Archive Film Material – A Novel Challenge for Automated Film Analysis

Matthias Zeppelzauer, Dalibor Mitrović and Christian Breiteneder

Vienna University of Technology,
Institute of Software Technology and Interactive Systems,
Favoritenstrasse 9-11, 1040 Vienna

1. Introduction

Automated indexing and searching of videos has become an important requirement, due to the availability of large amounts of media in that form held, for example, by broadcasting companies, museums, and on the Internet. The field of content-based video retrieval focuses on the analysis of a video's content, making it automatically searchable by computers. Typical tasks in video retrieval are the automated detection of shot boundaries and scene boundaries, and the detection of highlights. Most research in video retrieval focuses on particular types of video, such as news broadcasts, sports videos and commercials. Compared to the retrieval of these types of video content, the retrieval of *film* has received little attention by the research community. *Archive films* have been especially widely neglected in content-based retrieval. The automated analysis of archive films is more difficult than the analysis and retrieval of contemporary video and film material for several reasons. First, archive film is usually black and white, and thus color information cannot be exploited by automated analysis techniques. Second, archive film material usually has a lower material quality due to its old age, which impedes automated analysis. Third, archive film may exhibit stylistic aspects that are very different from contemporary material.

We have participated in the interdisciplinary research project Digital Formalism[1] that focused on the automated analysis and retrieval of historic film material. The project was a joint-effort between film scholars, archivists, and computer scientists. The project partners were the Department for Theatre, Film and Media Studies at the Vienna University, the Austrian Film Museum, and the Interactive Media Systems Group at the Vienna University of Technology.[2]

The goal of the project was to gain insights into the highly formalized style of filmmaking of the Soviet filmmaker Dziga Vertov (1896-1954). The film scholars at Vienna University provided their knowledge on the work of Vertov and his films. They manually analyzed the films and identified important stylistic aspects that should later be retrieved automatically from them. The Austrian Film Museum[3] provided the project with the historic film material and supported the partners in material-specific questions. The archivists generated comprehensive film annotations that later served as a basis for the quantitative evaluation of the developed retrieval methods. Furthermore, the archivists formulated requirements for the automated analysis from an archive's point of view. The responsibility of the computer scientists at the Vienna University of Technology (our team) was the development of automated retrieval methods based on the requirements of the film scholars and archivists in order to enable efficient access to the material and to support film scientists and archivists in their work. We first collected a comprehensive list of requirements at the beginning of the project and then evaluated the feasibility of the required tasks of analysis in the context of automated retrieval. The result of this process is a set of novel retrieval methods for the extraction of differently complex stylistic aspects in the investigated films. The resulting set of retrieval methods includes techniques for shot boundary detection, scene segmentation, intertitle detection, and the analysis of visual composition and motion composition. During the work on these tasks, we learned that both the complex stylistic attributes in the films as well as the low material quality (artifacts) significantly impede the automated analysis.

In this article, we present the characteristics of the film material from the perspective of computer science. Thereby, we set two emphases. First, we present stylistic aspects that characterize the films to point out the high complexity of the films at the syntactic level (montage) as well as at the semantic level (composition). Second, we demonstrate the state of the film material and overview the artifacts present in the film material that particularly impede automated analysis and retrieval. Finally, we discuss the challenges for content-based retrieval that result from the novel film material and draw conclusions from our work on the project.

2. Stylistic aspects of the film material

The films under investigation are characterized by sophisticated stylistic attributes that were highly innovative for the early years of filmmaking. The films were made by the Soviet filmmaker Dziga Vertov, famous for his highly formalized style of filmmaking.[4] The first films edited by Vertov were newsreel series for the Soviet regime (*Kino-Week*, 1917 and *Kino-Pravda*, 1922). The newsreel series show political, social and economic events of the time and have a purely documentary character, which means that the series contain hardly any staged scenes. Different topics are usually separated by intertitles that give contextual information (necessary due to the absence of sound). In later series the style of montage becomes increasingly experimental in character, and Vertov progressively neglects the narrative structure.

After the newsreel series, Vertov focused on the production of feature-film length documentaries, such as *Kino-Eye* (1924), *The Eleventh Year* (1928), *Man with a Movie Camera* (1929), and *Enthusiasm* (1930). In these, he strictly avoided overt narrative elements and linked shots solely by semantic relationships. In addition, the filmmaker experimented with different stylistic devices and artificial effects. The results are artistic and experimental documentaries that were revolutionary for that time of filmmaking. The most important devices that were relevant in the context of our research project are reviewed in what follows.[5]

The documentaries by Vertov are characterized by sophisticated, and repeatedly occurring, *visual compositions*, comprising particular spatial arrangements of objects and camera perspective.[6] Typical composition types in Vertov's films are diagonal composition, symmetric composition, and compositions achieved by unusual camera perspectives, for example, showing a train passing by from below by a camera mounted between the rails. Examples of other typical visual compositions in the films are shown in Figure 1. In the course of the project we have developed a system for the automated location and retrieval of such visual compositions. For this the user provides an abstract example image that roughly describes the desired composition and the system automatically retrieves scenes from the films with a similar composition.[7]

