# MASTERARBEIT | MASTER'S THESIS

Titel | Title

AI in the Archive: A Computational Film Analysis of Jim Hubbard's AIDS Activist Videos

verfasst von | submitted by

Manuel Meschiari BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2025

| | |
|---|---|
| Studienkennzahl lt. Studienblatt \| Degree programme code as it appears on the student record sheet: | UA 066 647 |
| Studienrichtung lt. Studienblatt \| Degree programme as it appears on the student record sheet: | Masterstudium Digital Humanities |
| Betreut von \| Supervisor: | Rebecca Josina Kahn PhD |

# Abstract (English)

This master's thesis goal is to show how the employment of Computer Vision, Machine Learning and Artificial Intelligence tools and methods can benefit the audiovisual analysis of digital archives' large collections. Moreover, multimodality is also incorporated, with the computational text analysis of AI-generated descriptions. For this approach, Jim Hubbard's filmography from the AIDS activist videos archive and from his own online digital collection has been chosen for the analysis. The reason behind this choice is the urge of engaging with minorities' stories and neglected archives within the Digital Humanities scholarship. Furthermore, the experimental nature of Hubbard's films lends themselves perfectly for a computational film analysis, given its focus on the stylistic formalist elements that appear on screen. In my study I propose a film analysis pipeline that starts from shot boundaries detection and brightness levels through the functions included in Arnold and Tilton's Distant Viewing Toolkit. It is followed by an analysis of the movement and the colors through the use of computer vision libraries such as OpenCV. Then with the use of generative AI, Google's Gemini vision model goes through the detected shots and creates a description sheet. The analysis concludes with a text analysis of such description with Natural Language Processing methods. With this analysis my thesis aims at bridging the gap between the difficulties in conducting in-depth analysis of large and complex audiovisual archives, particularly those from underrepresented communities, and the potential for integrated computational methods to unlock and make their formal and aesthetic qualities accessible for in-depth studies.

# Abstract (Deutsch)

Das Ziel dieser Masterarbeit besteht darin zu zeigen, wie der Einsatz von Computer Vision, maschinellem Lernen und Methoden der Künstlichen Intelligenz die audiovisuelle Analyse von großen Sammlungen digitaler Archive unterstützen kann. Darüber hinaus wird auch Multimodalität integriert, und zwar durch die computergestützte Textanalyse von KI-generierten Beschreibungen. Für diesen Ansatz wurde die Filmografie von Jim Hubbard aus dem AIDS activist video Archiv und aus seiner eigenen digitalen Online-Sammlung für die Analyse ausgewählt. Der Grund für diese Wahl ist der Drang, sich mit den Geschichten von Minderheiten und vernachlässigten Archiven innerhalb der Digital Humanities-Forschung zu beschäftigen. Darüber hinaus eignet sich der experimentelle Charakter von Hubbards Filmen perfekt für eine computergestützte Filmanalyse, da der Schwerpunkt auf den stilistischen formalistischen Elementen liegt, die auf der Leinwand erscheinen. In meiner Studie schlage ich eine Filmanalyse-Pipeline vor, die mit der Erkennung von Bildgrenzen und Helligkeitsstufen durch die Funktionen des Distant Viewing Toolkit von Arnold und Tilton beginnt. Es folgt eine Analyse der Bewegung und der Farben durch den Einsatz von Computer Vision Libraries wie OpenCV. Dann geht Googles Gemini-Vision-Modell mit Hilfe von generativer KI die erkannten Aufnahmen durch und erstellt ein Beschreibungsblatt. Die Analyse schließt mit einer Textanalyse dieser Beschreibung mit Nautral Language Processing Methoden ab. Mit dieser Analyse zielt meine Dissertation darauf ab, eine Brücke zu schlagen zwischen den Schwierigkeiten bei der eingehenden Analyse großer und komplexer audiovisueller Archive, insbesondere solcher aus unterrepräsentierten Communities, und dem Potenzial integrierter computergestützter Methoden, um deren formale und ästhetische Qualitäten für eingehende Studien zugänglich zu machen.

# Table of Contents

# Table of Figures

# 1. Introduction

The latest advances in the field of Digital Humanities have brought new tools and methods to process, analyze and explore images and videos. A discipline mostly focused on text since its infancy, Digital Humanities is now shifting towards the audiovisual and the multimodality, not treating text and images as separate bodies anymore. This goes hand in hand also with the new developments in Machine Learning and Artificial Intelligence, that allow faster and more precise processing of the enormous amount of visual data that is created every second. This also allows GLAM (Galleries, Libraries, Archives and Museums) institutions to implement these new tools for their digital audiovisual collections, to assist in the cataloguing and creation of metadata, but also to create new engagement possibilities with laypeople and scholars. Moreover, this also bring new possibilities for new modes of access and engagement in archives that have been historically marginalized or underrepresented, including those documenting minority experiences, such as the AIDS activist video archive.

The AIDS activist video archive was assembled and establish by experimental filmmaker and activist Jim Hubbard, with the help of the New York Public Library, where the archive is currently located. The activists' videos have been first transferred all into VHS tapes and, in the latest decade, digitized. However, no online access is possible. Jim Hubbard, who was also an active activist video maker during the AIDS crisis and his videos are part of the archive, made his filmography publicly available on his website and Vimeo channel, advocating for more open access for the material.

With the research of my master's thesis, I want to show how scholars can engage with this type of archive and how it would benefit from more open accessibility. Given Hubbard's significance in creating and maintaining the archive and the availability of his material, I will conduct a computational and multimodal film analysis of his activist videos. I aim to fill the gap between the evolving field of digital scholarly methods and the critical engagement with activist, amateur, and queer archival materials.

Computational Film Analysis is an emerging new method in Digital Humanities, but has yet no established tools or procedures. The most prominent in the field are Andrew Arnold and Lauren Tilton with their Distant Viewing Toolkit (2023), which allows users to perform computational analysis on visual culture (both still and moving images) by applying computer vision algorithms.

With my thesis, I also wish to contribute to this new emerging field of computational film analysis. I propose a multimodal analysis pipeline that starts from Arnold and Tilton's Distant Viewing Toolkit and includes the use of computer vision libraries such as OpenCV and also a textual analysis using natural language processing.

The aim of my research is not simply to analyze Hubbard's archive material, but to call for more access to digital archives and online databases, which can provide an unprecedented democratization of research data and facilitate collaborative projects into underrepresented communities' histories, such as queer communities. I strongly believe that as scholars of Digital Humanities, coming from a new emerging field, it is important to not simply repeat the same traditional "analog" research of the past decades just with new computational methods, but to start from what has always been ignored, neglected or barely researched. This urge is what motivates my master's thesis.

I will start in Chapter 2 by outlining some historical context on ACT UP, Jim Hubbard and AIDS activist videos. I also explore and showcase some examples on the possibilities of the use of AI and ML in audiovisual archives. In Chapter 3 I focus on the methodology. Firstly, I delineate the tools and possibilities that computational film analysis offers, together with the showcase of some examples. With this approach, I also frame my research within digital formalist film analysis. In Chapter 4 I outline the steps of the analysis, using Hubbard's film *Elegy in the Streets* as an example. I analyze brightness, editing, color and shot type, and compile a film analysis sheet with the help of generative AI. Lastly, in Chapter 5 I will show the results obtained by analyzing the whole filmography, here I conduct an exploratory data analysis with visualizations, a statistical analysis and a textual analysis.

# 2. Historical Context and Theoretical Framing

In this chapter I will outline and contextualize the major theories and studies on the use of Artificial Intelligence in digital audiovisual archives; the use of computer vision methods in film analysis; the history of ACT UP and AIDS activist videos and their influence on queer cinema; and the emerging field of "queer DH". In this process I will also showcase some examples and focus on Jim Hubbard's filmic and archival practices, explain his motivations behind his filmography, and illustrate the tradition of experimental filmmaking that inspired his work.

## 2.1 Audiovisual Archives and Artificial Intelligence

### 2.1.1 The visual and multimodal turn

The Digital Humanities (DH) field has witnessed in the latest decade what has been defined as the "visual digital turn" (Smits and Wevers, 2019, p.194), i.e. the emerging research focus on visual material as a consequence of the developments and improvements of convolutional neural networks, or CNNs. Smits and Wevers additionally identify another potential turn in DH, the "multimodal turn" (2020, p.1268), where image and text are not considered as separated identities. The authors explain that new research possibilities in the humanities are emerging, given by Machine Learning (ML) and Artificial Intelligence (AI) multimodal models such as OpenAI's Contrastive Language Image Pre-training (CLIP). Thanks to these advancements in computer vision, these new research opportunities have made it possible to engage with audiovisual archive's digital collections. These collections host an immense amount of audiovisual material, and the need for new ways of exploring large corpora has become urgent, as Lise Jaillant, 2024 explains. The scale of these digital objects has brought new challenges to the categorization and creation of metadata of such large corpora, often resulting in missing information. Moreover, exploring these collections as a layperson or accessing and analyzing for scholarly purposes present some challenges as well, given the inability to thoroughly search through the archive or process at scale. That is why ML and AI models open audiovisual archives to new methods, such as visual data exploration, visual search based on image, video and audio annotations, and also offer automatization possibilities in cataloguing, automatizing annotation and updating outdated language, and new tools to

improve preservation, such as digital enhancing or noise removal. Thomas Padilla particularly highlights these applications of AI and ML for video, indicating that sematic metadata can be generated automatically, together with speech-to-text transcription for audio material (in Jaillant, 2022). For Jaillant the value of AI comes precisely from its capacity to process huge amount of data fast, and she adds that applying these capacities to archival collection will result in an increased accessibility.

To show an example of these new possibilities Randi Cecchine from the Dutch Sound&Vision institution explains in her blog post[1] from 2021 how AI can also be a way to engage with the public, like the web platform *Citizen DJ*, where users can create music from the Library of Congress audio and video collection. She also mentions one of the methods that will be used for the film analysis in my thesis: Distant Viewing[2], developed by Taylor Arnold and Lauren Tilton, which offers ML tools to engage with and explore large audiovisual corpora. For example, the authors were able to conduct a computational visual analysis of the editing and the colors, including the application of face and object recognition, of the whole *Bewitched* television series (Arnold and Tilton, 2023).

Being able to analyze such large visual corpora is what the media scholar Lev Manovich (2020) expresses as being of vital importance in our age were so much digital material is produced every minute. He founded the term *cultural analytics*, i.e. the exploration of large cultural datasets of visual content, including both digitized visual artifacts and modern visual and interactive media. He explains that given the scale of culture in the twenty-first century it is impossible to approach them with existing methods. New digital methods will give us a numerical representation and visualizations that would allow us to describe the culture with a new vocabulary, better suited to the large amount of today's cultural data. However, he explains, these new quantification measurements are not to be understood as a substitution of human abilities, but as an augmentation thereof, always to be accompanied with a critical examination of the data science methods and their assumptions. Manovich's points are important for us scholars of DH, as we not only utilize digital methods and tools to enhance our research, but we also critically engage with the tools, find flaws and remedies and we contribute and push for a more inclusive data science. Moreover, regarding AI, Bajhor in his most recent volume from 2025, states that "[i]f AI systems are trained on data that is historically and culturally specific, their outputs cannot claim generality, let alone universality" (pp. 11-

[1] https://www.beeldengeluid.nl/en/knowledge/blog/potential-ai-audiovisual-archives#:~:text=With%20the%20ability%20to%20annotate,for%20access%20and%20re%2Duse.
[2] https://distantviewing.org/

12). He also proposes that the relationship between AI and humanities research needs to go both ways, to not only critically engage with AI but also with what AI can reveal about humanities and what kind of questions can be raised. I apply this belief to my analysis as well, as I will discuss in the coming chapters, where the challenges and difficulties encountered are mostly due to how models where pre-trained and the mainstream reality they reflect. At the same time, the use of ML and AI also highlighted the limits of traditional analog film and video analysis, identifying gaps and potential ontological biases.

### 2.1.2 Case studies

One landmark example of how AI and ML can be used not only to explore and analyze data but to also to generate new content is the BBC Archive's project for an AI-generated television program. The BBC experimented with its own archive where AI was utilized to develop two evenings of primetime television in the style of BBC Four, extracting material from the extensive BBC Archive (Heras, 2024). One of the main Computer Vision (CV) application they put into practice was object detection together with automatic generated transcription, using DenseCap (short for "dense captioning") for the purpose. Johnson et al. (2015) explain that DenseCap requires to both localize and describe regions and objects in images using natural language. To do so it employs a Fully Convolutional Localization Network architecture that processes the image to jointly generate region descriptions. In the BBC experiment this resulted in DenseCap "walking" through the television archive and generated sequences following a thread of associational editing, juxtaposing similar images coming from different television programs (Heras, 2024). This generated an experimental primetime television show aired on September 5th 2018, on BBC Four, composed of the clips put together by the model, and presented by Dr. Hannah Fry and a virtual host called "BBC 4.1"[3].

Furthermore, the platform Hugging Face, which offers a vast repository of machine learning models, can significantly benefit GLAM institutions. These models facilitate the analysis and exploration of archival materials through tasks such as text classification, named entity recognition, and image analysis. For instance, Daniel van Strien (2023), in his showcase of example cases for the use of the Hugging Face Hub for GLAM institution, illustrates how the National Library of Norway's AI Lab has utilized the hub to share language models trained on Norwegian texts, aiding in the processing and understanding of their extensive archives.

[3] https://www.bbc.co.uk/programmes/b0bhwk3p

This shows how the engagement of a GLAM institution with Hugging Face can offer a direct benefit in improving its own work with their collection.

Lastly, Generative AI (GenAI) also offers useful ways to engage with audiovisual material. OpenAI's ChatGPT and Google's Gemini both have vision models that are able to generate automatically image captions and description that can help in the cataloguing and metadata creation. describe a picture or a movie frame. For example, Gemini 2.0 has been used by AI engineer Dale Markowitz in 2020 to create a searchable and interactive family video archive[4].

These examples show how CV, ML and AI can be powerful tools in the engagement with and analysis of audiovisual material from digital archives. I find that they also offer a way to keep these archives alive, resurrect old and forgotten ones, and help in the navigation of the numerous and dispersive object in these digital collections.

## 2.2 AIDS Activist Videos

### 2.2.1 ACT UP and AIDS media

ACT UP (AIDS Coalition to Unleash Power) was formed in New York City in March of 1987, after a lecture at the Lesbian and Gay Center by writer Larry Kramer (Schulman, 2021). As explained on their still active website[5], ACT UP was established due to the imbalance of power where health officials, government researchers, medical professionals, and pharmaceutical executives were overshadowing the perspectives of those directly affected. Individuals living with AIDS felt their lived experiences were disregarded, although being the actual true experts on the realities of the pandemic. From New York, ACT UP expanded throughout the United States and eventually all over the world, with branches such as ACT UP Paris and ACT UP Deutschland. Distinctive of ACT UP were their direct-to-action practices or "zaps", such as the die-ins, where activist would lie on the ground pretending to be dead in public spaces[6]. Moreover, the interruption and disruption of public events was another distinctive tactic. A well-known example is the "Stop the Church"[7] protest in 1989, where ACT UP and WHAM! (Women's Health Action and Mobilization) activists barged a mass at St.

[4] https://www.youtube.com/watch?v=_IeS1m8r6SY
[5] https://actupny.org/indexfolder/NYC.html
[6] https://actupny.com/act-up-staged-new-years-day-die-in-at-de-blasio-inauguration/
[7] https://americanarchive.org/exhibits/hiv-aids-and-public-broadcasting/5-1989-1991-broadcasting-act-up-direct-action

Patrick's Cathedral in New York City. Other notable examples were the action of throwing ashes of people that died of AIDS into the lawn of the White House[8] or the interruption of the CBS news program in 1991[9] where a member barged into the studio during a live news segment on the Gulf War and screamed "fight AIDS not Arabs".

Many artists involved in AIDS activism were also active members of ACT UP, which notably influenced the group's creative visuals, with the involvement of prominent figures such as Keith Haring. As Douglas Crimp (1990) discusses in *AIDS Demo Graphics*, the creativity employed in ACT UP's posters, t-shirts, buttons, pamphlet and protest signs are characterized by bright colors, distinctive graphics, and provocative slogans, which helped establish a new aesthetic standard for political art during the AIDS crisis. The most recognizable and remembered iconic symbol of ACT UP was a pink triangle accompanied by the slogan "SILENCE = DEATH", as seen below in Figure 1.

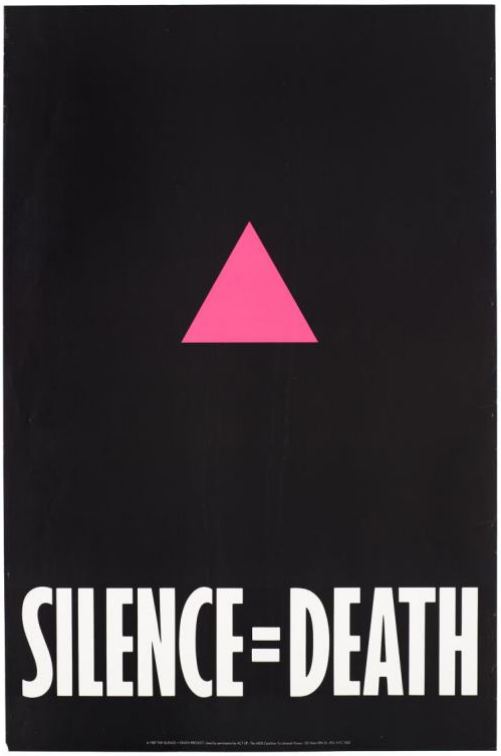



*Figure 1: ACT UP's slogan with the symbolic pink triangle.*
*Source: Wellcome Collection gallery: https://wellcomecollection.org/works/t3duhg9a*

The group deliberately drew on the historical context of the pink triangle, which had been used by the Nazis to stigmatize and identify homosexuals in concentration camps. Reclaiming the symbol, as gay activists had begun to do in the 1970s, ACT UP transformed it into a symbol of

[8] https://actupny.org/diva/synAshes.html
[9] https://actupny.org/diva/news-transcript.html

resistance and liberation that circulated between T-shirts, posters, buttons, stickers and signs. (Gerstner, Routledge, 2011).

