

Interactive Visualizations of Video Tours in Space and Time

Ana Jorge
LaSIGE, Faculdade de Ciências
Universidade de Lisboa
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
2829-516 Caparica, Portugal
ananunesjorge@gmail.com

Sérgio Serra
LaSIGE, Faculdade de Ciências
Universidade de Lisboa
1749-016 Lisboa, Portugal
sergioserra99@gmail.com

Teresa Chambel
LaSIGE, Faculdade de Ciências
Universidade de Lisboa
1749-016 Lisboa, Portugal
tc@di.fc.ul.pt

Videos and movies are increasingly being created, shared and accessed from different platforms and devices, that are supporting georeferencing as a form to enrich their contextualization, demanding for new and more powerful ways to search, browse and view them. An effective access to video is very challenging due to the amount of items available and their inner complexity. Interactive visualization, in complement to video cataloguing, can help to handle this challenge, making information more accessible and useful. Moreover, it is worth doing this navigation ludic, and aesthetically interesting, while providing new ways to search and filter through their properties and their impact on the users. In previous work we focused on the temporal dimension in movies and videos in a way that is possible to explore and access them through time, genre and rating criteria, and inside their contents regarding image, movement, audio, subtitles and emotions. We are now extending our goal by considering the spatial dimension, through trajectories like the ones we find in city tours. We present the design of the main interactive visualizations for navigating georeferenced videos, allowing the user to 1) overview videos shot in a given geographic location in a given moment of time; 2) zoom in the trajectories of individual videos, e.g. by length, speed, age and content of the shootings; 3) access and watch each video content in a specific trajectory, e.g. through color, sound, spoken words, emotions and neighbor connections.

Visualization, Movies, Videos, Content-Based Retrieval, Interaction, Sightseeing, Time, Maps, Tours, Space.

1. INTRODUCTION

The advances in video technology are allowing users to generate, share and access videos and movies in huge amounts, as enormous collections, demanding for new and more powerful ways to search, browse and view them. They are accessed from different platforms and devices, and increasingly video can be georeferenced, allowing to enrich its contextualization. All the richness that makes these video collections or spaces so interesting comes with a challenging complexity to handle - video is not a structured media type and it changes over time, so, perceiving and searching all the content of a video is often not an easy task, even more complex in the presence of a huge amount of videos being published over time. Interactive visualization can help to handle this challenge.

The focus of this work is spatio-temporal data in video trajectories, defined by (Peuquet 1994) as spatial events that consist of sequences of elementary spatial events relating time and space. We are extending our work on visualization of movies and videos by adding the spatial dimension to the temporal dimension, previously addressed through interactive visualizations that permit to explore and access videos and movies over their time of release (Jorge et al. 2012), and the time inside their contents

where their properties are weaved (Jorge & Chambel 2013). Our motivation is two folded: 1) we are further extending our work on MovieClouds (Gil et al. 2012), an interactive web application based on the tag cloud paradigm that allows to explore and access movies through the information conveyed in their content, mainly in the audio and the subtitles, where most semantics is expressed, and for which we already enriched the support for the visual and temporal dimensions (Jorge et al. 2012) (Jorge & Chambel 2013); and 2) we are extending the work on Sight Surfers (Noronha et al. 2012) (Ramalho & Chambel 2013), an interactive web application for sharing, visualizing and navigating georeferenced 360° interactive user-generated videos, as hypervideos, including city tours or more extreme activities like kart racing. These can be experienced in increased immersion and isolation, or synchronized with a map while being played. At the crossroads of these works, we are now addressing the spatio-temporal dimensions and the rich content of videos, both user-generated videos and movies, with a special focus on trajectories like the ones we find in city tours.

Scenarios for visualizations focused on trajectories are wide range in entertainment, learning, and the arts' field, and consider both professional and casual use cases. We believe it is interesting, useful and even

fun to be aware of information about events and trajectories, either in more professional and accurate uses (e.g. the most shot location in Lisbon by night, or the relation between the most shot area with the most noisy, and coloured); or in informal seeking for information (e.g. a tourist that is interested in the greener and calmer part of the city). We present a design study of interactive visualizations that allow to navigate and access video spaces, or collections, and individual videos, in trajectories, through criteria such as the amount of videos shot in a specific area, the speed along the trajectories, the age of the videos, and their content, in terms of colour, sound, spoken words and movement, around the concepts of time and map representations.

