CineSearcher

A Multimodal Film Exploration Workspace

Tobias Kreten (Institute for Digital Humanities, Göttingen, Germany)
Marta Kipke (Center for Humanities Computing, Aarhus University, Denmark)

SUMAC – ACM Multimedia 2025

Introduction: A Multimodal Medium

Film = image + color + sound + timeline (measurable modes)

// + style + narrative + themes (higher concepts)

Related Work

- Distant Viewing and Cultural Analytics quantify style and color, but remain static and descriptive.
- CLIP-based retrieval bridges vision and language, yet lacks an interface for interpretive understanding.

Distant Viewing (Arnold & Tilton 2023)

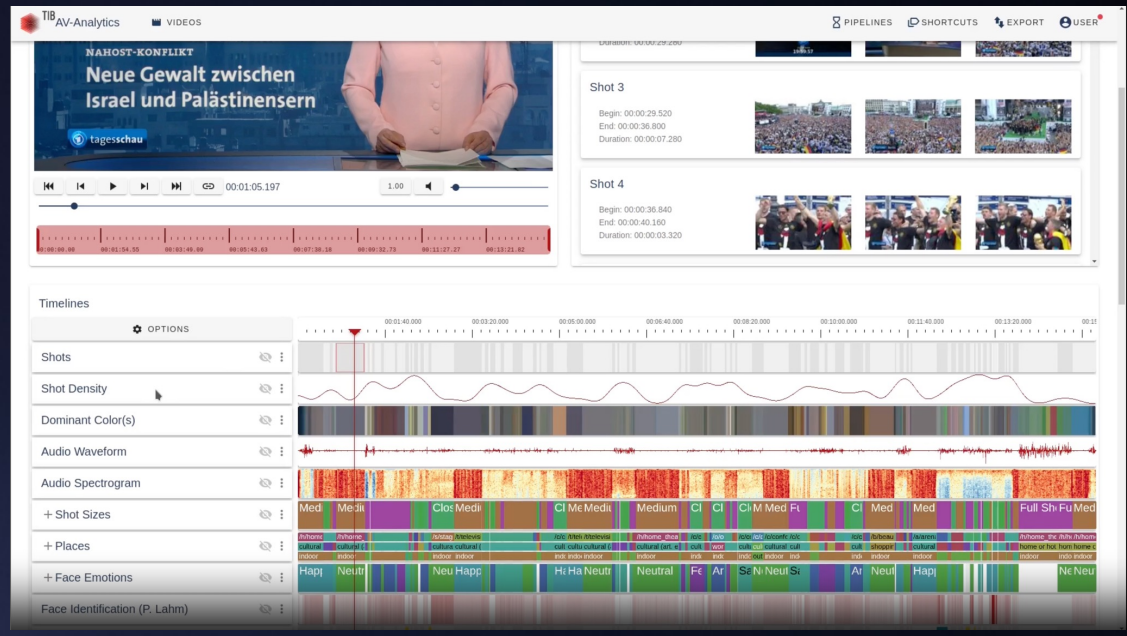
Cultural Analytics Lab (Manovich et al.)

MovieBarcodes (Burghardt et al.)

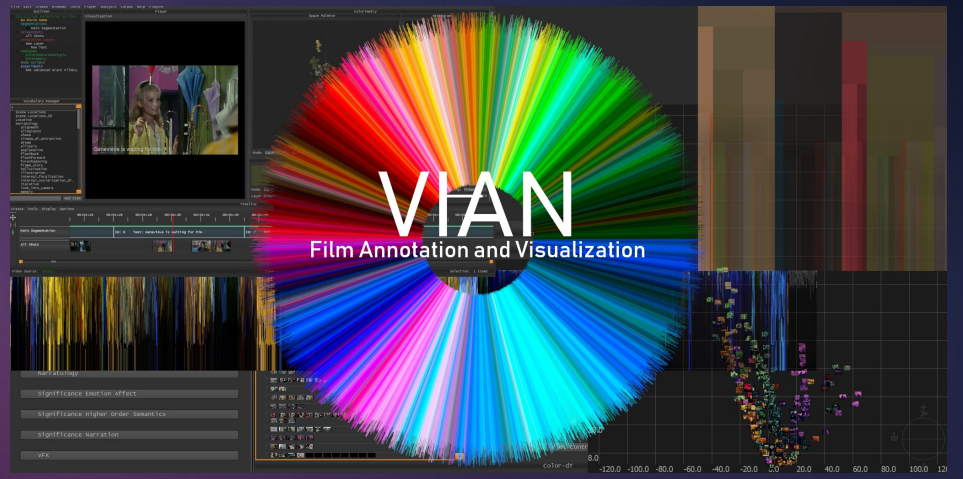
KALMUS (Chen et al. 2021)

