

VideoCycle: User-Friendly Navigation by Similarity in Video Databases

Christian Frisson¹, Stéphane Dupont¹, Alexis Moinet¹,
Cécile Picard-Limpens¹, Thierry Ravet¹, Xavier Siebert², and Thierry Dutoit¹

¹ Université de Mons, TCTS lab, Boulevard Dolez 31, B-7000 Mons, Belgium

² Université de Mons, MathRO lab, Rue de Houdain 9, B-7000 Mons, Belgium
numediart Institute, Boulevard Dolez 31, B-7000 Mons, Belgium

Abstract. VideoCycle is a candidate application for this second Video Browser Showdown challenge. VideoCycle allows interactive intra-video and inter-shot navigation with dedicated gestural controllers. MediaCycle, the framework it is built upon, provides media organization by similarity, with a modular architecture enabling most of its workflow to be performed by plugins: feature extraction, clustering, segmentation, summarization, intra-media and inter-segment visualization. MediaCycle focuses on user experience with user interfaces that can be tailored to specific use cases.

1 The MediaCycle Framework

The MediaCycle framework has been developed at the numediart Institute¹ of the University of Mons since 2008. Its scope is to provide tools to create applications for the navigation by content-based similarity into multimedia databases, currently audio, image, video, text, sensor signals and files. It focuses “by design” on fostering new artistic practices (the numediart Institute aims at improving new media arts technologies) and stresses on providing a tailored user experience.

MediaCycle yields to modularity through a plugin architecture, from media types to all the steps of the workflow of organization by similarity: feature extraction, clustering, segmentation, summarization, visualization. Similarly to tools for computer-aided design, the generic graphical user interface of MediaCycle provides a main “canvas-like” space with browser and timeline views, and control panels associated to plugins. Some panels are generated automatically from plugin parameter serialization. We will describe in the following subsections all steps of the workflow of organization by similarity as offered by MediaCycle.

1.1 Audio and Video Feature Extraction

MediaCycle provides several algorithms for the extraction of low-level features. For audio signals, wrapper plugins around the YAAFE library and the Vamp

¹ MediaCycle:

<http://www.mediacycle.org> and numediart: <http://www.numediart.org>

audio analysis plugin system² allow access to verified and acknowledged algorithms from the music information retrieval and audio signal processing communities.

We initially developed features specific to dance videos [1], aside generic features such as: global motion orientation, optical flow, blob pixels speed, color moments; all based on OpenCV³ Time-dependent features are summarized by the calculation of their mean and standard deviation.

1.2 Clustering and Neighborhoods

MediaCycle proposes two modes of navigation: “clusters”, where the whole database is displayed on the screen, and “neighbors” where a user-defined small scale number of nearest neighbors are presented to the user. For clustering, we have been using K-Means. MediaCycle suggests neighbors using either an Euclidian distance or Pareto ranking, as explained in [1].

1.3 Segmentation

MediaCycle provides several segmentation methods, based respectively on:

1. the self-similarity matrix to compute a signal of novelty (see [2]);
2. the Bayesian Information Criterion (BIC) with two variants : browsing the data frame-by-frame or through a “divide-and-conquer” approach (see [2]);
3. a third-party library, Johan Mathé’s shotdetect⁴, segmenting by shots using consecutive frame pixel-by-pixel threshold-based color comparison.

Methods 1 and 2 are media-agnostic, 3 requires only video content. Once pre-computed, segmentation profiles associated to each method can be applied on the fly and affect the visualization content.

1.4 Visualization: Browser and Timeline

MediaCycle offers two views for navigation into media content: the “browser” positions media nodes in a 2D space for inter-media navigation, the “timeline” is dedicated to intra-media navigation. Several visualization techniques can be chosen for the browser: in clustering mode, the “propeller” tries to spread efficiently a user-definable number of clusters, a scatter plot (one user-definable feature dimension per x and y axes), a polar representation (one on radius, another on angle); in neighbors mode, radial and flat tree views show hierarchical nearest neighbors of given media elements. A fisheye distortion highlights segments hovered in any of both views with propagation to the other view.