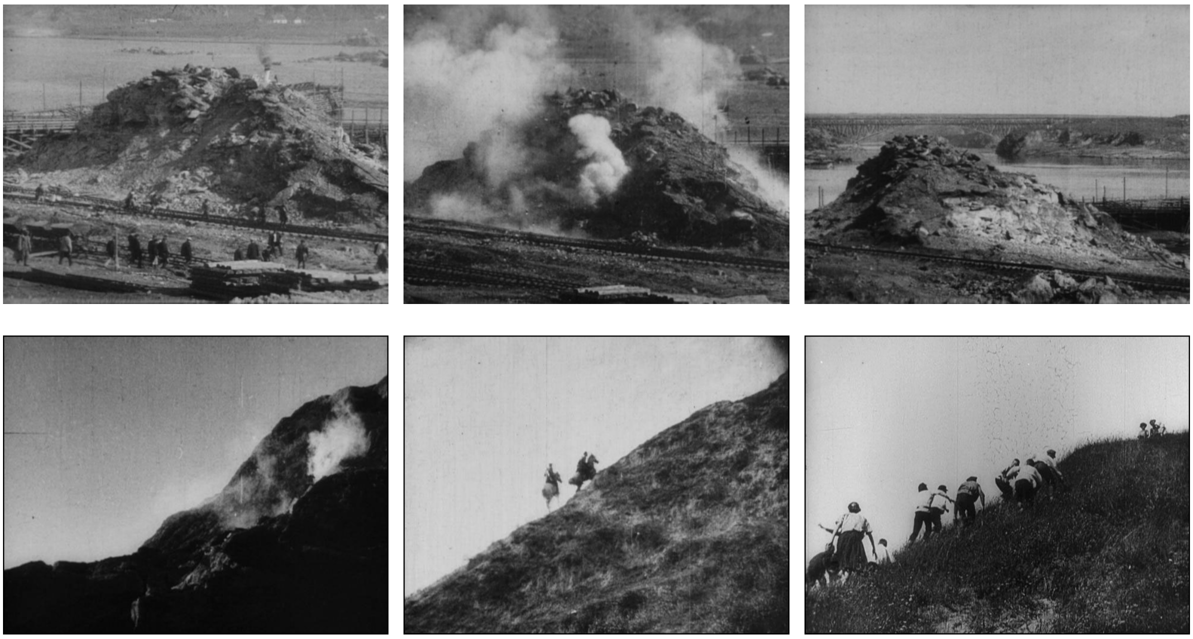
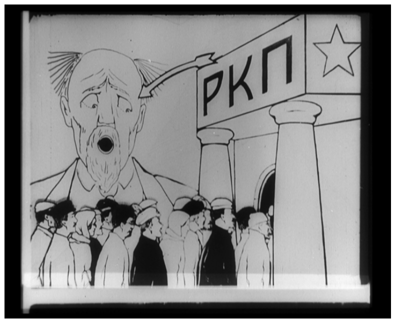


Figure 1: Two examples of three different shots sharing the same visual composition.[8]

(a) multiple images

(b) multiple exposure



(c) animated cartoon

Figure 2: Examples of different artificial effects in Vertov's films.

Vertov experimented with innovative techniques to achieve artificial compositions and special effects. The filmmaker frequently uses *multi-image* compositions (e.g. split-screen) and *multiple exposure* effects to merge several images together and to establish a semantic relationship between them. Examples of multi-image and multiple exposure compositions are shown in Figure 2. The first, a multi-image frame, shows four different images joined together: a piano player and three shots of ballet dancers. The multiple exposure frame in Figure 2b shows the cameraman with his camera in a glass of beer. Vertov further employed the *stop-motion* technique to animate real objects as well as hand-drawn cartoons and intertitles. Figure 2c shows an example of an animated cartoon.

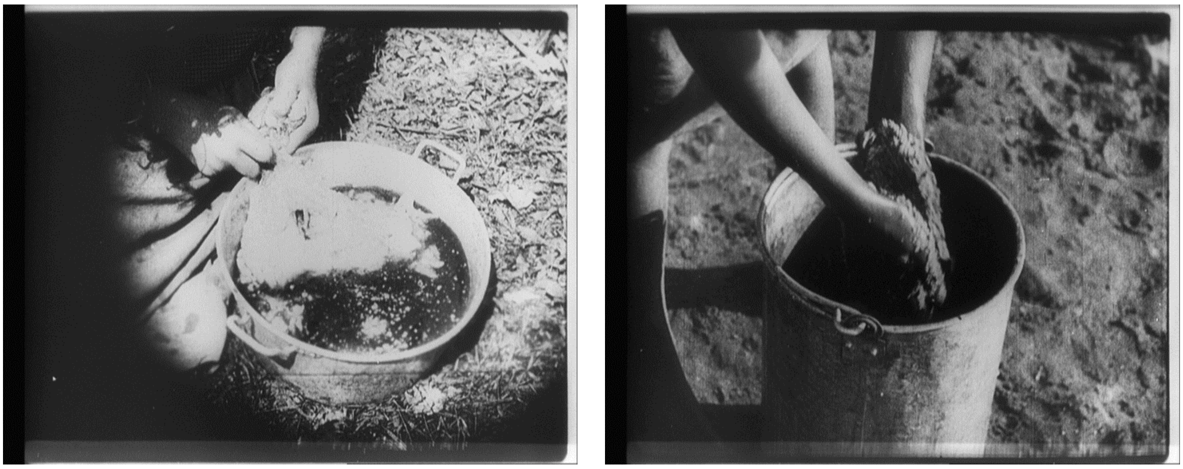
In addition to visual composition, Vertov experimented extensively with motion in his films. Camera and object motion and the interaction of both are used to create complex *motion compositions*. Typical examples in Vertov's films are the motion of machines and parts of machines (pistons, cogwheels, etc.) and typical activities of workers (e.g. hammering, sawing, drilling). The captured motions are often repetitive (cyclic, up/down, left/right) and are often emphasized by simultaneous rhythmical (partly contrapuntal) camera motions.

Figure 3: A combined gradual transition consisting of a fade-out and an iris-out, which is particularly difficult to detect automatically.

Additionally, Vertov shows chasing scenes from different perspectives and uses shaky hand-cameras to increase motion intensity. We have investigated the retrieval of motion composition and have developed an automated retrieval system that allows the user to *sketch* a desired motion composition (e.g. consisting of camera and several object motions).[9] The system analyzes the sketch and then retrieves scenes from the films with a similar motion composition.

A further stylistic device of Vertov's is the manipulation of the *temporal axis*. Vertov shows scenes in forward and reverse order, repeats scenes, and uses fast motion and slow motion. Additionally, Vertov employs the *freeze frame* effect where single frames are unexpectedly frozen for a particular amount of time. Vertov makes extensive use of time manipulation for example in the film *Kino-Eye* where he shows people jumping into the water in forward and reverse order at different speeds. Vertov puts special attention on the transitions between successive shots.

The filmmaker employs a wide range of different gradual transitions to create smooth transitions between shots. In addition to fade-ins, fade-outs, and dissolves, Vertov employs different types of wipes (e.g. bar wipes and iris wipes).[10] Furthermore, he overlays wipes with dissolves and fades to create more complex transitions. An example of a combination of fade-out and iris-out is shown in Figure 3. The intertitle is faded out and simultaneously an iris-out proceeds.[11] An additional means to create smooth transitions between shots used by Vertov are *match cuts*. Vertov employs *form cuts* where similarly shaped objects appear in two successive shots (see Figure 4 for an example), as well as *matching motion*, where the motion direction between two shots is kept consistent. However, he sometimes deliberately violates continuity rules, for example by the integration of abrupt *jumps* which yield a discontinuity in motion that appears unexpected and thus raises the attention of the viewer.



(a) first shot (b) second shot

Figure 4: Keyframes from two successive shots that are connected by a form cut. Note that the two shots originate from different scenes and show different objects and people.