In section 2, we present the conceptual framework, and in section 3 the most relevant work that relates to our own, followed by the presentation of our approach in section 4. The paper ends with conclusions and perspectives for future work in section 5.

2. CONCEPTUAL FRAMEWORK

In this work, we explore georeferenced trajectories in videos. We ground our work on a taxonomy that relates types of spatiotemporal data with types of analysis tasks (Andrienko et al. 2011), and on the conceptual frame from Peuquet (1994, 2002) that separates the components in spatiotemporal data, as follows: 1) space (where), 2) time (when), and 3) objects (what), concept that is disentangled by the following three questions: 1+2) 'when' plus 'where' that leads to 'what', and describes objects in a given location and time; 2+3) 'when' plus 'what' that leads to 'where', and describes locations that are occupied by given objects at given times; 1+3) 'where' plus 'what' which result is 'when', and describes the time that objects occupy in given locations.

Extending previous work in which we used abstract in time-based representations (Jorge et al. 2012) (Jorge & Chambel 2013), we now explore space and the map concept since they are the key when it is intended to visualize objects or phenomena with a spatial address (Peuquet 2001)(Kraak et al. 2010). They do not only situate objects and phenomena in some location but also relate to one another (Peuquet 2001). We believe that maps establish a direct and easy connection between a location of interest to the user, and captured on the videos being accessed, and its visual representation, helping to relate and localize the real and the represented space.

Interactive video visualization carries complex challenges, and even more when dealing with space and time that flow in synchrony. Our aim is to make possible the navigation through movies and videos, either collections or items, allowing the linking to their spatial location and thereafter the filtering of the information regarding more accurate findings.

3. RELATED WORK

The work with more similarities with our own follows the space-time concept whether, or not, based on map representations.

3.1 Time-Space Visualization

The static representations that follow represent space in an abstract way, and describe contents of video over time through summarization.

Flickr Flow (Viégas et al. 2009) represents the dominant colours in pictures shot throughout a year. The visual result is a plot of images disposed in a clockwise circular shape, with the different seasons shown in colours. Slit-scan imaging techniques (Levin 2005), often used in interactive art, are adopted in What Did I Miss? (Nunes et al. 2007), a timeline visualization system that allows to explore a video history trace to give mutual presence awareness of participants and their activity in collaborative scenarios, encouraging their connection while working. Last Clock (Angesleva et al. 2005) provides also a visual record of what happens in a given space, the pace and volume of motion, and the moment in time when it happened. Artifacts of the Presence Era (Viégas et al. 2004) presents the events in a museum over time through recorded video images that change in their height by the environmental sound.

3.2 Georeferenced Media Visualization

These works focus on georeferenced information, and some allow to access or navigate through collections and contents of movies and videos.

The work by (Cruz et al. 2011) presents a dynamic visualization of the evolution of georeferenced data captured by the speed of the traffic that runs through the roads being their thickness, colour and length altered by the number of vehicles, and their average velocity. For orientation, and based on maps, Google Street View allows a 360° photo view by using a spherical image projection with geolocalization. Panoramio (.com) allows to select places displayed by photos on a map, and it is possible to filter them by tags. But none of these applications provide video. Space and time in visualizations is also the focus of (Andrienko et al. 2011) though not specific for videos.

The Movie Mashup Application MoMa (Finsterwald et al. 2012) mashes up information from DBpedia and GeoNames, permitting to find movies through their location, but does not support trajectories. Hao et al. (2011) present user-generated videos that relate to geographic areas in a map interface. They focus on the automatic selection of keyframes to represent the videos, based on the popularity of the hotspot, and the determination of the location to place them on the maps. So they emphasize hotspots that are shot in the videos in front of the shooting spot, and not so much on their trajectories.

