.2

Usage of SMIL for data transfers

4.3. API

Most internet services focus on specific goals and real power emerges from the interconnectivity of multiple services. Application programming interfaces (APIs) open up new possibilities for improved experiences. As an example, a person shares videos from Vimeo[34], tweets about it in Twitter and embeds the same video to a Tumblr[35] blog while sharing the location at Foursquare. None of these replace another but rather complement each other. Data builds up in separate services, for instance videos in Vimeo, microblogs in Twitter, locations in Foursquare, pictures in Flickr, sounds in Freesound or SoundCloud, etc. The amount of these services increase day by day, thus the task of merging data from several of them in order to create richer experiences, is left to the users who can be considered as co-developers. While sharing data

34 A high quality, free video sharing service, http://www.vimeo.com

35 A flexible, popular blogging system, http://www.tumblr.com

through API enables developers to find new usages for provided data by creating new functions, it creates opportunities to merge functionalities.

There are several styles to implement an API system. For this thesis project, REST (Representational State Transfer) style that is proposed by Fielding (2000) will be used as it is based on HTTP architecture and it is quick to implement. API will provide access to the mean of the quantified data of each film. Therefore, developers will be able to make their own visualizations or applications without worrying about the data pollution (previously mentioned in 4.1.2. Data Pollution).

5. Reflections on Visualizations

In this chapter, first, design method for colorgram is evaluated while describing the algorithm. Later, several film groups formed by similar films are taken as examples to reflect upon the analysis and visualization methods described in chapter “3. Quantification of Films”. First, films that cause excelling performance for visualizations are presented and discussed. Later, shortcomings with other sets of films are described.

5.1. Colorgram Algorithm Evaluation

Various compression algorithms can be used as shown in Figure 5.1. While upscale algorithms are important, thus researched and improved by many, downscale algorithms are limited. In this context, anti-aliasing scale down algorithm is chosen for its reduced clutter; therefore its increased readability. Anti-aliased scaling takes averages of neighboring pixels to reduce distortion, producing average colors for smaller resolutions. Resulting image is slightly blurred but it is favored in this case due to reduced visual complexity.

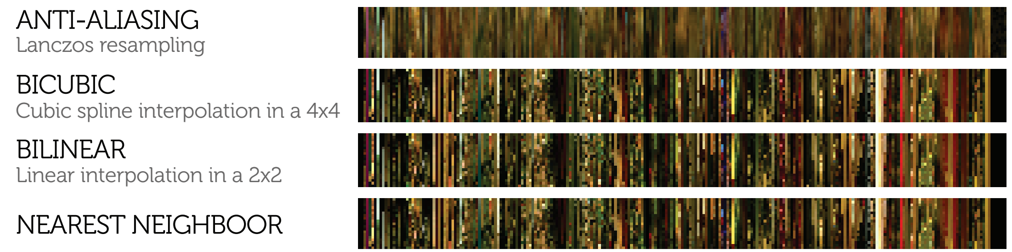


Figure 5.1
Different scaling algorithms for making Colorgram, Amelie (2001)

A 120 minute, 30 fps film has 120 × 60 × 30 = 216 000 frames. If it is HD, it would have 216 000 × 1920 × 1080 = 447 897 600 000 pixels. Obviously, it would be impossible and inefficient to represent each pixel in colorgram. Even if each frame is scaled down and represented with one pixel wide column, 216 000 pixels width would be still impractical to look at, considering a normal screen size as 1600 × 1200 pixels. Because of these visual space constraints, colorgrams must be made by selecting individual frames in intervals and by summarizing them further. Modifying interval length, changing the number of frames selected per minute, affects the colorgram pixel width as well as the delegation level. Bigger intervals result in smaller colorgrams that are easy to handle and grasp visually but reduce accuracy. Smaller intervals are best for accuracy but produce big images that are hard to analyze, understand and represent on a computer screen. In this case, an optimal interval length that produce a visually compact but efficiently representative colorgram has to be chosen. Colorgrams made with various interval lengths from same film are shown in Figure 5.2.

