

White Paper Report

Report ID: 104081

Application Number: HD5139411

Project Director: Michael Casey (michael.a.casey@dartmouth.edu)

Institution: Dartmouth College

Reporting Period: 9/1/2011-12/31/2013

Report Due: 3/31/2014

Date Submitted: 4/4/2014

White Paper

Grant Number: HD5139411

ACTION: Audio-visual Cinematics Toolbox for Interaction, Organization, and Navigation of film.

Michael Casey¹ and Mark Williams²

Dartmouth College, NH, USA

31st March, 2014

¹ Bregman Digital Media Labs, Departments of Music / Computer Science, Dartmouth College, USA

² Department of Film and Media Studies, Dartmouth College, USA

Table of Contents

[1. Narrative Description](#)

[2. Project Activities](#)

[Major Activities](#)

[Example Use Cases](#)

[Publishing and Dissemination](#)

[3. Accomplishments](#)

[Audiences and Evaluation](#)

[Continuation of the Project and Long-Term Impact](#)

[Appendix A: Summary of ACTION DB Films](#)

[Appendix B: Example Web Tutorial in Action](#)

1. Narrative Description

Introduction

ACTION is an open-source data set, software framework, and suite of example use-cases, that explore the possibilities for machine analysis in the study of film. Computation has a pervasive and central position in the creation and dissemination of film media via coding standards (MPEG-2, MPEG-4, H264), metadata standards (MPEG-7, MXF, SPARQL), and cataloging standards (Dublin Core, FIAF, METS). However, computational tools and methods for research using media archives are nascent. To address this gap ACTION sought to develop free and open-source computational tools, best-practice documentation, and new media-analytic methodologies for film and media scholars. ACTION utilizes machine vision, machine audition, and machine learning algorithms and builds upon open-source libraries such as [OpenCV](#), [PyMVPA](#), [OpenFrameworks](#), and the [Bregman Audio-Visual Toolkit](#). The ACTION toolkit currently includes: spatio-temporal color analysis, motion analysis, soundtrack analysis, automatic segmentation, visual and audio-visual structure segmentation, audio-visual stylometry, automatic classification by director, feature visualizers, film summarizers, and other tools in support of film.

The remainder of this document provides an overview of ACTION's data set and software framework and describes how they can be used in support of research into the development of film editing styles, scene composition, lighting, sound, and narrative construction. ACTION's use-case scenarios provide a foundation for future humanities research.

Statement of Innovation and Humanities Significance

Computational media analysis focuses on the syntactic, surface, and structural levels of features of the media, and cross-modal relationships between audio and visual features. Currently, little is known about what is possible to observe, from a humanist perspective, using surface-level computational analysis. Traditional methodologies in the humanities are predominantly based on semantics, semiotics, and intentionality, which are the sole preserve of human agency. ACTION, therefore, provides new knowledge in the humanities, otherwise unavailable by traditional analysis, regarding salient motifs, patterns, practices, and data relationships in audio-visual media. This affords subsequent research, informed by traditional humanities approaches and methods, but built upon the computational methods and practices documented in ACTION.

Technical innovations are in the application and evaluation of existing advanced machine-vision, machine-hearing, and machine-learning algorithms and software, and in the integration of such tools into a multimedia open-source framework designed for humanities researchers engaged in cinematic research. The project provides new knowledge on the application of such tools to humanities research via the evaluation and documentation of novel use-cases in cinema.

2. Project Activities

Major Activities

Michael Casey (Professor of Music, Professor of Computer Science), Mark Williams (Associate Professor of Film and Media Studies), Dr. Tom Stoll (Research Fellow).

Following are descriptions of major activities undertaken over the duration of the project (21 months: June 2012 - March 2014). The activities are divided into Data Set, Toolkit, Web Site, Example Use Cases, and Publications. Details on datasets and tools are appended.

ACTION Database: http://bregman.dartmouth.edu/action/action_db.html

We collected a set of 180 films divided into three data sets. The first, consisting of 140 films, is the core ACTION data set (`actionDB`), the remaining two databases consist of a set of documentaries (`actionDocumentariesDB`) and a paper print collection, consisting of an archive of early film stock that was provided by the Library of Congress (`actionPaperPrintDB`). These data sets were used for use-case development and testing of the ACTION Toolkit. The primary `actionDB` instance consists of 140 full-length films for which features, metadata, and summary statistics have been extracted and made available for download and further development, see Table 1, Figure 1, and Appendix A. The collection was sourced from the Dartmouth Film and Media Studies department, under the guidance of Professor Mark Williams. Films were chosen to represent a range of historical time periods, auteurs, locations, and genres, relating to a diverse pool of scholarship.

Table 1. List of 24 Directors, and their keys, in ActionDB core data set.

AH: Alfred Hitchcock	ChA: Chantai Akerman	HH: Howard Hawks	PSt: Preston Sturges
AK: Akira Kurosawa	DA: Darren Aronofsky	JF: John Ford	RBr: Robert Bresson
AT: Andrei Tarkovsky	DFr: David France	JLG: Jean-Luc Godard	SS: Steven Spielberg
AWe: Apichatpong Weerasethakul	DL: David Lynch	LB: Luis Bunuel	VsP: Vsevolod Pudovkin
CB: Coen Brothers	DzV: Dziga Vertov	MD: Maya Deren	WH: Werner Herzog
CTD: Carl Theodor Dreyer	GoR: Godfrey Reggio	MiN: Mikio Naruse	YO: Yasujiro Ozu

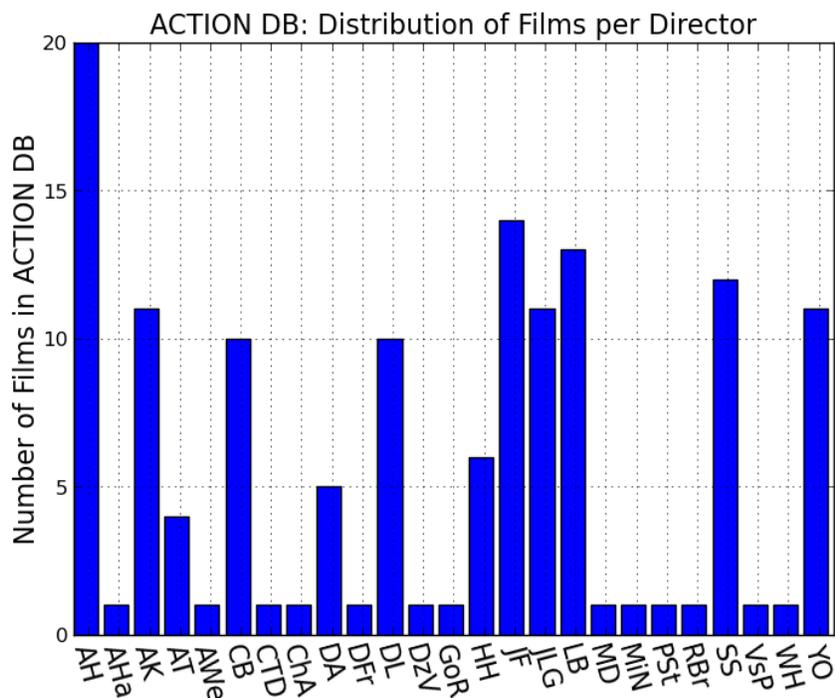


Figure 1. Distribution of 140 films by director in the ActionDB core data set.

ACTION Python Toolkit: <http://github.com/bregmanstudio/ACTION>

The ACTION Python Toolkit consists of a set of modules organized into a framework for access, navigation, and visualization of audio and visual features in a data set (`FilmDB`) instance. The code and corresponding Web site contain extensive documentation in the form of Python docs, examples, and tutorials. The ACTION Python Toolkit's modules consist of:

- `action_filmdb` - database instances reflected as a set of Python objects consisting of film metadata, content-based feature accessors, and visualizers. This module supports storage and retrieval of feature vectors contained within the database. The module is extensible to allow development of new film databases using the framework.
- `actiondata` - `ActionData` is a Python object encapsulating a collection of processing algorithms for analyzing `ActionData` data sets. The module and its classes support reading raw feature data, extracting summaries such as moving averages, windowing, histograms, clustering, and feature visualization.
- `color_features_lab`, `opticalflow`, `opticalflow_tv11`, `phase_correlation` - feature extractors for image-based features: color, optical flow, and visual phase correlation features.

- `bregman.features` - modules supporting audio feature extraction, accessors, visualizers.
- `segment-` module supporting segmentation: use for approximate shot detection and analysis via `VideoSegmenter` class. Based on visual feature cluster analysis. Scene detection and analysis based on hierarchical `VideoSegmenter` tool, hierarchical visual feature cluster analysis
- `distance` - module to support feature similarity analysis: visual display of segment similarity relationships, both intra-document and inter-document, for comparing styles, takes, edits, editions, audio-visual correspondence, and anomalies among documents in an archive.

ACTION C++ Toolkit <http://github.com/bregmanstudio/ACTION>

Supplementing the ACTION Python Toolkit are several cross-platform custom tools written in C++ for speed and memory efficiency. These applications support the heavy lifting of content-based feature extraction, content-based search and retrieval, and slit-scanning summary visualizations.

- `fftExtract` - C++ fast optimized audio content-based feature extractor support multiple features (FFT, CQFT, MFCC, and others) and multiple formats.
- `iMatsh` - C++ fast optimized content-based search engine for audio and visual features: audio and visual feature search and retrieval systems based on feature analysis and matching.
- `filmScapes` cross-platform C++ slit-scan movie visualizer and film summary generator based on OpenFrameworks for display of whole movie in static image view (supports browsing, visualization, and navigation of large film archives).

ACTION WEB: <http://bregman.dartmouth.edu/action>

- <http://bregman.dartmouth.edu/action> - dynamic HTML documentation, tutorials, and EXAMPLE use caseS demonstrating application of ACTION data sets and ACTION Python Toolkits to methods of film and authorship analysis.
- http://bregman.dartmouth.edu/action/action_data.html - location of all feature data available for download as a collection of compressed archives (.zip).
- http://bregman.dartmouth.edu/action/segmenter/action_db.html - location of ACTION DB with links images of segmented audio and visua features, plus similarity matrices for all 140 `actionDB` films.
- http://bregman.dartmouth.edu/action/tutorial_one_analysis.html - entry point for tutorials demonstrating how to use the ACTION Toolkit on the audio and visual feature data provided at the links above.
- http://bregman.dartmouth.edu/action/example_one_clustering.html - entry point for use case examples, showing clustering, classification, and visualziation strategies for the films in the `actionDB` `ActionData` instance.