Vertov strictly opposed any narration in his films. He used visual (and later also auditory) motifs that repeatedly appear in a film to introduce structure. Usually, each motif has an associated semantic meaning. Typical motifs are for example power poles and reservoir dams of hydroelectric power plants (as a symbol for electricity and thereby a symbol for economic and social development and prosperity), and factory chimneys (as a symbol for industrial progress). Figure 5 shows four examples of the typical power pole motif. The first three examples in Figures 5a-5c are additionally framed by a superimposed iris mask. The fourth example in Figure 5d is a multiple exposure shot. A flywheel is superimposed inside the tower at the right side of the frame (the arrow labeled "C" points at the center of the wheel). In all four examples we observe that the similarity between the instances of the motif exist mostly on a semantic level. The low visual similarity of the different instances makes the automated detection of a motif a difficult task.

With the establishment of sound film, Vertov incorporated also auditory motifs. Typical auditory motifs in his films are church bells and work sirens. He experimented with the co-occurrence and correlation of visual and auditory motifs to convey sophisticated messages to the viewer. For example, in the film *Enthusiasm* a central sequence shows several consecutive static shots of different religious and monarchal symbols (e.g. a tsarist monogram, statues of Christ, crucifixes). At each shot cut between two different symbols Vertov positioned the sound of a church bell in the soundtrack. The synchronous church bells increase the perceptual salience of the sequence and create a threatening and warning atmosphere. According to film literature, this scene is a key one in the film that expresses the rejection of religion and the tsarist regime by the communist regime.[12] An illustration of the scene with key-frames of the shown symbols is given in Figure 6.

(a) (b)

(c) (d)

Figure 5: Different instances of the power pole motif.

The co-occurrence of visual and auditory motifs is composed in a rhythmic manner. Vertov experimented with this kind of rhythmic montage in different ways. In the already discussed sequence, for example, the cutting rate doubles at the end of the sequence, while the frequency of the church bells remains the same. The doubling of the frequency leads to an increase of the tempo in the sequence towards the end. Vertov frequently correlates auditory and visual motifs to emphasize the importance of a scene. In the course of the project we developed techniques to extract and segment such scenes automatically from the films, which is of great interest for film scholars.[13] Vertov further employs *accelerated montage* where the shot frequency is increased successively. In such sequences the shot length is decreased down to a few frames or even only a single frame. Subsequent extremely short cuts visually merge and create a special type of flicker. Examples of such sequences can be found at the end of the film *Man with a Movie Camera*. This kind of sequence is especially challenging for automated shot cut detection due to the short shot lengths. Vertov further inserts completely black frames into sequences at regular intervals to create rhythmic patterns. An example of such a sequence is illustrated in Figure 7. The sequence shows a train ride where the director alternately shows black frames and shots of the rail track. This should evoke the impression of the train passing over expansion joints (responsible for the "clickety-clack").

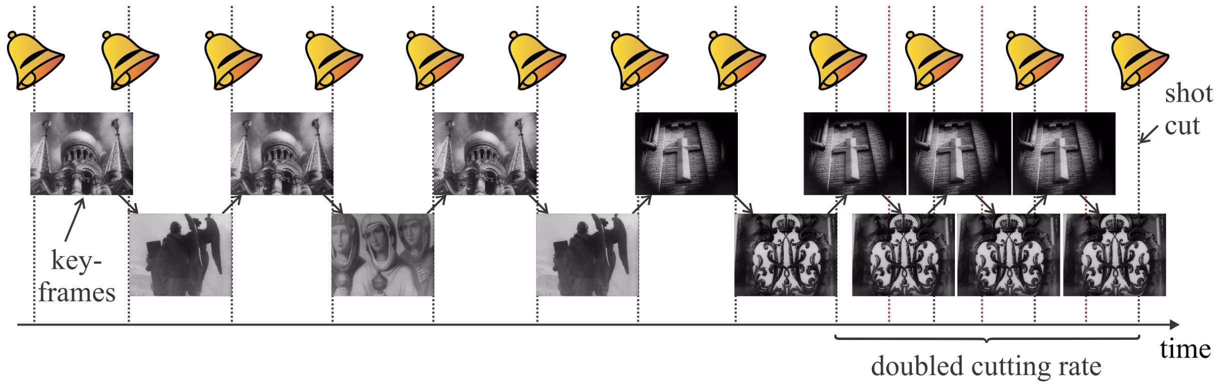


Figure 6: Synchronously intercut visual and auditory motifs.

Figure 7: Black frames cut in-between a series of frames showing rail tracks. The black frames visually evoke the otherwise auditory impression of passing over expansion joints.

The montage style of Vertov has a very systematic and formalistic character. The filmmaker arranges shots of different motifs systematically over time. This results in parallel montages where the visual motifs alternately appear in repetitive patterns. For one such sequence (for the film *Kino-Eye*) a plot sketched by Vertov has been preserved that shows how the filmmaker conceptualized the sequence, see Figure 8. Time is represented horizontally in the plot. Each row represents a motif and each column represents a shot. The entries in the cells represent the duration of each shot (in number of frames). The right-most column contains the sum of frames for each motif in the entire sequence. The plot indicates that both, the temporal composition of the motifs and their corresponding shot lengths have been systematically determined.[14] Key-frames (manually selected) of the first 15 shots of this sequence are shown in Figure 9. The sequence shows the hissing of a flag and alternately shows the flag and different people's faces that observe the hissing. The order of the motifs as well as the shot lengths are consistent with the plot in Figure 8, except for a few missing shots at the beginning (maybe lost due to film tears).

The distinct composition and montage of Vertov's films, then, make them a challenging material for automated film analysis. The large variety of employed stylistic devices and the strong formalistic structure raise novel requirements for automated film analysis and retrieval. However, the *state* of the material is challenging for automated analysis, as well. Due to their age the films contain numerous artifacts that interfere with automated analysis techniques. These artifacts impede the automated analysis and retrieval of the films.

Приложение

ЦИФРОВАЯ ЗАПИСЬ

монтажного куска из фильма Дзиги Вертова—момент подъема флага в день открытия пионерского лагеря.