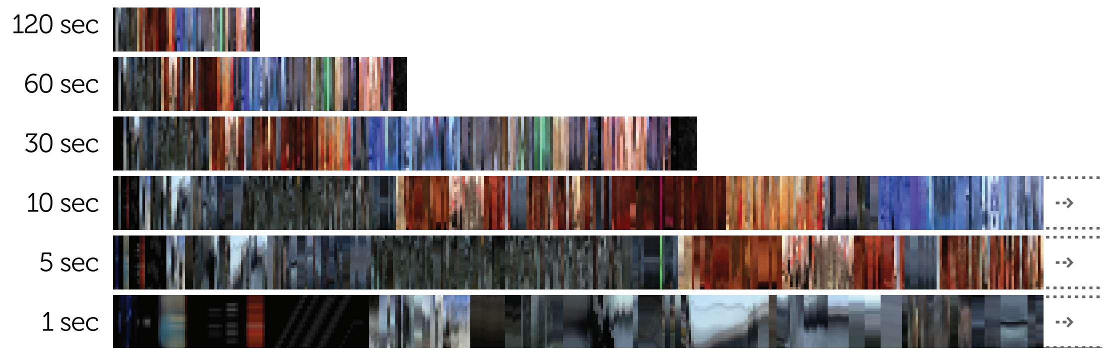


Figure 5.2
Various key-framing intervals for making colorgram, Hero (2002)

After varying levels of experimentation with intervals and scaling algorithms, I choose to produce colorgrams by selecting a single frame for each 10 seconds interval. Later the result is scaled to 1/3 in order to achieve usable widths.

5.2. Outstanding results with certain films

It is found that each visualization performs better with certain kind of films or with specific cinematographies. Films selected and presented in this section don't follow any other pattern than their outstanding efficiency for individual visualization results. It should be anticipated that this distinction is not related at all to the quality of the films.

5.2.1. Color

Films that incorporate color in visual storytelling stand out from others in colorgram and color energy analysis results.

In Hero (2002), one story is shown three times from three different characters perspectives. Each story is prominently color corrected by each character's perception: Red, Blue and White. Another dominant

color correction—green—is used for flashback scenes. Stories, their colors and timings can be recognized easily from its colorgram in Figure 5.3.

Figure 5.3
Colorgram for Hero (2002)

Traffic (2000) is also a color coded film, in which three different geographical locations and stories are told by three distinct color corrections. Yellowish hues, light blueish hues and natural hues can easily be seen from its colorgram in Figure 5.4.

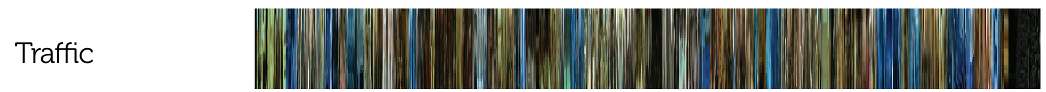


Figure 5.4
Colorgram for Traffic (2000)

In Shinboru's (2009) colorgram (see Figure 5.5), the scenes that take place in "brightly illuminated light blue room" are very apparent. They can be distinguished easily from the rest of the movie which has darker tones.

Figure 5.5
Colorgram for Shinboru (2009)

In Wizard of Oz (1939), color is used for various meanings. Apart from Dorothy's bright colored shoes, green Wicked Witch, bright landscape of the Oz, etc., one feature that can be easily spotted from its colorgram is that the scenes that takes place in Kansas are shot in black & white and later given sepia tone while the scenes in the are shot in vibrant colors (see Figure 5.6).

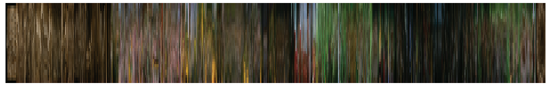


Figure 5.6
Colorgram for Wizard of Oz (1939)

For some films, the subject or the location dictates certain colors to be predominant. This becomes apparent especially in science fiction and when the intended location or situation is very different from usual. Sunshine's (2007) colorgram has full yellow columns as it takes place in a ship that travels to the Sun, Mission to Mars (2000) uses red hues to give the feeling of the Mars. THX 1138 (1971), Tron (1982) and Blade Runner (1982) create their own fictional worlds by using specific color corrections. Their distinctive colorgrams are shown in Figure 5.7.