Workflow

ACTION's workflow design consists of several stages from source media ingestion, through feature extraction, summarization, and machine learning, to visualization and final interpretation, see Figure 2. The ingestion stage implements methods to reliably extract high-quality motion JPEG (MJPEG) movie files and 16-bit 48kHz stereo WAV files from DVDs. A metadata file in JavaScript Object Notation (JSON) was created for each film giving basic details such as normalized title (primary key), length in seconds, length in frames, frames per second, and aspect ratio. This media metadata was combined with auteur and production metadata in a Python data structure. The ACTION Python Toolkit was used to further analyze the video and audio into color, audio, and motion features. To reduce the volume of data, and make computation on, and visualization of, large feature sets tractable, a second stage of feature extraction recruiting automatic segmentation via clustering was used.

The metadata, audio, and visual features for each film (~450MB per film, ~60GB all 140 films) are available for download from the ACTION Web site using the links provided above. For copyright reasons we cannot not make the source film media available for download, but we provide detailed instructions on the Web site on the DVD rip process so that other researchers may create video files from their own DVDs.

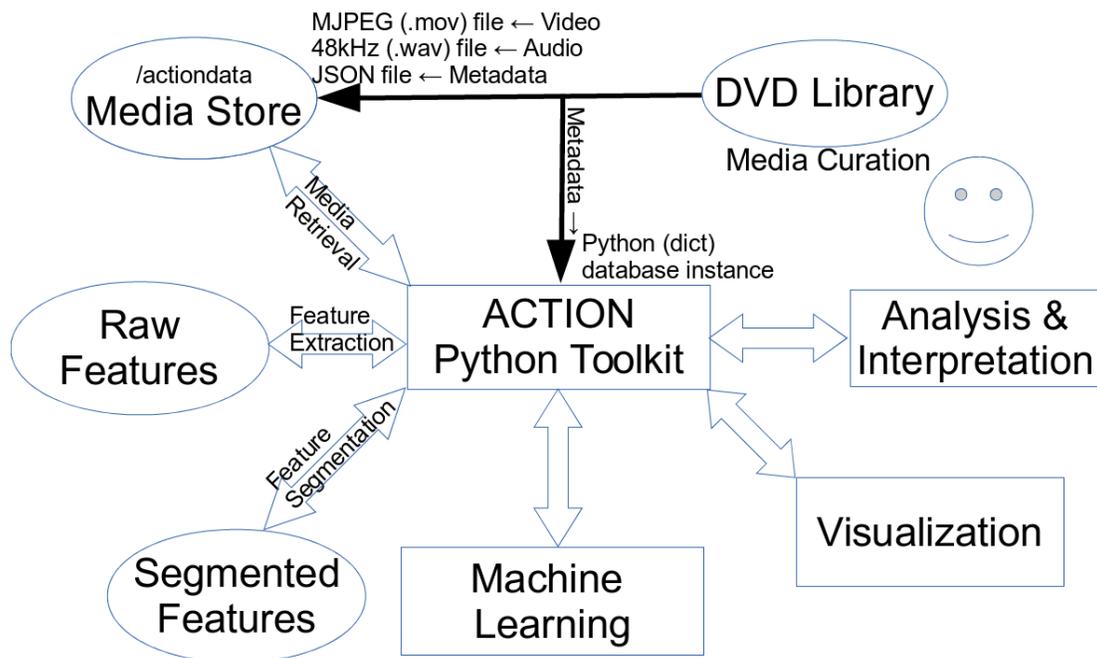


Figure 2. ACTION workflow, illustrating progression from ingestion of source media, through machine learning algorithms and visualizations, to human analysis and interpretation of processed data.

Color Features

ACTION provides a suite of feature extractors for gathering different types of surface-level information from media. The first of these extracts color features and arranges them into histograms representing color distribution in one second segments, overlapped with a hop of $\frac{1}{4}$ second. The temporal windowing parameters were determined by a significant amount of experimentation. Color information is encoded using the $L^*a^*b^*$ color space, where L^* is Luminescence (brightness) and the remaining two channels encode hue using opposing axes. Figure 3 illustrates the meaning of $L^*a^*b^*$ color-space dimensions. We chose this color space due to its correspondence with color perception by humans. Colors that are geometrically close in $L^*a^*b^*$ space are also known to be judged as perceptually close.

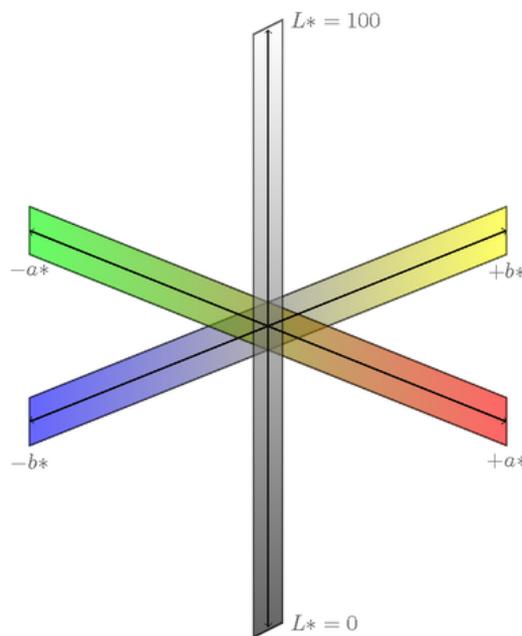


Figure 3. $L^*a^*b^*$ color space dimension. The vertical dimension is luminescence or brightness (L) and the remaining two dimensions are color opposition axes which encode hue.

Color histograms divide the luminance, a^* , and b^* , channels into 16 levels of quantization giving 48 bins per histogram. In addition to color histograms covering an entire image frame, we also divide the image into a 4-by-4 grid to capture the spatial distribution of color values within frames, see Figure 7 below. The combination of whole-frame and “gridded” histograms yields up to 17 histograms for each frame. After some experimentation, options were implemented for selecting parts of frames (such as the horizontal center two bands of the 4x4 grid) and for eliminating color (a^*, b^*) information when comparing color and black and white films.

Segmentation

In the first stage of feature extraction, raw color, motion, and audio features are extracted for every frame in a movie. This yields features that are too fine-scaled for visualization and interpretation. To illustrate, consider Alfred Hitchcock's *The Birds* which has a duration of 01h:54m:01.05s. The movie consists of 164,310 frames at 24 frames per second. To display the entire movie's features, with one feature vector per horizontal pixel, would require a monitor that had over 100 times the horizontal pixel resolution of a high definition television. Furthermore, if we wanted to compare the geometric distance between each feature vector against every other in the movie (a similarity analysis) that would require 13.5-billion comparisons, or more than 1 trillion operations, occupying around 100 billion bytes (100 Gigabytes) of data storage per movie.

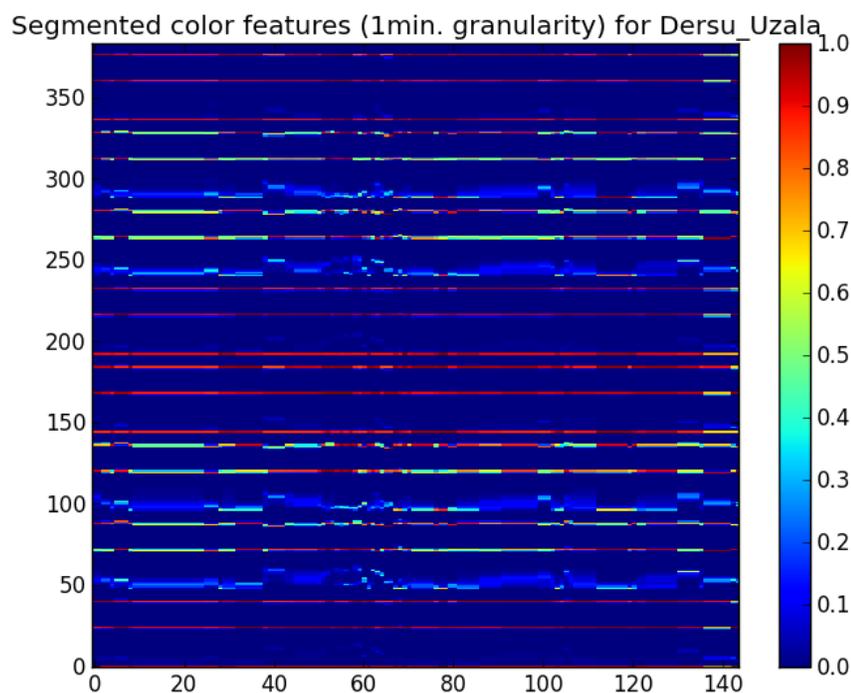


Figure 4. Segmented L*a*b* color histogram 1-min summaries for *Dersu Uzala* by Akira Kurosawa displayed as a multi-dimensional time series. The horizontal axis shows time, in minutes, and the vertical axis shows 1-min averages of 368 color features.

To reduce the amount of data to within practical limits we perform segmentation on the features, summarizing them over fixed durations of 1 minute, or over periods of relative stability via cluster analysis. We employed three types of segmentation: fixed-length windows (1-min summaries of features), k-means cluster analysis (automatic segmentation by similarity), and hierarchical cluster analysis. Figure 4 shows a time-series of 368-dimensional color features that have been averaged in time into a series of 1-minute fixed-duration summaries.

Motion Features

We experimented with, and provide, three different algorithms for collecting motion and optical flow data. The first algorithm is optical flow using the [Lukas-Kanade optical flow algorithm](#)³. This data is collected over an 4-by-4 grid, and within each region, angles are “binned” into 8 bins. The second algorithm, called phase correlation, and is collected as 2D vectors over an 8-by-8 grid over the entire screen, as well as a vector for the entire screen. The third algorithm is the [TV-L1 optical flow algorithm](#)⁴. This algorithm collects data over the entire screen, as well as over an 8-by-8 grid. The output data consists of a normalized histogram of raw angular magnitude motion data.