VIAN (Halter et al. 2019)

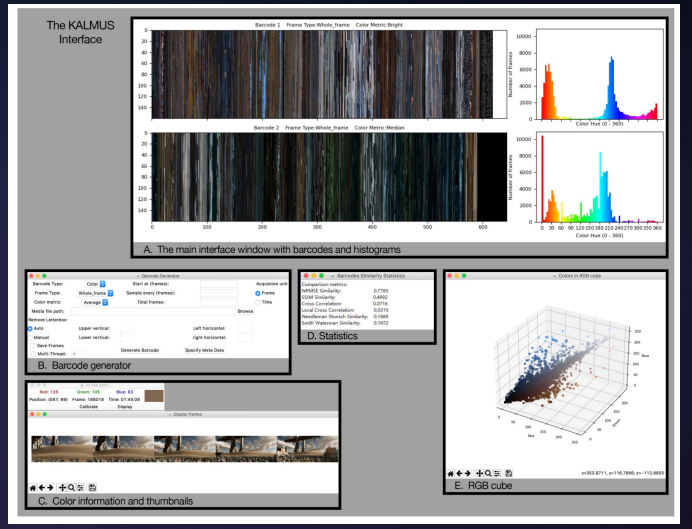
TIB-AV Analytics (Springstein et al. 2023)



TIB-AV Analytics
(Springstein et al. 2023)



VIAN (Halter et al. 2019)



KALMUS (Chen et al. 2021)

Gap

- Conventional metadata and text-based retrieval dissolve this complexity into fragments.

- No existing system links semantics, structure, and rhythm within a single exploratory surface.

CineSearcher

A MULTIMODAL EXPLORATION WORKSPACE

Change Corpus

religious iconography | Min. 0.26 | SEARCH | CLEAR | VISUAL | Min. 0.88 | Shot type...