In Sight Surfers (Noronha et al. 2012), designed and developed in our project, users can capture, publish, share and search 360° videos, along trajectories that can be viewed and navigated on a map, and synchronized with the video being watched at any time. A marker moves along the trajectory for the current location, and the user can click on any point in the trajectory to view the video from that location and time. But video trajectories are presented as colored lines on a google maps interface, without a representation for such properties as speed, age and contents, like we are doing now.

4. VISUALIZING VIDEO TOURS

In this section, we present the design rationale of our approach to the visualization of video tours and trajectories, and we present the interactive visualizations focused on spatio-temporal and content properties and conceived to navigate georeferenced videos along trajectories, from video spaces down to the individual videos, where they can be watched.

4.1 Design Rationale

Movies and videos have the power to affect us perceptually, emotionally, and cognitively by the story they tell and, in great extent by the colours, sound and movement they show. Some trajectories might not consider explicit storytelling concepts, nevertheless they render experiences that people like to share and access over time. On the other hand, maps can help to ease the understanding of the spatial dimension, and as models of reality, when linked with video, offer the viewer a new view of reality (Peuquet 2001) amplifying the power of movies and videos in affecting the perception of the user. We believe that contemplating the spatial and temporal dimensions in the visualizations of videos, can enhance and enrich user awareness when accessing and viewing videos.

In the visualizations that we propose, we focus on videos' trajectories in geographic space, represented in maps, and multidimensional attributes that change along the trajectories and throughout time.

We believe that visualization, with this spatial and temporal dimensions, is of great use at easing and enriching the overview, navigation and access of videos, and especially those that involve tours along trajectories that capture and record user experiences. Our aim is to explore the power of visualization in characterizing locations visually, generating an impression, an emotion, a thought in the viewer, in addition to the accurate analysis that can take place for more precise results. Information such as the number of shot trajectories might be needed e.g. to analyse tourist preferential tours in Lisbon, and if they prefer to shoot or access those by day, or by night, in low or higher speeds. They might want to know if the street they are walking in is one of the most popular or recorded ones, or more nature like or calmer than others nearby. A

student, or a teacher might need information about a city, and instead of watching all movies with the name of that city in the title or keywords, they can filter them by properties of their content.

Conceptually, this work is based on the footprints people leave on the ground when they walk, and thus we explore the marks' metaphor as if when shooting a location the user was leaving visible recorded emotions shown by the colours, movement, sound and words.

We present the design study of the main interactive visualizations conceived for navigating georeferenced videos, and movies, along trajectories, through quantitative and qualitative properties of the videos and their contents, from video spaces down to the individual videos, through: 1) overview collections; and 2) individual movie views on the video space; and 3) access individual videos to view details and watch them. These allow the user to visualize amount properties, video trajectories and speed, and video content along trajectories, and to visualize properties and watch the content in any selected video, as described in the next sections.

4.2 Visualizing Amount in Video Spaces

The amount of videos shot in a requested area is exemplified in the visualization of Fig. 1. This overview gives the user an idea about the amount of videos shot in a given limit of time and space. The brighter the color, the higher quantity of videos shot in that trajectory. The users might filter the search by criteria relating with time and space (e.g. *the most shot location in Lisbon, at 21p.m.*);. It is possible to know which part of the city has more videos shot in general, and to narrow the information, filtering it by some specific hour of day or night, if it is sunny or rainy, the age of the shooting, the speed of the tour, and through the most rated, or watched trajectories by the viewers. Criteria can be quantitative (e.g. higher quantity of videos shot, or fastest tours) or categorical (e.g. by day or in spring).

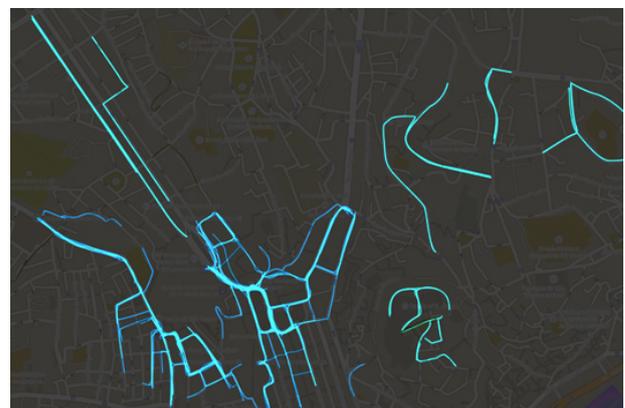


Figure 1: Trajectories shot in a requested area in a given moment of time.