This use of art for political causes also extended to video and media. As Alexandra Juhasz (1995) notes in her extensive study on AIDS media production, recording and broadcasting were chosen as a counteraction to the misinformation on how AIDS was represented on television, bringing a new form of alternative AIDS media. She identifies various main types of alternative AIDS media practices, from cable access talk show and experimental works by artists to educational tapes and activist tapes. The video collectives Testing the Limits and DIVA TV (Damned Interfering Video Activist Television) were the most prominent and closely connected with ACT UP (ibid.). Testing the Limits in an interview with Lorraine Kenney (1989) expressed how mainstream media are unlikely to recognize AIDS as a social crisis rooted in broader systemic issues, ignoring protests' reasons and motivations. Through these alternative media practices activists wanted to share their narrative and experiences of the AIDS crisis, but also develop media for education, empowerment, and community mobilization.

### 2.2.2 AIDS activist videos

A determined group of artists and activists from ACT UP New York began capturing their protests on video, taking advantage of the new cheap and broadly available video technology, and documented their struggles going against political and institutional neglect. These recordings, now labeled as AIDS activist videos, went beyond simple documentation, and evolved into a distinctive form of activist filmmaking characterized by subversive aesthetics and experimental techniques. The queer activist communities of ACT UP and Queer Nation (a group formed from ACT UP dedicated to fight homophobia) used whatever resources were at their disposal to spread their message. Typically, they used Hi-8 film and processed their recordings at public access media arts centers, AIDS organizations, schools, and late at night at commercial facilities for little to no pay[10]. Alexandra Juhasz (1995) argues that the "lining up of the new video technologies […] with the AIDS crisis is the founding condition upon which the alternative AIDS media is built" (p.2). This means that the technology itself played a crucial role in shaping the form of AIDS activist media, making it possible to produce and share their own narratives in almost real time. The style of some of these numerous tapes includes fast cuts and music to accompany presentations, mimicking the style of music videos.

[10] https://archives.nypl.org/mss/3622

Some other videos were also educational and other simply had the purpose of documenting various street marches protest of ACT UP. For example, there were movies dedicated to women as caretakers for AIDS patients, such as *We Care: A Video for Care Providers of People Affected by AIDS* (1990) or others were information on safer sex practices.

### 2.2.3 Jim Hubbard

An important figure in preserving the *AIDS Activist Videos* is Jim Hubbard, who, in collaboration with the New York Public Library (NYPL), took care of archiving and digitizing the whole video collection after ACT UP was starting to lose its spaces because of monetary problems (Hallas, 2010). As Juan Suárez (2024) also noticed, Hubbard (and other activist artists that used analog celluloid film and experimental style such as Barbara Hammer and David Wojnarowicz) has received little to no attention, even in extensive studies on AIDS media and AIDS art. My research wants to fill this gap and focus on his experimental short films on queer and AIDS activism. His digitized films are also part of the AIDS activist videos archive in the form of tapes. As he himself explains on his website, Jim Hubbard started filming pride marches and protests in 1979, with his first official short film being *May 21-22 1979* (1979), where he documented a protest in New York against the too-forgiving sentence given to Harvey Milk's killer. His videos can be clearly defined as experimental and avantgarde. For example, in *Gay and Lesbian Pride March 1985* (1985), in Figure 2, we can see how he manually processed the celluloid, tinting the emulsion in a blue dye, as it was typical practice in the silent film era. Furthermore, in *Elegy in the Streets* (1989) he captured this moment of people walking through the AIDS quilt, where the colors are negatively inverted and a pink/bluish tint is added, as seen in Figure 3.

*Figure 2: Still from Hubbard's Gay and Lesbian Pride March 1985 (1985). 17:04.*


*Figure 3: Still from Hubbard's Elegy in the Streets (1989). 09:45.*

Typically, in the filmic context, experimental practices mean "intentionally try[ing] to destabilize the characteristics that define classical Hollywood style, like casual logic, narrative linearity, and unobtrusive craftmanship" (Verrone, 2011, p.20). Verrone identifies certain main features of avantgarde films, which can also be observed in Hubbard's films; they contain non-narrative stories, unusual manners of representation, poetic or lyrical flourishes and alternative production/distribution/exhibition practices. Still using *Elegy in the Street* (1989) as an example, hand processed and filmed in 16mm, scenes of ACT UP's protest marches filmed in color are interspersed with images of flowers in black and white. Moreover, all of his films are silent on purpose, mainly for the reason that "it is a literalization of the phrase Silence=Death" (Hubbard, n.d., *Elegy in the Street* film description[11]). Through the use of direct modification and meddling with film material, montage and color he conveys the desperation, rage and urgency of the times. In an interview with Dagmar Brunow in 2018 Hubbard explains how a long and diversified heritage of alternative film can be directly traced to AIDS activist videos, for example Dziga Vertov's formalist work, the New American Cinema, cinema-vérité, feminist documentaries of the 60s and 70s, or newsreels. Juhaz (1995) adds Third Cinema, ethnographic film and the black film movement to the long tradition of inspiration, highlighting how these types of filmmaking share a commitment to non-fictional, experimental, and socially engaged approaches to cinema.

In her book on the history of ACT UP Sarah Schulman (2021) laments the lack of real stories of AIDS in mainstream media and in cinema. For this reason, she and Hubbard conceived and organized together the ACT UP Oral History Project, a video and audio

[11] https://www.jimhubbardfilms.com/films/elegy-in-the-streets

collection that includes more than 100 interviews with member of the activist group. They did not stop at that, they also co-founded MIX, the New York Queer Experimental Film Festival, an important cornerstone for the queer cinema landscape that provided platforms for new voices in queer cinema. The festival has seen movie premieres of the caliber of Todd Haynes, who showed his thesis film *Assassins: a Film Concerning Rimbaud* (1985) and Jennie Livingston, with the first screening of *Paris is Burning* (1990). In fact, AIDS Activist Videos are credited with inspiring the first wave of New Queer Cinema (NQC), which revolutionized the landscape of independent queer film, marking the beginning of the 1990s as a "watershed year for independent gay and lesbian film and video", a phenomenon introduced at the Toronto's Film Festival (Rich, 2013, p.16).

### 2.2.4 Influence and legacy

The influence that the AIDS activists videos had on NQC has been explored and analyzed in various studies. In their exhaustive survey on queer imagery Benshoff and Griffin (2010) explain how the main characteristics of AIDS activist videos can also be found in NQC, with the reappropriation and pastiche of mainstream images as well as their desire to provoke and challenge the viewers. They also underline how many of the directors and creatives who became leading names in the NQC movement were part of AIDS activism, such as Derek Jarman and John Grayson. Building on this approach, we can see how the influence of these activist videos on NQC highlights the interconnectedness of social movements and artistic expression, demonstrating how activism can shape and inspire new forms of cultural production in the landscape of queer cinema. Todd Verov (in MacKenzie, 2014) in his manifesto on NQC mentions how "we were all angry little art terrorist coming out of ACT-UP and Queer Nation and ready to take over the world" (p.393), underlining how the urge to bring queerness and AIDS stories to the big screen was strongly felt. Michele Aaron (2022) in his exhaustive NQC reader affirms that "NQC cannot be removed from the context of the AIDS epidemic", it is either lurking in the background or has a central absence (p.6). There was also a strong desire and vital need for telling the stories of the victims. In the context of AIDS, a new "queer paradigm had emerged: [...] non-normative status was then attached to you via the virus" (McCann and Monaghan, 2022, p.102).

Usually, mass media narration about the epidemic divided the people with AIDS only in two categories: the victim, for example those that contracted the virus through blood transfusion, and the deserving, such as drug addicts and gay men. As the author Julianne Pidduck (in Aaron 2022) explains, the activism was a key moment for the development of an

independent queer circle, not relying on commercial funding, which found then its way to museums, and independent film festivals. In the same volume by Aaron, Monica Pearl makes another important point, where she sees the chaotic, non-narrative, non- progressive nature of the AIDS virus as a way that depicted a shattered chronology, by reassigning responsibility, time and death into a new timeline. Basically, even if AIDS is not present as a theme of the movie or is not even present at all, it gets expressed through cinematic aesthetic: editing, lights and sound.

### 2.2.5 Towards a "vital nostalgia"

Most, if not all, of Film Studies research on AIDS activist videos focuses on their influence on NQC, meanwhile framing AIDS activist videos alone as being part of an experimental and avantgarde cinematic tradition is mostly never considered. Furthermore, traditional Film Analysis, let alone Computational Film Analysis, of the AIDS activist videos archive material has never been conducted. Juhasz (1995) cites Ruby Rich and her writing about feminist film practice of the seventies, where she states that the condition of production is an integral part of understanding feminist films. She states that Rich's questions are equally significant in order to make sense of the alternative AIDS media. This approach is central to my research, which seeks to analyze the formal elements of these videos and emphasize the technological and aesthetical elements that influenced their production. In doing so, I aim to demonstrate how the materiality of these works is inseparable from their activist function and value.

Furthermore, I also aim to bring focus and attention to the AIDS activist video archive and engage with it. Marika Cifor in her volume *Viral Cultures – Activist Archiving in the Age of AIDS* (2022), stresses the need to make archives less static and less about the past, since they are always conceptualized as dead and irrelevant in the public imaginary. She underlines how it is important to avoid making AIDS a thing of the past and be forgotten in archives. She states that framing AIDS and its archives as relics of the past runs the risk of marking HIV/AIDS as over. Her concept of "vital nostalgia" explores the idea of nostalgia not as a passive longing for the past, but as an active, dynamic force that shapes the present and future. "Vital nostalgia" recognizes the emotional power of archives and memory, treating them as living, evolving sites of engagement rather than static repositories (ibid.). She puts an emphasis on how nostalgia can inspire creative and critical interactions with history, fostering connections that reimagine the past in ways that are relevant and impactful for contemporary and future contexts. Jim Hubbard himself (n.d.) expresses his worries on how AIDS activist videos can be remembered

in the future. If removed or forgotten it could alter their intended significance, as they are not just archived recordings but represent a community's collective ideas about fighting and living with AIDS. This concept of "vital nostalgia" can provide a productive approach for analyzing Hubbard's work. By engaging with these videos as dynamic tools of remembrance and resistance, we can see how they enact the archival practice Cifor advocates for, one that insists on the continued presence and relevance of AIDS histories.

## 2.3 Film Analysis in Digital Humanities

### 2.3.1 Digital Film Analysis

As already stated, the field of DH offers new quantitative digital tools and methods that enable scholars to analyze large datasets, uncover patterns, and visualize information in innovative ways that would be difficult or impossible when using traditional methods alone. Nevertheless, as Erik Champion puts it: "Digital humanities is text heavy, visualization light and simulation poor" (2016, p.25). He highlights how the field has been for many years focused on textual and historical analysis, privileging mostly text-based research from the disciplines of literature, history and linguistics. However, as previously mentioned, in the last few years there has been a "multimodal turn" (Smits and Wevers 2023, p.1267) that has seen advancement and new studies in the audiovisual fields such as art history, music studies and film and media studies. This shift towards multimodal analysis reflects the recent recognition of the importance of visual and auditory elements in understanding cultural and historical contexts, and the possibilities it gives to DH in expanding the scope and impact of its research.

The field of Film Studies has not been left untouched by the introduction of digital tools and methods in humanities research. Contrary to what it might be expected, the application of quantitative, statistical and computational methods to film analysis has actually a long-standing tradition, dating back to the foundational work of Barry Salt (1983) with his statistical style analysis of motion pictures, where he admittedly took inspiration from the traditional statistical analysis of literature and music. As he explains in *Film Style and Technology*, "to carry out an analysis of a film in this way, it is necessary to run it on some sort of viewing machine" (p.141). In his first writings from 1983 it is not clear what sort of "viewing machine" he used for the analysis or if it was simply a suggestion. However, in more recent writing, such as *The Metrics in Cinemetrics* from 2011 he explains that he normally uses non-linear editing software, which

then evolved in creating together with Gunārs Civjans the Cinemetrics software[12]. His calculations typically focus on the scale of shot distribution. This means how many times a different type of shot (Close-Up, Long Shot, Medium Shot, etc.) are used and how they are statistically distributed. In this way he could identify stylistic tendencies and, for example, observe if a director or a certain film era favored close-ups for emotional intensity or long shots for establishing context. An example from one of his first experiments from the 1980s can be seen in Figure 4 below.

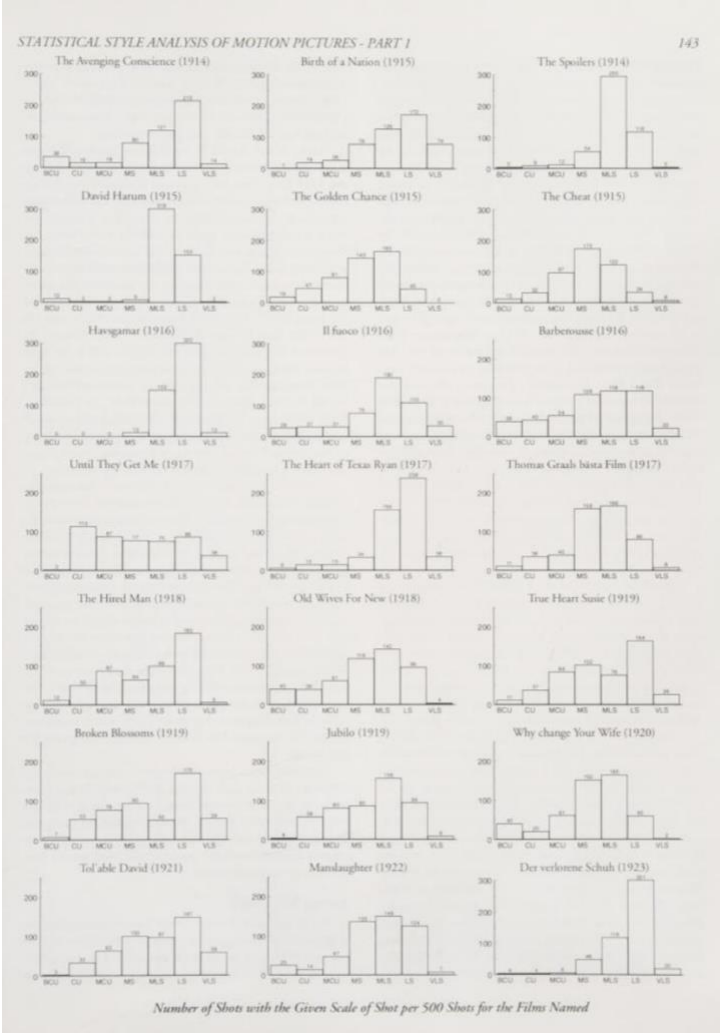



*Figure 4: Statistical shot distribution analysis from Barry Salt (1983, p.143)*

Looking at the present's possibilities for Film Analysis through computational methods, similar to the transformative impact of Franco Moretti's practice of *distant reading* in textual analysis (2013), AI and ML present new tools for engaging with audiovisual material

[12] https://cinemetrics.site/#/

and for studying form, aesthetics and artistic expression of images and video. Non-narrative avantgarde films perfectly lend themselves to this kind of qualitative analysis and are a preferable object of study of formalist film analysis. In the next section I will outline some examples.

### 2.3.2 Case studies

One seminal example is Heftberger's *Visualizing Dziga Vertov* (2018), a study where the collection of film from the Russian director, situated in the Austrian Film Museum, has been computationally and statistically analyzed. This process includes both manual and computer-assisted film annotation, resulting in empirical data that serves as a basis for statistical analysis and, more significantly, for creating visualizations. The study not only gave more insights into Vertov's style and filmography, but it was able to shed some light on the archival practices and policies in the Soviet Union of the 1920s.

Another example is the aforementioned Cinemetrics (Tsivian in Ross et al., 2015), which is an interactive data collection platform designed to assist with the manual annotation of films. The site serves as both a repository for a crowd-sourced database containing quantified information on thousands of films throughout history and a hub for several published articles that have made extensive use of this database. One interesting function is that the user is able to compare various movies based on action, movement and colors.

Furthermore, the tool Anvil offers the possibility of annotating audiovisual material with multimodal dialogue. Users can create annotations on multiple customizable tracks by adding time-anchored elements with attribute-value pairs. It supports hierarchical structures, cross-level links, and flexible annotation schemes. It also allows speech transcription imports and exports text or table data for further analysis (Kipp, 2001).

The scholar Nick Redfern (2021) conducted a full computational film analysis using the programming language R, where he analyzed color, editing and music. He was then able to apply statistical functions on the results and retrieving stylistic and aesthetics information from formulas and modeling.

Kevin Ferguson (2017) conducted an experimental analysis of Walt Disney movies with the image processing software ImageJ. With the summed z-projection function, which is a method for creating a 2D image from a 3D stack of images, he "added" all the frames of a movie together, which in some cases revealed what the main focus of the film was. One interesting result is, for example, that using this function on *Alice in Wonderland* revealed the silhouette of the protagonist Alice as prominent.

### 2.3.3 Digital Formalist Film Analysis

Heftberger (2018) underlines how these types of studies can contribute to the exploration of visual representations of time-based processes, which is an area that originated from the Russian formalist movement of the 1920s. Furthermore, neo-formalist film analysis, where I situate my research, is rooted in Russian formalism and expanded by renowned film scholars and theorists David Bordwell and Kristin Thompson (Thomson-Jones, 2008). It focuses on deriving meaning from the elements explicitly presented on screen, or more precisely said, in the *mise-en-scène*. These elements include visual composition, editing, sound design, and other cinematic techniques that are in front of the camera. This approach emphasizes analyzing the film as a constructed artifact, with its meaning emerging through the deliberate arrangement of its formal elements. By focusing on the formal aspects of film, this method allows for a detailed and systematic analysis of how cinematic techniques contribute to the overall impact and meaning of a piece of work.