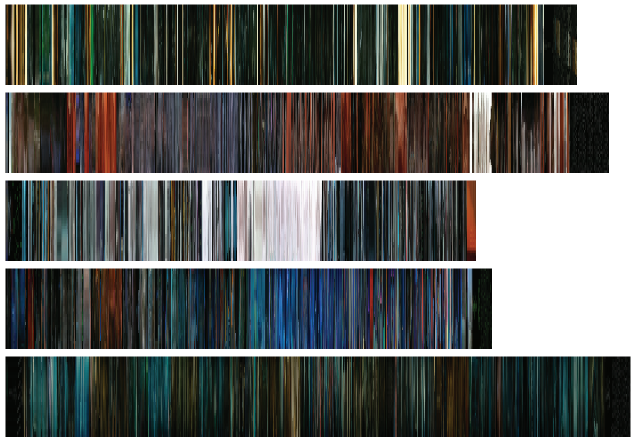
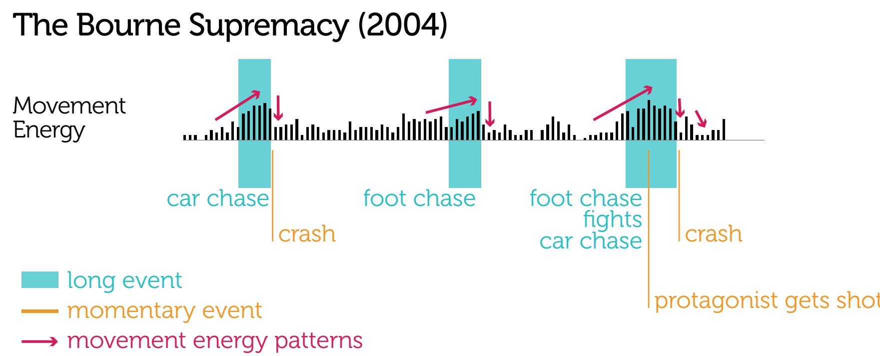


Figure 5.7
Colorgrams from science fiction films

5.2.2. Movement

As movement energy is calculated by measuring overall movement between each frame in film, it is possible to distinguish fast action films from the more static ones. It is not only possible to understand general high levels of motion in The Bourne Supremacy (2004) but three diverse tension buildup and cool down acts can be distinguished easily (see Figure 5.8). In the first one, movement, therefore action, increases until the fast car chase scene ends with a crash. The second buildup, which is more gradual than others is mostly caused by a foot-chase scene that incorporates handheld camera shakes and very wide pans for increased action and immersion. Movement energy decreases gradually afterwards. Several other chases, followed by immediate cool downs are also easily spotted. The final and concluding action in the film is also visible, starting to buildup from eighty-second minute. Movement peaks when protagonist Bourne gets shot and a foot chase starts. The same scene, which includes guns and several fights, continues with a high level of motion until it concludes with a car crash, after which the setting changes.

Figure 5.8
Movement energy examined in detail for The Bourne Supremacy (2004)



A contrasting result for movement energy is visible in Ozu's Sanma no Aji (1962). Ozu is known for his static camera work which causes less overall movement between frames. Most scenes take place indoors where characters are sitting, eating and talking therefore the film's movement energy visualization only contains three marks, which represent 10 percent average of movement at the corresponding minutes (around 43, 51 and 70). (see Figure 5.9)

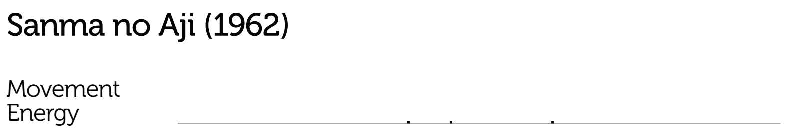


Figure 5.9

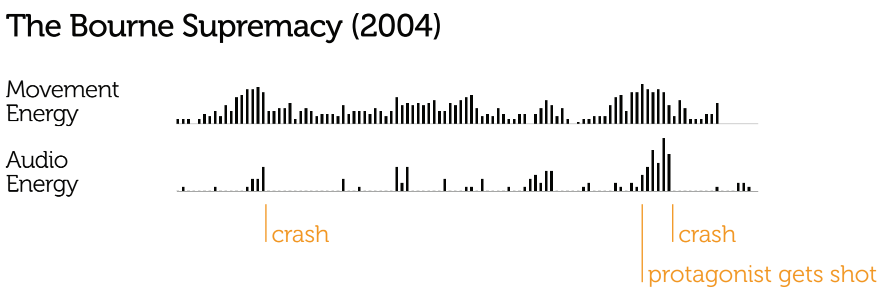
Movement energy for Sanma no Aji (1962)