The visualization in Fig. 1 represents space on the map, and time is inherent to the trajectories but not made explicit in this visualization. Time can also be used for the video selection (e.g. shot in the current

month) It is possible to zoom in and out and to change the view of the representation by dragging. The users can also change to other perspectives of the visualization (Fig.1-Fig.2).

4.3 Visualizing Video Trajectories and Speed

A 3D representation was conceived to allow knowing the speed of the trajectory, and the age of the shooting in a given geographic location (illustrated in Fig. 2). Any trajectory can then be selected in order to individualize the information, i.e. be aware of these properties in one particular trajectory (Fig. 3).

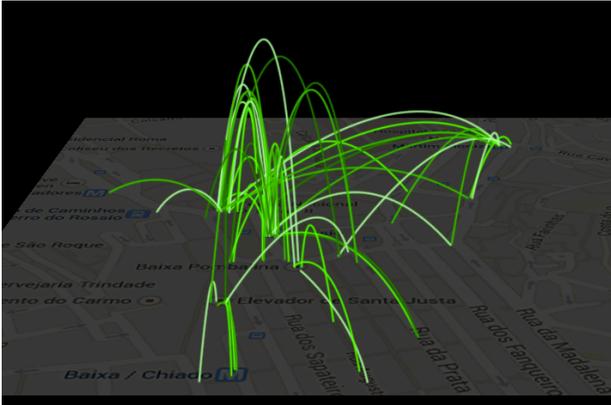


Figure 2: 3D visualization of video tours speed and age.

The trajectory speed is represented by different heights corresponding the highest, and longer, to the slower tour, since it gathers more frames shot than the lower and shortest ones. Green gradients were used in the ageing of the shootings, and similarly to the effect of the passing of time on colored objects, also the shooting representations loose their brightness. Therefore, brighter green means more recent shots. By dragging, the user can choose the angle view in order to see the higher (slower tours) and lower (faster tours) curve lines, and the greener (recent) or darker (older shootings) lines. In this view the start, and end of the trajectory is more important than the specific streets that constitute it, since this information is already in focus of the overview (Fig.1), and can be seen in detail, on demand (Fig.3).

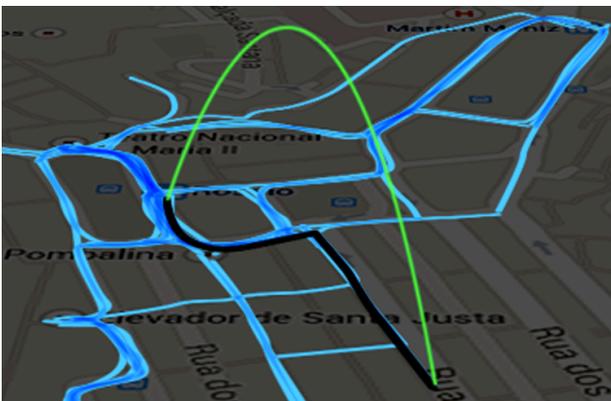


Figure 3: 3D visualization of an individual tour. It allows a clearer and more accurate perception of the trajectory streets (black line), its duration (line height), and age (green brightness gradient).

Each trajectory has its georeferenced coordinates on the map, and accurate information about quantities can be switched on/off, depending on the user's interest. They might prefer to turn the information invisible to allow a clearer idea of the relations between the length, and age of the videos. Clicking the trajectory of interest, leads the user to a more detailed visualization to observe more accurately the chosen trajectory (Fig.3). On double-clicking that trajectory, it is possible to navigate to a visualization that weaves information about the video content (Fig.4).