SINGLE | CORPUS

SAVE
LOAD
EXPORT

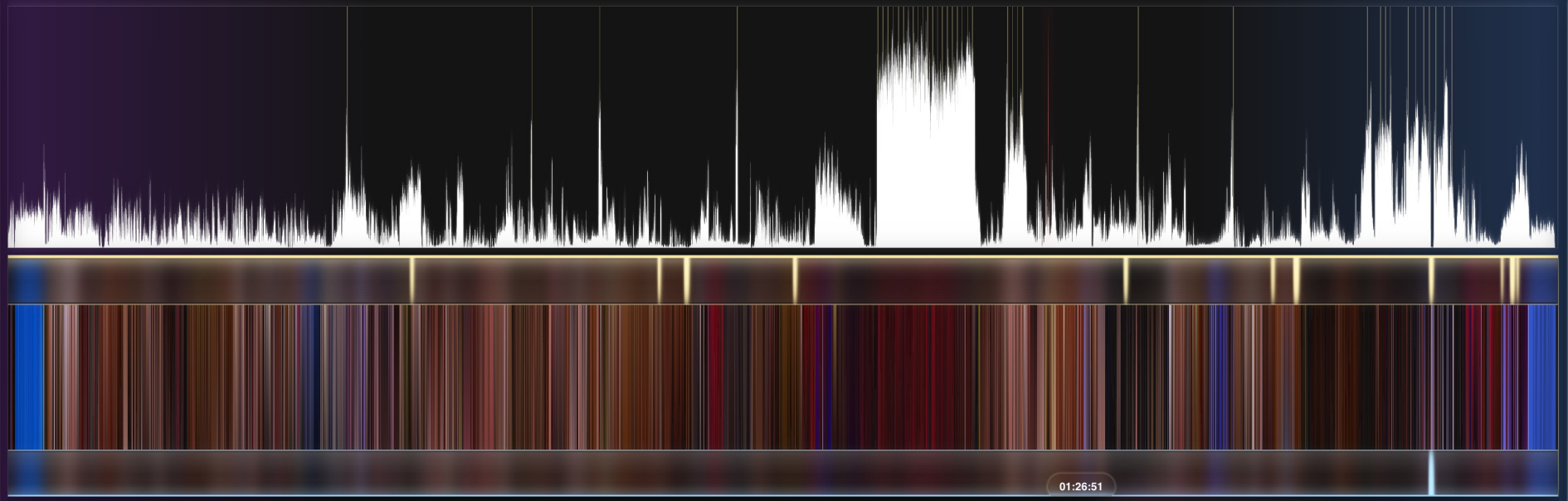
TUTORIAL

BARCODE
CLUSTER
SCATTERPLOT

CINEVIEW

CUTS
SHOT BANDS
RESULTS

Extreme Close-... ON
Close-Up ON
Medium Shot ON
Full Shot ON
Wide Shot ON



00:23:38 ×
00:42:34 ×
01:26:39 ×
01:41:26 ×

Twin Peaks: Fire Walk with Me by David Lynch

SEND FRAME TO CORPUS SEARCH

OPEN RESULTS

Metadata | Results | Chatbot

Film Statistics

Number of Cuts: 1738
Number of Shots: 1739
File Name: TwinPeaks.mp4

Avg. Shot Length:
4.47 s

Metadata

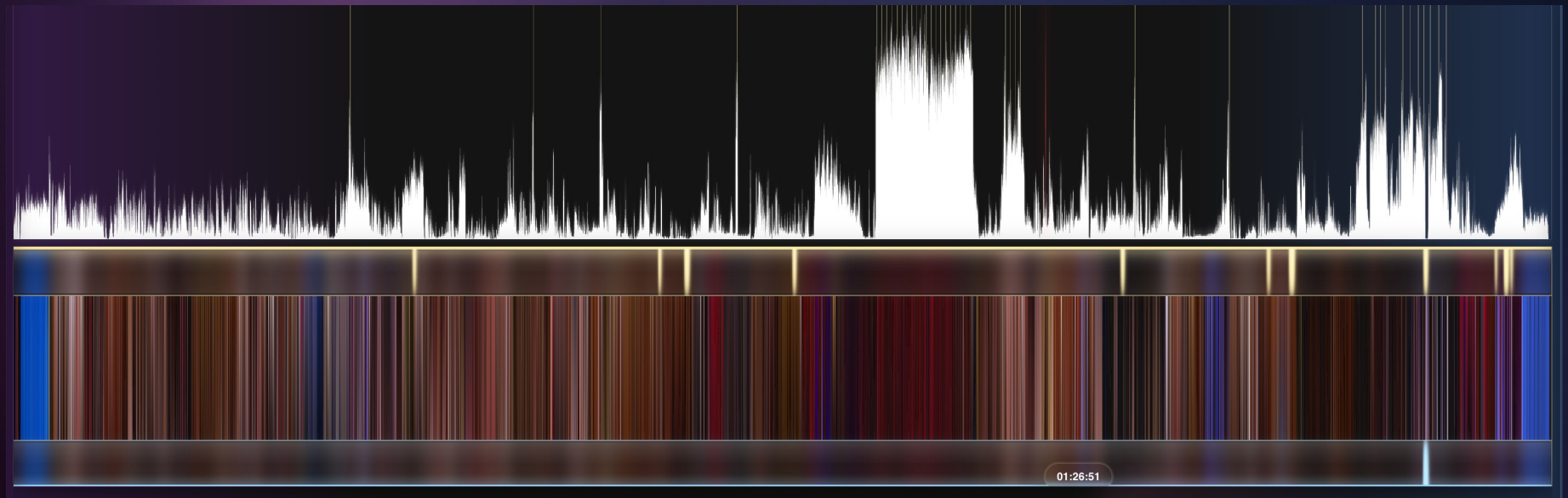
Load JSON
Create/Edit Metadata

Twin Peaks: Fire Walk with Me
Director David Lynch
