Eurographics Conference on Visualization (EuroVis) 2019
M. Gleicher, H. Leitte, and I. Viola
(Guest Editors)

Volume 38 (2019), Number 3

VIAN: A Visual Annotation Tool for Film Analysis

Gaudenz Halter[†], Rafael Ballester-Ripoll[‡], Barbara Flueckiger[†], and Renato Pajarola[‡]

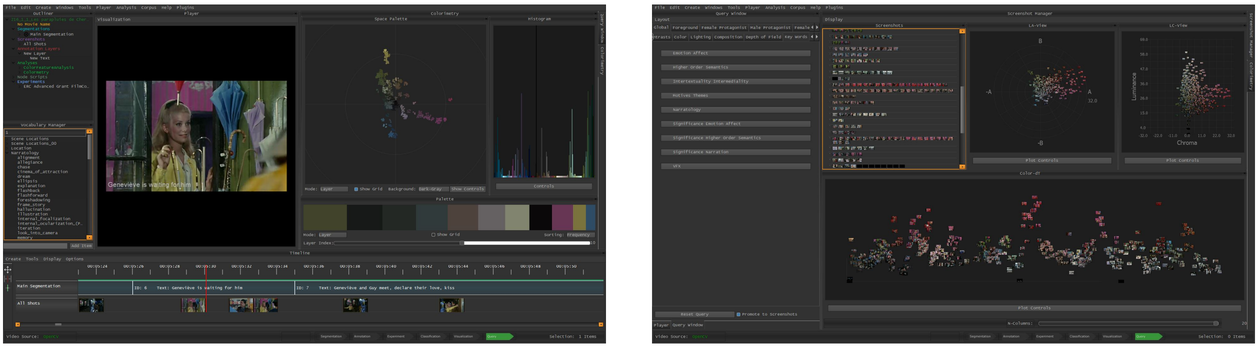


Figure 1: *Left*: Single segment analysis in VIAN showing its timeline and perceptual color widgets. *Right*: Global view showing a vocabulary navigator for interactive semantic analysis and annotation, keyword-filtered ensemble of screenshots, and visualization of the film's frames over time according to their chromatic qualities.

Abstract

While color plays a fundamental role in film design and production, existing solutions for film analysis in the digital humanities address perceptual and spatial color information only tangentially. We introduce VIAN, a visual film annotation system centered on the semantic aspects of film color analysis. The tool enables expert-assessed labeling, curation, visualization and classification of color features based on their perceived context and aesthetic quality. It is the first of its kind that incorporates foreground-background information made possible by modern deep learning segmentation methods. The proposed tool seamlessly integrates a multimedia data management system, so that films can undergo a full color-oriented analysis pipeline.

CCS Concepts

•***Human-centered computing*** → *Visualization systems and tools;* •***Applied computing*** → *Media arts;*

1. Introduction

Digital analysis and visualization techniques have become central in film and media studies. Since the emergence of computer-assisted film analysis, numerous design and exploratory systems have been proposed that focus e.g. on network analysis, average shot length [Tsi05], or framing. There is a growing interest among film scholars in extracting and analyzing color features in the context of film visualization based on digital methods. A number of those visualization systems are specifically tailored for color analysis and aesthetics in films, including those by Brodbeck (*Cinemetrics* [Bro11]) Ferguson [Fer16], Burghardt et al. [BKW16, BHE*17] or Olesen [OMG*16]. Such visualizations produce significantly more meaningful results when backed by a theoretical framework for visualization (see [Hef16, Stu16, Ole17] for an extended investigation).

Unfortunately, existing color-oriented visualization systems are usually based on features computed for a complete frame. This coarse granularity results in a loss of semantic information that limits its potential scientific value for various applications. In addition, the visualizations are often not matched to color perception. Finally, even if they provide enough information for a film scholar, they are usually not implemented flexibly enough to adapt to existing, well established workflows.

Such workflows typically include a manual segmentation stage that partitions a given film into temporal segments, then classifies them based on one or more sets of keywords. This task is usually performed in a film annotation software. Obviously, these tools vary greatly in their target audience, platform/implementation and supported functionalities. Typical use cases include

† Department of Film Studies, University of Zurich, Switzerland
‡ Department of Informatics, University of Zurich, Switzerland

textual annotation, speech analysis, and transcription or gesture analysis. While traditional tools were mainly desktop applications, browser-based solutions have recently become more popular [CLA, Lig, SAT]. While all these annotation tools were developed for film analysis, they do not feature color visualizations (see [Giu10, Giu14, MEHK*17, Flu17] for an overview).

After the segmentation step, the second stage in the typical film analysis research flow, classification, is usually based on *vocabularies*. Annotation tools may follow different approaches for this task. Many tools *hardcode* and attach classification labels to all annotation strings, which is arguably an error-prone task. More elaborate tools implement the concept of a so called *controlled vocabulary*, which restricts the terms that are allowed for each string, for example using a collection of checkboxes. While this is definitely sufficient in many applications, it does not scale well in practice. For instance, some segments may not find any matching term within the controlled vocabulary, or the GUI design may be overwhelmed by the sheer number of terms contained in large vocabularies.

Contributions

We present *VIAN* (VIsual ANnotator), a novel film annotation software that places emphasis on visual aspects of film style and its color aesthetics. We take a step towards the *qualitative data analysis* (QDA) paradigm, where classification is an integral part of design. Semantics between classification and their target are more explicit in VIAN than in any prior annotation tools.

VIAN has been developed in collaboration with an extensive film studies project that has qualitatively analyzed a corpus of approximately 400 films through all periods of film history. As part of this project's manual labeling stage, a total of over 17,000 temporal segments were classified using approximately 1,200 keywords organized in vocabularies that produced more than half a million summations. Our goal was to develop a comprehensive tool that can be used both to set up new projects and to enrich datasets that are available. By incorporating semantic segmentation, we improve the issue of semantic preservation of film features in visualizations and allow users to establish a link between their verbal classification and the visual content of a frame. As an evaluation, we conducted interviews with three current users of VIAN (see Sec. 6.2), who assessed the system usability for a range of tasks of interest. Specifically, VIAN's main contributions include:

- A system featuring integrated annotation and film exploration that allows for both temporal and spatial selection, several degrees of time granularity (film/segment/screenshot), colorimetric analysis, 2D/3D navigation among several possible color spaces, image plots, and an array of related user interface widgets.
- To the best of our knowledge, VIAN is the first such system that supports deep learning-driven segmentation for spatially-aware color analysis. VIAN goes beyond basic foreground-background segmentation and refines it further into key semantic and aesthetic elements of film analysis: male/female protagonists vs. supporting characters, lighting, mood and contrast, etc.
- We have packaged the proposed system into a comprehensive desktop application that currently supports a group of film analysis experts, and it is scheduled to be released as open-source software.

2. Related Work

In this section we review literature related to qualitative film analysis and media annotation in the field of digital humanities, then discuss existing visualizations for film material.