4.4 Visualizing Video Content along Trajectories

This visualization (Fig. 4) presents the content of one trajectory through the following properties:

- 1) georeferenced position on the map permits the user to be aware of the precise location, the name of the streets and the start and end of the tour;
- 2) the colours of the shot trajectory are presented by the summarization of the video, allowing e.g. to know if the tour is bright or dark, possibly outdoors or indoors, nature or city like, and relate to the emotions that are associated with colours;

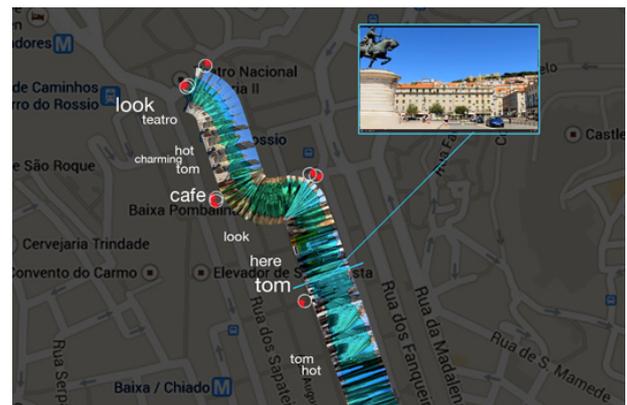


Figure 4: Visualizing content in a video tour.

3) amplitude of video sound is represented through colour contrast over the summarized trajectory. It is not much evidenced due to the primordial importance given to the coloured frames but it might be selected by the user in order to make it more visible, through the increasing of the opacity until it is coloured in blue. We believe this information can add an idea of the mood of that trajectory, i.e. if it is noisy, or quiet;

4) spoken words are captured along the shooting and gathered in tag clouds that represent their frequencies either by the person who shot the trajectory, or the people that surround them nearby. It might characterize the tour, if words relate to the shot location. These tags are represented aside the trajectory, in white;

5) neighbour connections that connect this specific tour and follow outward through other streets are shown through small red circles specifying also the direction, i.e. a street which the user can choose to navigate (to access videos shot in that direction);

6) frames of one moment of the video sequence for a real awareness of the trajectory. On mouse over the user can see the frame of interest displayed, and on double-clicking, the video plays.

This visualization can be zoomed in for more detailed information, i.e. zooming on the summarization trajectory makes visible a major part of the frames. All the represented properties might be switched on, and off depending on the intention of the user, and as referred previously, the transparency that represents the audio is adjustable, being possible to highlight that information by colour opacity. Transparency has the affordance of letting visible the colours of the tour, and carry aesthetic qualities, and therefore it is emphasized, nevertheless only when needed. It is possible to change views on this visualization by dragging, turning it to a 3D representation, allowing the adding of information about speed, and age of the chosen trajectory (Fig.5).

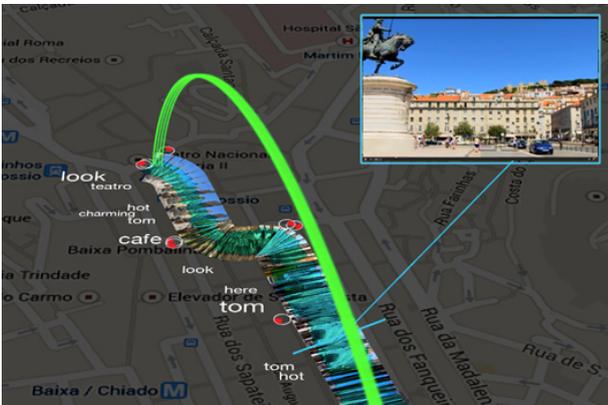


Figure 5: 3D visualization of an individual tour with content, speed, and age.

4.5 Visualizing Properties and Content in Video

This final visualization (Fig. 6) allows the overview, browsing and watching of the video content with a higher focus on the video content.

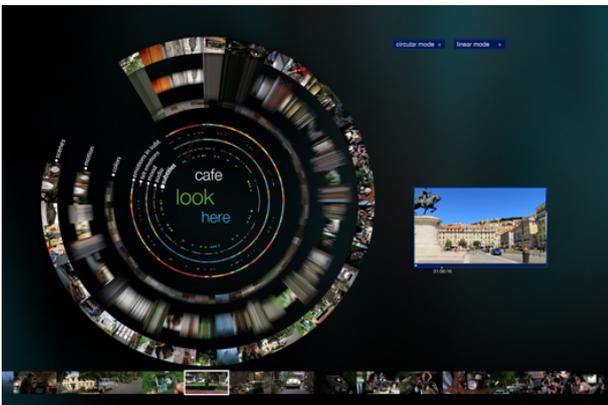


Figure 6: Visualizing content in the movie, or video.