Runtime 2:09:32

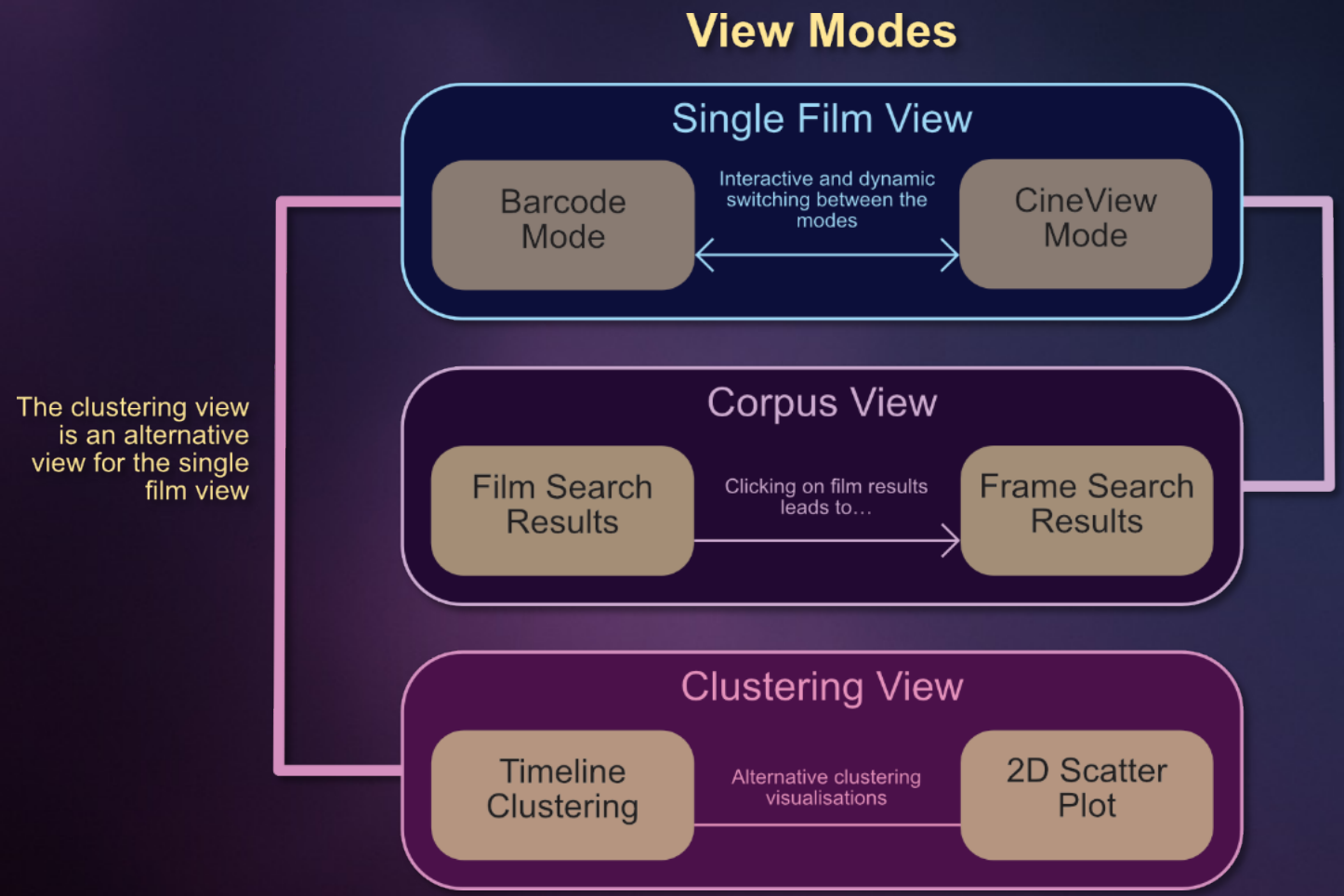
Key Contribution

1. Multimodal retrieval: CLIP text & image search + shot-type detection + audio peaks

2. A unified workspace allowing for a seamless exploration between close and distant viewing



Search paradigm



Features

- CLIP Embedding Retrieval
- Image Similarity Search
- Meta Data Search
- Cut Detection
- Shot Type Detection

Single Film View

Barcode Mode

Interactive and dynamic switching between the modes

CineView Mode

CineSearcher

A MULTIMODAL EXPLORATION WORKSPACE

Change Corpus

religious iconography | Min. 0.26 | SEARCH | CLEAR | VISUAL | Min. 0.88 | Shot type...