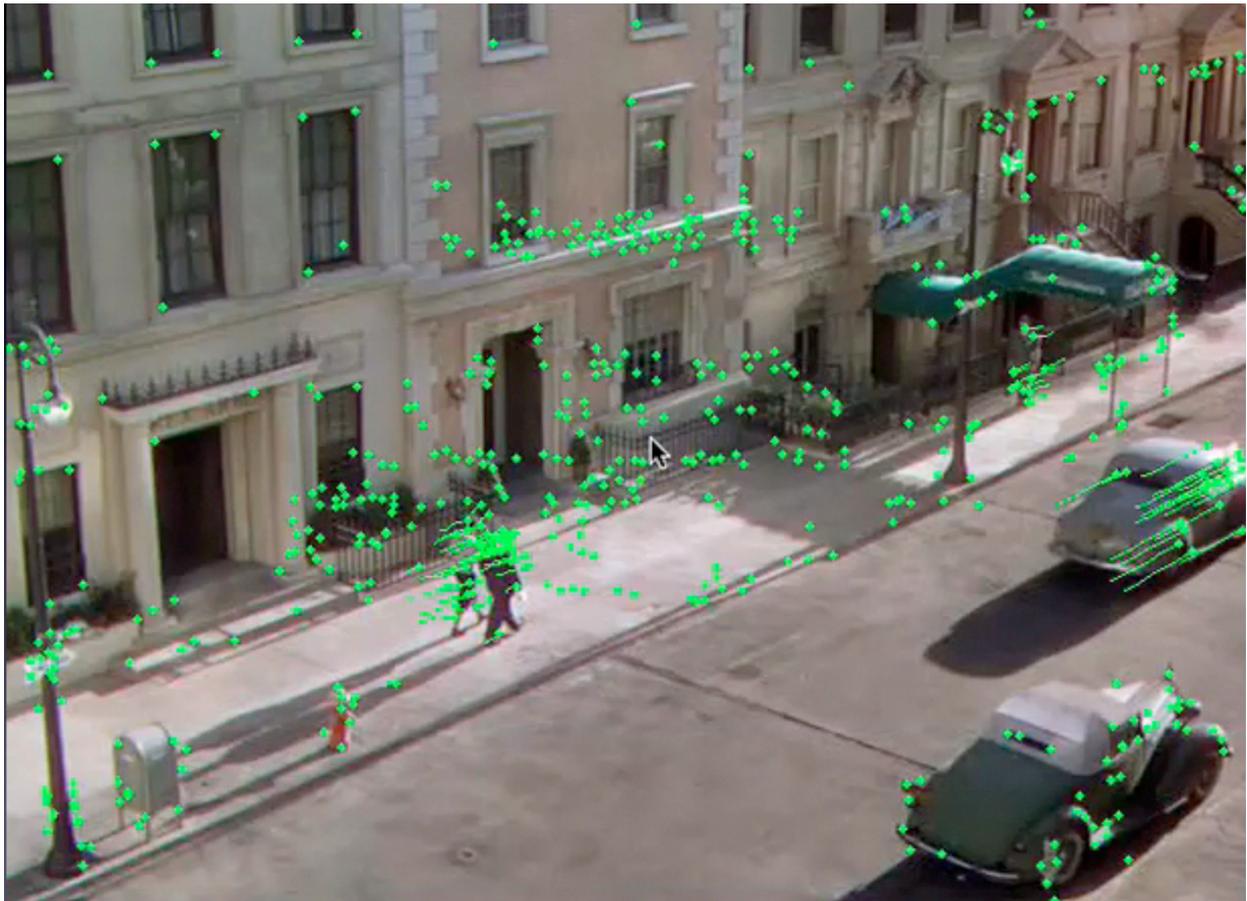


Figure 5. ACTION motion vector feature visualization. The visualizer illustrates the regions of the image that are moving, and the direction and magnitude of the motion.

³ B. D. Lucas and T. Kanade (1981), *An iterative image registration technique with an application to stereo vision*, Proceedings of Imaging Understanding Workshop, pages 121–130

⁴ C. Zach, T. Pock, and H. Bischof. A Duality Based Approach for Realtime TV-L1 Optical Flow. Pattern Recognition (Proc. DAGM), Heidelberg, Germany, 214-223, 2007

Audio Features

In addition to color and motion features, ACTION implements audio features and these were extracted for each film. Each vector consists of a list of numbers representing feature values for one second of film. Feature vectors were generated at the rate of one-quarter second per stride, for the entire duration of each film. Specifically, audio features consist of constant-Q Fourier Transform (95 log-spaced frequency bands per feature vector) representing a perceptual encoding of the power spectrum; Mel-Frequency Cepstral Coefficients (13 cepstral features) representing broad spectrum envelope, such as formants in speech; and Chroma (12 pitch-class features) representing energy per musical pitch-class.

Segmentation and Similarity Matrix Summarization of ACTION DB

Further processing on visual and audio features yielded whole-film structure plots in the form of similarity matrices, called distance matrices or dissimilarity matrices depending on the computational metric used. These visualizations were designed to allow exploration of the temporal structure of films via concepts of self-similarity and recurrence, first introduced in the 1990s by Jonathan Foote⁵ as the S-Matrix for analyzing musical structure from music audio recordings, the representation has become an increasingly important visual tool for time-based media data analytics. The ACTION Web site hosts (dis)similarity matrix visualizations, using audio, visual, and joint audio-visual feature spaces, of all 140 films: http://bregman.dartmouth.edu/action/segmenter/action_db.html

Computation of similarity matrices is relatively straight forward given a time series of feature vectors. For every pair of time points in the series we compute the vector Euclidean distance between them and place the resulting distance value into a two-dimensional grid that is indexed by the two time-points (t1,t2).

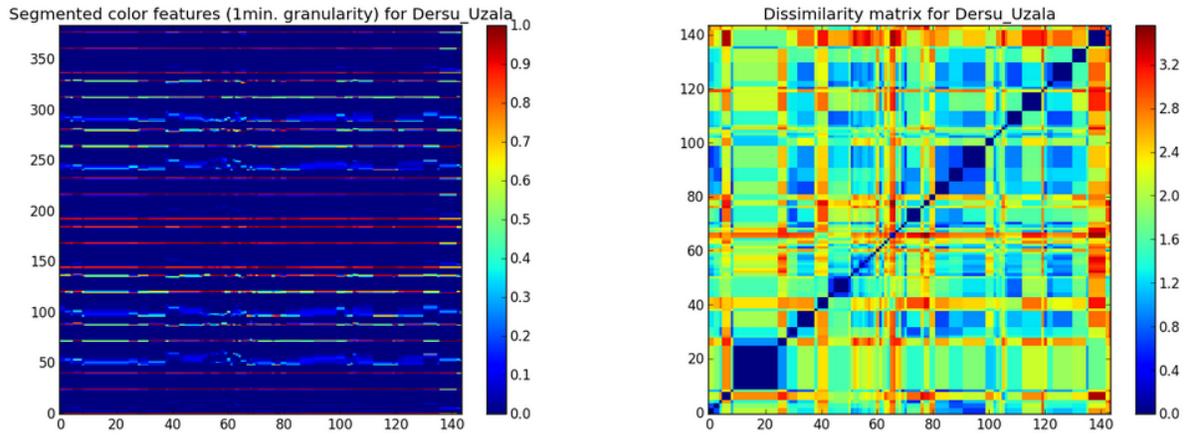
Figure 6 shows two similarity matrices, on the right hand side, computed using the summarized features shown on the left. Blue regions (lower values) correspond to smaller distances showing time locations where self-similarity occurs in the feature, and red regions (higher values) correspond to dissimilarity. The patterns of repeating blocks of blue and red are informative about recurrence of color and spatial layouts of shots and scenes, or recurring textures in the sound.

This structural view of a film can be used to illustrate important aspects of style and narrative construction. We illustrate several use cases for within-film and between-film similarity evaluation in the section entitled *Use Case Scenarios* below.

⁵ Jonathan Foote and Matt Cooper, "Media Segmentation using Self-Similarity Decomposition," in *Proc. SPIE Storage and Retrieval for Multimedia Databases*, Vol. 5021, pp. 167–75, 2003, January 2003, San Jose, California.

Film title: Dersu Uzala
Director: Akira Kurosawa
Year: 1975
Color: Color

Color Features



Audio Features

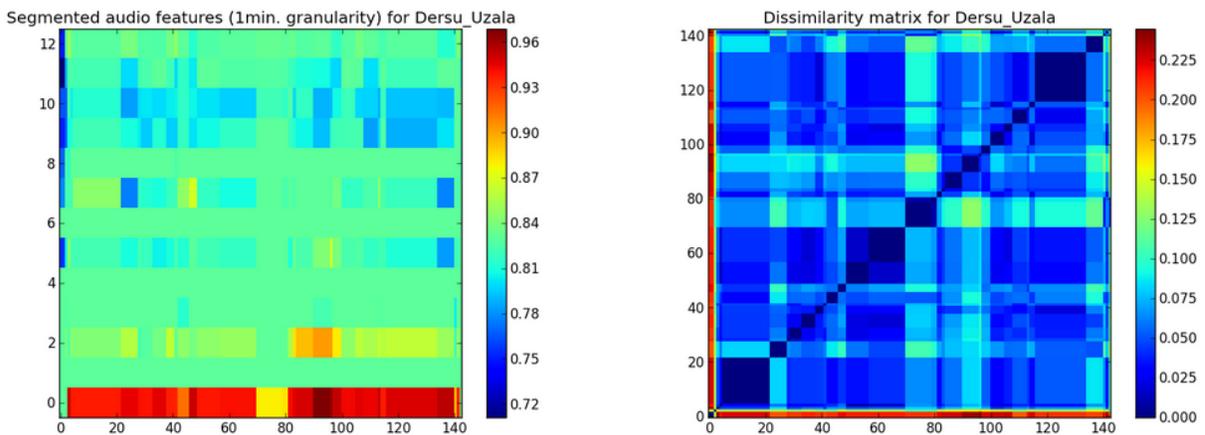


Figure 6. The left plots show segmented and summarized color features (top) and audio features (bottom) and the right plots show their distance (dissimilarity) matrices.

Visualization

In addition to the scientific plotting tools provided within the Python environment (matplotlib), ACTION provides specialized methods for visualizing specific film features. Each of the feature extractors in the ACTION Toolkit provides a custom method to visualize, i.e. play, the extracted features alongside the film. This capability is essential for researchers to get acquainted with the relationship between a film's contents and their feature-space manifestations. Figure 7 shows the L*a*b* color histogram feature visualizer applied to a segment of Alfred Hitchcock's *Rope*. Color histograms are extracted for screen regions corresponding to a 4 x 4 tiled-grid, and one

histogram for the entire screen. This yields 17 histograms per shot segment, as shown in the figure.