However, when we apply these kinds of methods to digital tools, we always have to keep in mind that they are still fairly new and experimental, so we always have to go back and forth between new and traditional methods. On this topic, the scholar Daniel Chávez Heras (2024) in *Cinema and Machine Vision* calls for a more critical and reflexive engagement with computational tools, advocating for methods that incorporate both the analytical power of machine vision and the interpretative depth of traditional film studies. He appears to be skeptical of letting a machine "view" a film, since it is unable to capture the affective dimension of the action . I find this as another good reason why experimental films lend themselves perfectly for more formalist or "colder" analysis. Since the focus is on the most formal technical elements, technical tools are able to extract and process this kind of information faster and more precise than a human can. Heras also indicates that a computational analysis reflects the same biases and perspectives of the present society. This is a well-known problem when dealing with data. Also, CV is typically used in the medical field, in the military and for video surveillance. Training on cultural data and especially with an aesthetical focus in mind is still on its infancy in the mainstream CV research. Social, historical and cultural context needs always to be integrated, as the author argues, otherwise the insights may end up being too shallow or misleading. Overall, my position on this is that, as of the state of computational humanities research is now, a human still needs to always accompany the machine, aligning with what is already known as the "human-in-the-loop" model. As Wu et al. explain (2022) Human-in-the-Loop integrates human knowledge and experience, with human input and

feedback aiding AI and ML systems' lifecycle. I think this is necessary when integrating AI and ML tools in humanities research, since the humanities tend to emphasize context, interpretation, and critical reflection, which a machine alone hardy fully grasps or replicates, as of now. Computational methods can reveal patterns or anomalies at scale, but it is still the human scholar who must interpret these findings, situate them within a cultural framework. Without human supervision, there is always the risk of reinforcing biases, mis-readings, or over-relying on quantification in a field fundamentally grounded in meaning-making.

## 2.4 Queering DH

There have also not yet been many studies that converge DH and Queer Studies. Pioneer scholars in the field such as Bonnie Ruberg, Jason Boyd, and James Howe in their 2018 publication have been advocating for a "queer DH", underlining the shared objectives between queer studies and the DH, with their challenges to traditional paradigms and the potential for social critique and the development of practices that embrace non-normative and diverse perspectives. Another important influence in "queer DH" comes from Chun and her article *Queering Homophily* (2018). She critiques the principle of homophily, i.e. the "similarity breeds connection" principle, which underpins much of network science. Chun argues that homophily, while seemingly innocent, perpetuates segregation and inequality by embedding biases into network structures and algorithms.

Uniting these fields can undoubtedly offer exciting opportunities for interdisciplinary research, challenging conventional approaches and highlighting the importance of inclusivity and diversity in DH research. I find that this interdisciplinarity can bring opportunities to re-activate queer archives, often ignored and left unexplored. An archive such as the AIDS Activist Video archive needs to be opened and be accessible. In an interview with Dagmar Brunow in 2018 Jim Hubbard has also urged that the digitized videos be available online and accessible to everyone. However, the videos where donated as physical objects and the makers retained the copyright. Because of this the NYPL is preventing online digital access, feeling they do not have the permission to show the videos to the public due to the lack of agreement from the owners. Hubbard also laments that the NYPL did not make any attempt to get the rights to themselves even just for the purpose of opening the archive. For this reason, he created the ACT UP Oral History Project, so that it was born digital, and made his own videos available both on his personal website and on his Vimeo channel.

There are however some successful examples of digital archives and online databases that have democratized access to primary sources, successful stories that facilitated collaborative projects that explore the histories of underrepresented communities. For instance, the Queer Archives Project at Lafayette College[13] integrates oral histories, archival materials, and undergraduate research to explore queer histories, utilizing platforms like Scalar to present interconnected content. Furthermore, some studies that challenge the status quo and push for diversity and inclusivity within DH platform are emerging. A study published by Joseph Nockels et al. in 2024 in *Digital Scholarship in the Humanities* questions whether current DH programs adequately facilitate diversity in research by looking at the Transkribus Scholarship program. The study suggests that more effort is needed to enhance the potential benefits, ensure the sustainability of the and support inclusivity. Furthermore, initiatives like the European Association for Digital Humanities (EADH)[14] are actively exploring the productive potential of diversity within DH, seeking visions and initiatives that cultivate and build upon diverse practices.

Therefore, while the potential for interdisciplinary research uniting queer studies and Digital Humanities is significant for promoting inclusivity and diversity in DH research, realizing this potential requires addressing issues of accessibility to ensure that marginalized voices and histories are not further obscured and closed in an inaccessible archive.

[13] https://queerarchivesproject.lafayette.edu/
[14] https://eadh.org/

# 3. Methodology

In this chapter I will outline the methodology used for conducting the film analysis of Jim Hubbard's videos. I will explain how the videos were selected and the data collected, and how I divided my workflow in 3 main areas and how each area is meant for different task: Distant Viewing for the analysis of brightness and editing, and for dividing the film in shots; OpenCV for analyzing movement and color and create a shot grid; Generative AI (Hugging Face, Gemini 1.5 pro) for identifying the shots type, create a description and formatting the final film analysis sheet for each film. Then lasty the results culminate together in an Exploratory Data Analysis (EDA) and textual analysis with natural language processing (NLP). I propose a pipeline for computational film analysis as seen in Figure 5.

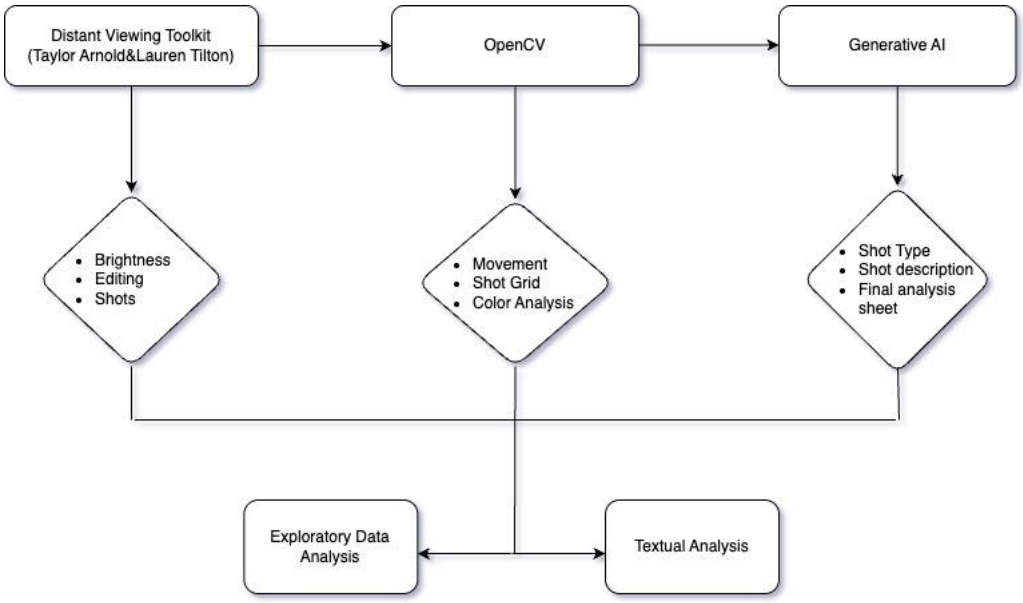



*Figure 5: Computational Film Analysis Pipeline*

## 3.1 Data Collection

The AIDS activist video archive is situated in the New York Public Library and consists of 630 VHS cassettes[15]. The videotapes have been fully digitized and are also available on DVD format (Hubbard in Brunow, 2018); however, the entire collection is not available online, the only way to have access to it is to send a request to the NYPL and physically work with it there. Aware of this lack of availability, Jim Hubbard collected his own digitized videotapes

[15] https://archives.nypl.org/mss/3622#controlaccess

and made them available online on his website[16] and on his Vimeo channel[17], together with the rest of his filmography. The videos all have a CC license "attribution, non-commercial". For the purposes of my research, I downloaded the video directly from his Vimeo channel as .mp4 files. Given my aim of focusing on activists' protest and pride marches, the main subject of Hubbard's filmography, I excluded any other video that did not fit the thematic, in order to create a thematically coherent corpus. I also chose videos that were filmed during the timeframe of the beginning of the HIV/AIDS epidemic to the years of ACT UP activism. This timeframe is crucial as it ensures that the analyzed films are deeply engaged with the core themes of this study, focusing on activism filmmaking during that time, 1979-1991. Moreover, for copyright reasons his feature film *United in Anger: A History of ACT UP* was not included. However, the film consists mostly of clips of his other already available videos with some interviews with ACT UP members. Jim Hubbard's film in my corpus were all filmed analogically with 16mm and 8mm film, and I downloaded the videos at 720p resolution. The numbers refer to the number of pixels horizontally, indicating the height of the picture, and the "p" stand for "progressive", as opposed to "interlaced" a video format only used on television broadcasting (Brown, 2015). As Giovanna Fossati extensively explains in her book from 2008 on digital film archives, up from 2k resolution (1080p) has become the accepted required resolution for 35mm color film digitization, as the number of digital pixels roughly correspond to the analog grain. 16mm would fall below that with 8mm being exactly around 720p, meaning that there is not that much visual information loss taking place if the digital copy is 720p. While higher digitization resolutions were technically possible, since the Vimeo stream of some of the videos can go up to 4K (2160p), 720p offers a sufficient level of visual fidelity to accurately analyze the key aspects of the footage for my thesis. Analyzing at a higher resolution would not automatically mean an increase in relevant information in the analysis, given the inherent limitations of the original film stock. Furthermore, 720p strikes a practical balance between visual detail and manageable file sizes, a "sweet spot" that ensures that the essential visual information is preserved for the analysis without contracting the potentially unnecessary computational power and storage demands of higher resolutions.

A dataset with the videos that were collected can be observed in Figure 6. This is the dataset I started the analysis with, with the values Title, Year, Resolution, Format and Duration. In the course of the analysis and based on the results other values were added.

[16] https://www.jimhubbardfilms.com/films
[17] https://vimeo.com/unitedinanger

| Title | Year | Resolution | Format | Duration |
|---|---|---|---|---|
| ACT UP Albany Action | 1988 | 720p | mp4 | 11m 55s |
| ACT UP Kiss-In | 1988 | 720p | mp4 | 02m 59s |
| Elegy in the Streets | 1989 | 720p | mp4 | 27m 21s |
| Gay and Lesbian Pride March 1985 | 1985 | 720p | mp4 | 17m 54s |
| Gay and Lesbian Pride March and Gay Pride Run 1986 | 1986 | 720p | mp4 | 08m 31s |
| Gay Pride Run 1983 | 1983 | 720p | mp4 | 21m49s |
| Gay Pride Run 1985 | 1985 | 720p | mp4 | 09m 43s |
| International March for Lesbian and Gay Freedom | 1984 | 720p | mp4 | 11m 05s |
| June 12, 1982 | 1982 | 720p | mp4 | 08m 52s |
| Lesbian & Gay Pride New York 1983 | 1983 | 720p | mp4 | 11m 46s |
| Lesbian & Gay Pride New York 1987 | 1987 | 720p | mp4 | 10m 02s |
| Lesbian & Gay Pride New York 1988 | 1988 | 720p | mp4 | 08m 33s |
| March on Washington 1979 | 1979 | 720p | mp4 | 24m 29s |
| May 21-22, 1979 | 1979 | 720p | mp4 | 11m 21s |
| Stop the Movie (Cruising) | 1980 | 720p | mp4 | 11m 20s |
| Two Marches | 1991 | 720p | mp4 | 09m 38s |
| Wall Street II | 1988 | 720p | mp4 | 04m 52s |

*Figure 6: Film corpus data frame*

Most of the videos contain opening and closing credits inserted after the digitization process, such as the title of the video and of the company that conducted the digitization, called Pro8mm. I edited these out with the non-linear trimming function included in QuickTime Player. Although texts such as opening and closing credits are usually considered part of the film in Film Studies and should not be ignored in a film analysis, I would argue that in this case it would only skew the analysis as they were not really an aesthetic choice, were added later and are not part of the original "world" of the film. In fact, credits made at the time by Jim Hubbard were left intact in the videos.

## 3.2 Neo-Formalist Film Analysis

As mentioned in the previous chapter, I will conduct a neo-formalist film analysis, as experimental films lend themselves perfectly for this type of analysis, given their structuring around technical film elements. Neo-formalist film analysis takes inspiration from Russian film formalism. The main tradition of Russian film formalism emerged with names such as psychologists Hugo Munsterberg and Rudolf Arnheim, as James Dudley Andrew (1979) explains. They saw technology as the main driver of the evolution of motion pictures and argued that film should be an art form because of its formal elements. Moreover, Dziga Vertov and Sergei Eisenstein, the names mostly associated with formalism both as film directors and as film theorist, focused on the editing and the meaning it conveys through association of the images (ibid.).

From this formalist base, in the 1980s, neo-formalism evolved, the main voices being Kirstin Thompson and David Bordwell. In their praised and renowned volume *Film Art: An Introduction* (2020, first ed. 1993), they analyze how films construct meaning and evoke responses in viewers through their formal systems. Rather than focusing on social, political, or authorial interpretations, neo formalism emphasizes the objective study of film form, including narrative structure, cinematography, editing, sound, and *mise-en-scène* and how these elements interact to create specific aesthetic effects and guide the spectator's understanding. The goal is to understand the intrinsic logic of cinematic construction and the perceptual and cognitive processes involved in film viewing. Particularly, Thompson (1989) identifies two main assumptions for a formalist film analysis: firstly, film is an artificial construct, with the artist being a rational agent making aesthetic choices to an end; secondly, film "defamiliarize[s] conventional devices of narrative, ideology, style, and genre" (p.36). These two assumptions are true also for the films of Jim Hubbard. For example, the intentional omission of sound in most of his movies, which "is a literalization of the phrase Silence=Death […] and also in New York City [...], silence is a great luxury" (Hubbard, n.d., *Elegy in the Streets* film description[18]). Moreover, Thompson also identifies various types of "motivations" (p.15) behind films: compositional, realistic, transtextual and artistic. Compositional is when the settings are planned in a way that has narrative purposes; realistic refers to those elements that make the world more real and help justify actions that would be unfeasible in the real world; transtextual are the references to other elements in history recognizable by the spectator. The most relevant for my research is the "artistic motivation" (p.19), i.e. when the film puts the aesthetic qualities and technical artistry on the foreground and as the main (if not the only) motivation for creating it. The technical and aesthetic decisions of Jim Hubbard in his films lead to the conclusion that his intentions were artistic. We can observe by looking that the use of colors and editing that he focused on the aesthetic and technical aspects of filmmaking, creating a work that prioritizes artistic expression over more conventional narrative, in order to convey the urgency and reality of the ACT UP protests. For example, in the film *Elegy in the Streets* he juxtaposes vibrant color scenes of protest marches against black-and-white images of flowers (01:50-02:10).

I will adapt this neo-formalist methodology to experimental films, which means leaving out narrative structure. Focusing on these formal filmic elements means to look at what is shown on the screen, this is normally called in Film Theory *mise-en-scéne*. The *mise-en-scéne* refers to "the director's control over what appears in the film frame" (Bordwell and Thompson,

[18] https://www.jimhubbardfilms.com/films/elegy-in-the-streets

2020, p.113). The term is borrowed from theatre and includes setting, lightning, costume and makeup, staging and performance. Adapting Bordwell and Thompson's definition to my research purposes I will analyze:

- Settings: in Jim Hubbard's films the settings are the protests, pride marches and sit-ins of ACT UP. I will analyze, through the textual descriptions, what appears in these settings and let the GenAI model transcribe the protest signs. In this way, the analysis of the settings will show common patterns within the visual and spatial representation.
- Lightning: the use of lights in my case refers to the use of brightness, as the lights in the cinematic sense are always the outside natural light. This will reveal how Hubbard plays around with the brightness levels in some scenes, going from lower to higher levels intermittently.
- Movement: movement is part of the staging. Analyzing movement in Hubbard's film will help shed some lights on the meanings it can convey, such as, for example, chaos and uprising.
- Color and range of tonalities: after brightness, color is another value that Hubbard tackles with, going from negative colors inversion to oversaturation, creating unsettling effects. I am interested in focusing on dominant colors and color palettes to highlight the change in color and the somber tone of the tinting.
- Frame dimension and shape: the dimension of the frame can help tell which type of film material was used, either 16mm or 8mm.
- Editing: another way Hubbard experimented in order to convey the urgency of the time was using dynamic editing. This will reveal ow the rapid succession of shots and the rhythm created through editing contribute to a sense of immediacy, tension, and the pressing need for action that characterized the ACT UP movement.

Lastly, for the analysis it is also important to keep in mind that the films went through the process of digitization but not that of restoration, so scratches, dirt and noise are present in the films and may sometimes mislead the analysis.

## 3.3 Computational Film Analysis

### 3.3.1 Computer Vision

As explained in the previous chapter Digital Humanities has revolutionized numerous fields, starting from textual analysis, and expanding to visual fields like art history and film studies, where the development of computer vision has opened up new possibilities for

audiovisual research. Computer vision is a research field that employs computation to understand and analyze images, relying on algorithms to detect shapes and objects (Smits and Wevers 2019). Computer vision is a branch of Artificial Intelligence, and its ultimate goal is to emulate human vision, even reacting and taking action to visual inputs (Gonzales and Woods 2018). For the scope of my research, the branch of computer vision that I am interest in is focused on both image and video analysis and processing. As Gonzales and Woods explain, a digital image can be defined as a two-dimensional array of pixels, each representing a discrete unit of color and brightness, and a digital video is essentially a sequence of such images, or frames, displayed in rapid succession. These frames are traditionally reproduced at a rate of 24 frames per second (fps), creating the illusion of motion by presenting 24 distinct images within a single second.

The computer vision libraries that are used in my studies are: OpenCV[19], PyTorch[20] (through the Distant Viewing Toolkit), Pillow[21], and Moviepy[22]. The most used computer vision library is OpenCV (Open Source Computer Vision Library), it consists in a collection of algorithms that process and analyze images and video. It can also conduct object detection, feature extraction, camera calibration, and even 3D reconstruction. It can be used with various programming languages. It can be easily used with Python, my preferred choice of programming language for my thesis, and it can perform various video processing tasks, such as reading and writing video files, extracting frames, applying filters and transformations, detecting and tracking objects, recognizing faces, and directly modify the footage, like noise removal. For the purposes of my research, I will use this library to analyze colors, range of tonalities and movement. In the Distant Viewing Toolkit, the main library used by Arnold and Tilton in their model is PyTorch, which provide the necessary infrastructure for building and training more complex neural networks tailored for visual tasks such as image classification, object detection, and semantic segmentation. In particular, PyTorch's torchvision library specifically offers essential datasets, pre-defined model architectures, and image transformations that streamline deep learning workflows for visual data. Complementing this, Pillow, a fork of the Python Imaging Library (PIL), serves as a fundamental tool for opening, manipulating, and saving images in various formats. Lastly, Moviepy was used to access the

[19] https://opencv.org/
[20] https://pytorch.org/docs/stable/library.html
[21] https://pillow.readthedocs.io/en/stable/index.html
[22] https://pypi.org/project/moviepy/

video directly on the notebook after loading it, so to not having to use external players when there was the need to double-check something manually.