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. Голова вожатого	11			7		9				14												
2. Пионерка у мачты		14					28					10				10			8			
3. Мачта и флаг			31					24			10			10			10					26
4. Надпись: флаг поднять!					24																	
5. Трубачи									16													
6. Лицо № 1													10									
7. Лицо № 2															10							
8. Лицо № 3																		10				
9. Лицо № 4																				10		
10. Рука, защищающая от солнца																						
11. Поднятые руки крестьянских ребят																						
12. Крестьянка с поднятой рукой																						
13. Лицо № 5																						
14. Отряд в строю																						
15. Тень флага																						
16. Ноги пионерки на посту																						

№	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	Количество кадров каждого действующего лица.
1. Голова вожатого					11								10																		62
2. Пионерка у мачты		20																			93										183
3. Мачта и флаг				10		10		10		9				10																48	208
4. Надпись: флаг поднять!																															24
5. Трубачи			14									10																			40
6. Лицо № 1																10								7					5		32
7. Лицо № 2																	10								7						27
8. Лицо № 3																		10								5					25
9. Лицо № 4	15																		10								5				40
10. Рука, защищающая от солнца							10																								10
11. Поднятые руки крестьянских ребят									28																						28
12. Крестьянка с поднятой рукой											13																				13
13. Лицо № 5															10													5			15
14. Отряд в строю																				73											73
15. Тень флага																						65									65
16. Ноги пионерки на посту																							60								60

ИТОГО .... 905 кадров
т. е. около 17½ метров

Примечания:
1. №№ 1, 2, 3 и т. д. до 52 обозначают порядок следования кусков — смен друг за другом
2. Цифры в клетках обозначают число кадров каждого куска (каждой смены)
3. Слева таблицы с 1-го по 16-й перечислены действующие лица и вещи
4. Справа таблицы в столбике цифр указано число кадров для каждого действующего лица
На 17½ метров здесь приходится 52 смены при 16 действующих вещах или людях

Figure 8: The visual representation of the montage of a sequence.[15]

Figure 9: Keyframes of the first 15 shots of the sequence that shows the hissing of a flag. Each row shows one of the motifs that are specified in the plot in Figure 8.

Figure 10: A 35mm silent film strip.[16]

3. State of the film material

Vertov produced the films that we are investigating in the Soviet Union in the 1920s and 1930s. Most of these archive films are silent, only the late films contain a sound track. The material is 35mm black-and-white film made of cellulose triacetate (see Figure 10). The films are played at non-standard frame rates of 16 to 20 frames per second for silent film and 24 frames per second for sound film. In sound films, the soundtrack has been optically stored on the filmstrip (sound-on-film technique)[17] alongside the visual content of the frame, see Figure 11. Note that the introduction of sound film had an effect on silent film material as well. In sound film there is some reserved space for the soundtrack at the left of each frame which reduces the area reserved for the visible content of a frame. When a silent film was copied with a copy machine configured for sound film, the area reserved for the sound track remained unexposed and the coinciding image content from the silent film was lost. This results in frames that are cropped at the left side. In the filmstrip in Figure 10 we can observe this copy artifact. An overview of the available archive films in our project is given in Table 1. For two of the films different versions exist. For *Enthusiasm* there is a slightly varying restored version of the original one.[18] For *Man with a Movie Camera* there exist a version from the Vertov Collection of the Austrian Film Museum in Vienna and a version from Amsterdam that was generously provided by the EYE Film Institute Netherlands.[19]

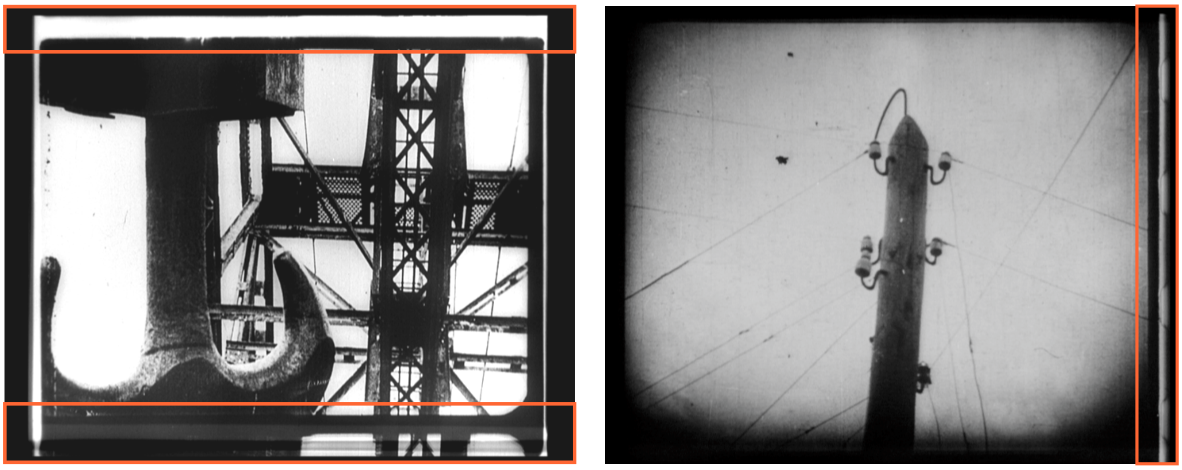
Prior to automated analysis we have digitized the analog filmstrips. For this purpose we scan the films *frame-by-frame* in PAL quality (720x576 pixels) and with 256 gray values (8 bit). The result of digitization is an image sequence that represents each frame of a film. The frame-by-frame digitization avoids the introduction of interpolated frames, which usually occur during digitization at standard frame rates (e.g. 25 fps for PAL) when the projected film is, for example, captured by a digital camera. Since interpolated frames represent information that does not exist in the original material they would interfere with automated analysis. It is crucial to note that Vertov employed the exact number of frames for a given shot as a stylistic device (as can be observed from the plot in Figure 8 above. Thus interpolated frames would tamper with the intended statements of the films.[20] Additionally, we skip (lossy) compression to avoid the introduction of additional artifacts in the digitized stream. If an audio track is available in a film, the track is scanned as well and converted into an uncompressed PCM[21] encoded file.