In addition to feature visualizers, ACTION provides a novel slit-scan summarization viewer that creates visual summaries of sequences. Slit scanning is a well-known summarization method that has been used by numerous media user interface researchers to illustrate temporal structure, shot lengths, and narrative devices used in film⁶. Figure 8 shows the slit-scan viewer in ACTION applied to several minutes of Alfred Hitchcock's *The Birds*.

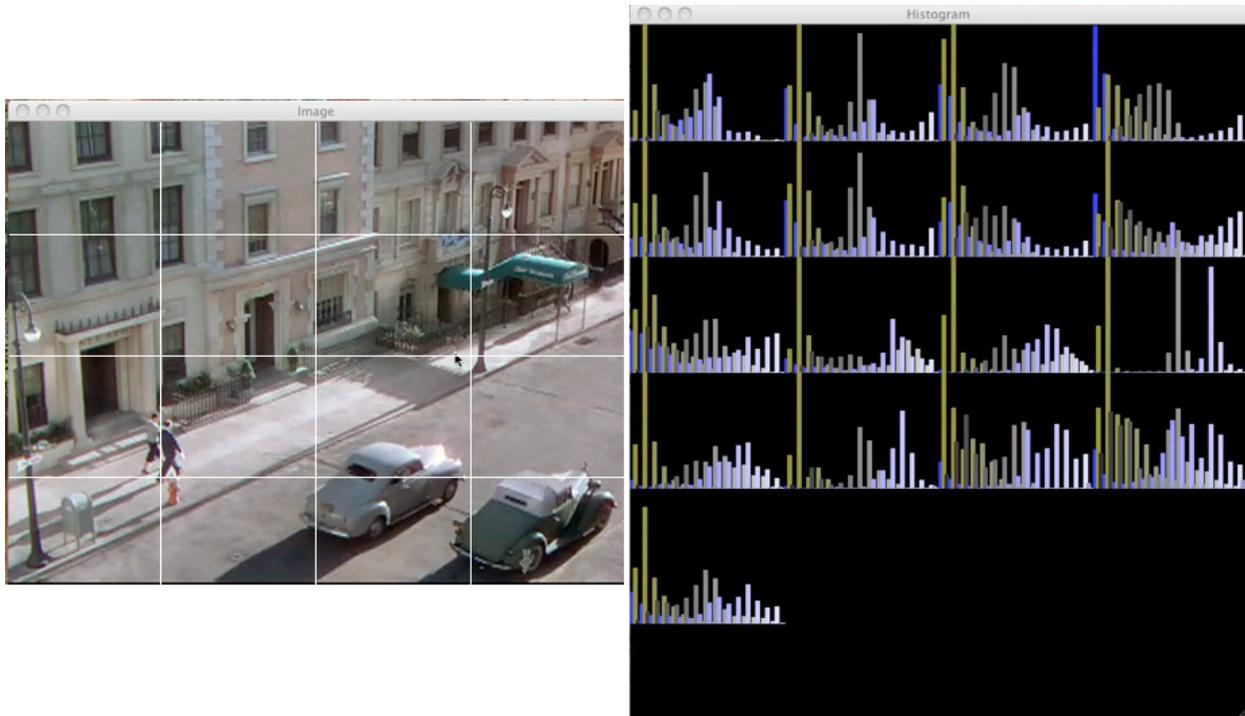


Figure 7. ACTION Feature Visualizer (L*a*b Color Histograms, 16 tiles + whole screen)

⁶ Yeung, M.M, Video visualization for compact presentation and fast browsing of pictorial content, IEEE Transactions on Circuits and Systems for Video Technology, 7:5, 1997

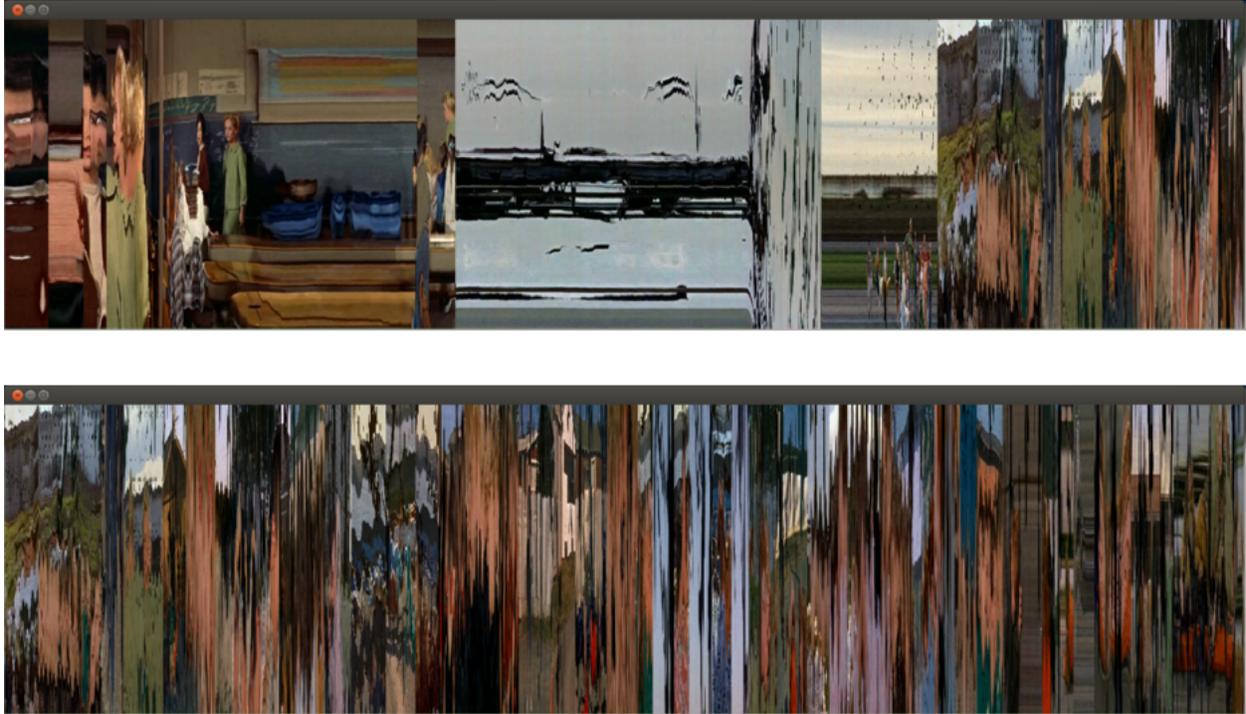


Figure 8. ACTION slit-scan visual summarization showing relative shot durations and center-of-screen subject in a 3-minute segment of Alfred Hitchcock’s *The Birds*.

Example Use Cases

To test-drive our features, and to speculate on possible new methodological avenues for Film scholarship, we devised a number of example use-cases, with each use case asking a specific question relating to style and authorship across multiple directors, genres, and time periods.

USE CASE 1: Computational Stylometry (Auteur Classification by Color Use)

Stylometry is the expert practice of uncovering distinguishing features of an artist’s output in relation to other artists. In visual art stylometry is used to establish authenticity of works. We do not expect to be questioning authorship of films in our dataset, however, we do want to ask if directors have distinguishing features that enable their films to be accurately attributed to the correct author when no information other than visual or audio features are made available to the analysis. In a sense, computational stylometry is a test of the power of color, audio, or motion features for uniquely characterizing a director’s style.

The stylometry use case selects all films from eight directors, those with 10 or more films in the ActionDB data set, for model learning and automatic attribution. The data are divided into a training set and testing set, with one movie from all movies for eight directors held out of the training set for testing. We can consider the held-out movie as an ‘unknown’ movie, for the purposes of explanation, even though we know the attribution of all of our movies.

From each of the training movies we randomly sampled a representative set of 1000 segmented color features. This representative data set was labelled by director and a support vector machine (SVM) classifier was trained using the libsvm implementation available inside the PyMVPA machine learning toolbox in Python. Figure 9 illustrates the processes of dividing a data set into training and testing sets, sampling, model training, and testing (predicting director label for ‘unseen’ data).

Results of the 8-way supervised director task are shown in Figure 10. The graph shows the full 8 x 8 confusion matrix derived from averaging 16 different samples / models over all movies for each director. This repetitive process allowed us to estimate a mean performance over all films and all directors, and to compute standard error rates as a measure of reliability of the methods. In the confusion matrix, correct classification results lie on the main diagonal and mis-classifications are in the off diagonal entries.

The mean accuracy for classification of unseen movie samples among eight directors was 39.2% with baseline random classification 12.5%. The result illustrates that significant information about the director of a film can be gleaned from summary information about the spatial distribution of color features alone. This surprising result suggests that important autoreal signatures reside within low-level color features.

Further to classification by directors, we also attempted to classify based on historical time periods and film genre. These different partitionings of the data set did not reveal results that were significantly above chance. Again, we were surprised that historical time periods were not discriminating and yet directors were. Further analysis will need to be undertaken to seek features that could discriminate alternative partitions of the data set beyond division by director.

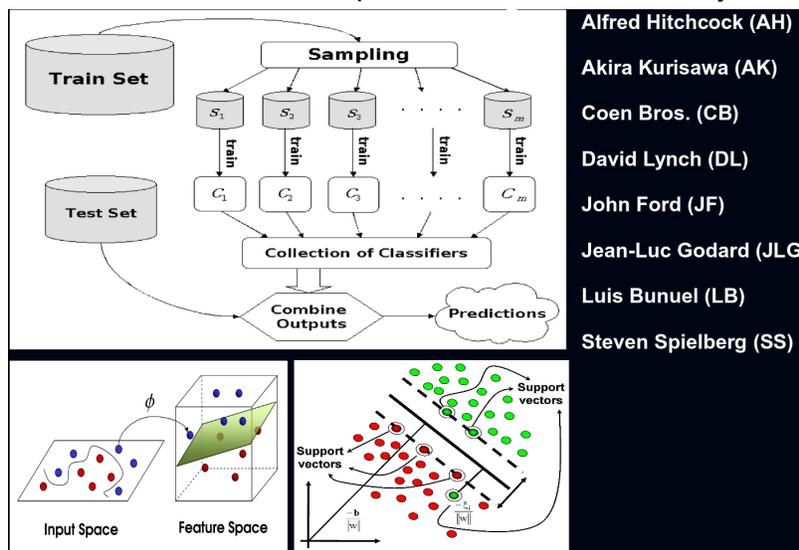


Figure 9. Support vector machine (SVM) training and classification on eight directors’ films.