² YAAFE:

<http://yaafe.sourceforge.net> and VAMP: <http://www.vamp-plugins.org>

³ OpenCV: <http://opencv.org>

⁴ shotdetect: <https://github.com/johmathe/Shotdetect>

1.5 Interaction and Devices

MediaCycle allows server-client application through the Transmission Control Protocol (TCP) [1] and external control through the OpenSoundControl (OSC) protocol [3] extended to TUIO, a protocol for Tangible User Interfaces. In short, MediaCycle applications can be displayed on touch-screen interfaces [1] and controlled via USB/HID devices such as jog wheels and multi-touch trackpads [3].

2 Our Approach to Known-Item Search with VideoCycle

We believe that our user interface VideoCycle built upon the MediaCycle framework can be used to browse single video files for known-item search as follows:

- Once the video file is imported (features are extracted and the video file is segmented), segments of the video are displayed in the browser view and organized by similarity, the user can choose the weights of the features to base the similarity on (here by checkboxes) and the positioning methods.
- The timeline view allows to navigate inside the imported video file, with a video player playing back the file, a “slider” view with an overlaid visualization technique that the user can choose among several (keyframes evenly-distributed in time, slit-scan, none), and a focus slider whose time interval is determined by the span of the aforementioned slider.
- A trackpad offers translation/rotation/zoom/reset gestures for navigation in the browser and skip/swipe gestures for the timeline. We pair it to a jog wheel, whose dial scrubs the timeline precisely and spring-loaded wheel adjusts the playback speed. Its keys are assigned to: previous/next segment, cue in/out and send the found “known item”, select features and visualizations.

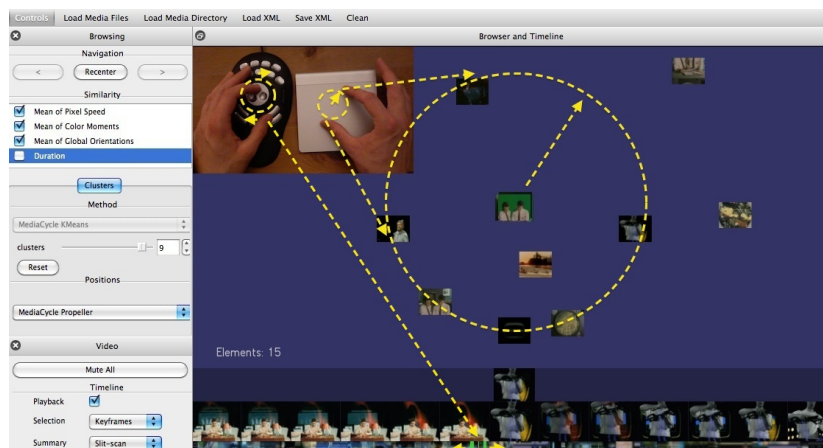


Fig. 1. *The Remote Controller* (2003) by People Like Us alias Vicki Bennett in VideoCycle: zooming the inter-segment “browser” (up), scrubbing the intra-media “timeline” (down) with the jog shuttle wheel (down). Yellow dotted lines/arrows are annotations.

References

1. Tardieu, D., Siebert, X., Mazzarino, B., Chessini, R., Dubois, J., Dupont, S., Varni, G., Visentin, A.: Browsing a dance video collection: dance analysis and interface design. *Journal on Multimodal User Interfaces* 4(1), 37–46 (2010)
2. Dupont, S., Frisson, C., Urbain, J., Mahmoudi, S., Siebert, X.: Mediablender: Interactive multimedia segmentation. In: Dutoit, T. (ed.) *QPSR of the Numediart Research Program*, vol. 4, pp. 1–6 (2011), <http://www.numediart.org>
3. Frisson, C., Dupont, S., Siebert, X., Tardieu, D., Dutoit, T., Macq, B.: DeviceCycle: rapid and reusable prototyping of gestural interfaces, applied to audio browsing by similarity. In: *Proc. of the New Interfaces for Musical Expression, NIME* (2010)