SINGLE | CORPUS

SAVE

LOAD

EXPORT

TUTORIAL

BARCODE

CLUSTER

SCATTERPLOT

CINEVIEW

CUTS

SHOT BANDS

RESULTS

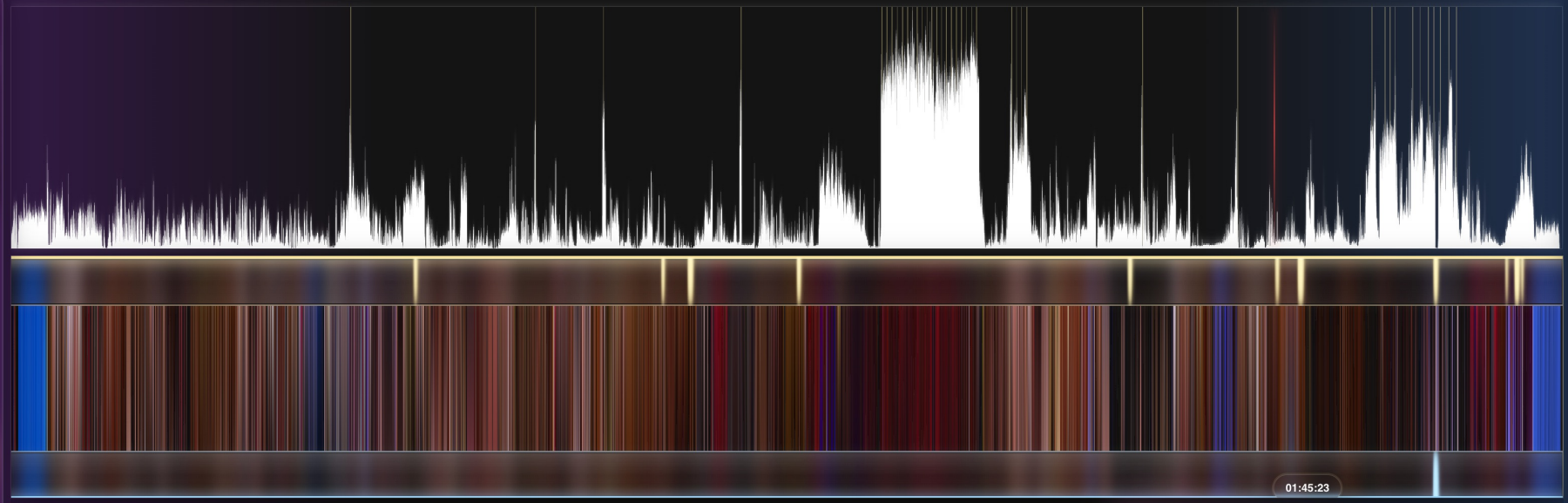
Extreme Close-... ON

Close-Up ON

Medium Shot ON

Full Shot ON

Wide Shot ON



00:23:38 ×

00:42:34 ×

01:26:39 ×

01:41:26 ×

02:05:43 ×

Twin Peaks: Fire Walk with Me by David Lynch

SEND FRAME TO CORPUS SEARCH

OPEN RESULTS

Metadata | Results | Chatbot

Film Statistics

Number of Cuts: 1738

Number of Shots: 1739

File Name: TwinPeaks.mp4

Avg. Shot Length: 4.47 s

Metadata

Load JSON

Create/Edit Metadata

Twin Peaks: Fire Walk with Me

Director David Lynch

Runtime 2:09:32

Single Film View

Barcode Mode

Interactive and dynamic switching between the modes

CineView Mode

FrameView TileView

Text Search

Corpus View

Film Search Results

Clicking on film results leads to…

Frame Search Results

‹ Selection

Change Corpus ›

CORPUS-SEARCH

Text | Visual

Min. Score: – 0.25 +

religious iconography

Search

Here you can carry out cross-film text searches.
The min. score impacts the amount and accuracy of the results.

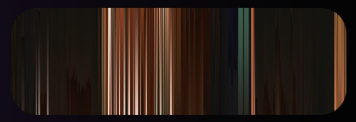
Download

WildAtHeart
44 Results

Eraserhead
36 Results

TwinPeaks
16 Results

Dune
15 Results

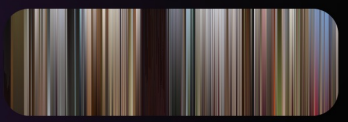


InlandEmpire
13 Results

BlueVelvet
10 Results

ElephantMan
7 Results

MulhollandDrive
2 Results



TheStraightStory
2 Results

LostHighway
1 Results

Clustering View

Timeline Clustering

Alternative clustering visualisations

2D Scatter Plot

Change Corpus

religious iconography | Min. 0.26 | SEARCH | CLEAR | VISUAL | Min. 0.88 | Shot type...