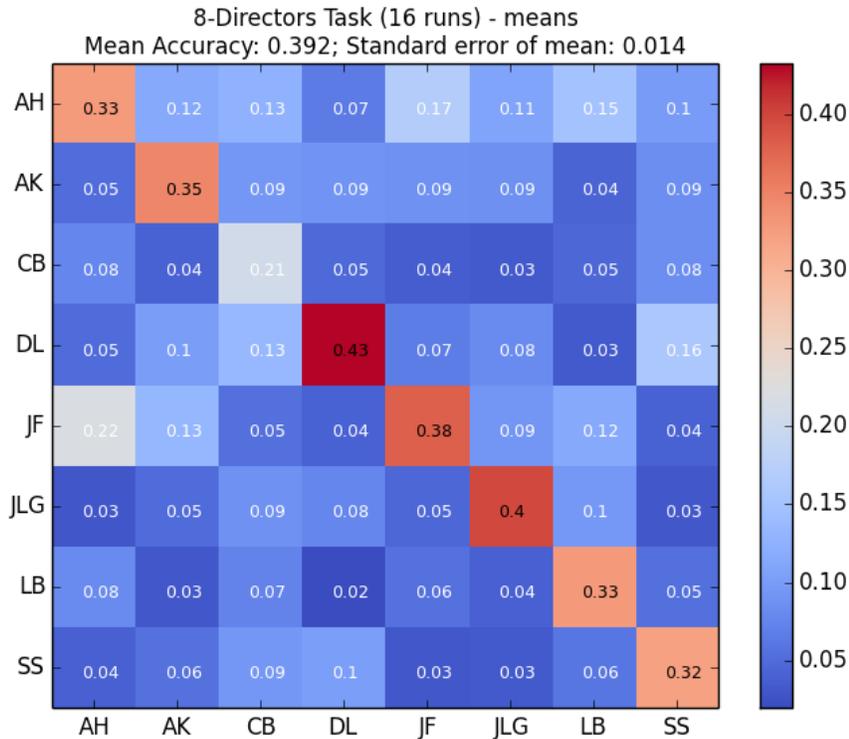


Figure 10. Results of director classification task showing classification accuracies in the main diagonal entries and between-director confusions in off-diagonal entries. Accuracy:39.2%, std error:1.4%, random baseline: 12.5%.

USE CASE 2: Auteur Audio-Visual Feature Correlation

The second use-case example shows how joint audio-visual feature similarity analysis can reveal the degree of cross-modal design within a director’s filmography. We analyzed features for the films of eight directors for whom we had at least 10 films resident in the ActionDB data set. Audio and visual features were segmented separately and a similarity analysis for each feature set generated, two similarity matrices for each film, one for audio and the other for visual features, see Figure 6.

To measure how much the structural aspects of a director’s films were jointly manifest in the audio and visual materials we calculated the correlation for audio and visual similarity matrices using the `corr_coeff` function available within the `actiondata` Python module. The correlation measure used was the pearson product-moment correlation coefficient. The correlation distance function requires similarity matrices to be of the same dimensionality (same number of time points), so audio and visual features were jointly segmented.

To determine if a director’s audio-visual materials were correlated we devised empirical null hypothesis (H0) models (labeled using a ‘0’ at the end of the director key) that computed correlations between similarity matrices of different films. Intuitively, we do not expect the audio

and visual structure to correspond between different films by the same director, so this constituted a reasonable null model.

Correlation coefficient distributions for audio-visual correlation analysis of eight directors are shown in Figure 11. The t-statistic and p-values are shown above each director's result. The second instance of each director shows the correlation coefficient distribution for the null model (shown with a '0' appended). The figure shows that three directors had correlation coefficients that were significantly different than the null model with $p < 0.05$. Those were Alfred Hitchcock, Jean-Luc Godard, and Darren Aronofsky.

From the cross-modal similarity matrix correlation analysis we conclude that these three directors employ a joint audio and visual schema for sound, music, lighting, cinematography, and visual effects. This example shows how measures made over multiple sets of films, and compared between directors, can highlight significant differences in approach to the medium.

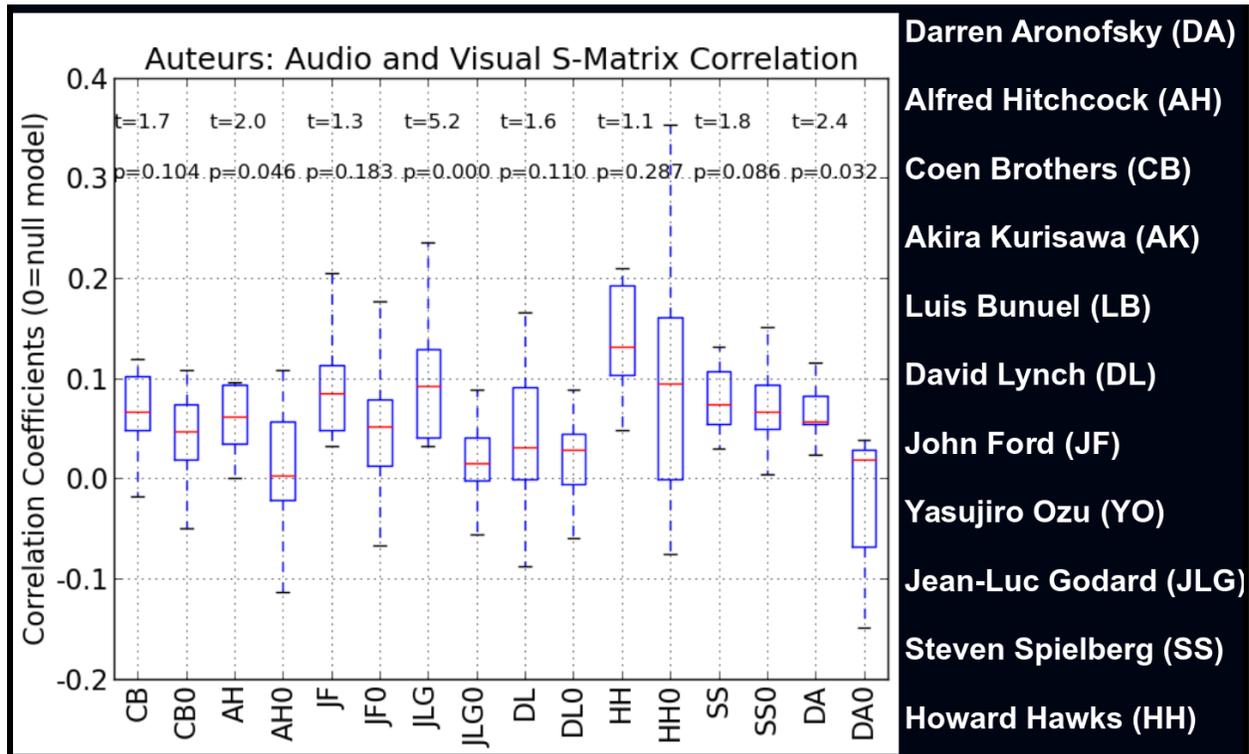


Figure 11. Box and whiskers plot showing audio-visual correlation coefficient for eight directors (e.g. CB=Coen Brothers) against the per-director null model (e.g. CB0). Results show significant audio-visual feature correlations for three directors: Alfred Hitchcock, Jean-Luc Godard, and Darren Aronofsky.

The third use case we explored was a longitudinal study of the filmography of a single director, Alfred Hitchcock. Between-film similarity was computed by sampling and averaging the distances between color histogram features for each pair of films in the filmography. We first arranged the between-film distances by alphabetical ordering by film title, see Figure 12. No obviously discernible pattern emerged in this view, which is to be expected as we would not expect films whose titles start with the same letters to be related. We then arranged the same information by year of production, see Figure 13. In this view a clear pattern emerged. Two distinct periods, the first occurring between *The 39 Steps* (1935) and *Strangers on a Train* (1951), and the second occurring between *Rear Window* (1954) and *Frenzy* (1971). *Rope* and *Psycho* appear to be outliers, the former to both time periods.

A straight-forward interpretation is that Hitchcock first experimented with color film in *Rope* but did not begin his color period until *Rear Window*, with *Psycho* and *The Wrong Man* being the only black and white films after *Rear Window*. It is easy to see from Figure 13 that *Psycho* belongs to the first group, however, the same is not true for *The Wrong Man*, nor is it apparent that *Rope* belongs to the second group which leads us to conclude that a simple distinction between color vs black and white does not explain the two periods.

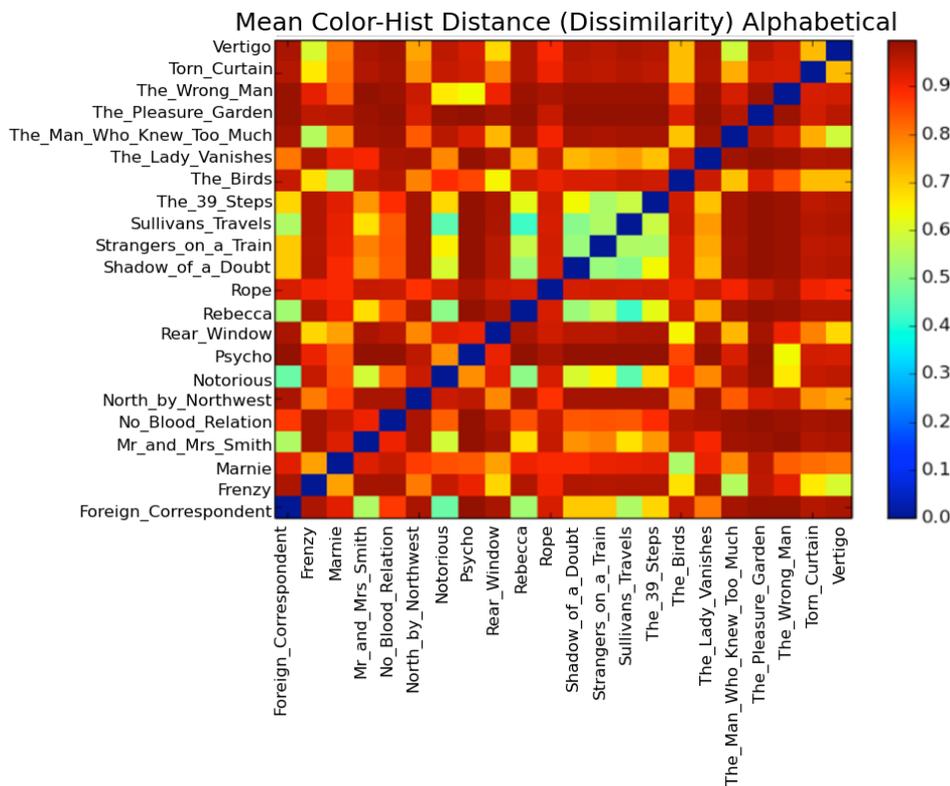


Figure 12. Between-film similarity analysis for 21 Alfred Hitchcock films (plus Sullivan's Travels by Preston Sturges).

