SIMILARITY IN MEDIA CONTENT: DIGITAL ART PERSPECTIVES

Christian Frisson, Stéphane Dupont, Xavier Siebert & Thierry Dutoit

This essay examines how media content navigation by similarity can foster new practices in digital arts, blurring the boundaries between composing/performing, curating/authoring, creating/interpreting. With MediaCycle, a framework for browsing media databases by similarity, we created several prototypes: a website for browsing dancers' identities through video recordings, a collaborative dancefloor for music creation.
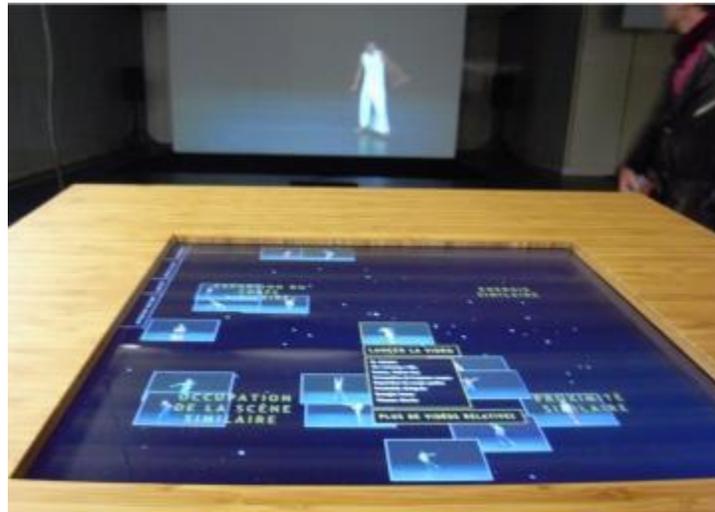
Fig 1. The DANCERS! Relational Browser seen on a multitouch booth.
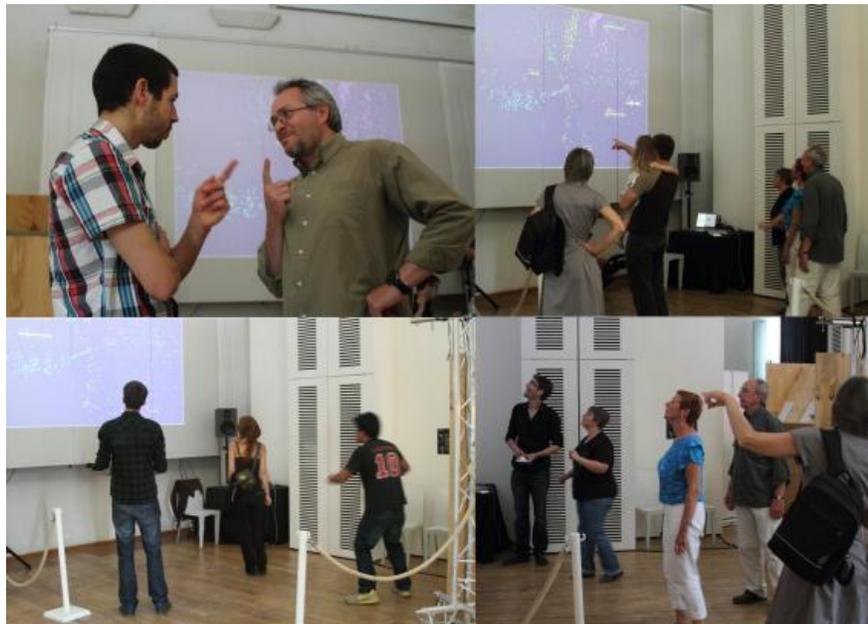
Fig 2. LoopJam, a collaborative dancefloor for music creation

Introduction

Art and science cycle in loops: digital artists have been appropriating themselves technological advances in scientific fields such as telecommunication, signal processing and information visualization; while engineers and scientists are often inspired by concepts imagined by science fiction authors or fueled by artists trying to push the limits of these technologies when applied in their new media artworks. Beyond storage and transmission, one current limitation in technologies regarding media content is how to maintain an overall understanding of the information provided by each new media recording against the profusion and proliferation of newly-created and user-generated content. In his book *The Language of New Media*, [1] Lev Manovich stresses the importance of the "representation" of media objects in a database, how these are organized together as a mirror to the knowledge that these provide. When hypermedia links attempted to connect media objects one another into the world wide web through metadata and labels, the intervention of humans to annotate the massive data produced continuously is becoming more and more difficult. Recent progress in automated content-based analysis of media recordings opens new perspectives for structuring and navigating media databases, for instance by organizing the media objects by similarity.