SINGLE CORPUS

SAVE

LOAD

EXPORT

TUTORIAL

BARCODE

CLUSTER

SCATTERPLOT

CINEVIEW

CUTS

SHOT BANDS

RESULTS

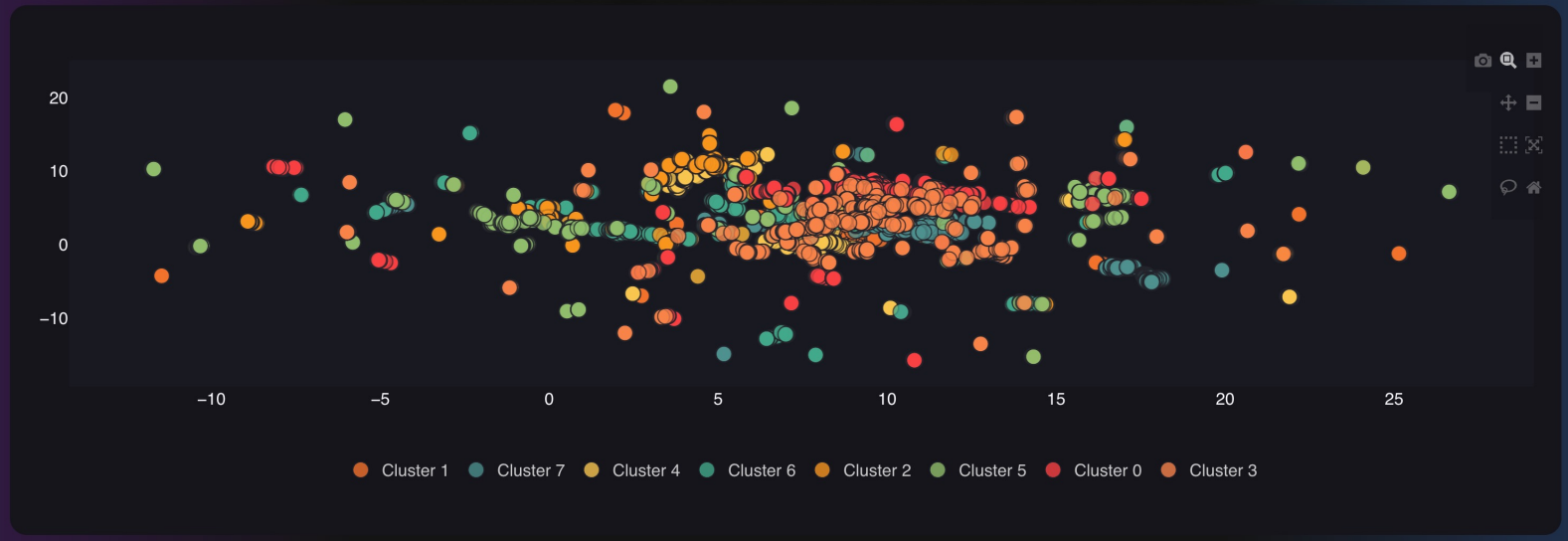
Extreme... ON

Close-Up ON

Medium ... ON

SCATTERPLOT (INTERACTIVE)

Clusters: 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20



Each point is a frame. Colour = Cluster.
Clicking opens the image in a new tab. The number of clusters (k) can be changed above.

Metadata | Results | Chatbot

Film Statistics

Number of Cuts: 1738
Number of Shots: 1739
File Name: TwinPeaks.mp4

Avg. Shot Length: **4.47 s**

Metadata

Load JSON

Create/Edit Metadata

Twin Peaks: Fire Walk with Me
Director David Lynch
Runtime 2:09:32

Clustering View

Timeline Clustering

Alternative clustering visualisations

2D Scatter Plot

Search by text... Min. 0.27 SEARCH CLEAR VISUAL Min. 0.88 Shot type...