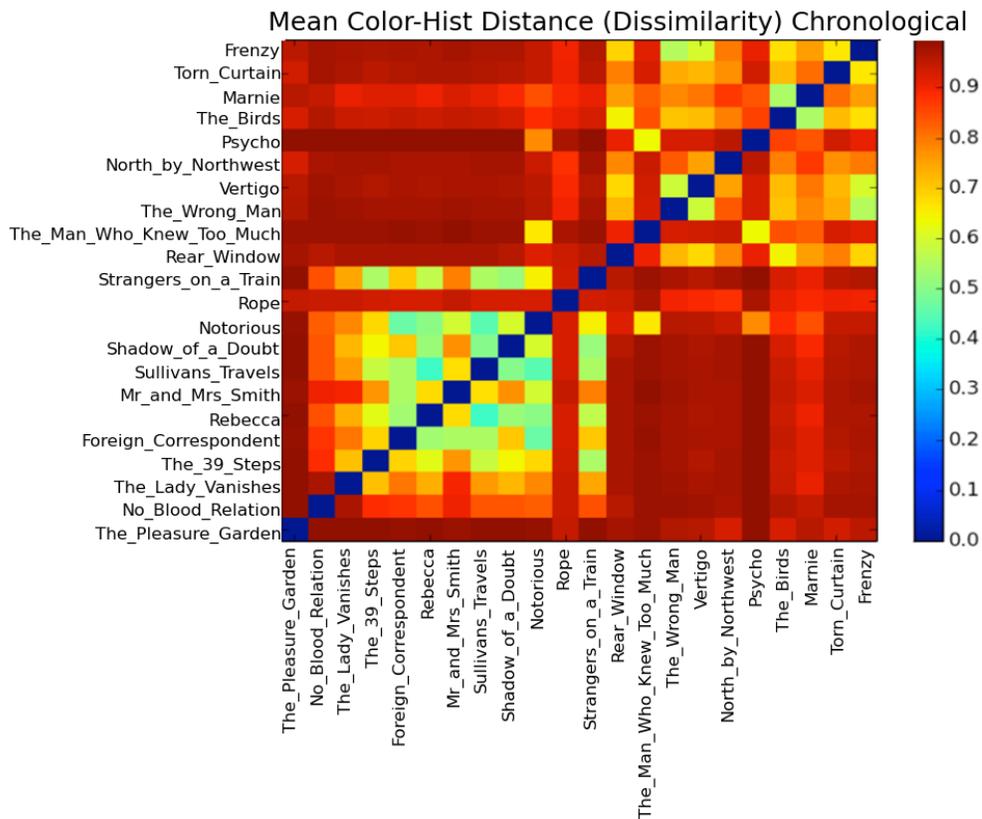


Figure 13. Longitudinal sort of between-film similarity analysis for 21 Alfred Hitchcock films (plus Sullivan’s Travels by Preston Sturges).

This use-case illustrates the need to find ways to navigate the data from multiple viewpoints, which can be achieved by sorting according to different fields in the data set. Extensions of viewpoints would be to include production information, such as location and crew, technical information, such as camera and color technologies, in addition to year of production.

Omissions and changes in project activities

There were two planned activities that we did not investigate due to time and technology constraints. The first of these was shot-type classification. We studied several previous works that performed such automatic shot-type classification which require training supervised machine learning models, such as support vector machines, on labelled motion vector data. Labels would have to be human mark-up of many hours of different shot types, camera motions, and scene compositions. Whilst possible in principle, we determined that such labelling of the data set was intractable given the limited human resources available in this startup phase of the project.

The second omission was speech keyword spotting. We intended to evaluate the degree to which keywords could be identified in the sound track of films, using a phoneme lattice representation such as that found in the Spoken Content part of the MPEG-7 Content Description Interface Standard. However, for similar reasons to shot-type detection, we determined that the process of gathering training data, developing expertise in use of speech recognition tools, and developing evaluation strategies, were beyond the available resources of the project in its current phase. In the future, we hope to evaluate the degree to which keyword spotting in a film sound track can be used to identify narrative topics and specific script locations.

Matching Funds

Dartmouth College contributed significant additional resources that were not directly in the scope of the funded part of the project. Examples are : supercomputing resources (the *Discovery* high-performance computing center at Dartmouth); and large-data storage, backup and, temporary cluster node storage in support of large scale computations across the entire film database (RSTOR). The Bregman Audio and Music Research Laboratory at Dartmouth College contributed these resources to the project from its operating costs, thereby enabling large data analytics as part of the startup phase of ACTION.

Additional funds were made available in the form of support for symposia on themes related to the scope of ACTION, such as the Media Ecology Project symposium at Dartmouth in May 2013, and planning for future extensions of ACTION, such as pilot studies to explore connections between ACTION and the MediaThreads scholarly collaboration platform.

Software Development

The development of Action proceeded roughly as planned, building on the Python numerical computing and visualization modules (numpy, matplotlib), the Bregman Toolkit for audio feature analysis, and the OpenCV framework for computer vision. A number of design challenges needed to be met along the way, not least among them was to develop schema and formats for metadata, features, database structured information, and a logical organization of software modules. After several iterations we settled on a heterogeneous solution employing JSON for storing metadata about media, python structured arrays to represent information about film authorship and production data, and python's built-in binary formats to store color, sound, and motion features. Whilst this multi-faceted approach is usable, and practical for the purposes of our use cases and examples, we would like to develop a more integrated format to contain the multiple types of data. Such integration is the subject of ongoing development as we seek to eliminate unnecessary programming complexity from the work flow.

Publishing and Dissemination

In addition to the open-source data and software available at the Web sites, listed at the top of this document, aspects of the ACTION project have been publicly presented in the following national and international conferences and symposia:

Mark Williams, ACTION overview, *Dartmouth Media Ecology Project Symposium*, Dartmouth College, USA, May 17-18, 2013.

Mark Williams, Audio-visual Cinematic Toolbox for Interaction, Organization, and Navigation (ACTION) and the Media Ecology Project, Panel presentation: *Digital Humanities: New Opportunities for Funding, Research, and Access*, Association of Moving Image Archivists (AMIA) Conference, Richmond, Virginia – November 6-9, 2013

Tom Stoll, Mark Williams and Michael Casey, Action: Cross-modal cinematics for analyzing the joint audio-visual structure of film, poster presentation, *Digital Music Research Network Symposium (DMRN+8)*, London, UK, Dec. 17, 2013.

Michael Casey and Mark Williams, Investigating Film Authorship with the ACTION Toolbox, *Archives and Algorithms* Panel Presentation at *Society of Cinema and Media Studies Conference*, Seattle, WA, March 18th - 23rd, 2014.

Michael Casey, *One Million Seconds: Time and Motion Picture Study*, a generative film computed by searching and matching thin slices of audio in the 140 films in ActionDB. Submitted for presentation at the *Digital Musics and Digital Arts Expo*, Dartmouth College, May 2014. Submitted for installation at major international venues: <http://bregman.dartmouth.edu/~mcasey/OneMillionSeconds> [Mar 31, 2014].

In development are a journal article to be submitted to a Film and Media journal publication and second technical article to be submitted a multi-media technical publication, such as *ACM Multimedia*. Other products will be forthcoming as we realize the longer-term implications of the work started in ACTION.

3. Accomplishments

One of the more significant themes of this project that became evident as we developed ACTION is the creative tension that can exist in the Digital Humanities between what Alan Liu⁷ and others have termed close versus distant readings.

Most of the critical traditions in the Arts and Humanities are grounded in some form of close reading, especially when there is an attention to aesthetic form as regards textual analysis. Most humanists have experienced a specific text or even a portion of a text that has proven to be

⁷ Lie, A., "The Meaning of the Digital Humanities." *Proceedings of the MLA*, 128 (2013): 409-23

inspiring in some regard and has led to new directions of thought and research. Developing a thesis about the systematicity of a text, or the avant-garde or idiosyncratic aspects of a text, or even non-systematic methods and contexts of critical inquiry, are typically motivated and driven by a close reading of at least a portion of a specific text. This has been extremely productive for many generations.

But for machine-reading purposes, such a small sample size or dataset would be considered insignificant and unacceptable. One of the advantages to pursuing a new approach to auteur analysis via machine reading is that individual directors of note are generally recognized as such because they have worked reasonably often. If a director's filmography contains at least 8-10 titles, they were considered to be candidates for possible research in ACTION. Factors such as whether a good representative sample of this work was readily available on dvd also played a role in the selection process. It made great sense to focus on the films of Alfred Hitchcock at the center of our research, since his career was so renowned for many decades and because his films have attracted such a sustained attention in film criticism and theory. Selecting additional directors who could represent varied alternative aesthetics from the "classical" form that Hitchcock arguably defined was part of the subsequent selection process.

The feature histograms, similarity matrices, and other visual representations that were produced regarding the Hitchcock canon were very instructive regarding the close versus distant reading paradigm. Viewpoints proved to be important, with plotting of similarity matrices of Hitchcock films in alphabetical order yielding no conclusive result, which was an expected result. But the same visual representation of the relationships between films presented in chronological order revealed what appears to be different patterns of consistency between the earlier films in his career and the latter films, during what many consider to be his "mature" and more experimental period. The dividing line between these two periods is *Rope*, his most experimental film in certain formal regards. This mapping alone is strongly suggestive of the potential for ACTION as a tool for broad historical analyses via "distant" machine-reading techniques..