(a) (b)

Figure 11: Two examples of sound-on-film in historic material from the film *Enthusiasm*. The waveform of the sound is optically stored at the left side of the frame.[22]

Title	FPS	Duration	#Frames	#Shots	Sound?
The Eleventh Year	18	00:58:26	63123	660	no
A Sixth Part of the World	18	01:04:03	69182	1017	no
Enthusiasm (Original)	24	01:07:26	97116	604	yes
Enthusiasm (Restored)	24	01:07:27	97134	612	yes
Im Schatten der Maschine*	18	00:23:43	25622	420	no
Kino-Eye	18	01:17:54	84132	1304	no
Kino-Pravda 21	18	00:32:27	35060	413	no
Man with a Movie Camera (Vienna)	22	01:12:33	95768	1782	no
Man with a Movie Camera (Amsterdam)	22	01:11:01	93743	1781	no
Stride, Soviet!	18	01:12:28	78272	1110	no
Three Songs of Lenin	24	01:01:49	89023	817	yes

Table 1: Films of Vertov analyzed in the project in chronological order.
*** The film is a compilation film by Adalbert Viktor Blum and reuses content produced by Vertov.**

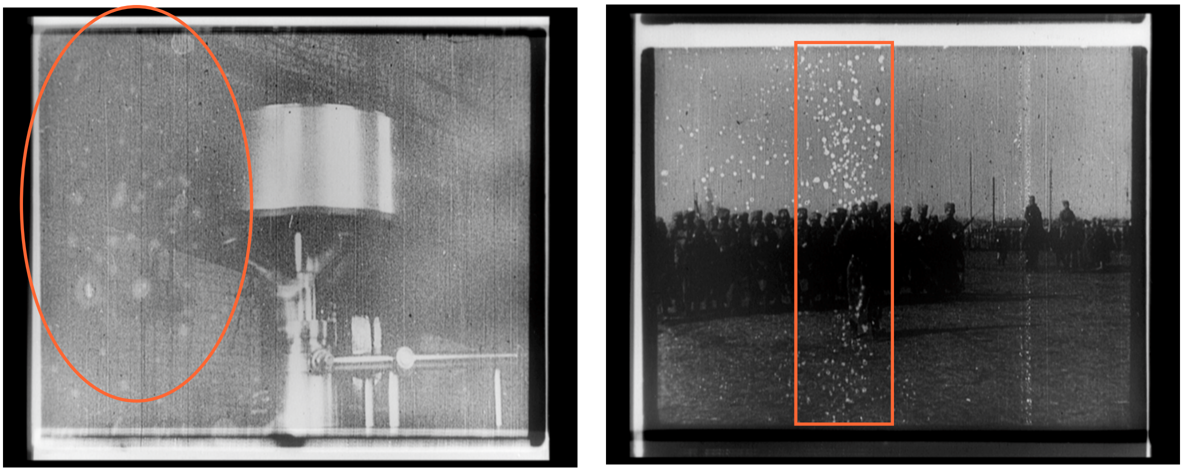


(a) vertical shrinking (b) horizontal shrinking

Figure 12: The effect of the shrinking and inappropriate copying of filmstrips. (a) framelines become visible at the top and the bottom of the frame, as well as image content of the next frame (at the bottom). (b) horizontal shrinking causes the perforation of the filmstrip to become visible (at the right side of the frame).

The provided filmstrips are multiple-generation copies that were never intended to be used for other purposes than backups. Due to this fact, these copies were not handled with much care in the past. Today, the original filmstrips do not exist anymore, hence the available backup copies are the only existing source material left. The state of the material has degraded significantly, during storage, copying, and playback over the last decades. The filmstrips are made of organic material (cellulose triacetate). Storage over the last decades resulted in *shrinking* of the filmstrips due to chemical processes in the base support. Because of this the filmstrips physically contracted horizontally and vertically. Shrinking, in combination with inappropriate copying (see below), results in frame displacements and non-linear geometric distortions. The framelines (the area between two successive frames on a filmstrip) become visible and sometimes also the content of the next frame, see Figure 12a. The shrinking in horizontal direction is usually less disturbing. However, if the horizontal shrinking exceeds a certain limit, the perforation of the filmstrip becomes visible in the frame, as in Figure 12b. Further artifacts originate from storage under suboptimal conditions. During the long storage time, mould and humidity harmed the filmstrips. Typical artifacts and distortions originating from mold and humidity are shown in Figure 13.

The archive films are multiple-generation copies. When copying is performed under suboptimal conditions dirt and dust may be copied into the films. With each generation of copy dirt, dust and previously existing artifacts (e.g. scratches and shrinking artifacts) accumulate. This results in a broad spectrum of distortions as demonstrated in Figure 14. Additionally, repeated copying leads to a degradation of contrast with each generation of copy. This results in low-contrast images as shown for example in Figure 13b. When the filmstrips are not exactly aligned to each other during copying the frames of the original filmstrip are incorrectly mapped to the frames of the new filmstrip. This results in frame displacements and visible framelines. For examples see Figures 13 and 14 where all frames suffer from these artifacts. The frame displacements from copying together with the displacements caused by film shrinking introduce a significant *shaking* in the films.



(a) (b)

Figure 13: Artifacts originating from liquids and mold.

(a) wrong exposure (b) dirt (c) scratch

(d) a fingerprint (e) dirt, vertical scratches (f) blurring

Figure 14: Artifacts introduced from (repeated) copying (dirt, scratches, and distortions in brightness, blurring).

Further artifacts have been introduced during playback of the films in old projectors. Dirt in the mechanics of the projectors (in the film transport) introduces vertical scratches that cover many subsequent frames. Vertical scratches are shown for example in Figure 14e. Playback and presentation of the films further bear the risk that a filmstrip tears. Each time, a filmstrip tears a few frames of the strip are destroyed. When the film is then glued together the absence of the destroyed frames produces abrupt jumps in the movies. Additionally, the splice becomes a visible artifact at the position where the filmstrip is glued together. An example of a film tear is shown in Figure 15.

(a) frame 1 (b) frame 2 (c) frame 3

Figure 15: Three successive frames of a sequence. The film has torn after frame 1 and several frames are missing between frame 1 and frame 2 which results in a discontinuity in motion. In frame 2 artifacts from gluing the film together (with adhesive tape) are visible in the upper part.

(a) (b) (c)

Figure 16: Recording with a historic movie camera. The cameraman rotates the crank manually to transport the filmstrip.

Additionally to artifacts from storage, copying and playback the films suffer from limitations of the recording technique of the early 1920s and 1930s. In the early years of filmmaking, the film transport was controlled manually. For this purpose, the cameraman moved a crank at one side of the camera to move the filmstrip and to control the shutter of the camera, see Figure 16. The manual film transport yields permanent variations of the frame rate across the filmstrip. As a consequence the exposure along the filmstrip varies which in turn generates frames of different brightness. The resulting effect during playback of the films is strong and fast alternating flicker shown in Figure 17.