Similarity in media content

THE CONCEPT OF SIMILARITY THROUGH (DIGITAL) ARTS

Besides the catchphrase "Similarities and differences..." used as a leitmotiv to begin the title of some essays dissecting specific art genres or styles, not much literature is centered around the concept of "similarity" in (digital) arts. Instead, an understanding of this concept can be grasped by focusing on specific uses of this term and peripheral vocabularies.

Literally, "similarity" names one of the Gestalt Laws in cognitive psychology: the human visual perception tends to discriminate outsider elements from groups in a visual collection. It has high implications in human-computer interaction and objector graphic design, such as how people can remember the content of a scene by gaining a structured knowledge of it, and how people can be attracted by objects or visuals that remind them of features of other objects they might have liked previously.

More specifically, humans have ever been fascinated by the complexity of natural phenomena showing repeating patterns: lightning bolts, clouds, tree-shaped vegetables, viscous flows. This concept of "self-similarity" has been inspiring fractal art, and is salient in paintings from Escher and compositions from Bach, as examined by Hofstadter. [2]

Each of both definitions, similarity as Gestalt law and self-similarity, underline one major aspect of similarity in media content: similarity can be used to characterize and compare several elements of a database (*inter-media*), and focus on the structure and contents of one single element (*intra-media*).

Several aspects of similarity can be used to describe the nature of artworks. First the relation of art to people: who originated it versus who keeps it alive in people's memories (the issues of authoring, composition, interpretation, performance, appropriation, inspiration, creativity, emulation, reproduction, recomposition, curation, restoration, preservation); second the relation of art works between them-

selves: what the work represents and how it place itself in a context of other works (identity, authenticity, singularity, resemblance). These concepts emerge in movie remakes (for instance re-performed or "sueded" movies in Michel Gondry's *Be Kind Rewind*), montages (Orson Welles' *F is for Fake*), collages (works of Jennifer and Kevin McCoy, Vicky Bennett aka People Like Us), cover bands and song covers (from re-interpretation to resampling with John Oswald's *Plunderphonics*).

How can organization by similarity of media content serve digital arts? Two tracks can be elicited from the aforementioned definitions and connotations: the first would consider the analysis, comparison, classification of existing art pieces; the second would focus on the generation of new art pieces based on content organization and navigation.

COMPUTATIONAL SIMILARITY

Computational similarity analysis consists in providing a machine interpretation of the salient characteristics of media objects, by applying feature extraction algorithms that downsize the data contained within the media into threads of specific information, and comparing the distribution of these features over a database using adapted distance metrics. Features can be content-based, that is extracted directly from the digital representation of the media object; or semantic-based, provided by manual annotation that labels specific elements of a scene. The book chapter from [3] provides a recent and detailed overview on the state-of-the-art of interactive representation of media databases, focusing on image browsing applications. It concludes mainly that most applications still miss a proper user-friendly interaction. Examples of higher-level content-based interpretation are identification of cover songs and style and genre detection for paintings.

Artistic works focusing on media similarity

Several end-user applications using content-based similarity have flourished recently, particularly recommender systems such as LastFM (http://www.last.fm), however artistic installations or works are more seldom.

Most notably, George Legrady's Pockets Full of Memories (2001-2007) [4] (see http://www.georgelegrady.com) make use of semantic-based similarity: both installation and formerly accessible online as a website, it uses the Kohonen self-organizing map algorithm to organize on a screen snapshots of everyday objects scanned by visitors in the installation, based on the textual description typed by visitors, thus proposing an emergent ordering since each individual induces her/his own perception of the object entered in the database that might differ from the visual similarities.

Martin Wattenberg's The Shape Of Song (2001) (see http://www.bewitched.com/song.html) proposes abstract visualizations in arc diagrams of the musical structure of hundreds of songs. Starting from MIDI transcripts of the musical pieces (sequences of notes defined by pitch, onset and duration), summaries are computed using the maximal matching pair algorithm and other rules to reduce the complexity of this algorithm. These summaries are visualized using overlaid semi-circular arcs whose thickness corresponds to the duration of the repeated musical passages, therefore incidentally underlining their relevance. Several non-interactive printouts of these arc diagrams have been exhibited at the Generator.x 2005 art event in Oslo, Norway.

MediaCycle (http://www.mediacycle.org) is a software framework for organizing media content by similarity. Since 2008, it has been developed towards a modular architecture, supporting several media types (so far: audio, image, video, text), various media-specific algorithms for content-based low-level feature extraction, plugins for clustering (particularly the K-Means algorithm) and positioning media elements in a 2D space. Designed deliberately cross platform using open-source libraries, it was initially targeted for major computer operating systems (Apple OSX, Linux such as Ubuntu and Microsoft Windows), and more recently for mobile platforms (Apple iOS and Google Android). It provides exemplar single-media standalone applications for desktops and laptops, and server/client applications for mobile devices and servers in the cloud. OpenSoundControl (OSC) networked communication support has also been added (see [5]) so as to control the navigation in media content using off-the-shelf devices such as jog wheels and 3D cameras.