2.1. Qualitative Data Analysis and Media Annotation in DH

Qualitative data analysis is a fundamental framework for digital humanities (DH) scholars, and the landscape of software packages for this task is commensurately large. We refer the reader to Melgar et al. [MEHK*17] for a broad survey that includes an extensive comparison of over 50 existing tools related to film analysis and their purpose within DH, grouped by functionality type. Three of those important functionalities are partially covered in VIAN: professional video annotation, qualitative analysis, and automatic video analysis. Schoeffmann et al. [SHH15] investigated common and newly arising ways to interact with multimedia content. We largely follow the *Web Annotation Data Model* [YSC17]: each annotation consists of a target and an annotation content (the *body*). Since annotations often only refer to parts of a media resource (for example, time intervals or spatial regions), *selectors* are used to define the annotated region. Notably, the nature of selectors depends on the type of resource that is annotated.

2.2. Film Visualization

Several visualization methods for color distribution in films have been proposed, see [Stu16] and [Ole17] for a more derailed discussion. Some fundamental methods have been developed for the visualization of artworks and other media [Man12, Man15, RG17, KL17], Frederic Brodbeck arranged color schemes in circles to express palette changes over time, Kevin Ferguson used the so-called *z-projections* to compute a mean frame of the complete film by calculating the mean along the temporal axis for each pixel location and normalizing the final frame. Such approaches use elements from temporal data visualization [AMST11] and have been created using existing tools like ImageJ [SRE12] or ImagePlot [Man13, Hef16, OMG*16].

A well known type of visualization to summarize film material are *MovieBarcodes* [BKW18], which are produced by either computing the average color for each frame or reducing their dimension by one, and plot the result as a color barcode [BKW16, BHE*17] or columns of sorted pixels over time [Cut16]. While these movie barcodes are of good use for distant reading tasks, they expressiveness depends heavily on the film material's color composition and due to the averaging process often overemphasize brown tints. Casey and Williams [CW14] divided the input frames in a 4×4 grid and computed color histograms for each cell to compare temporal segments in films and visualized their Euclidian distance using a similarity matrix. Hohman et al. [HSSS17] developed an interactive timeline visualization to explore the connection between color scheme, extracted using a modified median cut algorithm, and dialogue representation in film. However, these methods are generally not detailed and accurate enough for the investigation of film color aesthetics and style. They rather provide an overview that complies with basic requirements for distant reading [Mor13].

Apart from solely relying on visualizing low-level color features,

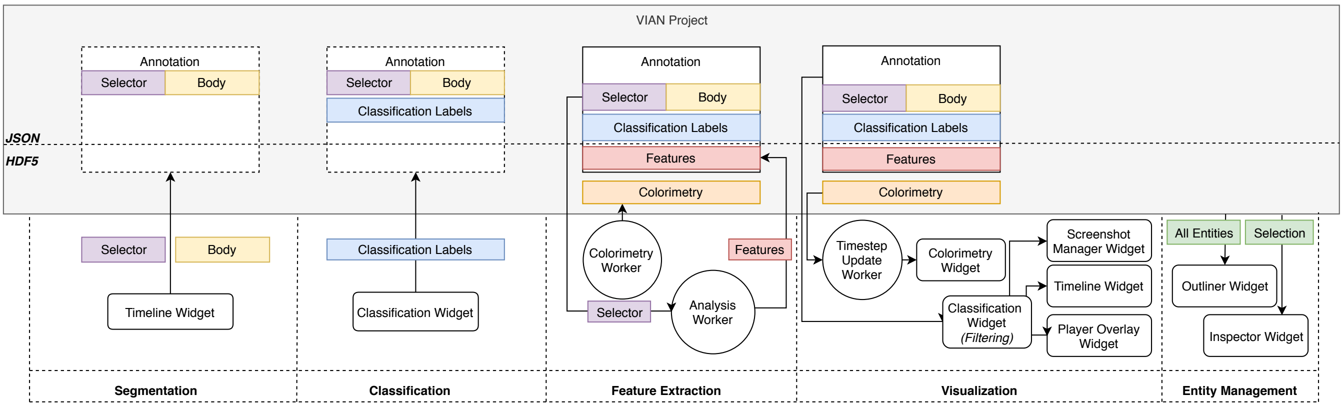


Figure 2: VIAN's overall system design and main functional components.

additional semantic features can be used for film visualization and exploration. Yang et al. [YFH*06] proposed a semantic-based image retrieval tool. Kurzhals et al. [KJH*16,JKKW17] analyzed film script and subtitle text to extract semantic tags, which where then displayed on a timeline.

Tab. 1 compares VIAN with prior film analysis systems in terms of seven functionalities.

	Speech segmentation	Interaction	Semantic annotation	Color feature extraction	Color visualization	Foreground detection	Corpus visualization
ACTION [CW14]				✓			
ELAN [SW08]	✓	✓	✓				
ANVIL [Kip14]		✓	✓				
Cinemetrics [Bro11]		✓		✓	✓		
MovieBarcodes [BKW16]		✓		✓	✓		
VIAN		✓	✓	✓	✓	✓	✓

Table 1: Scope and features supported by related film visualization and analysis tools, as compared to VIAN.

3. Design and Methodology

Next, we describe the main components and architecture of VIAN and justify the design choices made during the development process (see also Fig. 2).

3.1. Required Tasks

In close dialogue with fellow film scholars we identified five primary tasks that should be supported within VIAN:

T1 Creating annotations;
T2 Managing and modifying these annotations and their selectors;
T3 Classification of temporal selectors based on existing vocabularies;
T4 Visualization of selectors' film color features;
T5 Segmentation of frames into figure and background.

Since tasks T1 and T2 are commonly shared by film annotation software, we first considered to restrict VIAN to the visualization and classification tasks T3 to T5. However, existing tools do not implement a screenshot annotation and creation of temporal segments, and screenshots are usually performed in an iterative manner, hence the user would have to switch between different tools during the process. Therefore, we decided to integrate all five tasks in one single application.

VIAN has been developed through an iterative design process. After each cycle, a group of film scholars tested the tool and provided feedback for the further development. We first implemented T1 and T2 using existing tools as a guideline. This included embedding a media player, timeline, screenshot manager and an overlay widget of the player to draw SVG graphics. In the second step, we focused on feature extraction and corresponding visualization (T4), explained in-depth in Sec. 5. Subsequently, classification and vocabularies were implemented together along with their corresponding widgets (T3). Last, we integrated semantic segmentation into VIAN and modified the visualizations accordingly (T5).

3.2. Functional Components

As outlined in Fig. 2, VIAN's software framework and interface is structured around the following main components.

The *VIAN Project* is the root data structure VIAN operates on. It consists of a film file, a human readable JSON file describing all entities within the project and a corresponding HDF5 file structure that is used to store numeric data. Since data in digital humanities is often stored long-term within archives, a traditional requirement is to store the annotations in a human-readable file-format for better interoperability with other annotation software. However, since VIAN computes and operates on large feature vectors, memory management and runtime constraints make it necessary to store numeric arrays in a more efficient data structure. Such data is thus only referenced from the more expressive JSON file but actually stored compactly in an HDF5 file structure.

The *TimestepUpdate* worker listens for any event indicating that the current timestamp of the timeline changed, either because the user is scrubbing through the video or because the film is being played, and performs subsequent conditional tasks. Currently these include (i) updating visualizations in the colorimetry and overlay widget during play mode, plus additional tasks such as reading the exact frame from the film source and (ii) updating spatial visualizations that are executed during scrubbing.