5.2.3. Audio

If audio energy visualization of The Bourne Supremacy is examined together with movement energy, it is possible to see that patterns in the overall motion of the film are also present in the audio. It is not surprising as audio is used as an important mood-setting element in films. In this case it seems be a tension builder. Both of them are shown in Figure 5.10 along with timing of some important events in the film.

Figure 5.10

Similarities between movement energy and audio energy visualizations for The Bourne Supremacy (2004)



5.3. Shortcomings

For certain films, individual visualizations fail to produce distinguishable abstractions. It is an expected outcome as none of the visualizations are designed to be sole representations of films. They are aimed to only master in their own context.

Colorgrams for certain films are unreasonably close to each other, revealing not much info about single films. Colorgrams made for black and white films and early color films are especially vulnerable (see Figure 5.11). Resulting colorgrams are visually not distinguishable at all.

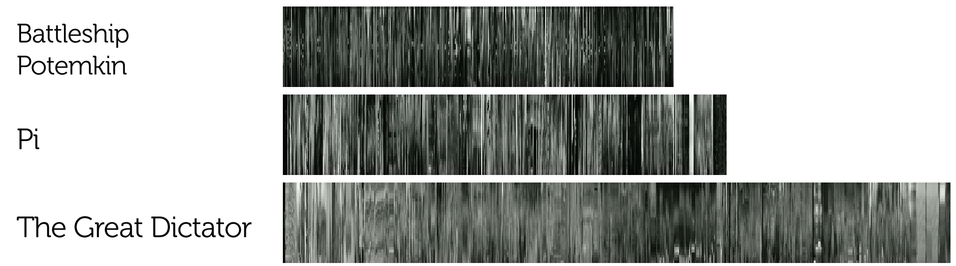


Figure 5.11

Colorgrams for black and white films

Trends in film color correction, especially in Hollywood, also affect efficiency of colorgrams. A recent popular style where most dark areas are colored with warm tones and where light areas are tinted with cooler tones produce visually similar films, thus colorgrams, that would normally be very different from each other. It is very surprising to observe that, a vampire movie, Daybreakers' (2009) colorgram is not that distinct from the colorgram of A Serious Man (2009) and the colorgram of Blade Runner(1982) (see Figure 5.12). It is an unwanted result as all of them have totally different setting and mood.

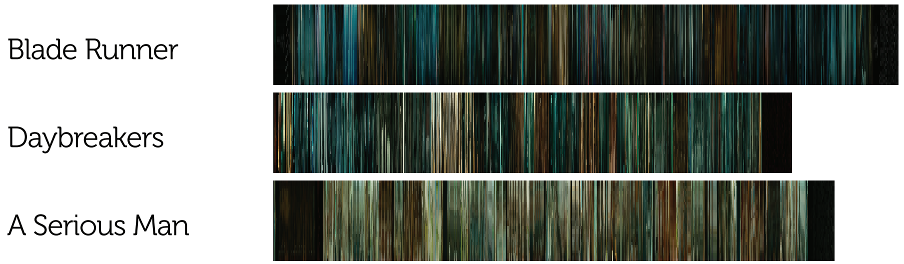


Figure 5.12
Similar Colorgrams from three dinstinctive films

Audio energy analysis presented in this thesis is in a very primitive stage and it is error prone. Audio energy visualization produced for Amelie, (2001) which frequently has non-diegetic narrator's voice, is imperfect. This is caused by high peak levels of speech in audio. Other speech heavy films produce similar results as well (see Figure 5.13). To overcome this problem, instead of analyzing waveform, more sophisticated audio feature extraction methods should be implemented.