The DANCERS! Relational Navigator

Choreographer Bud Blumenthal's DANCERS! project traveled in Belgium and France in 2009 and scheduled shootings to audition dancers by recording them with top and front cameras while they were asked to improvise dance moves on music played back without choreography. The DANCERS! installation is composed of a multitouch booth where the public can select dancers videos to be displayed on a video projector at the original body scale, with surround sound. Dancers videos are selected using the DANCERS! Relational Browser that is also available online (http://www.dancersproject.com/interactive/) and is powered by the MediaCycle framework. Videos were automatically analyzed so as to provide a content-based 2D representation of dancers groups sorted by features such as: position (mean, standard, max), speed, ratio of the dancer bounding box and contraction index, space occupation and trajectory (small/large, compact/sparse, proscenium/rear as preferred zone). This installation provides an interactive and alternative way of browsing through dancers videos beyond retrieval through standard metadata such as artists names. This work has been described in more detail in Tardieu et. al. [6]

LoopJam, a collaborative dance floor for music creation

In 2011, we proposed LoopJam (see http://www.numediart.org/demos/loopjam), an interactive installation that features a sound map with audio loops organized by similarity of timbre using MediaCycle and a Microsoft Kinect depth-sensing camera so as to map the visitors' positions to cursors on the map hovering sounds with audio feedback, loops being synchronized in terms of tempo by the sound engine. A few people from the audience can thus carefully select sounds and collaboratively create an "improvised music composition". Organization by similarity proves itself to be useful since audio loops from similar instruments tend to be grouped together, hence small movements of visitors would provide a slight variation in terms of content in the sound rendering. This installation revisits the artist to audience relation and interaction since the DJ or curator of the installation responsible with choosing the sound library to be browsed within the installation can select her/his trademark sounds conveying a personalized musical identity.

Conclusions and perspectives

This paper provided a brief overview of artistic works making use of content- or semantic-based organization by similarity of media databases. Two main disciplines are seen as beneficial from requiring media content similarity: media preservation and documentation, and (realtime) media recomposition.

The described works use one single media type for the content-based analysis. Combining features from multiple media should enhance the robustness of the representation, for instance using both the video and soundtrack in the case of film documentation so as to analyze the relation between image and sound, in parallel to expert interpretation models, [7] using interactive summaries. Similarly, when classifying music albums, the visual artworks from the album, the textual description, and the lyrics might add be relevant as well to sort and organize a collection.

Regarding media recomposition, alternative and engaging interaction methods might be investigated to browse media collections in realtime (particularly for DJs and VJs), such as query by sketching to retrieve visual elements from a database while drawing, or query by whistling or beatboxing to create musical content. These practises will certainly take advantage from advances in scientific fields such as multimedia information retrieval.

Acknowledgments

References and Notes:

1. *L. Manovich, The Language of New Media (Cambridge, MA: The MIT Press, 2001), ISBN: 978-0-26-213374-6.*
2. *D. Hofstadter, Gödel, Escher, Bach: an Eternal Golden Braid (New York: Basic Books, 1979), ISBN: 0-465-02685-0.*
3. *S. Marchand-Maillet, D. Morrison, E. Szekely, and E. Bruno, "Interactive Representations of Multimodal Databases," in Multimodal Signal Processing - Theory and Applications for Human-Computer Interaction, ed. J. Thiran, F. Marqués. and H. Bourlard, 279-307 (San Diego, CA: Academic Press, 2010), ISBN-13: 978-0-12-374825-6.*
4. *V. Vesna, ed., Database Aesthetics: Art in the Age of Information Overflow(Minnesota: University of Minnesota Press, 2007), ISBN-13: 978-0-8166-4118-5.*
5. *C. Frisson, S. Dupont, X. Siebert, T. Dutoit and B. Macq, "DeviceCycle: rapid and reusable prototyping of gestural interfaces, applied to audio browsing by similarity," in Proceedings of the New Interfaces for Musical Expression++ (NIME++), 2010.*
6. *D. Tardieu, X. Siebert, B. Mazzarino, R. Chessini, J. Dubois, S. Dupont, G. Varni, and A. Visentin, "Browsing a dance video collection: dance analysis and interface design," in Journal on Multimodal User Interfaces 4, No. 1 (2010): 37–46.*
7. *M. Chion, Audio-vision: sound on screen (New York: Columbia University Press, 1994), ISBN: 0-231-07898-6.*