Bringing together these computer vision applications with traditional film analysis has a lot of advantages. The most significant is its capacity to analyze entire films, while traditional film analysis is limited to the analysis of individual scenes or sequences, often attributing an overall meaning to the entire film or to the director's intention only based on those single scenes. Not only entire films, but even entire filmographies at a time can be analyzed through computer vision. This ability to process vast amounts of visual and auditory data gives us the possibility to better identify aesthetic patterns and gain a deeper understanding of the cinematic elements the compose a film, which we may have not been able to see by looking only at a few scenes.

### 3.3.2 Distant Viewing Toolkit

Developed by Taylor Arnold and Lauren Tilton, data scientist and cultural studies scholars respectively, Distant Viewing applies computer vision methods to the analysis of digital images and video. The name is a direct reference to Franco Moretti's (2013) distant reading, dedicated to the analysis of large corpora of texts.

The distant viewing method goes one step further than just simply analyzing an image. Arnold and Tilton explain in their *Distant Viewing* book (2023) that computer vision algorithms need to be seen as "active participants in the process of knowledge production through the act of decoding" (p.27). They underline how the labels and annotation created by the algorithms are themselves a product of a social and cultural context and not intrinsic of the images themselves. They updated a typical data science pipeline to the distant viewing purposes as seen in Figure 7.

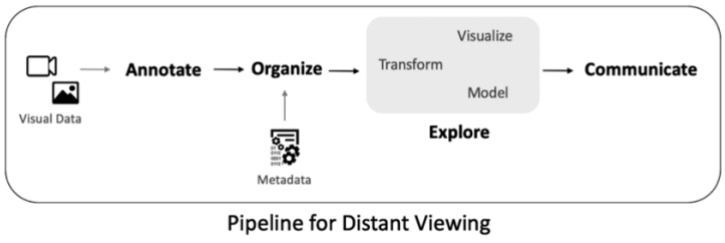



*Figure 7: Distant Viewing pipeline (Arnold&Tilton 2023, p.36)*

As they thoroughly explain in their book, the first step is to annotate the data and then combining them with metadata information. It is an iterative process, and it requires re-engaging directly with the source images and videos. Moreover, the annotation and metadata

chosen for the analysis need to be guided by the elements we are interested in studying. Metadata can come from various sources, for example in my study some of the metadata come from the video themselves (resolution, duration, height and width in pixels, frame number and frame rate). Moreover, on the respective films' Vimeo page descriptions, Hubbard offers more details on the protest or march that is the subject of the film, the filming conditions, the film material he used, the people present and usually also some personal anecdote. For the last steps, they state that unlike hypothesis-driven research, Exploratory Data Analysis (EDA) focuses on exploring data characteristics and developing assumptions based on those observations.

Arnold and Tilton made available online on their GitHub a distant viewing toolkit with ready-made Jupyter Notebooks to edit based on one need[23]. For example, one Notebook presents their entire analysis of the television series *Bewitched*[24], with the focus on editing and the character detection. I will adapt their Python code to the purposes of my analysis. Furthermore, I will extend their methodology using the library OpenCV and GenAI, as detailed in the following section.

### 3.3.4 Generative Artificial Intelligence

Before landing on a final choice for how to use GenAI, I conducted some trial-and-error experiments. Firstly, I tried to use OpenAI's ChatGPT 4o mini model to describe a couple of film stills and identify the type of shots. The model mostly ended up overinterpreting the images (for example, adding an extra sentimental layer and description to some of them) and was having troubles contextualizing the protests as protests (sometimes it would describe them as "people are gathered for a meeting together"). I then moved to Google AI Studio[25], where users are able to upload a video and give it pre-written prompts to analyze it. The results were really good, however, the limit to analyze the video was 4 minutes. It also appears that at the moment of writing the video analyzer application is not available anymore. Satisfied with Gemini capabilities I decided then to use Gemini 1.5 pro (more details in the second part of this section) and I was overall satisfied with the description, as the model seemed to contextualize the stills pretty well, and it gave objective descriptions and managed to correctly transcribe the protest signs' text. However, it did not manage to identify correctly most of the shots. Double checking the results manually made me notice some errors. For this reason, I

[23] https://github.com/distant-viewing/dvt
[24] https://github.com/distant-viewing/dvt/blob/main/tutorials/Distant_Viewing_Tutorial_2_Network_Era_Sitcoms_and_Visual_Style.ipynb
[25] https://aistudio.google.com/app/apps

decided to select a pre-trained transformer-based image classification model from the Hugging Face library[26]. The choice was motivated by the demonstrated ability of a transformer architecture to learn intricate visual representations and contextual cues, crucial for distinguishing between nuanced shot compositions (Vaswani et al., 2017). I also have already and successfully used this specific model during my studies for analyzing film stills. It achieves an accuracy of 62.59% on the evaluation set with a loss of 0.9864. While not perfect, this level of performance provides a valuable foundation for initial analysis, considering the inherent complexity of shot classification and the subtle visual variations present in cinematic language. The training process utilized standard hyperparameters, including a learning rate of 2e-05 and a batch size of 4, which is common practices for fine-tuning transformer models. After some experimenting, it appeared that the results were more accurate than the Google's GenAI model. Moreover, Film Analysis is not a precise science, the terms for the shots, the editing and the settings all come from film production, where there is not fixed agreement and differences are present through cultures and countries. The model used includes the class labels: ambiguous, close-up, detail, extreme long shot, full shot, long shot, medium close-up, medium shot. Future work may explore fine-tuning strategies or alternative architectures to further enhance classification accuracy.

As already briefly mentioned, to obtain concise textual representations of the identified film shots, Gemini 1.5 pro, a multimodal model from Google, was employed. Also, in this case some trial-and-error experiments were conducted. I firstly used the object detection algorithm YOLO, which gave poor results. For example, it kept identifying cellphones where there were none, farm animals like cows and for the majority of the shots the only detections resulted in "person" or "people". Another problem is that, even if the detection gave correct and satisfying results, it could not really generate a description, but only a list of objects. Gemini's vision model's inherent ability to understand and integrate visual and textual information makes it highly suitable for generating descriptive summaries of image content. Given the objective of obtaining broad, one-sentence descriptions for a large number of shots, I used zero-shot prompting with the straightforward prompt instruction: "describe what is visible in the shot in one sentence". This approach leverages Gemini 1.5's extensive pre-training on diverse image-text data, enabling it to generate relevant descriptions without requiring task-specific fine-tuning, which would be resource-intensive. While zero-shot descriptions may not capture every subtle cinematic detail, the level of information provided in a single sentence aligns with the

[26] https://huggingface.co/pszemraj/dinov2-small-film-shot-classifier

analytical goals of my study, prioritizing efficiency and the identification of overarching visual patterns. Furthermore, unlike OpenAI or other models that rely on specific pre-training techniques like CLIP for aligning image and text embeddings, Gemini models are fundamentally multimodal by design (Gemini Team, Google, 2023). Their architecture is engineered to natively ingest and process information across various domains, including text, images, audio, video, and code. This means that a model like Gemini is trained on multimodal datasets, allowing it to learn the complex relationships, patterns, and correspondences between different modalities directly within their foundational structure. Lastly, the accessibility and cost-effectiveness of Gemini 1.5 pro make it a practical and strategic choice for this research, allowing for comprehensive analysis within the available resources. As Smits and Wevers (2023) explain while standard computer vision models need extra training for new tasks, multimodal models perform well "in the wild" on different datasets without prior optimization. This "zero-shot" capability is particularly helpful for analyzing complex multimodal concepts in large visual datasets such as Hubbard's film corpus.

### 3.3.5 Textual Analysis

One final step typical of film analysis, especially in the Italian- and German-speaking tradition (see, among others, Casetti 1994, Bertetto 2014, Sayed 2019) is to create a film analysis log sheets with the shots chosen for the analysis, as seen in Figure 8.

| Subsequenz | Zeit | Inhalt |
|---|---|---|
| 1. Sequenz: Familie Grotowski | | |
| 1 | 00:00:00 | Vorspann: Produktionsinformationen |
| 2 | 00:03:20 | Des Nachts liegt Hank wach im Bett, übergibt sich im Badezimmer, zieht sich an und fährt in ein Diner, wo eine Kellnerin ihm eine Schale Schokoladen-Eis reicht und Smalltalk mit ihm macht. Gleichzeitig hat Hanks Sohn Sonny in einem Motelzimmer mit einer wasserstoffblonden Prostituierten mit makellosem Körper kurzen, sterilen Sex. Er fragt sie, ob sie etwas mit ihm essen möchte, doch sie lässt ihn abblitzen. |
| 3 | 00:07:19 | Tagsüber fährt Hank Heim, wo sein Vater Buck eine Zeitungsanzeige über den Gefangenen Lawrence Musgrove herausschneidet und in ein Buch klebt. |

*Figure 8: Example of Film Analysis Sheet (Sayed 2019, p.313)*

Creating a film analysis sheet with GenAI that goes through an entire filmography is an additional necessary step that is allowed only thanks to digital tools and methods, usually

not possible in traditional film analysis. It will also be useful in this case to extract textual data from the protest signs, automatically generate descriptions and captions and recognize symbology.

As Smits and Wevers (2023) argue in their paper on the multimodal turn in DH, images and text are not separate entities, and they highlight that "we need text in we want to understand what images mean" (p.1268). After analyzing each movie and having the results of the film analysis sheet at hand, I conducted the textual analysis. For this purpose, I used the libraries Spacy[27] and NLTK[28]. They are both advanced Natural Language Processing libraries in Python, designed for tasks like tokenization, part-of-speech tagging, named entity recognition, and dependency parsing. I used the libraries to generate a clean, tokenized representation suitable for the analysis. Following the tokenization, I aggregated all the extracted words and then filtered out stop words and punctuation. This allowed me to calculate the frequency of the remaining words and visualize their prominence using a word cloud to highlight the most frequent terms. Moreover, I used the same techniques to analyze the protest signs and visualize the word frequency on a word cloud as well. Word frequency analysis is a valuable tool for analyzing film shot descriptions as it allows for the identification of the "aboutness" (Archer, 2016) of the text. By generating frequency lists, I am able to highlight the most salient words within the descriptions, pointing to the key elements that are most commonly mentioned which reveal information about style, themes and patterns in Hubbard's films.

[27] https://spacy.io/
[28] https://www.nltk.org/

# 4. Analysis

In this chapter I will outline the steps of the computational film analysis. First, I will demonstrate how each one of the films is programmatically analyzed with Python and then I will introduce the organization and arrangement of the results for the final interpretation of the filmography as a whole. The film *Elegy in the Streets (1989)* will serve as a muster example in the entire section. This film was chosen since it is the one where Hubbard put a lot of creative effort and various stylistic choices, and therefore it can offer strong end results and examples to showcase. The Jupyter Notebooks for the film analysis of *Elegy in the Streets* and for the final data exploration are available on the master's thesis GitHub repository[29], together with the results from all the films.

## 4.1 Steps of the computational film analysis

To quickly summarize from the previous chapter, I collected the video data directly from Jim Hubbard's Vimeo channel and included all of his activist video filmography. I created a dataset with the film titles, duration, video quality and film material. Conducting the analysis allowed me to also collect the total number of frames, the frames per second (fps), the number of shots, the average shot length (ASL), and the correlation value between movement and brightness (Spearman's p).

### 4.1.1 Preparation

I conducted the analysis with Python on Google Colab. I chose Google Colab for its accessibility and ease of use, which provides a pre-configured Jupyter Notebook environment accessible through a web browser without requiring any local installation. Furthermore, Colab's providing of free GPU resources significantly accelerated the processing of the videos and the use of the various CV and AI models, making computationally intensive tasks feasible within the scope of my research. The platform's integration with Google Drive and GitHub also streamlined file management and ensured the secure storage of code and data, making Python on Google Colab a practical and efficient choice.

I start the analysis by uploading the video file. First, the Distant Viewing Toolkit (dvt) needs to be downloaded, installed and imported directly from the Distant Viewing Lab's GitHub[30]. I imported other necessary libraries such as numpy, pandas and matplotlib for data

[29] https://github.com/manuelm91/dh-master-thesis
[30] https://github.com/distant-viewing/dvt

analysis and visualization. Additionally, the computer vision libraries needed to work on the video files and shots' images were also imported (OpenCV as cv2, PyTorch as torch and moviepy). I then loaded the video file path in a variable called "video" and stored it for the analysis. I began by extracting information directly from the file metadata. From these results it can be observed, for example, that the film *Elegy in the Streets* has 39.366 frames, its pixels are 720 in height and 960 in width (which serves as the indicator of the video quality, usually referred in this case as 720p), and is reproduced 24 frames per seconds (fps). One thing to keep in mind is that this metadata of course refers to the digital copy obtained from the Vimeo platform and not the original analog copy and neither the digitized videotape in the NYPL archive. However, it still remains a useful reference for the analysis, and it offers a representative snapshot of the films' visual characteristics as they appear in its current, publicly accessible form.

### 4.1.2 Brightness

The dvt offers various functions to work directly with video files. One example is the yeld_video function, which goes through each of the frames individually and extracts brightness information. In *Elegy in the Streets,* it identifies all of the number of frames and the brightness levels, together with time in milliseconds, as seen in Figure 9.

| | frame | time | brightness |
|---|---|---|---|
| **0** | 0 | 0.000000 | 90.005419 |
| **1** | 1 | 0.041710 | 89.909845 |
| **2** | 2 | 0.083420 | 89.125366 |
| **3** | 3 | 0.125130 | 89.164228 |
| **4** | 4 | 0.166840 | 89.662299 |
| **...** | ... | ... | ... |
| **39347** | 39347 | 1641.168049 | 0.044057 |
| **39348** | 39348 | 1641.209760 | 0.044057 |
| **39349** | 39349 | 1641.251470 | 0.044057 |
| **39350** | 39350 | 1641.293180 | 0.044057 |
| **39351** | 39351 | 1641.334890 | 0.044057 |

39352 rows × 3 columns

*Figure 9: Elegy in the Streets brightness information*

These values can be plotted to visualize the different jumps in brightness values. This indicates not only the frequency of the editing cuts but also how much action happening on screen. In

*Elegy in the Streets,* we can observe a dynamic visual composition (Figure 10). This suggests quick scene changes, intense visual contrasts, flickering effects and inconsistent steady state.

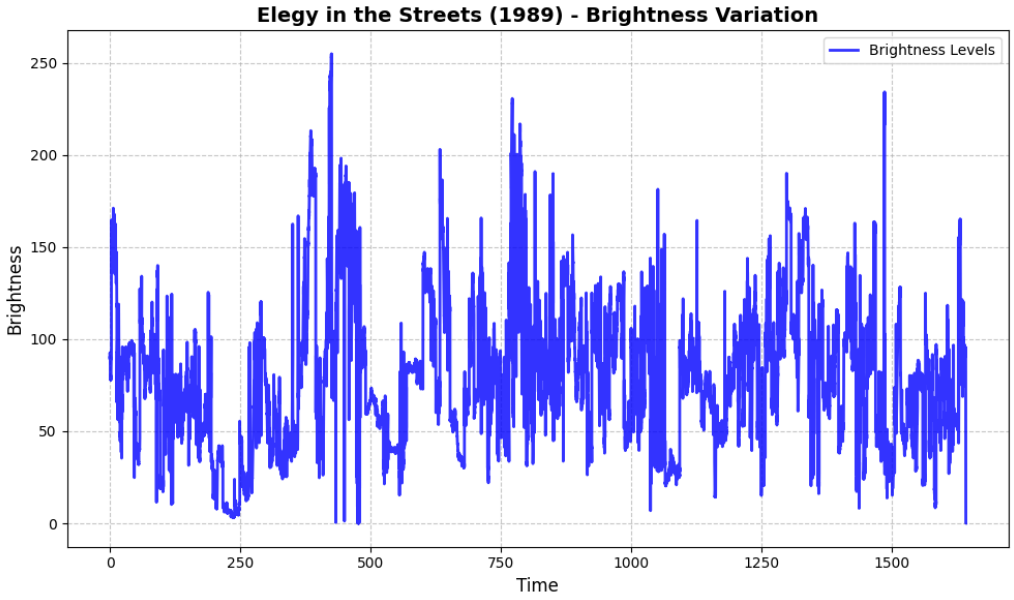



*Figure 10: Elegy in the Streets brightness levels*

### 4.1.3 Shots and Sequences

As Tilton and Arnold also suggest in their notebook[31], we cannot completely rely on brightness values to identify cuts and sequences. The dvt includes a neural-network-based shot boundary detection algorithm, called the AnnoShotBreak. Its function works directly on the video file and returns a dictionary object with the predicted boundaries. The algorithm is trained on narrative mainstream movies and series, and in some cases, it encountered difficulties in identifying exact sequences, since sudden change of brightness or colors can be mistakenly identified as a video cut. After processing each videoframe the data can be converted in a tabular dataset with runtime information such as the start and the end of the sequence. With these values it is now possible to calculate the average shot length (ASL) of the film. The ASL is a key statistic in film style, commonly used to indicate the pace of the editing[32]. Typically, a low ASL indicates a fast-editing style, while a high ASL reflects a slower cutting rate. The function calculates the shot duration by subtracting the start time from the end time. Then

[31] https://github.com/distant-viewing/dvt/blob/main/tutorials/Distant_Viewing_Tutorial_2_Network_Era_Sitcoms_and_Visual_Style.ipynb

[32] https://cinemetrics.uchicago.edu/article/2732e556-9e15-4d0b-b864-e08625fb6acf

statistical calculations process the mean, which will be the average shot length. It also finds the standard deviation and the middle value (Figure 11). The results for *Elegy in the Streets* indicate a fast-cutting rate.


*Figure 11: Elegy in the Streets Average Shot Length*

### 4.1.4 Movement

The dvt does not include a function to detect movement so I turned to the library OpenCV. Whit OpenCV I was able to define a function that analyzes the background by applying a mask and identify the changes in order to detect movement. It is also possible to remove noise, which comes in handy with analog films like these that were never restored. What the mask does is calculating the sum of non-zero pixels, which means pixels that have a value above zero. After this, the levels of color and light are summed up. This technique provides a straightforward metric for tracking visual motion, where the sum of non-zero pixels directly correlates with the intensity and extent of movement within a visual sequence. The results can be plotted and in *Elegy in the Streets* we can observe a highly dynamic and irregular movement patters, intense motion and periods of low movement intersected with intense motion (Figure 12). This suggests turmoil and a disjointed visual pattern.