Figure 5.13
Audio Fnergy for speech heavy films

Dialogue metric doesn't provide much information about a single film by itself, especially because of some common words that are present in most films. It becomes efficient only if analysis results are numerically compared among several films in order to discover statistical similarities. Visual representations for word frequencies in films are hard to grasp, especially hard to get familiar in film context as similar visualizations are mostly used in other contexts.

5.3.1. Language as Accuracy Issue

Another shortcoming of these analysis is the language. Since sources used for dialogue texts are subtitles, the problem of language selection emerges. If the subtitles in the original language are analyzed, only speaker of that language would be able to understand the visualization. Furthermore, if one common language, such as English is used, definitiveness of the analysis would be lost in translation. It is also possible to come across multi-language films where different scenes take place in remote locations. Therefore, quantifying dialogues for these films require more intelligent systems in order to output efficient and correct results.

Similar problem also emerges about audio if quantified films are dubbed. Depending on the dubbing, "new" audio for film might not inherit same characteristics, and most likely it would not. In this case, quality assessment described in previous chapter is critical as it should distinguish dubbed audio from the original audio without triggering false alarms with multi-language films. Audio related calculations made from dubbed films should not be used at all, as most often dubbing is done completely independent from the filmmakers.

5.3.2. Different Versions as Accuracy Issue

Films might have several edited versions, released in different years; release cut, extended cut, director's cut. Each would have similar analysis results as they are made from same material, however timings would be different. This is an important problem for averaging multiple quantification that are submitted by various users.

To overcome this problem, averaging algorithm should distinguish different versions of each film by creating different sets for each version to store and to make the calculations independently.

6. Conclusions

This project has started as an attempt to enrich a film's reach to the viewers before watching it. Summarization techniques and algorithms are borrowed from other fields; derivative abstraction and visualization models for films are proposed. A single film's abstraction requires that multiple analysis and summarization methods are used together to achieve efficient and correct representations. It is concluded that there is no best single method to abstract a film but that abstractions gain power from a combination of multiple methods. Each method is in continuous development as it is refined and fine-tuned.

Colorgram is the most mature of the quantification methods proposed in this thesis. Mostly because it is designed even before this thesis has started; colorgram is one of the first steps that initiated this research. Other quantification, analysis and visualization methods have to be developed further to reach same level of maturity. The audio feature extraction methods used in this thesis don't go beyond obvious basic approaches therefore these have to be expanded as well.

Color energy measurements presented are based on HSV color model, in future developments, other representational models (e.g. RGB, YUV, etc.) can be used for calculations to reveal different properties and patterns in color data.

Problems that will arise from quantifying the same film originating from different sources must be addressed properly. Timing is one of the most common problem; beginning credits or empty area can be different for same film that is quantified from different sources. Sometimes, one film can have different versions as well; studio's cut, director's cut, extended cut, etc. Methods to distinguish different versions and algorithms that properly average or separate these data also have to be researched.

Some methods, especially dense visual analysis algorithms are shown to be rather heavy to be calculated simultaneously which causes slowness with video playback. Therefore at this stage of the project, a hybrid video player-quantifier is inefficient; viewers (volunteer quantifiers) should not expect to watch the films at the same time. Improving algorithms to perform faster is a necessary step for future developments.

Experiments and methods presented in this thesis can be seen as starting points for derivative works which are not limited to but might include:

- A recommendation system which is more focused on the content: Most existing recommendation systems even though they are very complex, base their calculations on the user history and don't focus much on the content, therefore they become popularity measurements. This is especially true for automated film recommendation systems.
- Various stylistic film history researches: Researchers might benefit from the analysis results provided by the proposed quantification system.
- Official commentary from filmmakers: Filmmakers might be able to provide text based commentaries without needing to have distribution tools.
- Implementation for TV's, media centers or online streaming services: It is getting easier to have access to high quality video-on-demand. Such services might benefit significantly from the proposed system if they integrate it into their own platforms.