The *AnalysisWorkerPool* provides a worker crew for the execution of any task that can or should be executed in an asynchronous manner. This includes the extraction of color features, computation of clusterings as well as the automatic creation of selectors. An important general design criterion is to keep the implementation of new analyses and visualizations for scholars as simple as possible. Hence, they are all derived from one base class; we store the resulting data in an HDF5 file structure. Displaying the visualizations is handled by VIAN internally.

The *Timeline* widget is used to edit segments or SVG annotations. It allows the user to tweak the temporal boundaries of a selector and edit their textual annotation body. An additional merge-and-cut tool is implemented for respective modifications to temporal segments. VIAN belongs to the category of *tier-based* annotation software. That is, annotations are grouped in so-called tiers. The timeline visualizes these as horizontal strips where all annotations belonging to the corresponding tier are placed.

The *Screenshot Manager* consists of a set of different visualizations for screenshot selectors. Primarily, the screenshots are shown in a by-segment sorted fashion, focusing on the ones in the current segment given by the timeline. Additionally, if color analyses have been performed for the screenshots, these are displayed in three additional visualization panes.

The *Player* widget is an embedded VLC player instance, implemented as a transparent *Overlay* on top of the player. This allows SVG annotation and visualization of spatial features such as texture complexity.

The *Classification* widget displays a list of vocabulary keywords as checkboxes, sorted by classification object and user assignable vocabulary categories. It serves two purposes: in *classification mode* the user can activate certain keywords to attach them to the selected annotation. In *query mode*, it can be used to query project for all annotations tagged with the specific keywords. Classification in VIAN and vocabularies are explained in depth in Sec. 4.4.

The *Colorimetry* widget displays color features of the current frame while navigating through a film or watching it. The widget can convey the color scheme (as a palette or as a scatter plot), the color histogram, and the average spatial complexity.

The *Outliner* widget is a tree view representing all entities of the loaded VIAN project. It can be used to modify annotations and annotation groups (tiers), as well as perform selection and operation on multiple entities. While it has been one of the initial widgets implemented into VIAN, we have found that novice users may be overwhelmed by the sheer number of items contained in the outliner, especially in larger projects with hundreds of annotations. It is thus not included in the default layout, and most of the operations can be accessed via more illustrative widgets like the timeline. Nevertheless, it is shown for appropriate tasks such as finding a specific analysis or bulk modifying a list of VIAN's entities.

Last, the *Inspector* widget is used to view and edit a selected entity of a VIAN project. Similar to the outliner, the inspector was the primary widget for modifying entities during early development. However, many functionalities can be more efficiently operated through other more accessible widgets.

3.3. Implementation Details

VIAN is implemented in Python and exploits the PyQt GUI framework. We use VLC as an embedded media player, OpenCV for image manipulation, and Keras with a TensorFlow backend for semantic segmentation. Some further libraries such as NumPy and Scikit-learn are also used for data analysis and feature manipulation. A complete list of libraries is accessible in the appendix. Data generated in VIAN is stored in two files: annotations and entity-related information are stored in JSON files, whereas numeric features are stored in HDF5 file containers.

4. Segmentation and Annotation

As mentioned in the introduction, film scholars temporally segment films according to their research questions when aiming at a narrative or aesthetic analysis of films. The intention is to form portions of smaller and more homogeneous data. The collected analytical units are then enriched with annotations referred to as *coding* in DH [MEK18].

An annotation is a composition of a *selector* and a *body*. The selectors purpose is to define the region within a media resource to which the annotation is referring to. In film annotation software, typical selectors define a temporal segment or a single timestamp. The body can be anything related to and typically *about* the content of the media resource within the region defined by the selector [YSC17].

Unlike other tools, VIAN makes a clear distinction between two types of coding: Open natural language and vocabulary based annotation. The former is done by simply writing the annotation directly into the annotation body or its notes field. The latter is performed in VIAN's classification perspective. In this section we will assess the three different type of annotations available in VIAN, classification itself is explained later in Sec. 4.4.

4.1. Temporal Segments

The temporal segment annotations consist of a selector that defines a timestamp or range and a textual body. Such annotations could mark individual shots or sequences of shots sharing narrative units such as events, consistent location or time [CBC12]. In speech oriented tools, they are often used for transcription tasks. Our collaborators focused on segmentation based on coherent color schemes.

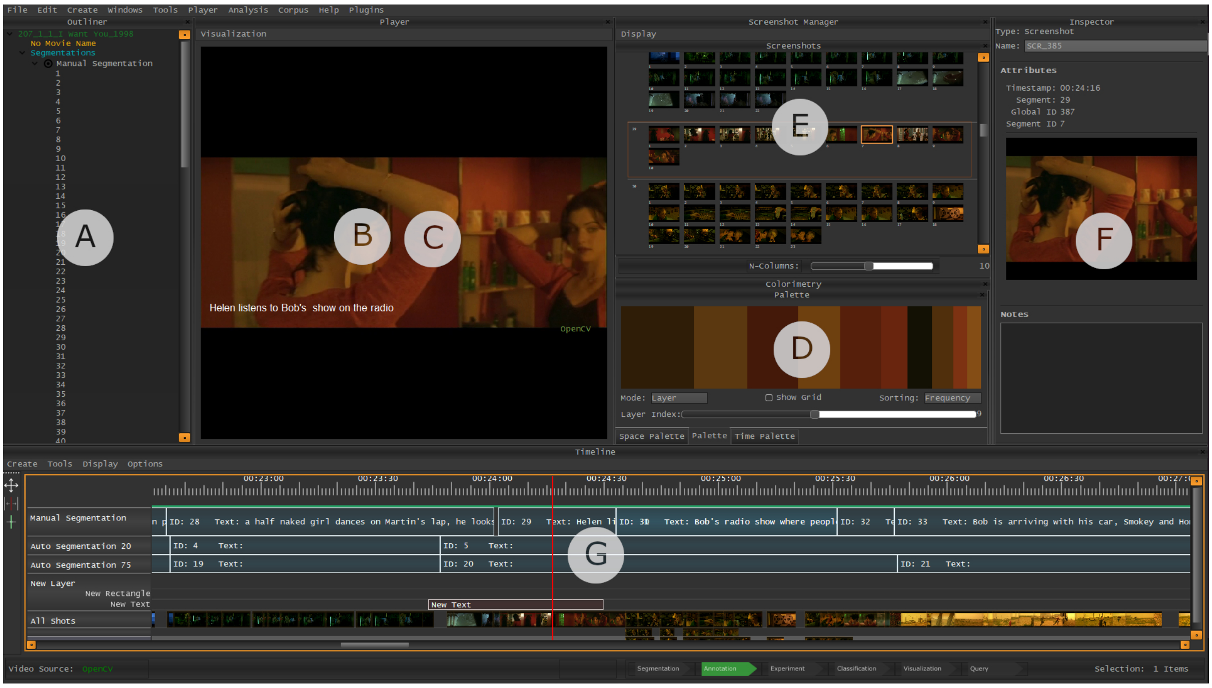


Figure 3: VIAN's most important widgets. The upper left showcases the outliner [A], player [B], and transparent overlay [C] widgets. The colorimetry widget [D] shows the current color palette. The screenshot manager [E] on top displays the screenshots sorted by the first segmentation. The inspector widget [F] gives details on the currently selected entity. Below is the timeline widget [G] showing two tiers of temporal segments and two tiers of screenshots (I WANT YOU, UK 1998, Michael Winterbottom).