The artifacts in the historic films impede the automated analysis. We distinguish between three different classes of artifacts in the context of automated analysis: global artifacts, local artifacts and temporal artifacts. An overview of the different types of artifacts and their observed instances is provided in Table 2. *Global* artifacts influence the entire area of a frame and comprise shaking, flicker, blurring, and low contrast. Shaking is most disturbing in motion analysis where motion vectors between pixels in the image have to be computed reliably. Due to shaking the resulting motion estimates are noisy and often the tracking of motion is not possible over longer time spans.[23] Together with the complex compositions of camera and object motion present in the films the analysis of motion is challenging for the material.[24]

(a) (b) (c)

Figure 17: Three successive frames from a shot in *Kino-Pravda 21*. Heavy flicker is introduced due to the manual and uneven film transport.

Global	**Local**	**Temporal**
Shaking	Visible framelines	Jump cuts, splices (film tears)
Flicker	Visible frame borders	Asynchronous visual and auditory tracks
Blurring	Scratches	
Low contrast	Dirt, dust	
	Liquids	
	Mold	

Table 2: Types of artifacts and their observed instances in the film material under investigation.

Flicker is problematic since it influences the overall brightness of frames and impedes similarity comparisons between frames which are an essential part in most visual retrieval tasks, such as shot boundary detection and scene segmentation. A common practice is to compare frames based on color and intensity histograms[25]. However, such comparisons are not robust in the presence of flicker because it distorts the histograms globally. Shot cut detection based on intensity histograms, for example, does not work because flicker produces abrupt changes in the histograms of subsequent frames, which are easily confused with shot cuts. The result is a large number of falsely detected shot cuts. Flicker further disturbs motion tracking. The goal of motion tracking is to track single points in a frame across subsequent frames. When flicker is present, the brightness of a point varies over time. This interferes with most motion tracking approaches, which require that a point has constant brightness over time. The same as for flicker applies to low-contrast images. Again analysis techniques that rely on brightness are not suitable because they cannot capture distinctive information from low-contrast frames. Additionally, the extraction of local image descriptors[26] is difficult due to a lack of distinct (salient) points in low-contrast images. Similarly, the extraction of local image descriptors is problematic in blurred images, which lack the distinct structures necessary for the identification of specific feature points.

In addition to global artifacts, there are *local* artifacts that affect only a part of a frame's area. Local artifacts comprise visible framelines and frame borders (e.g. perforation), scratches, dirt, dust, and artifacts from liquids spilled over the filmstrip and mould. Local artifacts represent misleading information that disturbs automated analysis. Visible framelines and frame borders can easily be removed by cropping the frame borders for an entire film. However, cropping with a constant offset may remove visual information since the position of the frame borders varies over time. Artifacts like dirt and scratches interfere with analysis techniques that operate on small scales (small analysis windows), such as block-based image features with small block-size and local descriptors of feature points. As a consequence, the description of fine structures in the frames is prone to errors. Additionally, the local artifacts generate abrupt visual changes that interfere with temporal film analysis, required for shot and scene segmentation.

The third class of artifacts is *temporal* artifacts, which comprise distortions of the temporal axis of a filmstrip. Temporal artifacts are, for example, abrupt jumps caused by film tears that introduce motion discontinuities. Such discontinuities especially interfere with motion analysis. Today, it is not clear whether Vertov employed such jumps on purpose as a stylistic device (jump *cuts*) or if these jumps indicate damages in the film material. Another temporal artifact stems from asynchronous visual and auditory information. The visual track and the auditory track have originally been stored on *different* filmstrips.[27] Due to external influences during copying (e.g. copying with wrong synchronization points) and playback (e.g. film tears) the visual and the auditory track may become asynchronous. Asynchronous visual and auditory information significantly changes the original intentions of the filmmaker and the meaning of a film. In automated analysis, asynchronous tracks for example impede the detection of co-occurring audio and visual motifs and events. We find asynchronous auditory and visual information for example in the film *Enthusiasm*.[28] An attempt towards re-synchronizing *Enthusiasm* has been performed by Peter Kubelka in 1972.[29]

The historic film material and its artifacts challenge automated analysis techniques and often make existing techniques inapplicable. A first step towards the analysis of archive film material is an automated restoration that removes the most disturbing global and local artifacts.

4. Automated film restoration

A straightforward approach to improve the quality of historic film material is to apply professional software for its restoration. However, professional software for film restoration is expensive and usually requires human interaction or supervision.[30] This makes high-quality restoration of large amounts of films costly and often unfeasible. Cost-effective alternatives to professional film restoration are algorithms that filter specific artifacts fully automatically. We exemplarily explored algorithms for deflickering, noise reduction, and image stabilization to remove flicker, dirt, and shaking. One might expect that such preprocessing improves the following content-based analysis and retrieval. In fact, most methods reduce the corresponding artifacts. However, they introduce new artifacts which are often more disturbing than the original ones. For example, the employed *deflicker* method based on histogram alignment[31] significantly dampens the brightness variations across the frames but fails when the brightness variations exceed a certain level, resulting in contrast distortions as depicted in Figures 18a-18d. Figures 18a and 18c show keyframes of two sequences where the deflicker filter does not work correctly due to large brightness variations. In Figure 18b noise is emphasized (mostly in the sky) and in Figure 18d noise is introduced in the background. We reduce *noise* by a temporal median filter. This removes most scratches and dirt but cancels out image details that are necessary for subsequent analyses (e.g. detection and tracking of feature points for motion analysis).

(a) original (b) deflickered (c) original (d) deflickered

(e) original (f) stabilized (g) original (h) stabilized

Figure 18: Artifacts that originate from automated restoration.

Stabilization aims at removing shaking from a sequence which is caused by repeated copying and film shrinking. The challenge is to remove shaking independently from the *intended* camera and object motions in the film. Stabilization methods work well for scenes with small moving objects or smooth camera operations.[32] In scenes with large moving objects or fast and non-uniform camera motion, stabilization methods often confuse unintended shaking with the intended motion. This behavior leads to unexpected results such as rotation and unwanted warping of the frames (see Figures 18e-18h). Figures 18e and 18g show keyframes of sequences where stabilization fails. Figure 18e shows a man who turns his head. The stabilizer fails to compensate for the object motion resulting in an unwanted rotation of the frame in Figure 18f. Figure 18g shows a train passing by. Since there is hardly any static background, the stabilizer fails to align the images and falsely translates and scales up the frame in Figure 18h.