- Combination with emotion based film research: If analysis methods proposed in this thesis are combined with research and systems that focus on measuring viewer's emotional reaction to films, new possibilities might emerge.
- Tools for filmmakers: Various tools can be developed for filmmakers to offer alternative ways to evaluate films before completing them.
- Film-based crowd sourcing projects: By taking temporal tags as starting point, other crowd sourcing projects might be developed. This way quantification methods can be expanded to include trivial metrics.

This thesis has provided me an opportunity to do formal research about a subject that I have been contemplating for a while now. It also expanded the possibilities far beyond my initial plans of a simple film website. During the process I have gained a better understanding of film and film theory through valuable theoretical and statistical research works. Meanwhile I found a chance to utilize my previous experience in a different way than usual; I also gained new insights by combining technical, visual, artistic, theoretical and analytical thoughts in a single project. The proposed system in this thesis is an ongoing work; it is in prototyping stage. It will be implemented in alpha and beta stages later on. I hope that it will attract attention from the film community and therefore with collaborative development, technology can be shaped to act as an intermediary between the audience and the film itself.

References

Films

A Serious Man, 2009. [Film] Directed by Ethan Coen, Joel Coen. US - UK - France

Adebar, 1957. [Film] Directed by Peter Kubelka. Austria

Amélie, 2001. [Film] Directed by Jean-Pierre Jeunet. France - Germany

Arnulf Rainer, 1960. [Film] Directed by Peter Kubelka. Austria

Asylum, 1975. [Film] Directed by Kurt Kren. Austria

Battleship Potemkin, 1925. [Film] Directed by Sergei M. Eisenstein. Soviet Union

Blade Runner, 1982. [Film] Directed by Ridley Scott. USA - Hong Kong

The Bourne Supremacy, 2004. [Film] Directed by Paul Greengrass. USA - Germany

Daybreakers, 2009. [Film] Directed by Michael Spierig, Peter Spierig, Australia - USA

The Great Dictator, 1940. [Film] Directed by Charles Chaplin. USA

Hero, 2002. [Film] Directed by Yimou Zhang. Hong Kong - China

Man with a Moving Camera, 1975. [Film] Directed by Dziga Vertov. Soviet Union

Mission to Mars, 2000. [Film] Directed by Brian De Palma. USA

No Country for Old Men, 2007. [Film] Directed by Ethan Coen, Joel Coen. USA

October, 1928. [Film] Directed by Sergei Eisenstein. Soviet Union

Pi, 1998. [Film] Directed by Darren Aronofsky. USA

Ponyo on the Cliff, 2008. [Film] Directed by Hayao Miyazaki. Japan

Psycho, 1960. [Film] Directed by Alfred Hitchcock. USA

Sanma no Aji, 1962. [Film] Directed by Yasujiro Ozu. Japan

Schwechater, 1958. [Film] Directed by Peter Kubelka. Austria

Shinboru, 2009. [Film] Directed by Hitoshi Matsumoto. Japan

Snatch, 2000. [Film] Directed by Guy Ritchie. UK - USA

Sunshine, 2007. [Film] Directed by Danny Boyle. UK - USA

The Great Dictator, 1940. [Film] Directed by Charles Chaplin. USA

THX 1138, 1971. [Film] Directed by George Lucas. USA

Traffic, 2000. [Film] Directed by Steven Soderbergh. Germany - USA

Tron, 1982. [Film] Directed by Steven Lisberger. USA

The Wizard of Oz, 1939. [Film] Directed by Victor Fleming. USA

Publications

Ballard, D.H. & Brown, C.M., 1982. Computer vision, Englewood Cliffs, N.J.: Prentice-Hall.

Benslimane D., Dustdar S. & Sheth A., 2008. Services mashups: The new generation of web applications. IEEE Internet Comput IEEE Internet Computing, 12(5), 13–15.

Bordwell, D., 2006. The way Hollywood tells it: story and style in modern movies, Berkeley: University of California Press.

Bosse, A. & Brisson, K., 2010. Cinemetrics. MIT TechTV. Available at: http://techtv.mit.edu/videos/7809-cinemetrics [Accessed February, 2011].

Brabham, D.C., 2008. Crowdsourcing as a Model for Problem Solving. Convergence: The International Journal of Research into New Media Technologies, 14(1), 75–90.