Manual creation of selectors is a time consuming process. However, as stated before, color analysis in film studies usually implies the creation of temporal segments with a coherent color scheme or narrative coherence. Fortunately, this process can be automated. VIAN uses an agglomerative clustering of evenly spaced histograms. Once the tree is computed, the result is visualized using a selection of evenly spaced screenshots together with a slider for selecting the merge iteration. Finally, the created segments can be fine-tuned using the *move*, *cut* and *merge* tools of the timeline. Fig. 4 compares a human made annotation with three auto-segmentations using different numbers of final clusters.

4.2. Screenshots

Screenshots are an integral part of visual assessment of films. By capturing a number of the most significant shots in a scene, the user generates a portfolio for individualized analysis of the film in connection to his or her research questions. Camera distance, depth-of-field, lighting, image composition or the representation of specific color schemes are typical dimensions that are best represented by screenshot piking. Beyond exemplary purposes, screenshots can be used as main data source for further qualitative studies. Because of their crucial role, VIAN implements a screenshot annotation with a selector pointing to a specific timestamp within a film. In addition to positioning in the timeline, screenshots are also visualized in the *screenshot manager* widget and organized in bins corresponding to the primary temporal segmentation as seen in Fig. 3. When the user watches or scrubs through a film, the screenshot manager automatically follows the player and displays the current temporal segment in the screenshot bin.

4.3. SVG Annotations

Temporal segments on their own do not provide information on the spatial location of target objects or certain areas within a frame. This problem becomes especially pressing in the assessment of spatial visual features of films, where numerical or learning methods should ideally be applied only to a region of interest (ROI) within the frame (e.g. the protagonist). For these cases, VIAN supports SVG annotations, where the selector defines both temporal as well as spatial regions. The SVG annotation body can include basic geometric shapes, text, images, and free-hand annotations. Additionally, the content of the text annotation can be driven by a wide variety of attributes from other entities within the project, allowing the user to visualize important information, such as the text body of the current temporal segment, the film timestamp or similar directly on screen while analyzing the film.

4.4. Classification

Annotations can be either classified on the fly as they are created, or globally at the end. In the latter case, the annotations to classify are successively displayed to the user with checkboxes for each keyword. Additionally, the player and screenshot manager always moves to the current annotation's selector in order to improve productivity during this process. Vocabularies in VIAN are hierarchically structured words. Each vocabulary belongs to a category, both

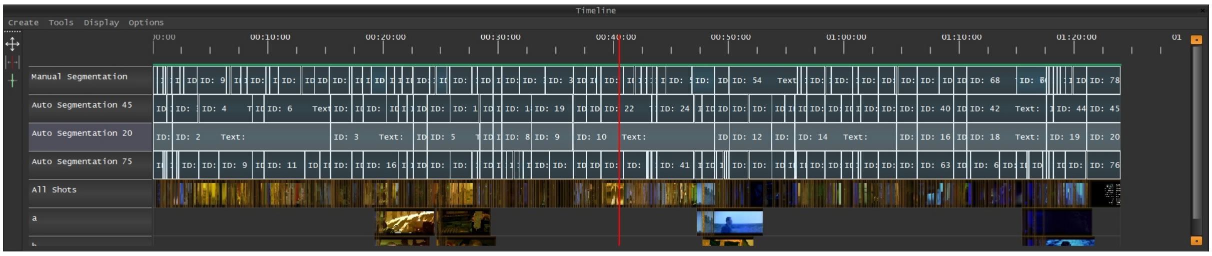


Figure 4: Comparative display between a manual segmentation (top row) and three automatic hierarchical segmentations (bottom rows) at three levels of aggregation in the timeline widget. (I WANT YOU, UK 1998, Michael Winterbottom)

each vocabulary and word may have an additional URL to a description and a visual representation of the concept.

Often there is a need for tagging different concepts within a video with the same vocabulary. For instance, specific types of textures to characters, objects or general locations/environments. In most tools such differentiation between the target and the vocabulary is done implicitly by the annotator by duplicating a group of annotations and naming it with its target concept. VIAN allows the user to define these targets as so called *classification objects* and attach vocabularies to them. For each classification object a set of tags is generated for all vocabulary words that are attached to them. These tags can than be added to the annotations described in the previous sections using the classification widget shown in Fig. 5. This yields several benefits over the traditional approach: first, there is no duplication of vocabularies (a vocabulary for textures is unique within a session in VIAN); second, the concept is defined explicitly, allowing also other functions of VIAN to access and operate on them (Sec. 5.1).

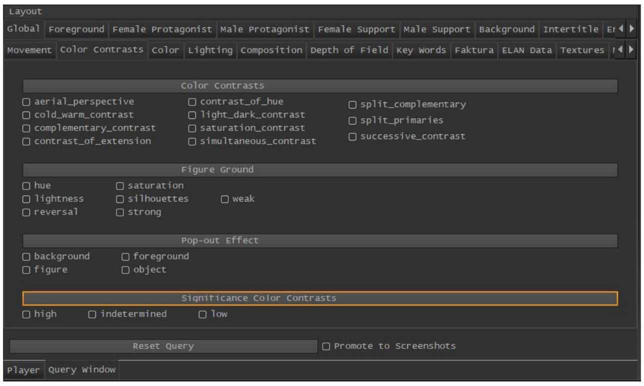


Figure 5: VIAN's classification widget. The upper tabs represent the classification object, the lower correspond to the vocabulary categories. For each vocabulary a set of checkboxes is shown in an expandable area.

5. Color Analysis

There are two main use cases for color analysis in VIAN: (a) using a selector-independent dataset of color features, which is computed in a user-defined resolution for the complete film (*film colorimetry*), and (b) using selector-dependent color features that are computed for a given selector. The visualizations for these two datasets are the same, but the film colorimetry is used to bring live-feedback to the user when watching or analyzing the film. All color-related computation is performed using the CIE-L*a*b* color space (*LAB*, for short) to enforce perceptually uniform results as much as possible. This section details both visualization types.

5.1. Deep Semantic Segmentation

As shown in Sec. 4.4, VIAN's classification objects allow the user to define conceptual targets which are then classified by vocabularies. Often, these classification object are represented by a subset of pixels in a given frame and the user is interested in the color features of those, rather than the complete frame. This may, for example, be to compare the color schemes of the characters and background within a temporal segment.

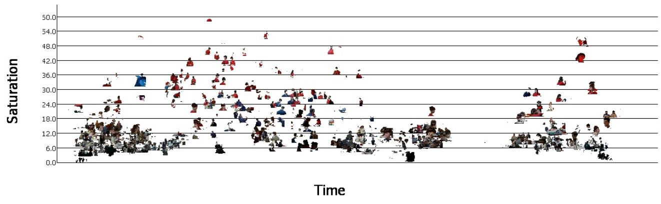
VIAN uses a semantic segmentation approach to solve this problem, namely the PSPNet [ZSQ*16]. Each classification object can be assigned one or more labels to connect them to their pixel representation within a frame, it is thus possible to compute all described color analyses for a specific classification object. In the outliner widget, the user can activate which features he currently wants to have displayed. Fig. 6 compares the saturation over time for two classification objects: *background* and *figure*.