We observe that fully automated algorithms introduce new artifacts and remove detail information that is necessary for automated processing. From these observations, we conclude that such an automated preprocessing is not advisable with the investigated archive film material and that intensive human interaction would still be required. In the course of our investigations, we observed that, for each retrieval task, there exists a specific subgroup of artifacts that interfere with the analysis while other artifacts do not. In shot cut detection, for example, shaking does not interfere since we operate on single keyframes and do not consider motion. Flicker however impedes similarity comparisons and thus must be handled carefully in shot cut detection. We concluded to skip automated preprocessing and instead focused on the development of retrieval methods that are robust to those artifacts that were actually interfering in the task.[33]

5. Conclusion

Archive film material is a novel and challenging type of film material, one that has not been subjected to automated analysis and retrieval prior to our project. The film material is challenging due to two reasons: First, the presence of sophisticated stylistic attributes (e.g. montage and composition) led to the formulation of innovative retrieval tasks by film experts. Second, the contained artifacts impeded automated analysis and required the development of particularly robust retrieval methods. The major goal of the project was the development of automated retrieval techniques that support film scholars in their work. In the course of the project, we have developed a number of retrieval techniques for the extraction of particular stylistic attributes from the films defined by the film scientists. Methods include, for example, the extraction of visual compositions by composition templates,[34] methods for the retrieval of user-defined motion compositions and motion continuity,[35] methods for temporal film segmentation (shot segmentation and scene segmentation),[36] and techniques for the analysis of the audiovisual composition.[37] Evaluations of the methods developed showed to which degree the automated extraction of these stylistic attributes from the films is feasible. Our research has taken original directions and we have developed tools that support film scholars in their analyses. Furthermore, several of the research needs of these scholars, ones that were not directly addressed in the project, open up space for future investigation, for example the retrieval of multi-image compositions, the detection of multi-exposure shots, and the automated detection of time axis manipulations.[38]

The contribution of this article is a first presentation and analysis of the collected stylistic and material-specific properties of the investigated film material. We have summarized the most important stylistic attributes in our overview, and we have classified the artifacts present in the historic material. Finally, we have discussed the effects of the artifacts on automated analysis and retrieval. Despite the low material quality of the archive films the retrieval methods we developed have been able to extract particular stylistic attributes from them. The methods have also allowed a much more efficient indexing of, and access to, the films and enable a much faster analysis of attributes whose manual extraction is tedious. During the project, then, we demonstrated that automated retrieval, and the specific techniques for this that we developed, are very well able to support the analyses of film scholars and archivists, as well as to aid them in gaining a original view on the historic films.[39]

References

Beaver, F., *Dictionary of film terms: the aesthetic companion to film art* (New York: Peter Lang Publishing, 2009).

Fisher, L., '*Enthusiasm*: From kino-eye to radio-eye,' In E. Weis and J. Belton (eds), *Film Sound-Theory and Practice* (New York: Columbia University Press, 1985), pp. 247–264.

Handzo, S, 'Appendix: A narrative glossary of film sound technology,' In E. Weis and J. Belton (eds), *Film Sound-Theory and Practice* (New York: Columbia University Press, 1985), pp. 383–426.

Lowe, D., 'Distinctive image features from scale-invariant keypoints,' *International Journal of Computer Vision*, 60(2):91–110, 2004.

Mitrović, D, Zeppelzauer, M., Zaharieva, M., and Breiteneder, C., ‚Retrieval of Visual Composition in Film,' *Proceedings of the 12th International Workshop on Image Analysis for Multimedia Interactive Services*, April 13-15, 2011, Delft, The Netherlands, 2011.

Pitie, F., Kokaram, A., and Dahyot, R., 'N-dimensional probability density function transfer and its application to colour transfer,' In *International Conference on Computer Vision*, 2:1434–1439, 2005.

Schallauer, P., Bailer, W., Morzinger, R., Furntratt, H., and Thallinger, G., 'Automatic quality analysis for film and video restoration,' In *IEEE International Conference on Image Processing*, 4:9–12, 2007.

Seidl, M, Zeppelzauer, M., Mitrović, D., and Breiteneder, C., ‚Gradual Transition Detection in Historic Film Material – A Systematic Study,' *ACM Journal on Computing and Cultural Heritage*, Vol. 4, No. 3, article 10, 2011.

Thalin, G., 'Deshaker - Video Stabilizer,'. Online at: http://guthspot.se/video/deshaker.htm.

Tode, T and Wurm, B. (eds), *Dziga Vertov: The Vertov Collection at the Austrian Film Museum* (Vienna: Austrian Film Museum/SYNEMA, 2006).

Zaharieva, M., Mitrović, D., Zeppelzauer, M., and Breiteneder, C., 'Film Analysis of Archive Documentaries,' *IEEE Multimedia*, 18(2):38-47, 2011.

Zeppelzauer, M, Zaharieva, M., , D., Mitrović and Breiteneder, C., 'A Novel Trajectory Clustering Approach for Motion Segmentation,' *Proceedings of Multimedia Modeling Conference*, Jan 6-8, 2010, Chongqing, China, 2010: pp. 433-443.

Zeppelzauer, M, Zaharieva, M., Mitrović, D., and Breiteneder, C., ‚Retrieval of Motion Composition in Film,' *Digital Creativity*, 22(4):29-234, 2011.

Zeppelzauer, M., Mitrović, D., and Breiteneder, C., 'Cross-Modal Analysis of Audio-Visual Film Montage,' In *Proceedings of 20th International Conference on Computer Communications and Networks*, 2011.

---

[1] See also: http://www.filmmuseum.at/en/_research__education/research_projects/digital_formalism_1

[2] We, the authors of this article, are all based at the latter institution.

[3] The Austrian Film Museum team included Adelheid Heftberger who has also contributed an essay to this issue of Frames. See also her publicatoons about the Digital Formalism project: Heftberger, Adelheid, 'Do Computers Dream of Cinema? Film Data for Computer Analysis and Visualisation', In David M. Berry (ed.), *Understanding Digital Humanities* (London, New York: Palgrave Macmillan, 2012), and Heftberger, Adelheid, 'Zerschnittene Bilder. Die drei Fassungen von Dziga Vertovs Tri pesni o Lenine (1934/35, 1938 und 1970)', In Georg Gierzinger, Sylvia Hölzl, Christine Roner (eds.), *Spielformen der Macht. Interdisziplinäre Perspektiven auf Macht im Rahmen junger slawistischer Forschung* (Innsbruck: Innsbruck university press, 2011), 259–275).

[4] Vertov was born as David Kaufman on January 15, 1896 in Bialystok, Poland, and he died on February 12, 1954 in Moscow, Russia. For further biographical information please see T. Tode and B. Wurm (eds). *Dziga Vertov: The Vertov Collection at the Austrian Film Museum* (Austrian Film Museum/SYNEMA, 2006).