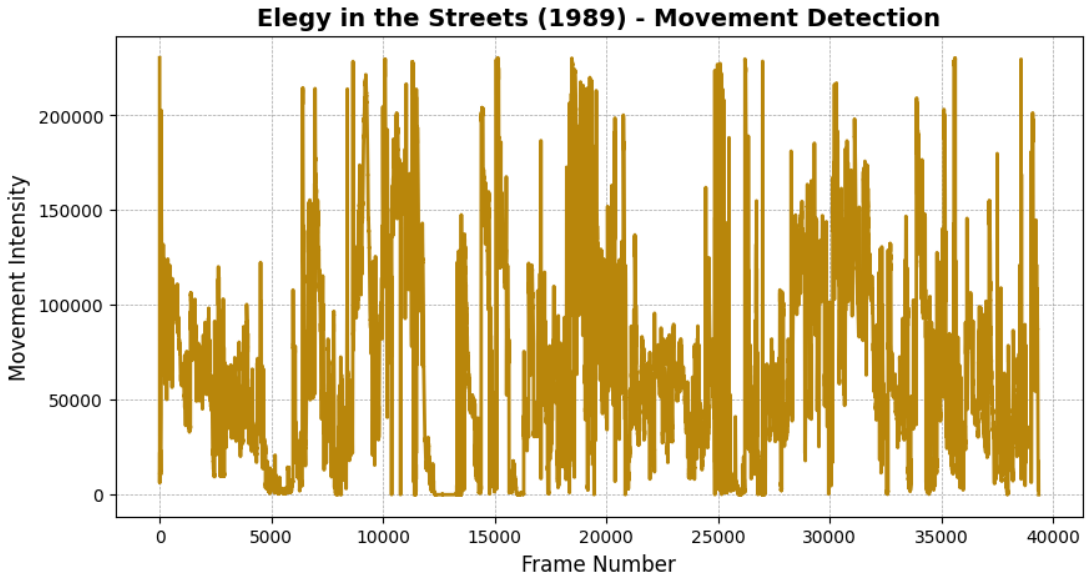



*Figure 12: Elegy in the Street movement intensitiy*

By comparing the visualizations of Figure 10 and Figure 12 it appears that movement and brightness could be correlated. It is possible to further investigate this with a statistical correlation measure such as Spearman's p. Spearman's p can measure monotonic relationships between variables when the data are not normally distributed (Dodge 2008, 502). In *Elegy in the Streets* Spearman's results is 0.79. A result above 0 corresponds to a positive Spearman correlation. This tells us that as one variable increases, brightness, the other variable tends to increase as well, movement, meaning there's a positive monotonic relationship between the two variables. In this case where the movie has more than 30.000 shots, the correlation scatter plot resulted in a dense and chaotic visualization (as it can be seen in the respective film notebook). For this reason, I decided to find and use two alternatives that visualize the correlation clearer: the Hexbin Plot and the Kernel Density Estimation (KDE)[33]. With the Hexbin plot it is possible to easily show the spot where the data are more condensed. Additionally, the KDE offers a smoother and rounded visualization. The visualizations in Figure 13 confirm the monotonic relationship, with the highest density in the middle range. It also appears that the most frames have moderate brightness and movement. This correlation can have various reasons. In a film like *Elegy in the Streets* that juxtaposes ACT UP protests to more intimate moments of Hubbard's relationship with Roger Jacoby[34], the main reasons lie in the camera exposure during the protests, as more movement requires more brightness to avoid motion-blur, and in the tint of the color purple used in low-lit intimate scene with less movement.

[33] https://dev.to/zmbailey/concentration-check-kdes-and-hexbins-1cn0#
[34] https://www.jimhubbardfilms.com/films/elegy-in-the-streets

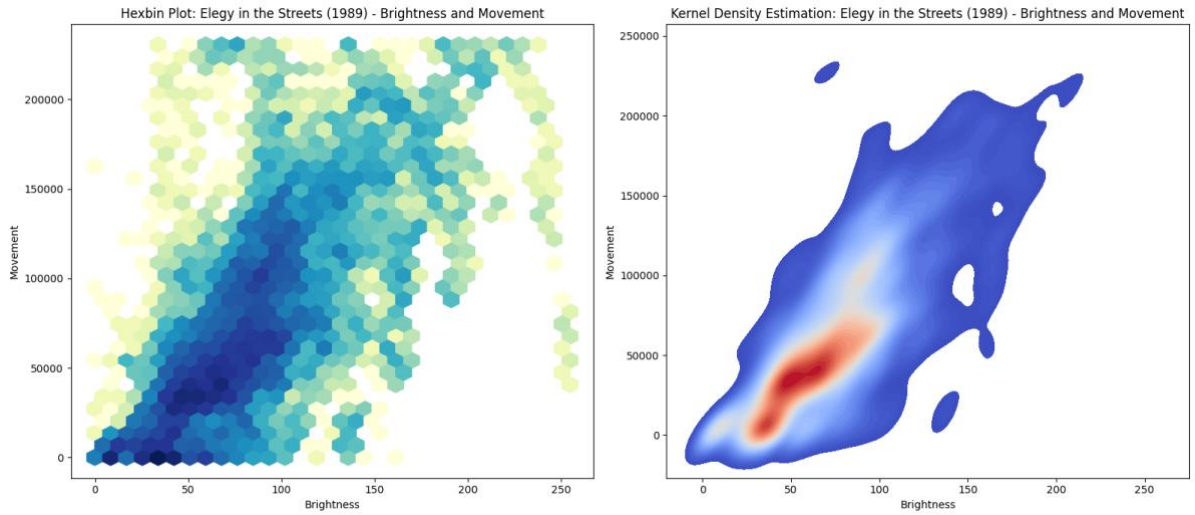



*Figure 13: Elegy in the Street Hexbin and KDE plots for correlation*

### 4.1.5 Color Analysis

To analyze color, I went back to the data frame created with the dvt's AnnoShotBreak function from which was then possible to extract one middle frame from each of the sequences, together with a timestamp. The sequences are formed by the frames between the shot boundaries. This gives the possibility to visualize the movie on a grid and observe the change of colors. What the OpenCV function does is, first of all, maintain the aspect ratio stable, given the experimental nature of the film, and after identifying each middle frame it applies the timestamp directly on the image. These frames can now be saved, and it is possible to work on them to facilitate the analysis, and the video is not needed anymore at this stage. Figure 14, on the next page, is an example of the grid from *Elegy in the Streets*.

*Figure 14: Elegy in the Streets shot grid*

I then follow the analysis with the calculation of the dominant color by applying a function from OpenCV that averages the colors of all the images in the folder. As seen in Figure 15 the result appears to be dark purple.


*Figure 15: Elegy in the Streets dominant color*

From the image stack it is also possible to visualize a color palette, to provide a quick summary of the film's overall color scheme from "above". For this, I proceeded to calculate the dominant color for each of the frames. I then proceeded to plot it, and for *Elegy in the Street* in Figure 16 the results can be seen, with a palette that goes from low saturated colors to more bright colors with the presence of purples, browns, reds, and occasional blues, to then come back to low saturation.


*Figure 16: Elegy in the Streets color palette*

### 4.1.6 Analysis Sheet with GenAI

As previously mentioned, one final step typical of film analysis is to create a film analysis log sheets with the shots chosen for the analysis, usually accompanied by a timestamp and the shot type.

For identifying the shot type I loaded the film shot classifier model from Hugging Face. To do so I imported transformers from pipeline and chose the model. The model iterated through each shot in the folder and classified it (Figure 17).

```
[ ]  from transformers import pipeline
     from PIL import Image
     import requests

[ ]  pipe = pipeline(
         "image-classification",
         model="pszemraj/beit-large-patch16-512-film-shot-classifier",
     )

     folder_path = "/content/shots"
     results = []

     for filename in os.listdir(folder_path):
         if filename.lower().endswith((".jpg",)):
             image_path = os.path.join(folder_path, filename)
             image = Image.open(image_path)
             prediction = pipe(image)[0]
             results.append({"shot": filename, "shot_type": prediction["label"]})


     df = pd.DataFrame(results)
```

*Figure 17: Film shot classifier*

For *Elegy in the Streets* it gave the results in Figure 18.

| | shot | shot_type |
|---|---|---|
| 0 | shot_0188.jpg | mediumShot |
| 1 | shot_0162.jpg | mediumShot |
| 2 | shot_0034.jpg | longShot |
| 3 | shot_0029.jpg | mediumShot |
| 4 | shot_0184.jpg | ambiguous |
| .. | ... | ... |
| 219 | shot_0053.jpg | fullShot |
| 220 | shot_0090.jpg | ambiguous |
| 221 | shot_0189.jpg | mediumShot |
| 222 | shot_0139.jpg | ambiguous |
| 223 | shot_0202.jpg | mediumCloseUp |

[224 rows x 2 columns]

*Figure 18: Film shot classifier results for Elegy in the Streets*

Furthermore, in order to accomplish the task of creating log sheets for the entire filmography, I integrated Generative Artificial Intelligence (GenAI) into my analysis flow. As previously

mentioned, I have chosen Google Gemini 1.5 pro[35] as it appeared to be the best image understanding model from my research purposes, also in terms of quality/cost range. Since the description of shots in film analysis is usually an objective and straightforward description of what it can be seen, ignoring author intentions and paratexts, I decided to apply zero-shot prompting, which means that the prompt used does not include examples or demonstration of the desired output[36]. I used the prompt "describe what is visible in the frame in one sentence. Use simple, objective language". It is possible to call the model programmatically in the Notebook through the Google GenAI API. Using GenAI also has the advantage of not running an object identification algorithm (such as YOLO) that takes too much time and computer power. The defined function (seen in Figure 19) goes through each image in the folder, creates the description through Gemini 1.5 pro and then saves them in a table as a .csv file.

```
[ ] folder_path = "/content/shots/"

    image_files = [os.path.join(folder_path, f) for f in os.listdir(folder_path) if f.lower().endswith(('.jpg',))]

    #function to generate descriptions
    def generate_description(image_path):
        try:
            with open(image_path, "rb") as img_file:
                image_data = img_file.read()

            model = genai.GenerativeModel("gemini-1.5-pro")
            response = model.generate_content([{"mime_type": "image/jpeg", "data": image_data}, "Describe what is visible in the frame in one sentence. Use simple, objective language. "])

            return response.text.strip() if response and response.text else "No description available"
        except Exception as e:
            return f"Error: {str(e)}"

    #creating the table
    data = {"shot": [os.path.basename(f) for f in image_files],
            "description": [generate_description(f) for f in image_files]}
    df = pd.DataFrame(data)
```

*Figure 19: Gemini 1.5 pro for film analysis sheet*

In Figure 20 is an excerpt of the analysis sheet for *Elegy in the Street.*

| shot | shot_type | description |
|---|---|---|
| shot_0000.jpg | mediumShot | A person in a white coat writes on the back of a person lying face down. |
| shot_0001.jpg | ambiguous | A bunch of white balloons floats in front of a building. |
| shot_0002.jpg | closeUp | A person with bare shoulders holds a bouquet of flowers close to their face. |
| shot_0003.jpg | mediumShot | A group of people march in a street, carrying banners and balloons. |
| shot_0004.jpg | fullShot | A group of people march in a parade, some holding a banner reading "THE LESBIAN & G BIG APPLE COR". |
| shot_0005.jpg | longShot | A blurry, low-angle shot shows the legs and feet of people walking in a line. |
| shot_0006.jpg | mediumShot | A blurry image shows a group of people, some holding signs, one of which reads "AGE 31". |
| shot_0007.jpg | mediumShot | A blurry image shows a group of people, some of whom hold small paper cups attached to strings. |
| shot_0008.jpg | mediumCloseUp | A person wearing a plaid shirt holds a small, lit candle in a paper cup. |
| shot_0009.jpg | closeUp | A close-up, dimly lit image shows part of a person's face with their eyes closed. |
| shot_0010.jpg | ambiguous | The video shows blurry, dark footage of what appears to be mold or fungus. |
| shot_0011.jpg | mediumCloseUp | A blurry, grainy, black and white image appears to show a piece of camera equipment in the lower left corner. |
| shot_0012.jpg | detail | Close-up of purple tulips. |

*Figure 20: excerpt from Elegy in the Street film analysis sheet*

[35] https://cloud.google.com/vertex-ai/generative-ai/docs/multimodal/image-understanding
[36] https://www.promptingguide.ai/techniques/zeroshot

## 4.2 Analysis of the Results

After all the analysis were done and the results collected, I was able to compare them, find common patters or differences, and give an interpretation.

I started with some simple exploratory data analysis, for example finding the longest movie by frame or visualizing all of the color palettes together to look at color from "above". Then I proceeded to conduct some statistical analysis such as looking at the ASL distribution of the whole corpus and exploring more in depth the correlation between brightness and movement.

Lastly, I conducted a textual analysis of the descriptions to give an interpretation of the shots and find common themes by looking at the most frequent words.
The result of this overall analysis will be presented and discussed in the next chapter.

## 4.3 Challenges

Some challenges and difficulties were encountered during the analysis. One of them is that the dvt AnnoBreak function is not always able to identify editing cuts. For example, in the film *ACT UP Kiss-In* (1988) the model is able to find only 4 shots, which when looking back at the movie it is clearly not the case. This may have different causes such as the amount of grain and dirt on the picture, which makes it difficult for the algorithm to "see" where the scene actually changes. In *ACT UP Kiss-In* Jim Hubbard intermittently cuts through different crowd, but being dark at night it is difficult for the model to differentiate between a cut or a panoramic. As seen in the shot grid in Figure 21 it can also already be difficult to identify scenes and details with the naked human eye.

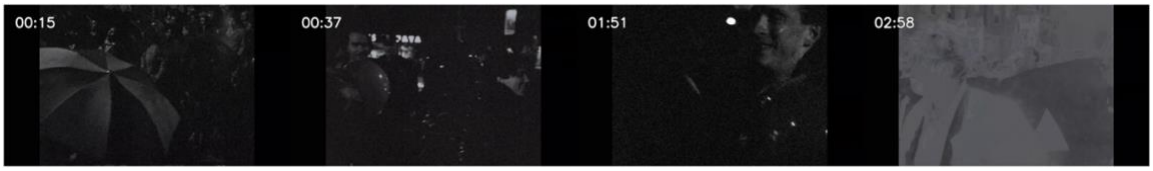



*Figure 21: ACT UP Kiss-In (1988) stills*

This is an issue that was noted also by Arnold and Tilton themselves together with Tim van der Heijden (2024) in their analysis of the Dutch Amateur Film Platform, where fade-in and fade-out were counted as cuts by the shot boundary algorithm. However, it can be argued that for an analysis of an experimental film this is exactly what we want. Since every color choice, burst of light, fades and all the changes are an active choice of the author, they need to be taken into

consideration individually. Differently to the analysis of a mainstream movie in traditional film analysis we do not need to divide the film based on when a scene thematically ends. Moreover, in film analysis when short sequences are analyzed, only one shot gets chosen for the final log sheet description. In my case they are all needed to highlight the stylistic choices of the author. In fact, despite these problems the results can be defined as satisfactory and will be discussed in the next chapter.

# 5. Results

In this chapter I will present the overall results of the films analysis and outline the investigation of the data. I start with an exploratory data analysis, followed by descriptive statistics and lastly a textual analysis. I will also connect these findings on the broader context of film analysis, queer temporality and materiality, and AIDS activism.

## 5.1 Exploratory Data Analysis

I started the exploratory data analysis by first looking at the dataset itself. The final dataset contains the films' titles, the year, the resolution, format, duration, fps, frames, shot, ASL, Spearman correlation of brightness and movement and the film material. This can be seen in Figure 22.

| Title | Year | Resolution | Format | Duration | fps | Frames | Shots | ASL | Spearman Correlation of Brightness and Movement | Filmed in |
|---|---|---|---|---|---|---|---|---|---|---|
| ACT UP Albany Action | 1988 | 720p | mp4 | 11m 55s | 24 | 17225 | 63 | 11.12 | 0.77 | 16mm |
| ACT UP Kiss-In | 1988 | 720p | mp4 | 02m 59s | 24 | 4331 | 4 | 36.02 | 0.34 | 16mm |
| Elegy in the Streets | 1989 | 720p | mp4 | 27m 21s | 24 | 39366 | 224 | 7.24 | 0.76 | 16mm |
| Gay and Lesbian Pride March 1985 | 1985 | 720p | mp4 | 17m 54s | 24 | 25813 | 96 | 11.04 | 0.52 | 16mm |
| Gay and Lesbian Pride March and Gay Pride Run ... | 1986 | 720p | mp4 | 08m 31s | 18 | 9012 | 39 | 12.76 | 0.17 | 16mm |
| Gay Pride Run 1983 | 1983 | 720p | mp4 | 21m49s | 18 | 23582 | 48 | 27.24 | 0.83 | Super 8 |
| Gay Pride Run 1985 | 1985 | 720p | mp4 | 09m 43s | 24 | 14001 | 28 | 20.70 | 0.33 | 16mm |
| International March for Lesbian and Gay Freedom | 1984 | 720p | mp4 | 11m 05s | 24 | 15994 | 69 | 9.34 | 0.72 | 16mm |
| June 12, 1982 | 1982 | 720p | mp4 | 08m 52s | 18 | 9619 | 27 | 18.99 | 0.54 | Super 8 |
| Lesbian & Gay Pride New York 1983 | 1983 | 720p | mp4 | 11m 46s | 18 | 12729 | 20 | 33.63 | 0.82 | Super 8 |
| Lesbian & Gay Pride New York 1987 | 1987 | 720p | mp4 | 10m 02s | 24 | 14454 | 28 | 20.72 | 0.52 | 16mm |
| Lesbian & Gay Pride New York 1988 | 1988 | 720p | mp4 | 08m 33s | 24 | 12344 | 48 | 10.69 | 0.81 | 16mm |
| March on Washington 1979 | 1979 | 720p | mp4 | 24m 29s | 18 | 26434 | 141 | 10.26 | 0.86 | Super 8 |
| May 21-22, 1979 | 1979 | 720p | mp4 | 11m 21s | 18 | 12300 | 25 | 26.19 | 0.61 | Super 8 |
| Stop the Movie (Cruising) | 1980 | 720p | mp4 | 11m 20s | 18 | 12286 | 64 | 10.38 | 0.87 | Super 8 |
| Two Marches | 1991 | 720p | mp4 | 09m 38s | 24 | 13868 | 91 | 6.30 | 0.52 | 16mm |
| Wall Street II | 1988 | 720p | mp4 | 04m 52s | 18 | 5280 | 25 | 11.19 | -0.27 | 16mm |

*Figure 22: Final film analysis dataset*

I then proceeded to create some visualization to explore the data further. In Figure 23-28 the results' plots are showed consecutively.