Hence using this method, all subsequently described color feature vectors can be targeted towards the complete frame or a specific classification object.

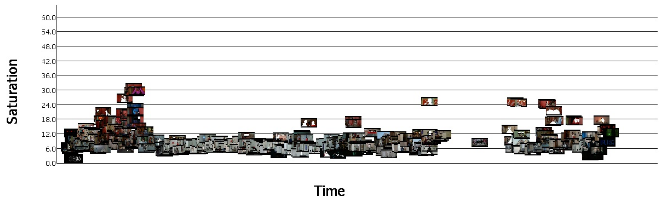
5.2. Average Values

The most basic feature vector that can be computed from a given selector in VIAN consists of the mean color in LAB or RGB as well as the LAB saturation [Lue10]. While such average values have limited informative significance, they are useful for summarizing visualizations. However, their discriminative power improves drastically if used together with semantic segmentation, allowing the user to compare color distribution or color evolution during a film for certain objects corresponding to the frame as a whole, or foreground vs. background respectively. We have implemented three visualizations to present such average values: features such as hue, saturation and luminance can be displayed with respect to time as e.g. shown in Fig. 6.

Another option is to place the selectors in the LAB color space,



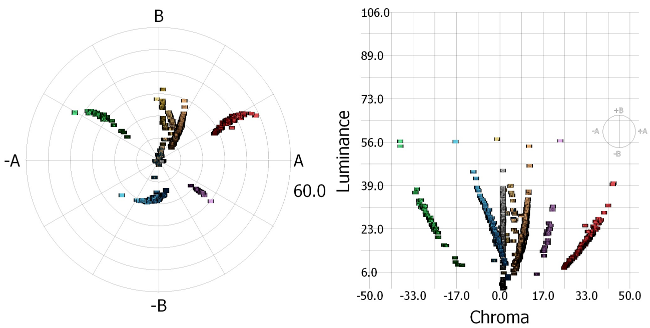
(a) Foreground (characters)



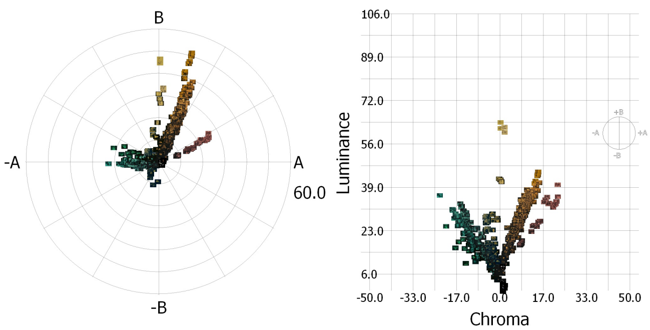
(b) Background

Figure 6: Saturation of screenshots extracted over time for figure and background (UNE FEMME EST UNE FEMME, France, 1961, Jean-Luc Godard).

for this we are using two views, one showing the a*b*-plane (top-down) and a second showing the luminance-chroma plane, where a slider can be used to rotate the view angle, which is visualized as a compass glyph as shown in Fig. 7.



(a) (L'INHUMAINE, France, 1923, Marcel L'Herbier)



(b) (DAS CABINET DES DR. CALIGARI, Germany, 1919, Robert Wiene)

Figure 7: Average screenshot color in the LAB color space.

5.3. Color Histograms

While average values of certain parts of a frame may yield insight into a film's general color scheme, they are usually not informative enough for more advanced processing such as finding matching selectors. Color histograms, on the other hand, are very suitable for such purposes as they can act as feature vectors [CW14]. However, color spaces are three-dimensional and so are color histograms. This makes their visualization challenging, especially if the user wants to compare two or more color histograms to each other. One way to ameliorate this, as done in VIAN, is to follow a 3D space-filling curve (the Hilbert curve, for instance) that maps the 3D color space histogram into a 1D feature vector. Another problem arises from the fact that the RGB gamut is only a non-axis-aligned subset of the LAB color space, and thus many LAB bins will always contain zero values.

We give the user the freedom to choose one of three histogram visualization modes: (i) plotting all bins, (ii) show only bins that lie within the RGB gamut, or (iii) show only non-zero values as shown in Fig. 8. We also observed that applying a logarithmic scale improves the perceptual consistency with the color content of a frame.

Figure 8: Hilbert-sorted color histogram visualization (LES PARAPLUIES DE CHERBOURG, France, 1964, Jacques Demy).

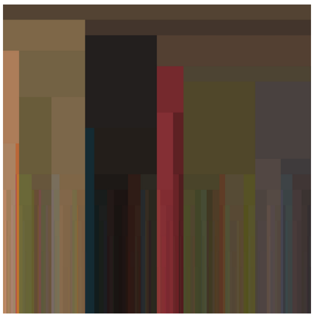
5.4. Color Palettes

Color palettes are an effective way of conveying a frame's color content. We compute them via agglomerative bottom-up clustering, which does not depend on random cluster initialization and is far more granular than e.g. k-means. Since computing a distance matrix for the complete set of frame pixels in a film is computationally infeasible, VIAN's color schemes are computed in two steps. First, the frame is simplified by creating *SEEDS superpixels* [BBRVG13]. We compute the average color of each superpixel, and use those averages as input elements for a subsequent agglomerative clustering. The resulting tree can then be visualized in VIAN either, whereby the merges in the tree and thus the resulting number of clusters can be tuned by the user using a slider. The superpixels and final clustering is shown in Fig. 9.

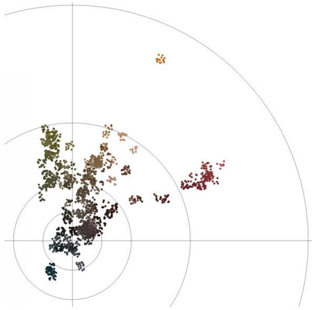
While this type of color scheme visualization is well established in the field and intuitive to read, comparing different color schemes can be simplified by plotting the clustering into the A-B plane of the LAB color space as shown in Fig. 9 d). We have found that using a jitter effect and plotting the number of dots according to the size of the color cluster within the palette allows to compare different color schemes better than the linear visualization. Palettes can also be used to visualize a film's color content over time as shown in Fig. 10.

(a) Input image

(b) SEEDS superpixels



(c) Visualization of the bottom-up hierarchical clustering of a palette.



(d) Palette color clusters in the AB plane of the LAB color space.

Figure 9: Stages in the generation of the color palette. (JIGOKUMON, Japan, 1953, Teinosuke Kinugasa).

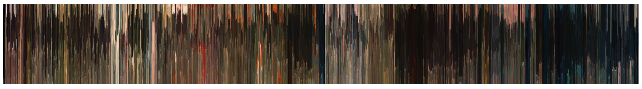


Figure 10: Barcode visualization of color palette changes over time (horizontal axis) (JIGOKUMON, Japan, 1953, Teinosuke Kinugasa).

5.5. Colorimetry

Apart from selector dependent analyses, the features can also be computed in a fixed resolution for the complete film. The reason for this is twofold: first, the user often wants to have direct visualization of the current frame; second, evenly spaced feature can be used to create selectors automatically as described in the next section.