The feature visualizations for individual films are also very instructive and compelling. On some occasions, the difference between the visual and the aural representations of a given film were remarkably dissimilar. This comparison is profound for example regarding *Rope*, which featured many innovative visual techniques to suggest one long but visually dynamic uninterrupted take. The film has a more regularized color palate than many classical films, since essentially shot in one apartment set. But the sound features exhibit uniformity, since the soundtrack consists of what is essentially the dialogue of a stage play, without music and featuring only a select number of effects. Comparison of the two modalities demonstrates this contrast decisively.

Understanding and learning to "read" the visual summaries of within-film structure and between-film relationships has produced its own series of reflective interventions in the close versus distant reading paradigm. Indeed, we would suggest that this process might be considered to avail the close reading tendencies of the Arts and Humanities back to something like a full circle. Whether representations such as similarity matrices with their variances of

intensity and hue make some kind of "sense" to the human reader is one key step among many in assessing the progress of the algorithms in producing qualitative machine-reading results. This iterative process is significant to underscore. As humanists, we know that vision is a biological and physiological activity. But we also know that vision is cultural: we learn how to see. The machines are just learning how to see. As humanists, we need to play a vigilant role in addressing and contributing to this important and historic process.

The feature visualizations in ACTION also contribute to a return to close reading in their distinctive and sometimes arresting and beautiful patterns. It will not be a surprise should these artifacts themselves inspire future work in the Arts and Humanities. Audiences at our presentations of this work are quite taken with the similarity matrices as objects of visual culture. Animators and experimental filmmakers may prove to be keen to work with these materials. Many audience members have speculated that the visual patterns might suggest physical manifestations in the form of textiles or quilts, etc. The legacy for close reading to inspire new and unforeseen critical and expressive capacities may well apply to the features themselves, as products of distant reading.

Hence, we consider that the key accomplishments of ACTION are the additional knowledge and understanding that is acquired by the transformation of long-duration time-based media into more static representations that can be read in new ways. The dissemination of these images, and their interpretations will likely be a factor in transforming the way that scholars, and lay persons, understand and digest large-scale narrative forms.

Audiences and Evaluation

We determined that the primary audiences for ACTION would be film and media scholars seeking new computational methods in research, media archive professionals, creative professionals seeking tools to support generative media projects, educators seeking new ways for students to learn about film and its relationship to other media, and students.

The materials and methods of ACTION are publicly available via the Web repositories listed at the top of the document. At this stage of the project there has not been sufficient time to do more than the most preliminary testing and evaluation. Among the tests that we have conducted are consistency checks between film data, features, metadata, and visual outputs. It is essential that the data that is disseminated have a high degree of accuracy and integrity. Ensuring these properties has been a fundamental goal of ACTION. It remains to be seen whether the resources will enjoy wide-scale uptake in the scholarly community. For our part, we will continue to promote and disseminate the materials among our respective academic and creative communities.

Continuation of the Project and Long-Term Impact

Both of the PIs are actively engaged with follow-on projects from ACTION.

Michael Casey is developing a series of works that explore large media archives from the perspective of generative film. *One Million Seconds* is one such work that has been submitted

for presentation at various film exhibitions, conferences, and festivals. A key contact person in this regard has been Carlos Cassas Martinez, an experimental film curator and archivist, currently living in Paris, France, who recently visited Dartmouth College to co-teach a new course entitled *Sonic Landscapes* in the department of music. Carlos expressed enthusiasm about the ACTION project, and the generative and Casey's interpretive explorations of its media repositories.

Mark Williams is leading the Media Ecology Project (MEP) which is a digital resource at Dartmouth that enables researchers to digitally access archival moving image collections and contribute back to the archival and research communities through the fluid contribution of metadata and other knowledge. The Media Ecology Project will enable new research capacities toward the critical understanding of historical media and facilitate a dynamic context of research that develops in relation to its use over time by a wide range of users. We intend MEP to support and advocate the essential work of media archives, including the important work currently underway to preserve Dartmouth's media collections.

One of the goals of the Media Ecology Project is to develop strong linkages between the ACTION Toolkit and various other platforms in support of film and media scholarship.

The specific platforms we have engaged and are working to bridge are 1) Mediathread, a classroom platform developed at Columbia University, that we are working to augment as a research platform; 2) Scalar, a digital publishing platform developed at The University of Southern California; and 3) onomy.org, a new online tool which was developed for MEP and will facilitate the creation of controlled vocabularies that can be assigned to online media files. For the time being, onomy.org is a stand-alone site, but this tool will soon be integrated with Mediathread and Scalar for MEP. The Media Ecology Project sits in between and in relation to these platforms and the participating archives, navigating the import, export, and production of metadata across participating archival content that has been engaged by a scholar or team of scholars. In this way we will contribute to the resultant capacities for search and discovery among these media elements in relation to others and realize new forms of research, scholarship, and publication.

We have three pilot projects currently in development:

In conjunction with the Library of Congress we are developing a project regarding early silent film era materials, with an emphasis on the historically significant Paper Print collection. The Paper Print collection is the equivalent of the Rosetta Stone for those who study moving image history in relation to visual culture. We have enlisted Prof. Tami Williams (University of Wisconsin at Milwaukee) and Philippe Gauthier (Universite de Montreal and Harvard University) as a core scholarly team who has engaged several additional members of the renowned DOMITOR research society for this pilot study. The Library of Congress has provided a first batch of 15 Paper Print media files with related metadata for use in this pilot study and will continue to supply additional titles as the project proceeds.

A second pilot study will focus on an important public television program called *In the Life*, which assays the history of gay and lesbian lived experience in the United States. The entire run of that program, plus all of the associated materials involved in the production of that program, will be provided and placed online by the UCLA Film and Television Archive. We have begun to assemble a group of prominent scholars to work on these materials, including Prof. Matthew Tinkcom (Georgetown University), Prof. Michael Bronski (Harvard University), and Prof. Stephen Tropiano (Ithaca College). The UCLA archive anticipates that the programming materials will start to be made available online during this calendar year.

The third pilot study involves the participation of multiple archives and is dedicated to providing more and better access to historical news materials: newsreels, news telecasts, news film, and other associated footage. Archives who will participate include WGBH in Boston, The UCLA Film and Television Archive, The University of South Carolina, The University of Georgia and the Peabody Award Archives, Northeast Historic Film in Maine, and the Library of Congress. A core group of scholars dedicated to this pilot project has been assembled, including Karen Cariani (WGBH Archive), Prof. Mark Garret Cooper (University of South Carolina), and Prof. Ross Melnick (University of California at Santa Barbara).

Appendix A: Summary of ACTION DB Films

Table 2: List of the 140 films in the Action data set

Title	Director	Color	Year
3 Bad Men	John Ford	B&W	1926
3 Godfathers	John Ford	COL	1948
A Serious Man	Coen Brothers	COL	2009
A Woman is a Woman	Jean-Luc Godard	COL	1961
Alphaville	Jean-Luc Godard	B&W	1965
Amistad	Steven Spielberg	COL	1997
Arrowsmith	John Ford	B&W	1931
Barton Fink	Coen Brothers	COL	1991
Belle de Jour	Luis Bunuel	COL	1967
Black Swan	Darren Aronofsky	COL	2010
Blood Simple	Coen Brothers	COL	1984
Blue Velvet	David Lynch	COL	1986
Bringing Up Baby	Howard Hawks	B&W	1938
Burn After Reading	Coen Brothers	COL	2008
Catch Me If You Can	Steven Spielberg	COL	2002
Cheyenne Autumn	John Ford	COL	1964
Death in the Garden	Luis Bunuel	COL	1956
Dersu Uzala	Akira Kurosawa	COL	1975
Detective	Jean-Luc Godard	COL	1985
Diary of a Country Priest	Robert Bresson	B&W	1951
Dreams	Akira Kurosawa	COL	1990
Drunken Angel	Akira Kurosawa	B&W	1948

Duel	Steven Spielberg	COL	1971
Dune	David Lynch	COL	1984
ET	Steven Spielberg	COL	1982
Early Spring	Yasujiro Ozu	B&W	1956
Early Summer	Yasujiro Ozu	B&W	1951
Enthusiasm	Dziga Vertov	B&W	1930
Equinox Flower	Yasujiro Ozu	COL	1958
Eraserhead	David Lynch	B&W	1977
Exterminating Angel	Luis Bunuel	B&W	1962
Fargo	Coen Brothers	COL	1996
Fata Morgana	Werner Herzog	COL	1971
Foreign Correspondent	Alfred Hitchcock	B&W	1940
Fort Apache	John Ford	B&W	1948
Frenzy	Alfred Hitchcock	COL	1972
Gentlemen Prefer Blondes	Howard Hawks	COL	1953
Grapes of Wrath	John Ford	B&W	1940
Hangmans House	John Ford	B&W	1928
High and Low	Akira Kurosawa	B&W	1963
His Girl Friday	Howard Hawks	B&W	1940
How Green Was My Valley	John Ford	B&W	1941
How to Survive a Plague	David France	COL	2012
I Was Born But	Yasujiro Ozu	B&W	1932
In Praise of Love	Jean-Luc Godard	COL	2001
Indiana Jones and the Last Crusade	Steven Spielberg	COL	1989
Indiana Jones and the Temple of Doom	Steven Spielberg	COL	1984
Inland Empire	David Lynch	COL	2006

Ivans Childhood	Andrei Tarkovsky	B&W	1962
Jeanne Dielman	Chantai Akerman	COL	1975
Kagemusha	Akira Kurosawa	COL	1980
Koyaanisqatsi	Godfrey Reggio	COL	1982
L'Age D Or	Luis Bunuel	B&W	1930
Las Hurdes	Luis Bunuel	B&W	1933
Late Autumn	Yasujiro Ozu	COL	1960
Late Spring	Yasujiro Ozu	B&W	1949
Le Petit Soldat	Jean-Luc Godard	B&W	1963
Les Dames du Bois de Boulogne	Maya Deren	B&W	1945
Los Olvidados	Luis Bunuel	B&W	1950
Lost Highway	David Lynch	COL	1997
Madadayo	Akira Kurosawa	COL	1993
Made in USA	Jean-Luc Godard	COL	1966
Marnie	Alfred Hitchcock	COL	1964
Meshes of the Afternoon	Alexander Hammid	B&W	1943
Millers Crossing	Coen Brothers	COL	1990
Mother	Vsevolod Pudovkin	B&W	1926
Mr and Mrs Smith	Alfred Hitchcock	B&W	1941
Mulholland Drive	David Lynch	COL	2001
Munich	Steven Spielberg	COL	2005
My Darling Clementine	John Ford	B&W	1946
My Name is Ivan	Andrei Tarkovsky	B&W	1962
Nazarin	Luis Bunuel	B&W	1959
No Blood Relation	Mikio Naruse	B&W	1932
North by Northwest	Alfred Hitchcock	COL	1959