Bradski, G.R. & Kaehler, A., 2008. Learning OpenCV, Farnham; Cambridge: O'Reilly.

Cahill, J.L., 2008. Anacinema: Peter Tscherkassky's Cinematic Breakdowns (Towards the Unspeakable Film). Spectator, 28(2), 90–101.

Chen, C. et al. 2008. A Brief History of Data Visualization. In Handbook of Data Visualization. Springer Handbooks Comp.Statistics. Springer Berlin Heidelberg, pp. 15–56.

Chen, H. et al., 2004. Action movies segmentation and summarization based on tempo analysis. In Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval. New York, NY, USA: ACM, pp. 251–258.

Cleveland, W.S. & McGill, R., 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Journal of the American Statistical Association, 79(387).

Croxton, F.E., 1932. Graphic Comparisons by Bars, Squares, Circles, and Cubes. Journal of the American Statistical Association, 27(177), 54–60.

Croxton, F.E. & Stryker, R.E., 1927. Bar Charts Versus Circle Diagrams. Journal of the American Statistical Association, 22(160), 473–482.

Cutting, J., DeLong, J. & Nothelfer, C., 2010. Attention and the Evolution of Hollywood Film. Psychological Science, 21(3), 432–439.

Danzico, L., 2008. Telling stories using data: An interview with Jonathan Harris. Available at: http://www.adobe.com/designcenter/thinktank/danzico2.html [Accessed February, 2011].

Dawes, B., 2004. Cinema Redux, Available at: http://daweso.me/b1KMxx [Accessed March, 2011].

Economist, 2010. Show me: New ways of visualising data. Economist (GBR) Economist, 394(8671).

Fielding, R.T., 2000. Architectural styles and the design of network-based software architectures. Ph. D. University of California.

Gershon, N. & Page, W., 2001. What storytelling can do for information visualization. Commun. ACM, 44(8), 31–37.

IMDB, 2011. Database Statistics. Available at: http://www.imdb.com/stats [Accessed February, 2011].

The Joy of Stats, 2010. [TV programme] BBC, BBC Four, 7 December 2010 21.00.

Kaltenbrunner, M. & Bencina, R., 2007. reacTIVision: a computer-vision framework for table-based tangible interaction. In Proceedings of the 1st international conference on Tangible and embedded interaction. Baton Rouge, Louisiana: ACM, pp. 69–74.

Kato, H. et al., 1999. ARToolKit, Hiroshima City University.

Larkin, J. & Simon, H., 1987. Why a Diagram is (Sometimes) Worth Ten Thousand Words. Cognitive Science, 11(1), 65–100.

Lillie, A.S. & Massachusetts Institute of Technology. Dept. of Architecture. Program in Media Arts and Sciences., 2008. MusicBox : navigating the space of your music.

Manovich, L., 2002. The language of new media, Cambridge, Mass.: MIT Press.

Mekas, J., 1967. Interview with Peter Kubelka. Film Culture, (44), 285–299.

Nathan, M. et al., 2008. CollaboraTV: making television viewing social again. In Proceeding of the 1st international conference on Designing interactive user experiences for TV and video. Silicon Valley, California, USA: ACM, pp. 85–94.

NUI Group. Natural User Interface Group. Available at: http://nuigroup.com/go/ [Accessed February, 2011].

O'Brien, C., 2005. Multiple Language Versions and National Films, 1930-1933; Statistical Analysis, Part I. Cinema & Cie : International Film Studies Journal, (6), 45–52.

Pfeiffer, S., Parker, C. & Pang, A., 2006. The Continuous Media Markup Language (CMML), Version 2.1. Available at: http://www.annodex.net/TR/draft-pfeiffer-cmml-03.html [Accessed February, 2011].

Playfair, W. et al. 2005. Commercial and political atlas and Statistical breviary, Cambridge; New York: Cambridge University Press.

Rasheed, Z. & Shah, M., 2002. Movie Genre Classification by Exploiting Audio-Visual Features of Previews. In Proceedings 16th International Conference on Pattern Recognition, 2, 1086–1089.

Ray-Ban, 2010. Virtual Mirror - USA. Available at: http://www.ray-ban.com/usa/science/virtual-mirror [Accessed February, 2011].