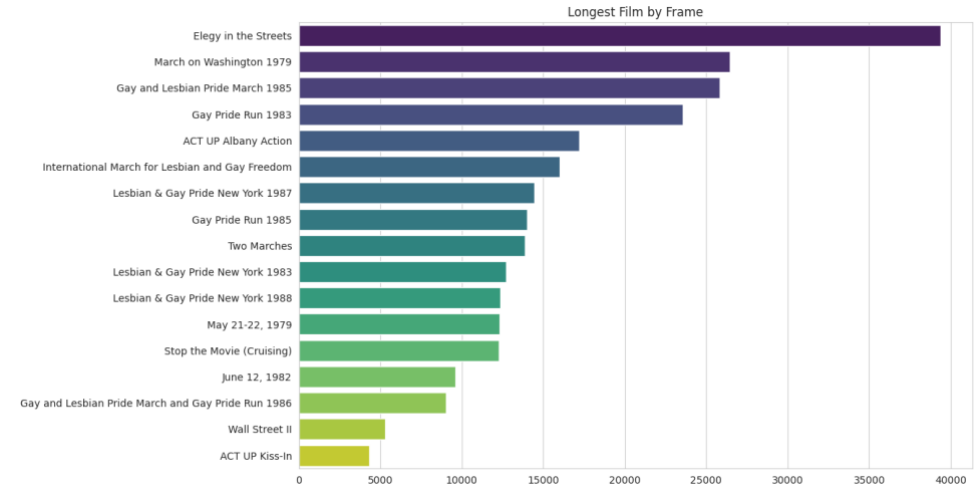



*Figure 23: Longest film by frame barplot*

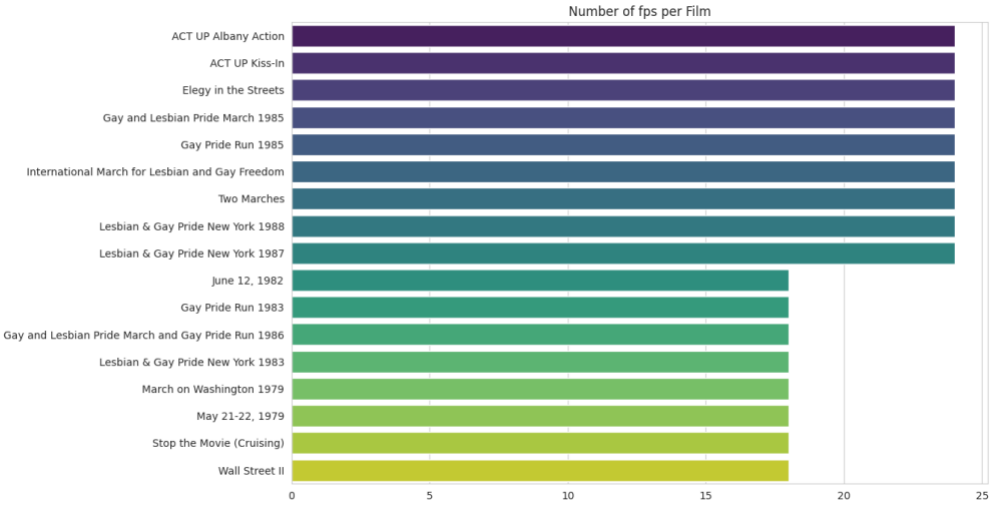



*Figure 24: Number of fps per film barplot*

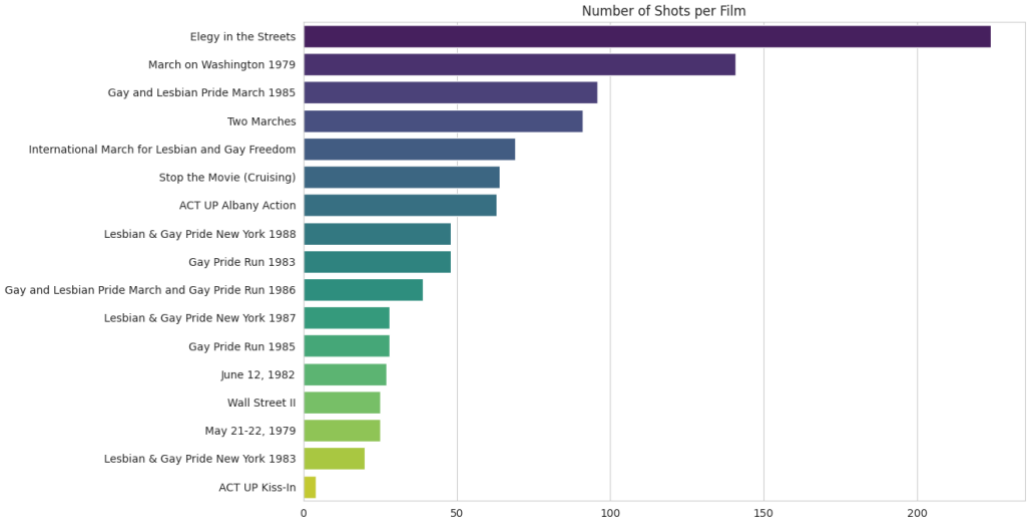



*Figure 25: Number of shots per film barplot*

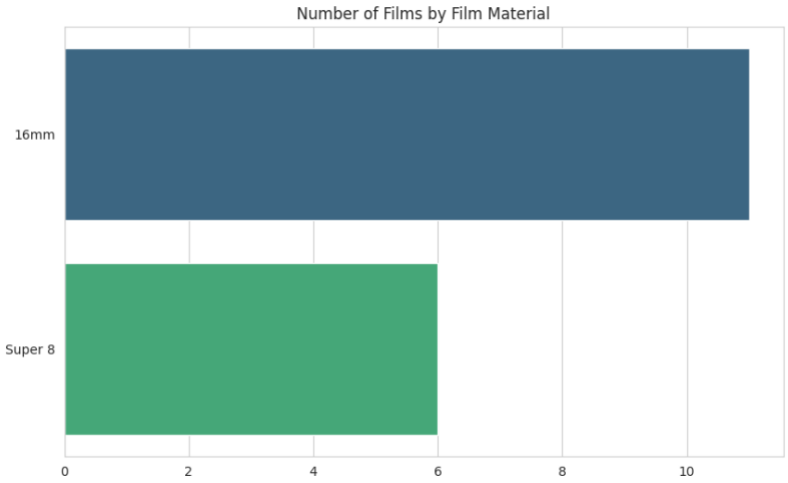



*Figure 26: Number of film by film material barplot*

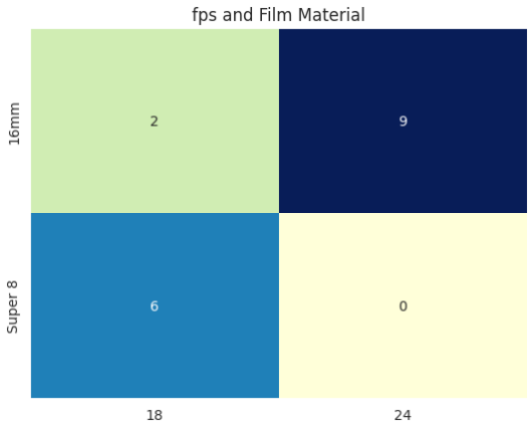



*Figure 27: fps and film material heatmap*

.

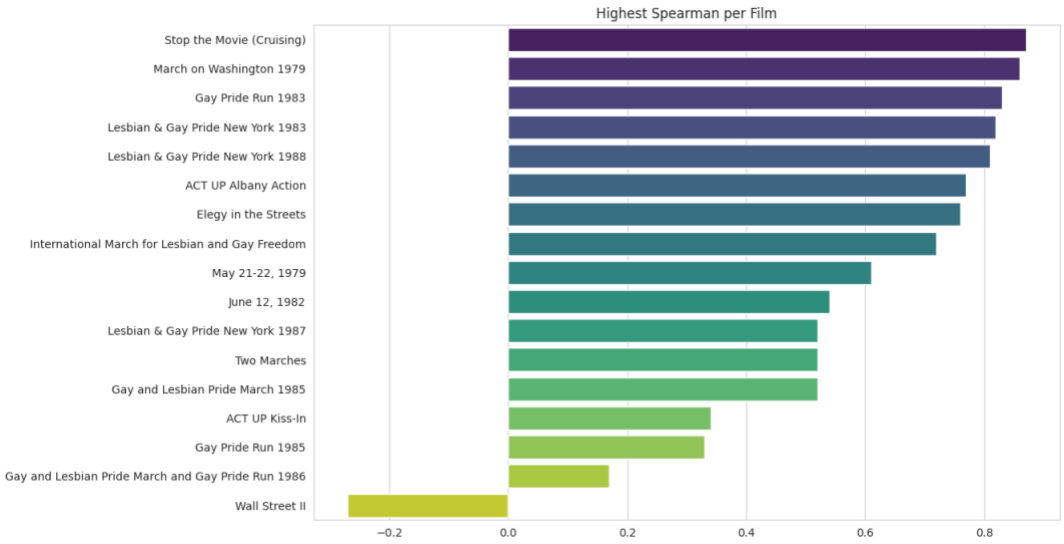



*Figure 28: Highest Spearman per film barplot*

Some general observations can be done from the visualizations. *Elegy in the Streets* is the film with the most frames, and therefore the longest, followed by *March on Washington 1979*. Moreover, the number of shots is also the highest in *Elegy in the Streets*. However, more shots do not necessarily mean a longer film. A film with a high number of shots might achieve its duration through a rapid succession of shorter scenes and perspectives. On the other hand, a film with fewer shots can be longer by employing extended takes, allowing scenes to unfold over a longer period without interruption. We can observe that, for example, *Two Marches* is at 4th place for the greatest number of shots, but 9th for the longest film. This mean that a film like *Two Marches* is composed of shots that, on average, are held on screen for a more extended duration. We can see this trend being in other films containing Pride Runs or New York Pride marches. From this and from the visualizations it can then be deducted that, except for a couple of exceptions, Hubbard used faster editing and pacing in films reporting ACT UP's actions and "zips", while for LGBTQ+ pride events he let the camera linger more and capture the moments. For the film *Lesbian and Gay Pride New York 1985* he writes in the film description on Vimeo "[…] there was a fluidity to the event and no clear distinction between observer and participant – everybody was part of the community."[37] (n.d.), and he was probably trying to convey this feeling of fluidity with longer takes. The faster cutting for ACT UP and protest actions focused film was then used to convey the urgency and energy of the protests and actions. Regarding the frames per seconds, it appears that the format most commonly used by Hubbard is 24 fps, followed by 18 fps. This is confirmed also by the fact that Hubbard's preferred medium appears to be 16mm analog film. In fact, this preference for specific materials generally also dictates the frame rate: 16mm is predominantly associated with 24fps in this corpus, while Super 8 with 18fps. However, there are two instances of 16mm films running at 18fps, potentially indicating a transfer from 24fps originals for technical and aesthetic reasons, such as achieving smoother motion during the filming of the protests and pride marches. This use of 16mm celluloid in a time where cheaper and more reliable technology such as camcorder were available, grounds the works in a distinct queer materiality with the grainy imagery. Juan Suárez defines granularity as representing "the old, the odd, and the queer" (2024, p.195), with the film material used by Hubbard being "both political and poetic, confrontational and experimental" (p.182). She also explains that this granularity represents queer vulnerability, since the medium is marked for eventual disappearance, given the precarious chemistry of celluloid. In this way, the chosen medium becomes inseparable from the message, being a testament to queer

[37] https://vimeo.com/309641596

resistance and memory, together with the experimental spirit of AIDS activist videos. Lastly, the Spearman correlation reveals a strong link between movement and brightness in *Stop the Movie (Cruising)*, likely due to its dynamic protest footage. While most films show a moderate to high correlation, *Wall Street II* exhibits a low relationship between brightness and movement.

I then followed the data exploration with an overall visualization of the colors, as seen in the following Figure 29 and Figure 30.

*Figure 29: Dominant colors*

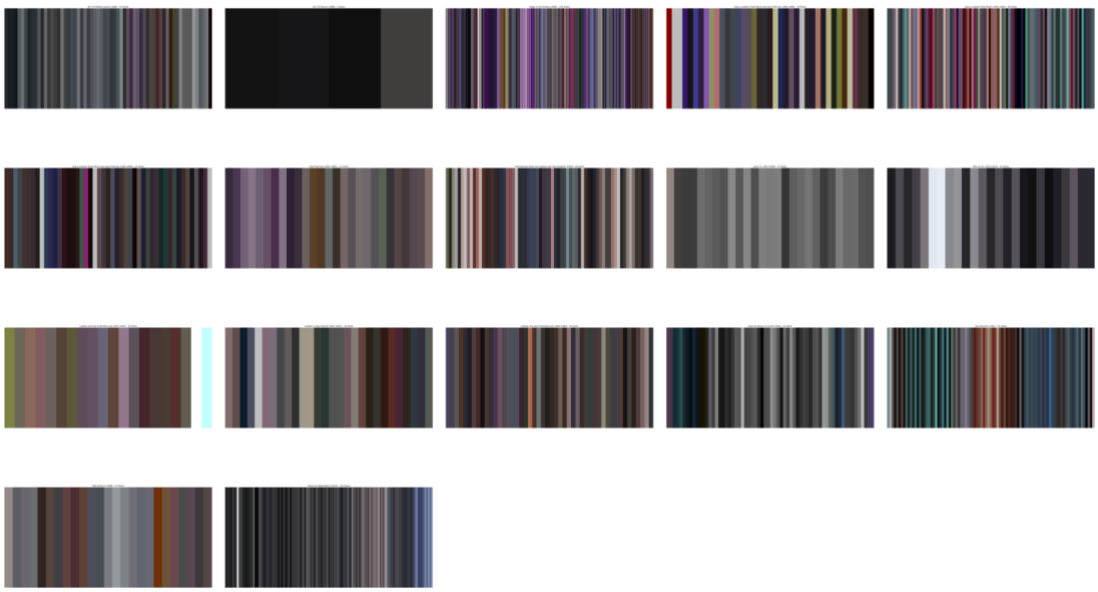


*Figure 30: Color palettes*

Looking only at the dominant colors they appear to be dark, muted, and desaturated. The lack of bright and vibrant color with a prevalence of dark greys, dark purples and blacks suggests a serious tone. This of course also reflects the visual aesthetic of documentary and archival footage from the film stock, which contribute to the desaturated look. The palettes largely confirm the observation from the dominant colors, where the overall feel is muted, desaturated, and often dark. Greys, blacks, dark blues, deep purples, and brownish tones are prominent across most of the films. However, a broader, though still limited range of hues can be observed. But even where brighter colors appear, they are usually part of a mix dominated by darker tones. This consistent characteristic of undersaturated and somewhat soft color are mostly due to the film material, which also contributes to documentary and realistic aesthetics. On the other hand, some films stand out for the high variety of color, for example for the video *Gay and Lesbian Pride March 1985* (the fifth palette from the left in Figure 30) Hubbard states in the Vimeo description that the film "wanted to explode into a rainbow of hues"[38] (n.d.). Differentiated hued color palette can also be achieved by the manual 16mm development, which allows for greater control over each stage, enabling intentional variations in temperature and chemical concentrations that can dramatically shift the final color rendition of the shots. Palettes dominated by grey strips (as in *June 12, 1982* and *ACT UP Kiss-In*) indicate black and white footage. All of the analog movies were hand-processed by Hubbard, this means that the brighter colors were added with the tinting and toning techniques typical of the silent film era.

## 5.2 Statistical Analysis

I then conducted some statistical exploration of the overall results. I started with the shot length distribution of all the films. In Figure 31 and 32 a density plot and a bar plot can be observed.

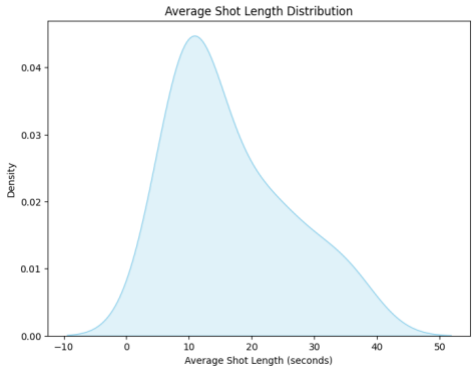



*Figure 31: Density plot ASL distribution*

[38] https://vimeo.com/394661685

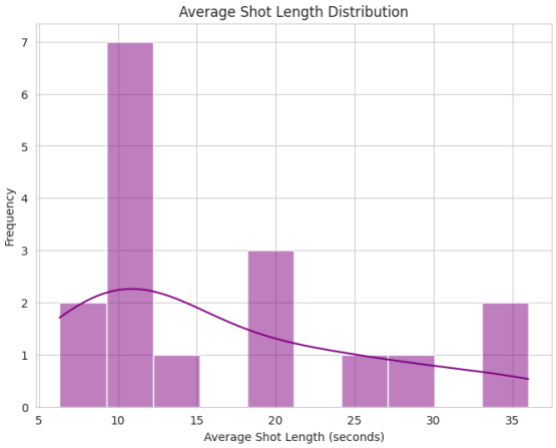
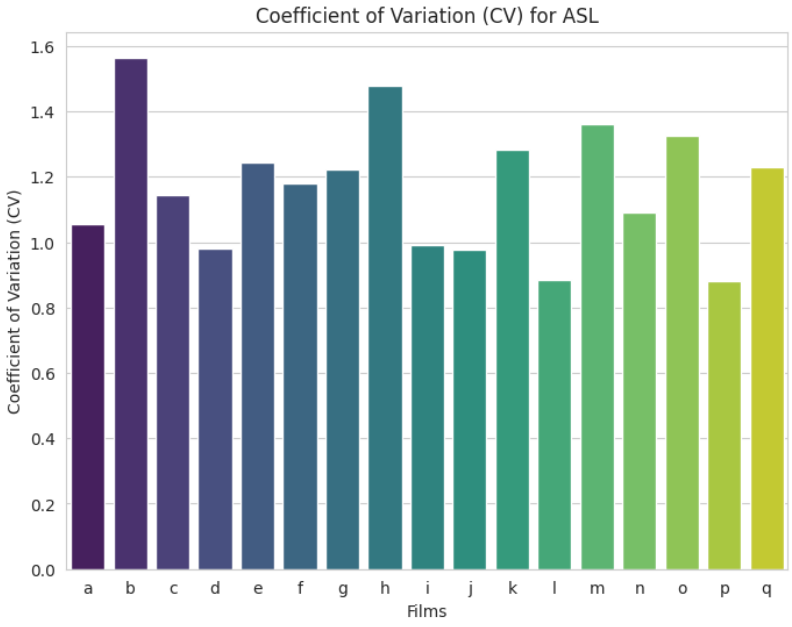



*Figure 32: Bar plot ASL distribution*

The density plot indicates that the average shot length of the overall filmography is around 10 to 11 seconds. Furthermore, the distribution is noticeably skewed, indicating the presence of a tail toward longer shot length. This means that occasional much longer average shots are present, but we know that films like *ACT UP Kiss-In* caused this skewedness because of the way they were interpreted by the Distant Viewing shot boundary algorithm. Looking at the bar plot we can further confirm that the average shot lengths are clustered around 10 to 11 seconds. The red line then further indicates that the distribution is positively skewed, meaning while shorter average shot lengths are most common, there are less frequent, but longer shots. From this, we can deduct that Hubbard tends to use very short shots, adding mostly dynamicity to his films.