SINGLE CORPUS

SAVE

LOAD

EXPORT

TUTORIAL

BARCODE

CLUSTER

SCATTERPLOT

CINEVIEW

CUTS

SHOT BANDS

RESULTS

Extreme Close-... ON

Close-Up ON

Medium Shot ON

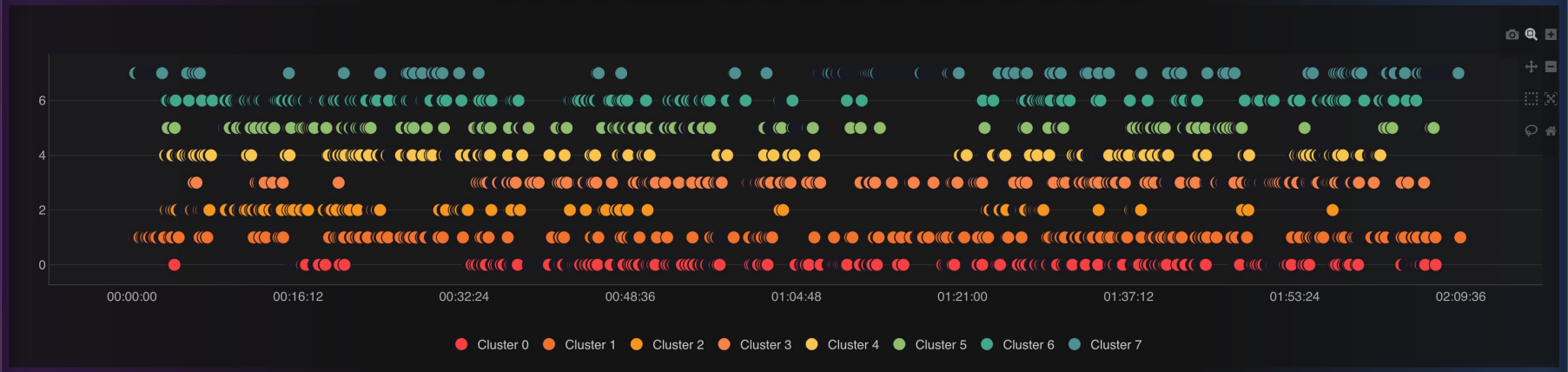
Full Shot ON

Wide Shot ON

Ultra Wide Shot ON

CLUSTERING (INTERACTIVE)

Clusters: 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20



Metadata Results Chatbot

Film Statistics

Number of Cuts: 1738
Number of Shots: 1739
File Name: TwinPeaks.mp4

Avg. Shot Length: **4.47 s**

Metadata

Load JSON

Create/Edit Metadata

Twin Peaks: Fire Walk with Me
Director David Lynch
Runtime 2:09:32

Conceptualization

- CineSearcher approaches film as a multimodal and temporal medium rather than a collection of discrete frames.
- Each movie is represented as a continuous sequence linking visual, acoustic, and semantic layers.
- By embedding these modalities in a shared representational space, the system preserves rhythm and atmosphere, the elements that define cinematic meaning.

Technical background

- CineSearcher consists of a modular architecture: a Python/FastAPI backend for feature extraction and indexing, and a React frontend for multimodal visualization.

- All components are synchronized through a REST API that manages metadata, embeddings, and retrieval results.

Backend: Feature Extraction and Indexing

- Each film is processed frame by frame at one frame per second to balance computational cost and temporal granularity.
- Visual embeddings are generated using CLIP (ViT-B/32) and stored in a Faiss vector index for efficient similarity search.
- Additional features include audio amplitude, color histograms, and structural cut/shot annotations.

Retrieval and Similarity Search

- Text queries are embedded into the same CLIP vector space, allowing semantic matching between natural-language descriptions and frames.
- The system supports both text-based and image-based queries, returning results ranked by cosine similarity.
- All retrieval results are mapped back to the film’s timeline for context-aware interpretation.

Performance

- Indexing a two-hour feature film requires approximately *14 MB* of storage for embeddings and enables sub-second retrieval on standard CPU hardware.

- The system can scale to multi-film corpora without GPU dependency, ensuring reproducibility in heritage or academic environments..