Notorious	Alfred Hitchcock	B&W	1946
Notre Musique	Jean-Luc Godard	COL	2004
O Brother Where Art Thou	Coen Brothers	COL	2000
Only Angels Have Wings	Howard Hawks	B&W	1939
Passing Fancy	Yasujiro Ozu	B&W	1933
Pi	Darren Aronofsky	B&W	1998
Pierrot le Fou	Jean-Luc Godard	COL	1965
Psycho	Alfred Hitchcock	B&W	1960
Raiders of the Lost Ark	Steven Spielberg	COL	1981
Raising Arizona	Coen Brothers	COL	1987
Ran	Akira Kurosawa	COL	1985
Rashomon	Akira Kurosawa	B&W	1950
Rear Window	Alfred Hitchcock	COL	1954
Rebecca	Alfred Hitchcock	B&W	1940
Requiem for a Dream	Darren Aronofsky	COL	2000
Rio Bravo	Howard Hawks	COL	1959
Robinson Crusoe	Luis Bunuel	COL	1954
Rope	Alfred Hitchcock	COL	1948
Saving Private Ryan	Steven Spielberg	COL	1998
Schindlers List	Steven Spielberg	COL	1993
Seven Samurai	Akira Kurosawa	B&W	1954
Shadow of a Doubt	Alfred Hitchcock	B&W	1943
Soigne ta Droite	Jean-Luc Godard	COL	1987
Stagecoach	John Ford	B&W	1939
Straight Story	David Lynch	COL	1999
Strangers on a Train	Alfred Hitchcock	B&W	1951

Sullivans Travels	Preston Sturges	B&W	1941
That Obscure Object of Desire	Luis Bunuel	COL	1977
The 39 Steps	Alfred Hitchcock	B&W	1935
The Big Lebowski	Coen Brothers	COL	1998
The Big Sleep	Howard Hawks	B&W	1946
The Birds	Alfred Hitchcock	COL	1963
The Color Purple	Steven Spielberg	COL	1985
The End of Summer	Yasujiro Ozu	COL	1961
The Fountain	Darren Aronofsky	COL	2006
The Hidden Fortress	Akira Kurosawa	B&W	1958
The Hudsucker Proxy	Coen Brothers	COL	1994
The Lady Vanishes	Alfred Hitchcock	B&W	1938
The Man Who Knew Too Much	Alfred Hitchcock	COL	1956
The Man Who Shot Liberty Valance	John Ford	B&W	1962
The Milky Way	Luis Bunuel	COL	1969
The Mirror	Andrei Tarkovsky	B&W	1975
The Pleasure Garden	Alfred Hitchcock	B&W	1925
The Quiet Man	John Ford	COL	1952
The Sacrifice	Andrei Tarkovsky	COL	1986
The Searchers	John Ford	COL	1956
The Wrestler	Darren Aronofsky	COL	2008
The Wrong Man	Alfred Hitchcock	B&W	1956
Throne of Blood	Akira Kurosawa	B&W	1957
Tokyo Chorus	Yasujiro Ozu	B&W	1931
Tokyo Story	Yasujiro Ozu	B&W	1953
Tokyo Twilight	Yasujiro Ozu	B&W	1957

Torn Curtain	Alfred Hitchcock	COL	1966
Tout Va Bien	Jean-Luc Godard	COL	1972
Tristana	Luis Bunuel	COL	1970
Twin Peaks	David Lynch	COL	1992
Twin Peaks Ep1	David Lynch	COL	1990
Un Chien Andalou	Luis Bunuel	B&W	1929
Uncle Boonme Who Can Recall His Past Lives	Apichatpong Weerasethakul	COL	2010
Vampyr	Carl Theodor Dreyer	B&W	1932
Vertigo	Alfred Hitchcock	COL	1958
Viridiana	Luis Bunuel	B&W	1961
War Horse	Steven Spielberg	COL	2011
Weekend	Jean-Luc Godard	COL	1967
Wild at Heart	David Lynch	COL	1990
Young Mr Lincoln	John Ford	B&W	1939

Appendix B: Example Web Tutorial in Action

Tutorial Two: Access Video Data — ACTION: Tools for Cinematic Information Retrieval - Mozilla Firefox

Tutorial Two: Access Video Da... +

bregman.dartmouth.edu/action/tutorial_two_access.html

ACTION: Tools for Cinematic Information Retrieval

PREVIOUS | NEXT | MODULES | INDEX

Tutorial Two: Access Video Data

Abstract

This tutorial will demonstrate methods to access, display, and generally use video data gathered through ACTION analysis function.

Prerequisites

- Data analyzed through ACTION, which can take several forms (see below).
- See the first Tutorial for more information.

Raw Data: Files

The raw analysis is stored in binary files. To summarize, these files, identified by descriptive extensions, are:

1. MOVIE_TITLE.color_lab - Normalized $L^*a^*b^*$ color space: histograms summarizing the distribution of color and luminescence (brightness) over each frame: 8-by-8 grid and full screen.
2. MOVIE_TITLE.opticalflow24 - Optical flow data: motion vectors based on an intermediate corner-detection step: 8-by-8 grid and full screen.
3. MOVIE_TITLE.phasecorr - Phase correlation data: 8-by-8 grid and full screen.
4. MOVIE_TITLE.tvl1 - TV-L1 optical flow data: 8-by-8 grid and full screen.

For a detailed list of the meanings of the various extensions, please see our [Overview](#).

Access

Import ACTION Classes, Create a Histogram Object, and Play a Movie

We will use `color_features_lab` for this first example. Let us assume that there are the following files: `~/Movies/Vertigo.mov` and `~/Movies/Vertigo.hist`. Execute the following commands in your Python interpreter:

```
from action.suite import *
vertigo_cfl = ColorFeaturesLAB('Vertigo')
vertigo_cfl.playback_movie()
```

You will see the analysis data as bar graphs in the Histogram window. When you have seen enough, press escape to exit the viewer. Below are still shots of the histogram visualization:

TABLE OF CONTENTS	
ACTION - data overview	
ACTION - raw data	
ACTION - interactive segmentation	
ACTION - source code and links	
Tutorial One - setup and analysis	
Tutorial Two - access to video data	
Tutorial Three - access to audio data	
Example One - simple clustering	
Example Two - centers of mass and random sampling	
Example Three - dissimilarity plots	
Example Four - viewing distributions of color data	
Example Five - in-depth example of segmentation	
Example Six - in-depth example of a simple director prediction ML task	
Example Seven - example of viewing relationships among films with multidimensional scaling (MDS) -	
color_features_lab - color and spatial frame-by-frame analysis and visulaization	
opticalflow - Lukas-Kanade optical flow/motion vector frame-to-frame analysis and visulaization	
opticalflow_tv1 - TVL optical flow frame-to-frame analysis and visulaization	
phase_correlation - phase correlation frame-to-frame analysis and visulaization	
segment - segmentation and container data structure	
actiondata - data analysis and view routines	

Tutorial Two: Access Video Data — ACTION: Tools for Cinematic Information Retrieval - Mozilla Firefox

Tutorial Two: Access Video Da... +

bregman.dartmouth.edu/action/tutorial_two_access.html

Google

which is indexed in the flattened output array like so:

```
X   X   X   X
X   0-47 48-95 X
X   96-143 144-191 X
X   X   X   X
```

Optical Flow

The same work flow applies to utilizing the optical flow data.

```
oflow = OpticalFlow('Vertigo')
myseg = Segment(60, duration=60)
oflow_data = oflow.middle_band_opticalflow_features_for_segment(myseg)
```

The optical flow data is collected for all 24 frames in each second, but our access functions have a default stride of 6 frames built in:

```
oflow_data.shape
>>> (240, 256)
```

Phase Correlation

Phase Correlation is identical to OpticalFlow for access...

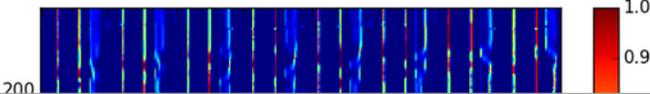
```
pcorr = PhaseCorrelation('Vertigo')
# etc...
```

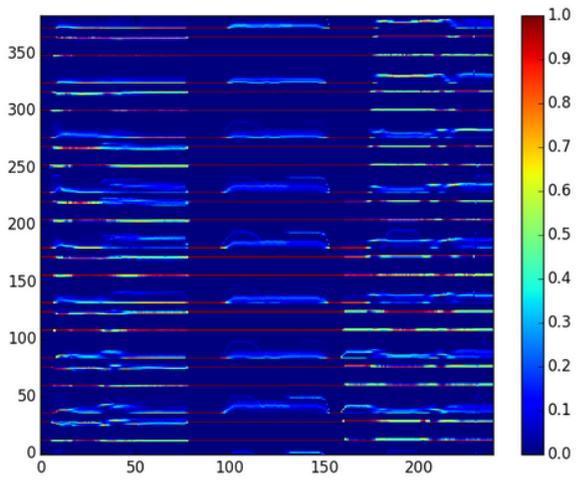
Visualizing the Data as Numpy Arrays

Finally, let us look at some examples of visualizing this data. Recall that `histograms_for_segment` will return a tuple: the full screen histogram data and 16 gridded histograms...

```
from action.suite import *
cfl = ColorFeaturesLAB('Vertigo')
myseg = Segment(60, duration=60)
mb_data = cfl.middle_band_color_features_for_segment(myseg)

# ACTION has has a function, borrowed from Bregman, which we use here to
imagesc(mb_data)
```





Now look at the all the data from the gridded histogram. You should see that there are now 16 histograms stacked one on top of the other.

```
gridded_data = cfl.gridded_color_features_for_segment(myseg)
imagesc(gridded_data.T)
```