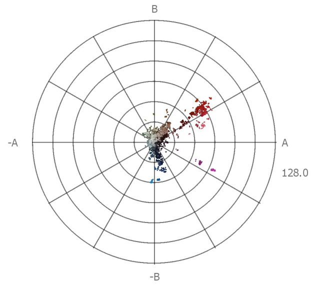
6. Results

6.1. A Use Case in Film Studies

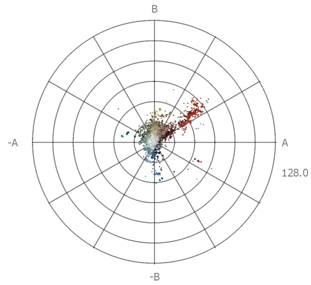
As we argue in the following use case, the visualizations offered in VIAN enable a number of crucial insights into films' color aesthetics.

A team of film scholars analyzed a large group of 414 films to investigate the relationship between film color technology and aesthetics. These films were produced in the first 100 years of film history, from 1895 to 1995. Based on a detailed verbal annotation, the temporal segmentation of the films and the acquisition of several hundred screenshots per film, the visualization methods aimed at three different goals:

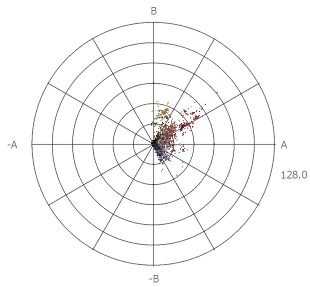
1. Represent visual impressions true to human perception;



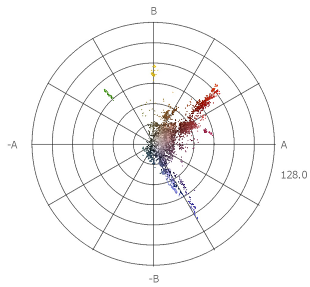
(a) Foreground (characters)



(b) Background



(c) Foreground (characters)



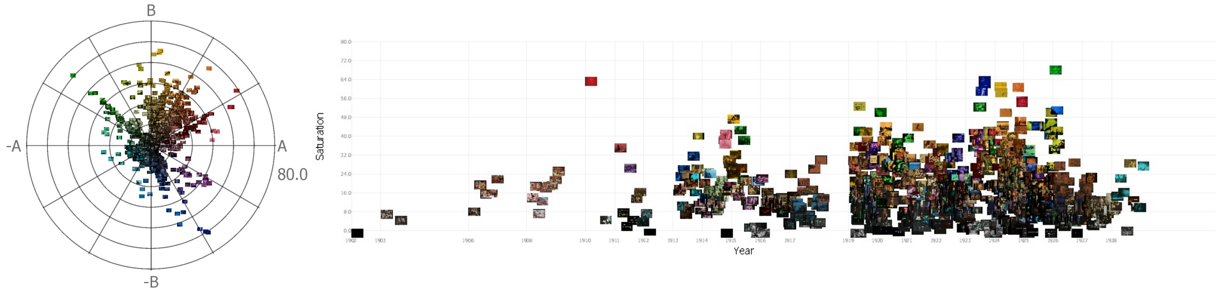
(d) Background

Figure 11: Palette plot comparison of UNE FEMME EST UNE FEMME, France, 1961, Jean-Luc Godard (a) and (b) and WEST SIDE STORY, USA, 1961, Jerome Robbins and Robert Wise (c) and (d). $n = 800$

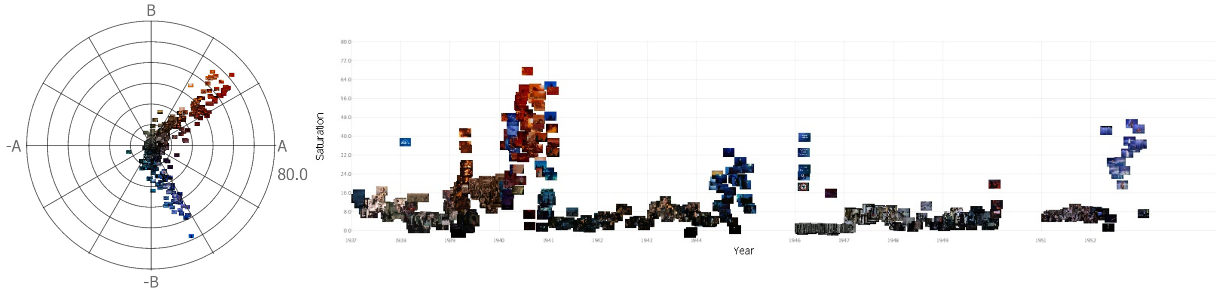
2. Represent subtle aesthetic nuances in figure and ground separately;
3. Visualize the films at the *micro* (screenshot, temporal segment), *meso* (individual film) and *macro* (corpus) levels.

Foreground/Background Saturation Fig. 6 shows a representation of the development of saturation levels over time for figure and ground separately. The film chosen for this type of visualization is typical of the French New Wave's sober style with non-saturated, mostly white backgrounds and color attribution to characters mostly in the primary colors red and blue as an ironic reference to the French tri-colore in red, blue and white. The strong figure-ground separation is connected to a style established in Pop Art. Also visible in these two plots are two colorful scenes in a music club where the female protagonist performs a chanson as a pastiche of musical numbers in Classical Hollywood films. This type of visualization, that we name *Color_dT*, gives an instant representation of a film's style with reference to its narrative unfolding that is much more nuanced and detailed than traditional MovieBarcodes.

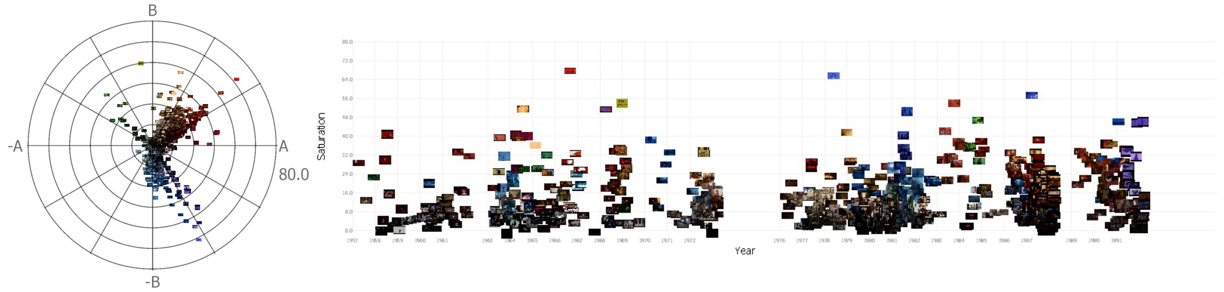
Screenshot Plane Plots CIE LAB plane plots show the overall color scheme of individual films' frames on the meso level; see for example Fig. 7 for two early films with applied colors in tinting and toning. Such plots provide an instant visualization of the films' color distribution in a perceptually uniform color space. By comparing films produced with the same or different color



(a) 1895 - 1930



(b) 1930 - 1955



(c) 1955 - 1995

Figure 12: Corpus visualization: VIAN can display large collections of films according to chroma, saturation, year of production, labels, etc. Here we compare and analyze aesthetic trends from 414 films produced over three historical periods.

processes, the scholars acquire an understanding of the influence of the technology on a film's palette. For instance the two films L'INHUMAINE (Fig. 7a) and DAS CABINET DES DR. CALIGARI (Fig. 7b) display the typical monochrome color schemes of these early technologies. But in the plot of L'INHUMAINE a problem in the digital reconstruction of these colors becomes instantly visible. In tinted films the saturation levels are supposed to evolve continuously from center to periphery (Fig. 7b) while the non-continuous saturation levels in Fig. 7a) display the artificial, non-authentic digital colorization applied to the scan of this film.