Multimodal Synchronization

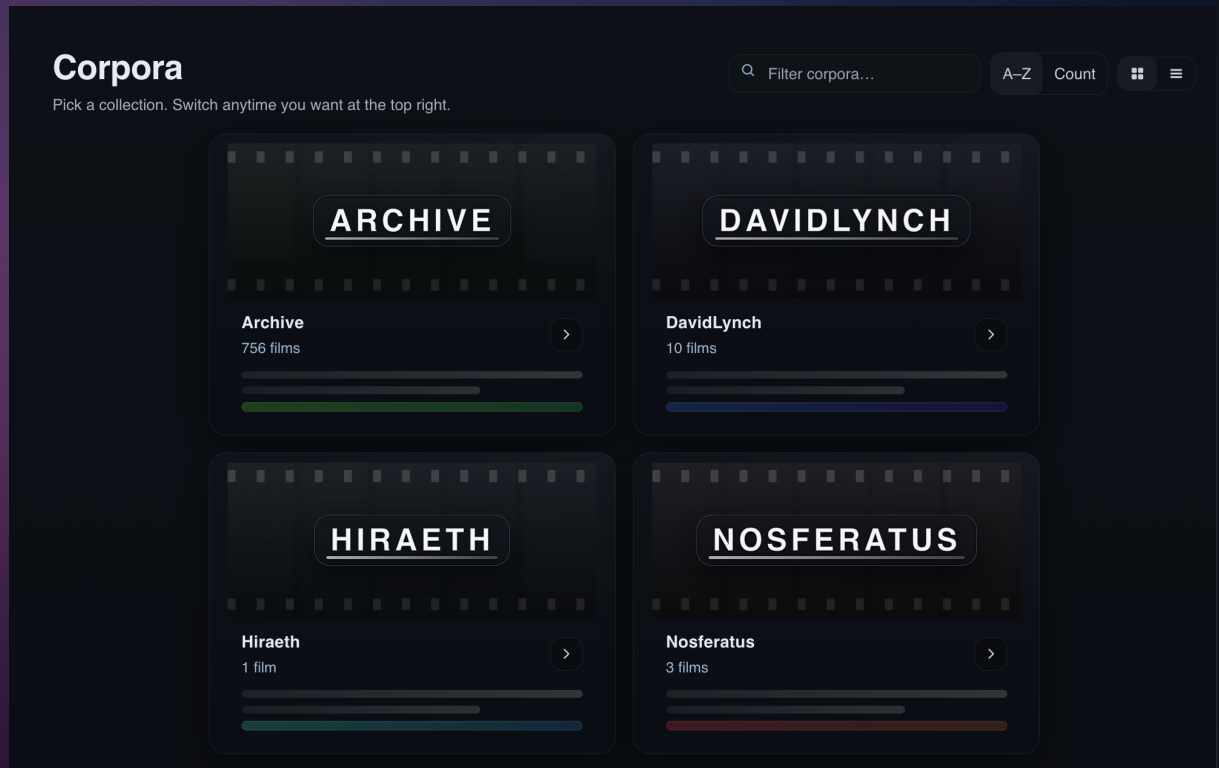
- The barcode is aligned with the audio waveform, creating a unified temporal view of visual and acoustic dynamics.
- Peaks in the waveform correspond to visual or narrative intensity, facilitating rhythm and pacing analysis.

Search and Navigation

- Search results are directly mapped onto the barcode, allowing users to explore semantic hits in context.
- Filtering thresholds and clustering visualizations support iterative refinement of results without re-querying the backend.
- Annotations and export functions ensure that exploratory processes remain transparent and reproducible.

Accessibility

- Instant corpus switching
- Color customization for the general mode and each individual cluster and shot overlay



Conclusion: From Data to Exploration

- The backend transforms a continuous audiovisual stream into a structured multimodal representation that integrates semantic, visual and acoustic dimensions.
- The frontend translates this representation into an analytical environment in which structure and interaction coincide. The interface transforms the processed data into an analytical space that can be explored intuitively.
- Every visual element corresponds directly to an underlying computational process. In this sense, the interface is not a secondary layer but an operational extension of the data model itself.

→ Combining exploration, transparency, interactivity and usability to create a workspace for scholars interested in video analysis.

Use Case: Historical Archives

Identification of interest points for research historical questions (e. g. the depiction and role of women, political propaganda and the instrumentalization of certain motifs / themes)

→ Evaluated on validation data provided by the challenge organizers

Corpus: Visual heritage data from the German Film Institute and Film Museum, the Federal Archives of Germany and The Netherlands Institute for Sound & Vision.

Use Case: Film studies

Case studies on film-analysis-specific terms and concepts, revealing patterns of visual language and themes within single films, filmographies and corpora.

→ Qualitative evaluation of domain experts (on-going)

Corpus: 58 horror movies / David Lynch filmography

Use Case: Film studies

„Religious Iconography“

Outlook and future work

Expanding the Multimodal Framework

- Integration of additional metadata and archival annotations to enrich contextual analysis.
- Development of comparative modes enabling stylistic alignment and motif tracking across multiple films.
- Extending multimodal retrieval to combine text, image, and sound into unified semantic queries.