Raymond, E.S., 2001. The cathedral and the bazaar : musings on Linux and Open Source by an accidental revolutionary, Beijing; Cambridge, Mass.: O'Reilly.

Romero, M et al. 2008. Viz-A-Vis: toward visualizing video through computer vision. IEEE transactions on visualization and computer graphics, 14(6).

Salavon, J., 2000. The Top Grossing Film of All Time, Available at: http://salavon.com/TGFAT/Titanic.shtml [Accessed March, 2011].

Salt, B., 1974. Statistical Style Analysis of Motion Pictures. Film Quarterly, 28(1), 13–22.

Salt, B., 2006. Moving into pictures: more on film history, style, and analysis, London: Starword.

Salway, A. & Graham, M., 2003. Extracting information about emotions in films. In Proceedings of the eleventh ACM international conference on Multimedia. Berkeley, CA, USA: ACM, pp. 299–302.

Segel, E. & Heer, J., 2010. Narrative Visualization: Telling Stories with Data. IEEE transactions on visualization and computer graphics., 16(6), 1139.

Shamma, D.A., Kennedy, L. & Churchill, E.F., 2009. Tweet the debates: understanding community annotation of uncollected sources. In Proceedings of the first SIGMM workshop on Social media. Beijing, China: ACM, pp. 3–10.

Shaw, D., 2004. Sergei Eisenstein. Senses of Cinema, (30). Available at: http://www.sensesofcinema.com/2004/great-directors/eisenstein/ [Accessed February, 2011].

Smith, J. & Schirling, P., 2006. Metadata standards roundup. Multimedia, IEEE, 13(2), 84–88.

Strobelt, H. et al. 2009. Document cards: A top trumps visualization for documents. IEEE Trans Visual Comput Graphics IEEE Transactions on Visualization and Computer Graphics, 15(6), 1145–1152.

Stump, D., 2008. Metadata & The Future of Filmmaking. Creative COW. Available at: http://magazine.creativecow.net/article/metadata-the-future-of-filmmaking [Accessed February, 2011].

Thompson, K.M., 2005. Herr Lubitsch goes to Hollywood German and American film after World War I, Amsterdam: Amsterdam University Press.

Tufte, E.R., 1983. The visual display of quantitative information, Cheshire, Conn.: Graphics Press.

Tufte, E.R., 1990. Envisioning information, Cheshire, Conn.: Graphics Press.

Tufte, E.R., 2006. Beautiful evidence, Cheshire, Conn.: Graphics Press.

Wallin, D., 2006. Touchlib: an open source multi-touch framework. Available at: http://www.whitenoiseaudio.com/touchlib/ [Accessed February, 2011].

Wang, H.L. & Cheong, L., 2006. Affective understanding in film. Circuits and Systems for Video Technology, IEEE Transactions on, 16(6), 689–704.

Zhang, D., Qi, W. & Zhang, H.J., 2001. A New Shot Boundary Detection Algorithm. Lecture notes in computer science., (2195), 63–70.

Zhou, H. et al. 2010. Movie genre classification via scene categorization. In Proceedings of the international conference on Multimedia. Firenze, Italy: ACM, pp. 747–750. Available at: http://dx.doi.org/10.1145/1873951.1874068.

Appendix A - SMIL Example

```
<smil>
  <head>
    <title>Hero</title>
    <meta name=”Director” content=”Yimou Zhang”/>
    <meta name=”Year” content=”2002”/>
    <meta name=”Actor” content=”Jet Li”/>
    <meta name=”Actor” content=”Tony Leung Chiu Wai”/>
    <meta name=”Actor” content=”Maggie Cheung”/>
    <meta name=”Language” content=”Mandarin”/>
  </head>
  <body>
    <par>
    <seq></seq>
    <smilText region=”c1”>
      <div>
        <metadata name=”author” content=”johdoe”>
        <metadata name=”context” content=”trivia”>
        <clear begin=”t1-5s” />
        Poster in the background is from another film
by same director
        <clear begin=”t1+15s” />
      </div>
    </smilText>
    </par>
  </body>
</smil>
```