To investigate further and also look at which movies are causing the skewness, and therefore use longer shots, I calculated the Coefficient of Variation, which quantifies the ratio of the standard deviation (Dodge 2008, 95). From the results (Figure 33) it can be seen that the films *ACT UP Kiss-In* (b), *International March for Lesbian and Gay Freedom* (h) and *March on Washington* (m) have the highest coefficient, confirming the shots in these films are more spread out compared to its average. Interestingly, among those only *ACT UP Kiss-In* is the one with very few shots (only four), and from this we can deduct that in this case the film deviates from the norm because of an error in processing the cuts, while for the others it means actually having longer shots.




| Alias | Film |
|---|---|
| a | ACT UP Albany Action |
| b | ACT UP Kiss-In |
| c | Elegy in the Streets |
| d | Gay and Lesbian Pride March 1985 |
| e | Gay and Lesbian Pride March and Gay Pride Run ... |
| f | Gay Pride Run 1983 |
| g | Gay Pride Run 1985 |
| h | International March for Lesbian and Gay Freedo... |
| i | June Twelve 1982 |
| j | Lesbian and Gay Pride New York 1983 |
| k | Lesbian and Gay Pride New York 1987 |
| l | Lesbian and Gay Pride New York 1988 |
| m | March on Washington |
| n | May 21-22 1979 |
| o | Stop the Movie (Cruising) |
| p | Two Marches |
| q | Wall Street II |

*Figure 33: Coefficient of variation with aliases*

Furthermore, I let Python analyze the Spearman correlation from the dataset and generate some descriptive statistics. In Figure 34 the results can be read.

```
1. Descriptive Statistics:
Count                16.000000
Mean                  0.524375
Median                0.520000
Standard Deviation    0.292050
Variance              0.085293
Minimum              -0.270000
Maximum               0.870000
Skewness             -1.311004
Kurtosis              2.447161
dtype: float64

2. Hypothesis Testing:
t-statistic                 7.182001
p-value                     0.000003
Significant at 0.05 level       True
dtype: object
```

*Figure 34: Descriptive statistics results*

Looking at the most important values, it can be observed that the standard deviation (0.29), tells us that the relationship between brightness and movement is generally positive, but it's stronger in some cases than in others. Moreover, the correlation is negatively skewed, that means skewed towards left. Also, the kurtosis being near the value 3 (but not above it) indicates that while there is a strong positive correlation, the values are spread out, and there are not many extremely high or extremely low correlations. In the hypothesis testing the “True" statement is the most crucial result. It indicates that the test found a statistically significant relationship between brightness and movement. The p-value results means that the probability of not being a correlation is <0.001. This means that it is so unlikely that it is possible to confidently reject the null hypothesis. The high t-statistic (7.182001) also supports this conclusion, indicating a greater difference between the correlation and the null hypothesis.
Furthermore, to interpret the correlation for each movie I visualized it on a bar plot as seen in figure 35.

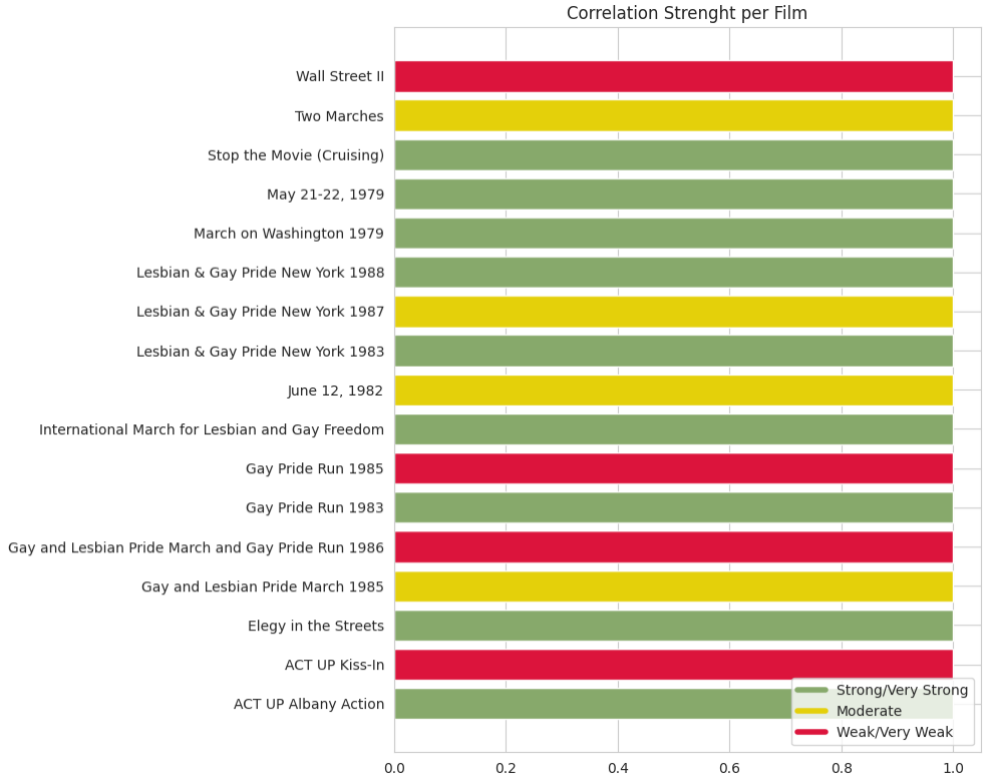



*Figure 35: Correlation strength per film*

We can observe that *Wall Street II*, *Gay Pride Run 1985*, *Gay Pride Run 1986* and *ACT UP Kiss-In* have the weakest correlations. These are also the films with the fewer shots (excluding *ACT UP Kiss-In*, they have all around 20 shots), suggesting once again that most of the films shot during pride events are less dynamic. Some other cases, as in the already mentioned

*Lesbian and Gay Pride New York 1983*, composed of only 20 shots but with a higher correlation, may also simply indicate a more crowded event or a higher use of close-ups.

Furthermore, the strong, statistically significant correlation found between brightness and movement in many of Hubbard's films, can also be interpreted as an experimental approach to capturing the chaos and rage of the scenes. Quick editing is not only a feature of experimental films, which have the goal to destabilize the classical Hollywood Style (Verrone 2012, p.20), but also of what Karl Schoonover and Rosalind Galt call queer cinematic time (2016, p.259). It consists of a warped non-synchronized time that reflect the displacement of the queer experience. Linearity and closure are instead part of "heterosynchrony" (p.268). Techniques like rapid intercutting and, as in *Elegy in the Streets*, the juxtaposition of past and present, contribute to queer cinematic time.

## 5.3 Textual Analysis

Before starting the textual analysis, I grouped all the film analysis sheets together in one data frame. What this allowed me to do  is to firstly look at the most used shot by Hubbard. The results in Figure 36 show medium shot as the most used, however the most frequent value is " ambiguous".

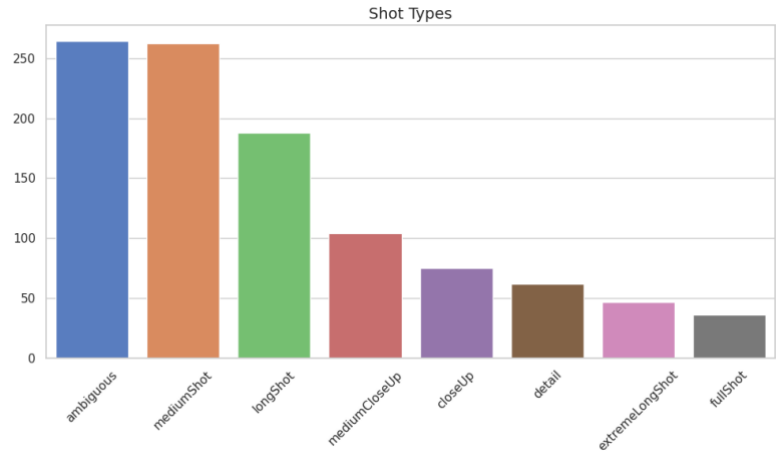



*Figure 36: Bar plot of shot types*

The fact that most shots were identified as ambiguous demonstrates the experimental nature of Hubbard's film. Of course, it might also be due to the fact that the algorithm, by choosing the middle frame of the detected shots, is, in the end, extracting a random image that may present

the classifier model with difficulties in identifying what it “sees”. I will show two different examples from *Elegy in the Street* that can explain this.

*Figure 37: shot 184 and 001 from Elegy in the Streets*

Both of these shots in Figure 37 were identified by the film shot classifier model as ambiguous shots. Shot number 184 is very difficult to identify also without a machine, this is due the experimental practices in the films. This can be a transition, a burst of light or Hubbard playing with colors during the hand processing. On the other hand, shot number 001 is clearly a shot of white balloons. The issue here is that the classifier model does not really understand who the subject of the shot is. This may lie on the way it was trained, mostly on mainstream films and probably using a human subject as the main reference. If we consider the balloons as the subjects this is clearly either a close-up shot, or a detail shot.

After analyzing the shot types I proceeded to the textual analysis with a word frequency analysis. I started by processing the text, first by cleaning it, removing stop words and transforming all into lower-case, then tokenizing it to make it more easily readable for the machine. With the list of tokenized words, I then calculated the frequency of words and created a word cloud. Looking at Figure 38 we can observe that “blurry”, “grainy” and “people” seem to be the most frequent words, followed by “person”, “white” and “black”. We can observe from this that the aesthetic of Hubbard’s film tends to a darker and grainy cinematography, given by the nature of the film stock and by the protests. The high frequency of "people" and "person" indicates a focus on human subjects, which aligns with the context of protests and actions. Furthermore, the significant presence of "white" and "black" points towards a potentially higher contrast and a notable use of monochrome and black-and-white shots in the films.

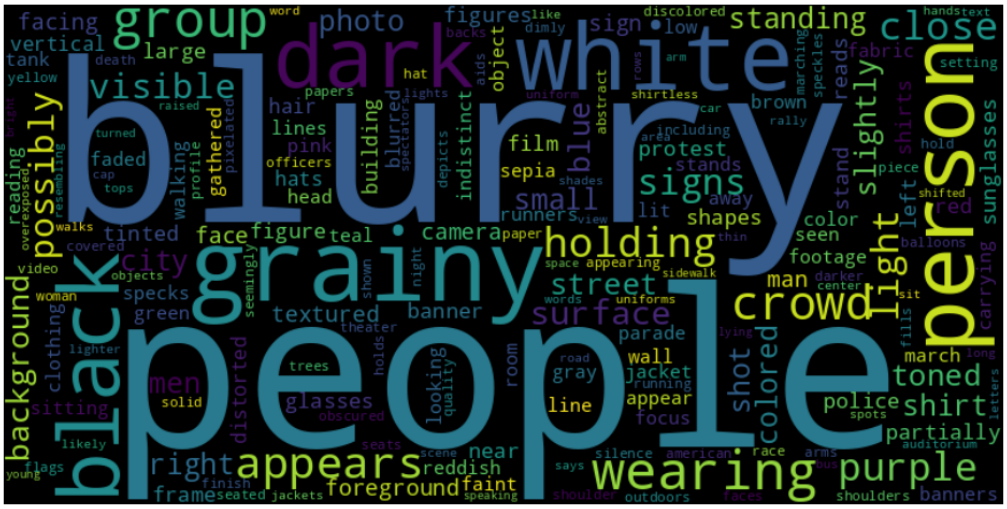



*Figure 38: Word cloud shot descriptions*

This is further confirmed by looking at the top 20 most frequent words as seen in Figure 39.

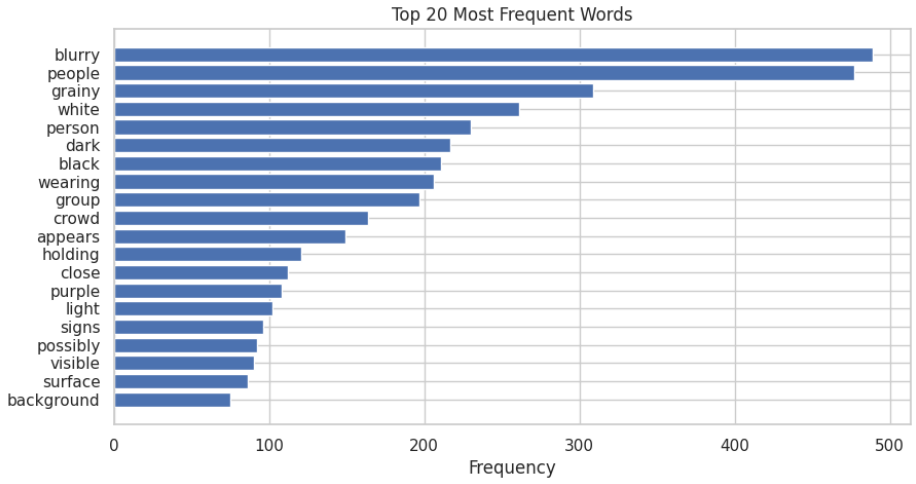



*Figure 39: Top 20 frequent words in shot descriptions*

Furthermore, since the protest signs were transcribed by Gemini 1.5 pro in all-caps, I was interested in looking at the most frequent words on the signs. This was possible by filtering the list and leave all the capitalized words. In Figure 40 we can observe that AIDS and SILENCE = DEATH are very frequent words.

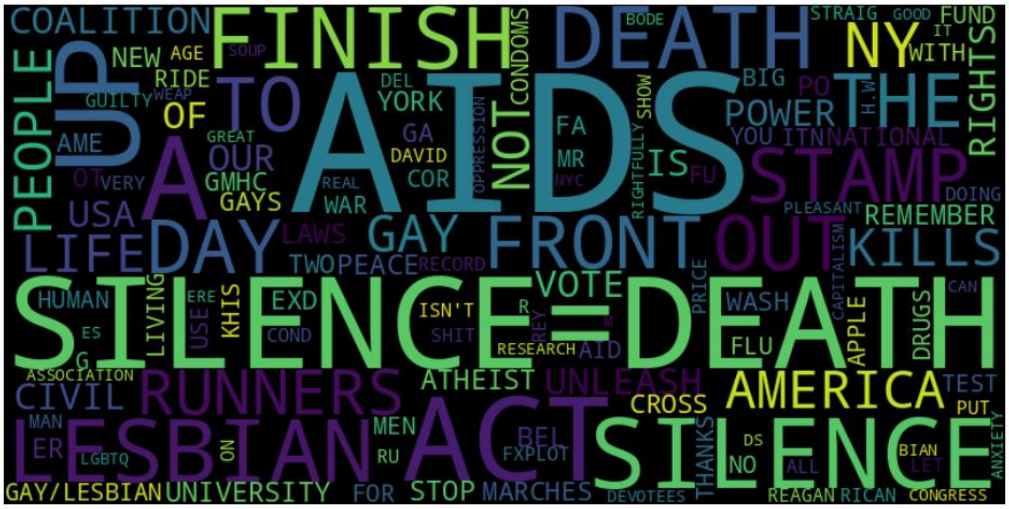



*Figure 40: Word cloud protest signs*

The strong presence of ACT UP's slogan and the word AIDS suggest that most of Hubbard's film corpus focused on ACT UP and the fight against AIDS. Other words such as "LESBIAN" and "GAY" indicates a direct engagement with the impact of AIDS on the LGBTQ+ community and the urgent need to break the silence surrounding the epidemic. Words like "KILLS", "FINISH" and "DEATH" indicate the use of strong words in the protest signs, reflecting the urgency, anger, and gravity of the AIDS crisis. The use of these words is further confirmed when looking at the list of top 10 most frequent words in Figure 41.

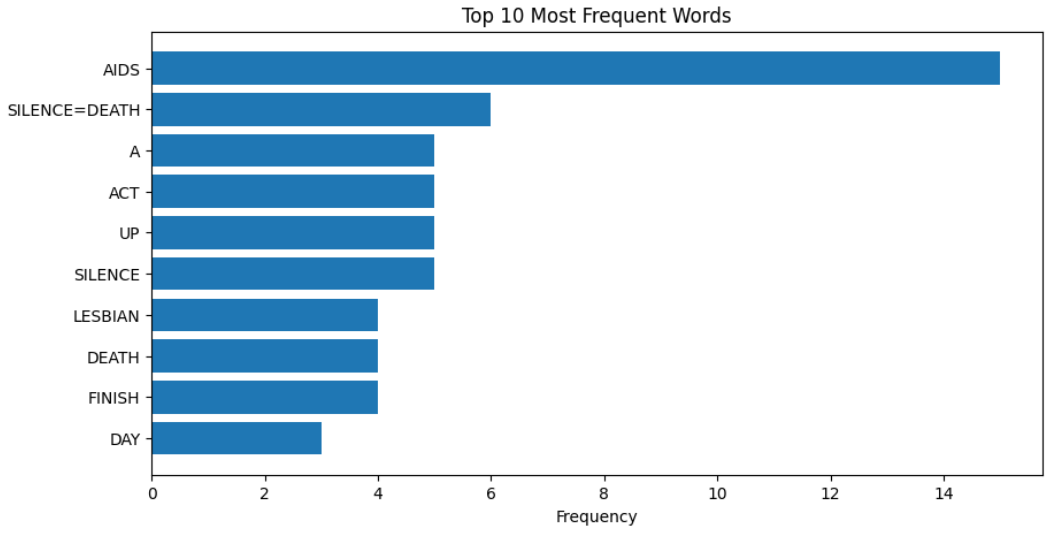



*Figure 41: Top 10 frequent words in protest signs*

## 5.4 Key takeaways

Looking at the main findings of the overall final analysis, we can see how the exploratory data analysis revealed some insights into the shot length and the use of film celluloid in Hubbard's activist videos. First of all, it was deducted that Hubbard uses longer shots in videos where he films LGTBQ+ pride parades and related events such as the pride runs. Meanwhile, shorter shots are used in the depiction of ACT UP disruptive direct actions, and in more chaotic protests such as "Stop the Movie". This was further confirmed by looking at the correlation between brightness and movement, which suggested a more dynamic editing in films with longer shots. Secondly, it was observed how Hubbard preferred 16mm celluloid film as medium, probably to convey a sense of realism, nostalgia and to have more room to experiment with lights and colors. This was also confirmed by the analysis of the colors, that revealed a higher use of unsaturated hues and muted colors.