[5] A more comprehensive presentation of Vertov's work and his filmmaking methods is, for example, set out in Tode and Wurm, op. cit.

[6] D. Mitrović, M. Zeppelzauer, M. Zaharieva and C. Breiteneder, 'Retrieval of Visual Composition in Film,' *Proceedings of the 12th International Workshop on Image Analysis for Multimedia Interactive Services*, April 13-15, 2011, Delft, The Netherlands, 2011.

[7] We provide details on the system and a user study in D. Mitrović, M. Zeppelzauer, M. Zaharieva and C. Breiteneder, op. cit.

[8] Similar compositions were employed in the user study for the evaluation of our retrieval system in Mitrović, M. Zeppelzauer, M. Zaharieva and C. Breiteneder, op. cit.

[9] We have investigated the retrieval of motion composition extensively in the following publications: M. Zeppelzauer, M. Zaharieva, D. Mitrović and C. Breiteneder, 'Retrieval of Motion Composition in Film,' *Digital Creativity*, 22(4):29-234, 2011; and M. Zeppelzauer, M. Zaharieva, D. Mitrović and C. Breiteneder, 'A Novel Trajectory Clustering Approach for Motion Segmentation,' *Proceedings of Multimedia Modeling Conference*, Jan 6-8, 2010, Chongqing, China, pages 433-443, 2010.

[10] These categories of wipes and other film transition and editing techniques, here and throughout our discussion, are taken from F. Beaver, *Dictionary of film terms: the aesthetic companion to film art* (New York: Peter Lang Publishing, 2009).

[11] Further examples of gradual transitions and a detailed investigation of gradual transition detection in the historic material can be found in M. Seidl, M. Zeppelzauer, D. Mitrović and C. Breiteneder, ‚Gradual Transition

Detection in Historic Film Material – A Systematic Study,' *ACM Journal on Computing and Cultural Heritage*, vol. 4, no. 3, article 10, 2011.
[12] See, for example, L. Fisher, '*Enthusiasm*: From kino-eye to radio-eye,' In E. Weis and J. Belton, editors, *Film Sound-Theory and Practice*, (New York: Columbia University Press, 1985), pp. 247–264.
[13] M. Zeppelzauer, D. Mitrović and C. Breiteneder, 'Cross-Modal Analysis of Audio-Visual Film Montage,' In *Proceedings of 20th International Conference on Computer Communications and Networks*, 2011.
[14] See Tode and Wurm, op. cit.
[15] The material is provided with kind permission of the Austrian Film Museum, 2011.
[16] The material is provided with kind permission of the Austrian Film Museum, 2011.
[17] See S. Handzo, 'Appendix: A narrative glossary of film sound technology,' In E. Weis and J. Belton (eds), *Film Sound-Theory and Practice* (New York: Columbia University Press, 1985), pp. 383–426.
[18] According to Tode and Wurm, op. cit.
[19] Restoration in April 2010 by EYE Film Institute Netherlands. Theroject was headed by Senior Curator Mark-Paul Meyer in collaboration with the Austrian Film Museum (Adelheid Heftberger), where a nitrate print dating from Vertov's time has been restored in the original full frame format (1:1.33 aspect ratio).
[20] See Tode and Wurm, op. cit.
[21] See also: http://en.wikipedia.org/wiki/PCM

[22] The material is provided with kind permission of the Austrian Film Museum, 2011.
[23] See Zeppelzauer, Zaharieva, Mitrović and Breiteneder, 'A Novel Trajectory Clustering Approach for Motion Segmentation', op.cit.
[24] We investigate motion analysis and the retrieval of motion compositions in Zeppelzauer, Zaharieva, Mitrović and Breiteneder, 'Retrieval of Motion Composition in Film,' op. cit.
[25] See also: http://en.wikipedia.org/wiki/Color_histogram#Intensity_histogram_of_continuous_data
[26] D. Lowe, 'Distinctive image features from scale-invariant keypoints,' *International Journal of Computer Vision*, 60(2): 91–110, 2004.
[27] See Tode and Wurm, op. cit.
[28] See Tode and Wurm, op. cit.
[29] See Tode and Wurm, op. cit.
[30] See P. Schallauer, W. Bailer, R. Morzinger, H. Furntratt, and G. Thallinger, 'Automatic quality analysis for film and video restoration,' In IEEE International Conference on Image Processing, 4:9–12, 2007.
[31] F. Pitie, A. Kokaram, and R. Dahyot, 'N-dimensional probability density function transfer and its application to colour transfer,' In *International Conference on Computer Vision*, 2:1434–1439, 2005.
[32] G. Thalin. Deshaker, 'Video Stabilizer.' Online at: http://guthspot.se/video/deshaker.htm. Last accessed April 2012.
[33] The major goal of this article was the presentation of the film material *itself* rather than a description of the retrieval techniques developed in the course of our project, or the retrieval results. A detailed discussion of the techniques is thus beyond the scope of this article. We give an overview of the retrieval techniques we developed in M. Zaharieva, D. Mitrović, M. Zeppelzauer and C. Breiteneder, 'Film Analysis of Archive Documentaries,' *IEEE Multimedia*, 18(2):38-47, 2011.
[34] Mitrović, Zeppelzauer, Zaharieva and Breiteneder, 'Retrieval of Visual Composition in Film,' op. cit.
[35] Zeppelzauer, Zaharieva, Mitrović and Breiteneder, 'Retrieval of Motion Composition in Film,' op. cit.
[36] Seidl, Zeppelzauer, Mitrović and Breiteneder, 'Gradual Transition Detection in Historic Film Material – A Systematic Study', op. cit.
[37] Zeppelzauer, Mitrović and Breiteneder, 'Cross-Modal Analysis of Audio-Visual Film Montage', op. cit.
[38] Zaharieva, Mitrović, Zeppelzauer and Breiteneder, 'Film Analysis of Archive Documentaries', op. cit.
[39] For an account of this by one of the scholars and archivists involved, please see Heftberger, Adelheid, 'Do Computers Dream of Cinema? Film Data for Computer Analysis and Visualisation', op. cit., and Heftberger, Adelheid, 'Zerschnittene Bilder. Die drei Fassungen von Dziga Vertovs Tri pesni o Lenine (1934/35, 1938 und 1970)', op. cit.

Frames #1 Film and Moving Image Studies Re-Born Digital? 2012-07-02, this article © Matthias Zeppelzauer, Dalibor Mitrović and Christian Breiteneder. This article has been blind peer-reviewed.