This type of visualization is also a useful tool to compare different versions of the same film, for instance various analog film prints with digitized versions, which is one of the most crucial tasks of research on film color, related to the assessment of source materials.

Palette Dots Dot plots (Fig. 11) show color distributions at the pixel level and thus eliminate distortions caused by averaging in image-based plots. In the comparison of the French film UNE FEMME EST UNE FEMME with the American musical WEST SIDE STORY the figure-ground inversion of the musical located in New York City becomes evident. The visualization shows that backgrounds are more saturated, often painted in red or scenes illuminated by red lights as opposed to the white achromatic backgrounds in the French film, already discussed with regard to Fig. 6. Costumes in WEST SIDE STORY are mostly reduced in saturation, with pastel colors dominating. The highly saturated backgrounds are important elements of the choreography that moves characters in relationship with changing environments to create expressive moods in line with characters' emotions.

Corpus Visualizations On the macro level, corpus visualizations (which are readily supported in VIAN) display diachronic historical developments connected to varying technologies, production practices, notions of taste, and aesthetic norms related to the broader cultural context. Fig. 12 showcases two types of corpus visualizations for monochrome color schemes identified in the verbal annotation of the film analyses for three periods:

- 1895–1930: early film technology;
- 1930–1955: standardization of film color production in Technicolor and early Agfacolor;
- 1955–1995: chromogenic film stocks.

The AB plots shown on the left hand side provide an overview of the color distribution. It becomes instantly evident that in early film we can observe a large and seemingly random distribution of hues in almost all sections of the LAB color space. This distribution is connected to the many dyes applied in early color system's tinting and toning that produced mostly monochrome color schemes. The results differ significantly in the Technicolor era with its heavily standardized production system, controlled by the company's own Color Advisory Service. The dominant hues in monochrome color schemes are now confined to the red and blue spectrum. Blue is applied mostly in night scenes and one water scene in MILLION DOLLAR MERMAID, whereas red is used in fire scenes, in sunsets, or in dream sequences as a visualization of dream sequences in VIAN has shown.

The second type of corpus visualization, we call it *Color_dY*, is shown on the right hand side. It provides a more detailed picture of the historical development by plotting the screenshots' saturation levels vs. year of production. In a) we notice increasing saturation levels starting in the 1920s when applied colors reached their peak and early two-color processes were used increasingly. Saturation levels are generally rather low in Technicolor / Agfacolor films in as shown in b), except for the animation film FANTASIA, produced by Disney in 1940. In the last period investigated, shown in Fig. 12c), color distribution becomes more varied in monochrome scenes and saturation levels increase due to the rising use of colored light by certain cinematographers and in certain genres, for instance Italian horror films (*gialli*). Many science fiction films are dominated by monochrome color schemes in the blue range, while red is increasingly associated to certain milieus (for instance, night clubs) or expressive styles of interior design.

6.2. User Feedback

We conducted interviews with three active researchers that have adopted VIAN to analyze films: one participant (P1) was primarily concerned with segmentation and screenshot acquisition, while the other two (P2 and P3) were additionally interested in the interactive visualization of the annotated film database.

In summary, all participants found VIAN to significantly improve their existing film color analysis pipelines. They reported an increased ability to engage in productive creation and management of annotations: they are able to scrub smoothly through their films and zoom in and out of timelines or color space visualizations as desired, which makes the task of screenshot selection and labeling truly an interactive experience. Participant P2's research involves comparing digitizations of several physical copies of a single film. She used VIAN to compare the copies' screenshot ensembles in the perceptually-uniform LAB color space and remarked the following:

> P2: *"I found the color space plots to be very valuable for my research, because I can easily compare color abundance varies between different copies of the same film"*

Participant P2 noted the following:

> P3: *"The main advantage of these visualizations is that I can see patterns in movies that I would not discover without them. Because the figure is separated from the background, I can answer questions about how a figure is staged in its [scenic] context."*

Participant P3 noted that VIAN initially has a somewhat steeper curve because of the sheer number of available tools. P2 and P3 had to become acquainted with the proposed feature extraction methods and their parameters, e.g. regarding how to operate and interact with the agglomerative clustering tool.

7. Conclusions

In this paper we have introduced VIAN, a novel comprehensive annotation and visualization system that supports interactive annotation, spatiotemporal selection, color analysis and classification of film material by large vocabularies. To this end, we categorized a set of meaningful color features and appropriate visualizations to convey color information and aesthetics contained in film, optionally distinguishing between key visual elements such as background, protagonists or supporting characters, etc. Thanks to this segmentation, VIAN reduces the loss of semantic information during the extraction of features and helps scholars interpret qualitative analysis in their research activities. The software was designed to be extendable and easy to operate with, and it can also be used without GUI as a Python API.

VIAN received a positive assessment during evaluation interviews: users reported increased flexibility and better-informed annotation power in their film analysis pipelines. Compared to previous solutions that they were using, they consider the visualization widgets and screenshot annotator to be the most impactful improvements. All in all, we believe VIAN to be a valuable and up-to-date addition to the existing landscape of film annotation software.

Limitations and Future Work

Some GUI elements have a significant learning curve: for example, setting up a classification experiment in VIAN requires the user to create or import vocabularies, define conceptual entities, and connect them to the semantic segmentation. This demands a certain level of expertise operating VIAN's multiple widgets and their semantics, although this is also highly dependent on the intrinsic task complexity.

VIAN will be further developed in specific areas, and we aim to extend its toolset in forthcoming releases. Our topmost priorities include character recognition for a more fine-grained feature analysis, neural network-driven texture assessment, and improved automatic creation of selectors.

References

[AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of Time-Oriented Data*. Springer Publishing Company, Inc., 2011. 2

[BBRVG13] BERGH M. V. D., BOIX X., ROIG G., VAN GOOL L.: SEEDS: Superpixels extracted via energy-driven sampling. *arXiv:1309.3848 [cs]* (Sept. 2013). 7

[BHE*17] BURGHARDT M., HAFNER K., EDEL L., KENAAN S.-L., WOLFF C.: An information system for the analysis of color distributions in moviebarcodes. In *Everything changes, everything stays the same? Understanding information spaces : Proc. Int. Symp. of Information Science*, Gäde M., (Ed.), vol. 70. Verlag Werner Hülsbusch, Glückstadt, 2017, pp. 356–358. 1, 2

[BKW16] BURGHARDT M., KAO M., WOLFF C.: Beyond shot lengths. using language data and color information as additional parameters for quantitative movie analysis. In *Digital Humanities Conference Abstracts* (Kraków, 2016), Eder M., Rybicki J., (Eds.), Jagiellonian University & Pedagogical University, pp. 753–755. 1, 2, 3

[BKW18] BURGHARDT M., KAO M., WALKOWSKI N.-O.: Scalable MovieBarcodes - an exploratory interface for the analysis of movies. In *IEEE VIS Workshop on Visualization for the Digital Humanities* (2018). 2