A statistical analysis cleared out that some outliers skewed the average shot length calculation due to the erroneous identification of cuts and shots by the shot boundary algorithm. Moreover, the hypothesis testing on the correlation between brightness and movement confirmed that it was statistically significant.

Furthermore, looking at the film analysis sheets it was revealed that the medium shot was the shot most used by Hubbard. This seems in agreement with the filming of crowds and protest events. This framing enables viewers to connect with the participants while still understanding their place within the larger demonstration. Also, most of the shot were classified as ambiguous, revealing both a difficulty in classifying shots for experimental films and at the same time the limits of the film classifier model trained on mainstream cinema.

Lastly, the textual analysis looked at word frequency and it was demonstrated that the focus of Hubbard's films was on the thematic of ACT UP, fighting against AIDS, and the urgent need of the protests. The protest signs more frequent words revealed the activists' main concerns and strong demands, often using strong and angry language that reflected the gravity of the crisis and the frustration with inaction from the government and the ongoing societal discrimination.

The use of CV, AI and ML has demonstrated that these tools are very powerful, however still present some flaws, mostly given by the way they were trained. This reason shows how it is strongly needed to include more diversity in the data sets used for training. Firstly, stylistically with more practices that distance themselves from the mainstream production such

as experimental and ephemeral film. Secondly, by including images and videos from minorities group and histories from impactful moments that tend to be forgotten, such as the AIDS crisis.

On the other hand, using these tools also puts a mirror in front of traditional humanities inquiry. In the case of film analysis, it was demonstrated how an entire filmography can be analyzed digitally, including the use of generative artificial intelligence. This is in contraposition with the limitation of traditional analog film analysis, which tends to choose shorter scenes and analyze only certain moments of a film. This brings some biases to light, such as how are some scene chosen and how can they be representative of the whole film interpretation? Computational film analysis “democratizes” the analysis process, resulting in more nuanced and objective results.

# 6. Conclusion

In my master's thesis I followed a pipeline for computational multimodal film analysis. I integrated and adapted for my purposes the tools offered by the Distant Viewing Toolkit, which processed directly the videos and enabled me to analyze brightness and editing in Hubbard's film corpus and identify and collect the shots. This was followed by the processing and retrieval of the film's shots as images with computer vision libraries such as OpenCV. This allowed me to analyze the colors, creating a shot grid and a color palette. I then used a film shot classifier model from Hugging Face to identify the shots in the films. I worked with Google's GenAI model Gemini 1.5 pro to describe each shot and compile a film analysis sheet for each of the films with description and shot type. Once each film went through those processes, I collected the results and conducted an exploratory data analysis with visualizations, which revealed patterns and tendencies in Hubbard's use of 16mm film and the use of longer shot. Moreover, a statistical analysis calculated the correlation between brightness and movement which resulted being statistically significant. Lastly, a textual analysis with natural language processing on the films' descriptions and on the protest signs revealed common patterns by looking at the word's frequency.

This workflow demonstrates of how AI and ML can revolutionize the analysis of archival materials, particularly visual media like the AIDS activist videos. Traditional archival research often involves time-consuming manual processes of cataloging, describing, and analyzing content. By applying computational methods, large video corpora can be efficiently processed to extract meaningful insights. Looking at my research, the automated identification and analysis of brightness, editing patterns, color palettes, and shot types provided quantitative data that revealed underlying stylistic choices that can reflect the emotional and political contexts of the films. Furthermore, the use of the film shot classifier from Hugging Face and the descriptive capabilities of Google's Gemini 1.5 Pro highlight the potential for AI to automate the annotation and interpretation of visual content at scale. This not only saves significant time and resources for archivists but also opens up new avenues for research inquiry questions that were previously infeasible due to the sheer volume of data, especially for film analysis. AI can transform archives from repositories of static documents into dynamic and explorable resources, enabling researchers to uncover hidden patterns and connections.

Looking ahead, the integration of AI and ML techniques bring new exciting promises to further evolve the field of film analysis, by applying similar pipelines to analyze vast archives of historical footage, documentaries, or even user-generated video content. Future

developments in computer vision, natural language processing, and multimodal analysis may lead to even more sophisticated tools for understanding the complex interplay of visual, auditory, and textual elements in film, which would go even deeper into the sentiment and rhetorical strategies employed in the films' descriptions or within the texts in film themselves. Further research on the topic can also introduce new models trained specifically for experimental films or for communities' archives such as the AIDS activist video archive or the ACT UP Oral History Project. This model can enable researchers to better understand the specific visual vocabularies, emotional registers, and historical contexts embedded within these materials. This would also benefit the field of film analysis, where more specific and nuanced research can be conducted, leading more contextualized understandings of cinematic works across cultural contexts.

# Sources

## Primary Sources

Jim Hubbard film collection:

https://www.jimhubbardfilms.com/films

https://vimeo.com/unitedinanger

Film corpus:

ACT UP Albany Action (1988)

ACT UP Kiss-In (1988)

Elegy in the Streets (1989)

Gay and Lesbian Pride March 1985 (1985)

Gay and Lesbian Pride March and Pride Run 1986 (1986)

Gay Pride Run (1983)

Gay Pride Run (1985)

International March for Lesbian and Gay Freedom (1984)

June 12, 1982 (1982)

Lesbian and Gay Pride New York (1983)

Lesbian and Gay Pride New York (1987)

Lesbian and Gay Pride New York (1988)

March on Washington (1979)

May 21-22, 1979 (1979)

Stop the Movie (Cruising) (1979)

Two Marches (1991)

Wall Street II (1988)

GitHub repository:

https://github.com/manuelm91/dh-master-thesis

## Secondary Sources

Literature:


Aaron, M. (with Aaron, M., Benshoff, H. M., Contreras, D. T., Davis, G., DeAngelis, M., Jennings, R., Leung, H. H.-S., Lominé, L., Pearl, M. B., Pick, A., Pidduck, J., Rich, B. R., Smelik, A., & Wallenberg, L.). (2022). *New Queer Cinema: A Critical Reader*. Edinburgh University Press. https://doi.org/10.1515/9781474463768

Andrew, D. (1979). *The Major film theories: An introduction* (repr. ed). Oxford Univ. Press.

Archer, D. (2016). 5. Data Mining and Word Frequency Analysis. In G. Griffin & M. Hayler, *Research Methods for Reading Digital Data in the Digital Humanities* (pp. 72–92). Edinburgh University Press. https://doi.org/10.1515/9781474409629-006

Arnold, T., & Tilton, L. (2023). *Distant Viewing: Computational Exploration of Digital Images*. The MIT Press. https://doi.org/10.7551/mitpress/14046.001.0001

Bajohr, H. (2025). *Thinking with AI: Machine Learning in the Humanities*. Open Humanites Press. https://www.openhumanitiespress.org/books/titles/thinking-with-ai/

Benshoff, H. M., & Griffin, S. (Eds.). (2010). *Queer images: A history of gay and lesbian film in America*. Rowman & Littlefield Pub.

Bertetto, P. (2014). *Metodologie di analisi del film*. Laterza.

Bordwell, D., & Thompson, K. (2020). *Film art: An introduction* (Eleventh edition). McGraw-Hill Education.

Boyd Jason, Howe, James, & Ruberg, Bonnie. (n.d.). *"CHAPTER 8" in "Bodies of Information" | Debates in the Digital Humanities*. Retrieved May 2, 2025, from https://dhdebates.gc.cuny.edu/read/untitled-4e08b137-aec5-49a4-83c0-38258425f145/section/e1a72933-2aef-4158-b21c-e6fe35afc091

Brown, B. (2015). *Cinematography: Theory and practice: imagemaking for cinematographers and directors* (2nd ed). Elsevier/Focal Press.

Brunow, D. (2018). Archiving AIDS activist video: A conversation with Jim Hubbard. In *A Visual History of HIV/AIDS*. Routledge.

Casetti, F., & Di Chio, F. (1994). *Analisi del film* (15. ed). Bompiani.

Champion, E. M. (2016). Digital humanities is text heavy, visualization light, and simulation poor. *Digital Scholarship in the Humanities*, fqw053. https://doi.org/10.1093/llc/fqw053

Chun, W. H. K. (2018). *Queerying Homophily*. meson press. https://doi.org/10.25969/MEDIAREP/12350

Cifor, M. (2022). *Viral cultures: Activist archiving in the age of AIDS*. University of Minnesota Press.

Crimp, D. (1990). *AIDS demo graphics*. Bay Press.

Dodge, Y. (2008). *The concise encyclopedia of statistics* (1st. ed). Springer.

Ferguson, K. L. (2017). Digital Surrealism: Visualizing Walt Disney Animation Studios. *Digital Humanities Quarterly*, *11*(1). https://www.digitalhumanities.org/dhq/vol/11/1/000276/000276.html

Fossati, G. (2018). *From Grain to Pixel: The Archival Life of Film in Transition*. Amsterdam University Press. https://doi.org/10.2307/j.ctv8bt181

Gemini Team. (2023). *Gemini: A Family of Highly Capable Multimodal Models* (arXiv:2312.11805). arXiv. https://doi.org/10.48550/arXiv.2312.11805

Gerstner, D. A. (Ed.). (2011). *Routledge international encyclopedia of queer culture*. Routledge.

Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing*. Pearson.

Hallas, R., Hubbard, J., & Levine, D. (2010). Moving Pictures. *GLQ: A Journal of Lesbian and Gay Studies*, *16*(3), 429–449. https://doi.org/10.1215/10642684-2009-038

Heftberger, A. (2018). *Digital Humanities and Film Studies: Visualising Dziga Vertov's Work*. Springer International Publishing. https://doi.org/10.1007/978-3-030-02864-0

Heras, D. C. (2024). *Cinema and Machine Vision: Artificial Intelligence, Aesthetics and Spectatorship*. Edinburgh University Press. https://doi.org/10.1515/9781399514736

Heijden, T. van der, Arnold, T., & Tilton, L. (2024). Distant Viewing the Amateur Film Platform. In *Doing Digital Film History* (pp. 239–270). De Gruyter Oldenbourg. https://www.degruyterbrill.com/document/doi/10.1515/9783111082486-012/html

Jaillant, L. (Ed.). (2022). *Archives, access and artificial intelligence: Working with born-digital and digitised archival collections*. transcript.

Jaillant, Lise. (n.d.). *Introduction to the Special Issue: Using Visual AI Applied to Digital Archives*. https://www.digitalhumanities.org/dhq/vol/18/2/000752/000752.html

Johnson, J., Karpathy, A., & Fei-Fei, L. (2015). *DenseCap: Fully Convolutional Localization Networks for Dense Captioning* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1511.07571

Juhasz, A. (2020). *AIDS TV: Identity, Community, and Alternative Video*. Duke University Press. https://doi.org/10.1515/9780822396079

Kipp, M. (2001). ANVIL - a generic annotation tool for multimodal dialogue. *7th European Conference on Speech Communication and Technology (Eurospeech 2001)*, 1367–1370. https://doi.org/10.21437/Eurospeech.2001-354

MacKenzie, S. (2014). *Film manifestos and global cinema cultures: A critical anthology*. University of California Press.

Manovich, L. (2020). *Cultural analytics*. The MIT Press.

McCann, H., & Monaghan, W. (2022). *Queer theory now: From foundations to futures* (Reprinted by Bloomsbury Academic). Bloomsbury Academic.

Moretti, F. (2013). *Distant reading*. Verso.

Nockels, J., Gooding, P., & Terras, M. (2025). Are Digital Humanities platforms facilitating sufficient diversity in research? A study of the Transkribus Scholarship Programme. *Digital Scholarship in the Humanities*, *40*(Supplement_1), i46–i65. https://doi.org/10.1093/llc/fqae018

Redfern, Nick. (2022). *Computational Film Analysis with R* (Version v0.9.000) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.7074521

Rich, B. R. (2013). *New Queer Cinema: The Director's Cut*. Duke University Press. https://doi.org/10.2307/j.ctv11hpp0s

Ross, M. (with Grauer, M., & Freisleben, B.). (2015). *Digital Tools in Media Studies: Analysis and Research. An Overview* (1st ed). transcript.

Salt, B. (1983). *Film style and technology: History and analysis* (Third edition). Starword.

Sayed, L. (2019). *Weiße Helden im Film: Der »White Savior Complex« - Rassismus und Weißsein im US-Kino der 2000er Jahre*. transcript Verlag.

Schoonover, K., & Galt, R. (2016). *Queer cinema in the world*. Duke university press.

Schulman, S. (2021). *Let the record show: A political history of ACT UP New York, 1987-1993* (First edition). Farrar, Straus and Giroux.

Smits, T., & Wevers, M. (2019). The visual digital turn: Using neural networks to study historical images. *Digital Scholarship in the Humanities*. https://doi.org/10.1093/llc/fqy085

Smits, T., & Wevers, M. (2023). A multimodal turn in Digital Humanities. Using contrastive machine learning models to explore, enrich, and analyze digital visual historical collections. *Digital Scholarship in the Humanities*, *38*(3), 1267–1280. https://doi.org/10.1093/llc/fqad008

Suárez, J. A. (2024). *Experimental Film and Queer Materiality*. Oxford University Press.

Thompson, K. (1988). *Breaking the glass armor: Neoformalist film analysis*. Princeton University Press.

Thomson-Jones, K. (2008). Formalism. In *The Routledge Companion to Philosophy and Film*. Routledge.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need* (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

Verrone, W. (2011). *The avant-garde feature film: A critical history*. McFarland & Company, Inc., Publishers.

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). *A Survey of Human-in-the-loop for Machine Learning*. https://doi.org/10.48550/ARXIV.2108.00941

Online Resources:

*ACT UP Staged New Years Day "Die-in" at De Blasio Inauguration | ACT UP NY. (n.d.).* *https://actupny.com/act-up-staged-new-years-day-die-in-at-de-blasio-inauguration/*

*AIDS Activist Video and the Evolution of the Archive. (n.d.). Jim Hubbard.* *https://www.jimhubbardfilms.com/writing/aids-activist-video-and-the-evolution-of-the-archive*

*archives.nypl.org—AIDS Activist Videotape Collection. (n.d.).* *https://archives.nypl.org/mss/3622#controlaccess*

*BBC Four - Made by Machine: When AI Met the Archive. (n.d.). BBC.* *https://www.bbc.co.uk/programmes/b0bhwk3p*

*CINEMETRICS. (n.d.). https://cinemetrics.site/*

*Concentration Check: KDEs and Hexbins. (2019, April 28). DEV Community.* *https://dev.to/zmbailey/concentration-check-kdes-and-hexbins-1cn0*

*Distant Viewing Lab. (n.d.). https://distantviewing.org/*

*Diversity and Inclusivity | EADH - The European Association for Digital Humanities. (n.d.).* *https://eadh.org/about/diversity-and-inclusivity?utm_source=chatgpt.com*

*GitHub—Distant-viewing/dvt: Distant Viewing Toolkit for the Analysis of Visual Culture. (n.d.). https://github.com/distant-viewing/dvt*

*Google AI Studio. (n.d.). https://aistudio.google.com/*

*Google Cloud Tech (Director). (2020, June 17). I created an AI-powered video archive for searching family videos [Video recording].* https://www.youtube.com/watch?v=_IeS1m8r6SY

*Image understanding | Generative AI on Vertex AI | Google Cloud. (n.d.).* https://cloud.google.com/vertex-ai/generative-ai/docs/multimodal/image-understanding

*Jim Hubbard. (n.d.). Jim Hubbard.* https://www.jimhubbardfilms.com

*Moviepy · PyPI. (n.d.).* https://pypi.org/project/moviepy/

*NLTK :: Natural Language Toolkit. (n.d.).* https://www.nltk.org/

*OpenCV. (n.d.). OpenCV.* https://opencv.org/

*Pillow. (n.d.). Pillow (PIL Fork).* https://pillow.readthedocs.io/en/stable/index.html

*Pszemraj/dinov2-small-film-shot-classifier · Hugging Face. (2023, July 18).* https://huggingface.co/pszemraj/dinov2-small-film-shot-classifier

*Queer Archives Project · Digital Scholarship Services · Lafayette College. (n.d.).* https://dss.lafayette.edu/queer-archives-project/?utm_source=chatgpt.com

*spaCy · Industrial-strength Natural Language Processing in Python. (n.d.).* https://spacy.io/

*The ACT UP Historical Archive: ACT UP New York City Information. (n.d.).* https://actupny.org/indexfolder/NYC.html

*The ACT UP Historical Archive: ACT UP's CBS News Zap ~ NPR Transcript. (n.d.).* https://actupny.org/diva/news-transcript.html

*The ACT UP Historical Archive: Ashes Action '92. (n.d.).* https://actupny.org/diva/synAshes.html

*The HIV/AIDS Epidemic and Public Broadcasting. (n.d* https://americanarchive.org/exhibits/hiv-aids-and-public-broadcasting/5-1989-1991-broadcasting-act-up-direct-action

*The Potential for AI in Audiovisual Archives | Sound & Vision. (2021, May 10).* https://www.beeldengeluid.nl/en/knowledge/blog/potential-ai-audiovisual-archives

*torch.library—PyTorch 2.7 documentation. (n.d.).* https://pytorch.org/docs/stable/library.html

*Zero-Shot Prompting – Nextra. (2025, April 24).* https://www.promptingguide.ai/techniques/zeroshot