[Bro11] BRODBECK F.: Cinemetrics – film data visualization. http://cinemetrics.fredericbrodbeck.de/, 2011. 1, 3

[CBC12] CUTTING J. E., BRUNICK K. L., CANDAN A.: Perceiving event dynamics and parsing hollywood films. *Journal of Experimental Psychology Advance online publication* (2012). doi:10.1037/a002773. 4

[CLA] CLARIAH Media Suite [online]. URL: http://mediasuite.clariah.nl/. 2

[Cut16] CUTTING J. E.:. Perception, attention, and the structure of Hollywood film [online]. 2016. URL: http://people.psych.cornell.edu/~jec7/curresearch.htm. 2

[CW14] CASEY M., WILLIAMS M.: *ACTION (Audio-visual Cinematic Toolbox for Interaction, Organization, and Navigation): An Open-source Python Platform*. White paper HD5139411, Dartmouth College, 2014. 2, 3, 7

[Fer16] FERGUSON K. L.: The slices of cinema. digital surrealism as research strategy. In *The Arclight Guidebook to Media History and the Digital Humanities*, Acland C. R., Hoyt E., (Eds.). Reframe Books, 2016, pp. 270–299. 1

[Flu17] FLUECKIGER B.: A digital humanities approach to film colors. *The Moving Image 17*, 2 (2017), 71–94. 2

[Giu10] GIUNTI L.: *Problemi Dell'analisi Del Testo Di Finzione Audiovisivo. Verifica E Sviluppo Di Un Modello Analitico E Interpretativo Con Strumenti Digitali*. PhD thesis, Università degli Studi di Pisa, 2010. 2

[Giu14] GIUNTI L.: L'analyse du film a l'ère numérique. Annotation, geste analytique et lecture active. *Cinéma & Cie 14*, 22/23 (2014), 127–143. 2

[Hef16] HEFTBERGER A.: *Kollision Der Kader: Dziga Vertovs Filme, die Visualisierung ihrer Strukturen und die Digital Humanities*. No. 2 in Film-Erbe. Edition Text + Kritik, München, 2016. 1, 2

[HSSS17] HOHMAN F., SONI S., STEWART I., STASKO J.: A viz of ice and fire: Exploring entertainment video using color and dialogue. In *IEEE VIS Workshop on Visualization for the Digital Humanities* (2017). 2

[JKKW17] JOHN M., KURZHALS K., KOCH S., WEISKOPF D.: A visual analytics approach for semantic multi-video annotation. In *IEEE VIS Workshop on Visualization for the Digital Humanities* (2017). 3

[Kip14] KIPP M.: ANVIL: A universal video research tool. In *Handbook of Corpus Phonology*. Oxford University Press, 2014, ch. 21, pp. 420–436. 3

[KJH*16] KURZHALS K., JOHN M., HEIMERL F., KUZNECOV P., WEISKOPF D.: Visual movie analytics. *IEEE Transactions on Multimedia 18*, 11 (Nov 2016), 2149–2160. 3

[KL17] KING L. M., LEONARD P. S.: Processing pixels. towards visual culture computation. 2

[Lig] Lignes de temps IRI [online]. URL: http://www.iri.centrepompidou.fr/outils/lignes-de-temps/. 2

[Lue10] LUEBBE E.: *Colours in the Mind. Colour Systems in Reality. A Formula for Colour Saturation*. Books on Demand, 2010. 6

[Man12] MANOVICH L.: How to compare one million images? In *Understanding Digital Humanities*, Berry D., (Ed.). Palgrave Macmillan UK, London, 2012, pp. 249–278. 2

[Man13] MANOVICH L.: Kino-Eye in Reverse. Visualizing Cinema. In *Cinematicity in Media History*, Geiger J., Littau K., (Eds.). Edinburgh University Press, Edinburgh, 2013, pp. 211–234. 2

[Man15] MANOVICH L.: Data Science and Digital Art History. *International Journal for Digital Art History 1* (2015). 2

[MEHK*17] MELGAR ESTRADA L., HIELSCHER E., KOOLEN M., OLESEN C. G., NOORDEGRAAF J., BLOM J.: Film analysis as annotation: Exploring current tools. *The Moving Image: The Journal of the Association of Moving Image Archivists 17*, 2 (2017), 40–70. doi:10.5749/movingimage.17.2.0040. 2

[MEK18] MELGAR ESTRADA L., KOOLEN M.: Audiovisual media annotation using qualitative data analysis software: A comparative analysis. *The Qualitative Report 23*, 13 (Mar. 2018), 40–60. 4

[Mor13] MORETTI F.: *Distant Reading*. Verso Books, June 2013. 2

[Ole17] OLESEN C. G.: *Film History in the Making. Film Historiography, Digitised Archives and Digital Research Dispositifs*. PhD Thesis, University of Amsterdam, 2017. 1, 2

[OMG*16] OLESEN C. G., MASSON E., GORP J. V., FOSSATI G., NOORDEGRAAF J.: Data-driven research for film history: Exploring the Jean Desmet collection. *The Moving Image 16*, 1 (2016), 82–105. 1, 2

[RG17] REYES-GARCIA E.: *The Image-interface. Graphical Supports for Visual Information*, 1 edition ed. Wiley-ISTE, Hoboken, NJ, Nov. 2017. 2

[SAT] The semantic annotation tool [online]. URL: https://sites.dartmouth.edu/mediaecology/. 2

[SHH15] SCHOEFFMANN K., HUDELIST M. A., HUBER J.: Video interaction tools: A survey of recent work. *ACM Computing Surveys 48*, 1 (Sept. 2015), 14:1–14:34. 2

[SRE12] SCHNEIDER C. A., RASBAND W. S., ELICEIRI K. W.: NIH image to ImageJ: 25 years of image analysis. *Nature Methods 9* (June 2012), 671. 2

[Stu16] STUTZ O. K.: *Algorithmische Farbfilmästhetik. Historische Sowie Experimentell-digitale Notations- Und Visualisierungssysteme Des Farbfilms Im Zeichen Der Digital Humanities 2.0 Und 3.0*. Master's thesis, Universität Zürich, 2016. 1, 2

[SW08] SLOETJES H., WITTENBURG P.: Annotation by category: ELAN and ISO DCR. In *Proceedings International Conference on Language Resources and Evaluation* (May 2008). 3

[Tsi05] TSIVIAN Y.: Cinemetrics Database. http://www.cinemetrics.lv/database.php, 2005. 1

[YFH*06] YANG J., FAN J., HUBBALL D., GAO Y., LUO H., RIBARSKY W., WARD M.: Semantic image browser: Bridging information visualization with automated intelligent image analysis. In *2006 IEEE Symposium On Visual Analytics Science And Technology* (Oct 2006), pp. 191–198. 3

[YSC17] YOUNG B., SANDERSON R., CICCARESE P.: *Web Annotation Data Model*. W3c Recommendation, W3C, Feb. 2017. 2, 4

[ZSQ*16] ZHAO H., SHI J., QI X., WANG X., JIA J.: Pyramid Scene Parsing Network. *arXiv:1612.01105 [cs]* (Dec. 2016). arXiv: 1612.01105. 6