The wheel presents in the middle a tag cloud of the spoken words during the shooting. Around, from inside out its circular timelines represent the different content tracks: 1) spoken words during the shooting; 2) audio events; 3) mood captured by audio;

4) emotions felt by the viewers when they watched the video, or movie, 5) emotions expressed on the spoken words; 6) dominant colors; 7) movement; and 8) scene thumbnails. It is possible to select the properties of interest, and watch the video synchronized with the content represented around the wheel and along the scene thumbnails in the film strip in the bottom.

From this overview (Fig.6), it is possible to get the main idea and visual properties of the video. For more details about this visualization see (Jorge & Chambel 2013).

4.6 Exemplified Interaction

Users might be visiting Lisbon, and want to have an idea of the places that were most shot in this city, and that might correspond to the most interesting visits (Fig.1). They might want to know which ones are the longest, and newest videos shot by night (Fig.2). Intending to narrow the information they pick a tour, and access information about colours, sound, and spoken words along the video, as well as neighbour connections that can lead them to other places of interest (Fig.4). They might be interested in knowing the emotions of the viewers when they watched that video (Fig.6), and eventually, to watch the video and have a feeling of how it is like to be there.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we presented most recent work towards interactive visualization of georeferenced trajectories. On previous work we addressed the time component combining information about movies, and videos, in chronological spaces, when they were released, and inside their content. We now considered geolocation trajectories, and thus presented the design study of three main interactive visualizations for navigating georeferenced movies and videos allowing the user to, 1) overview movies and videos' collections, 2) individualize them, and 3) navigate throughout their contents. Firstly, it is possible to overview amounts of movies, and videos shot in a given geographic area in a given moment of time, their trajectories and speed, and to detail the criteria of interest about their selection. Secondly, to select and zoom in the trajectory of interest to access more properties about their trajectories, speed, and content, in terms of colours, sound, spoken words; about neighbour tours; and to watch the video. Thirdly to detail those content properties of the selected videos relating spoken words, audio events, mood in audio, felt emotions, emotions in subtitles, dominant colours, movement and tour thumbnails, synchronized with the video being watched.

As next steps, we intend to refine this work. A user evaluation will aim at learning about the efficacy and preferences in each visualization and to provide directions for possible improvements, through a sustained progress. The focus will then be on developing effective interactive visualizations from a

user perspective, that can be obtained or enriched by video processing techniques, for a faster and dynamic creation, following our work on MovieClouds (Gil et al. 2012), and collected data about users sharing and access activity (Noronha et al. 2012) (Ramalho & Chambel 2013). The ultimate goal is to provide users with visualizations that allow them to get overviews and insights about videos and the places where they were shot, and to access and watch them based on relevant spatio-temporal properties and enriched perspectives of their content. Complementary directions include exploring further the visualization of immersive videos, like the ones we have been addressing (Noronha et al. 2012) (Ramalho & Chambel 2013) through georeferenced 360° panoramic videos, and the integration with multimodal interfaces towards more flexible and effective interactive content access (Serra et al. 2014), through natural interaction with shape, speed and time.

6. ACKNOWLEDGMENTS

This work is partially supported by FCT through funding of LaSIGE Strategic Project, ref. PEst-OE/EEI/UI0408/2014, the ImTV research project ref. UTA-Est/MAI/0010/2009, and an FCT fellowship under the UT Austin|Portugal Program (SFRH-BD/51798/2011).

7. REFERENCES

- Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S., and Wrobel, S. (2011). A Conceptual Framework and Taxonomy of Techniques for Analyzing Movement. *Journal of Visual Languages & Computing* 22 (3): 213–32.
- Angesleva, J., and Cooper, R. (2005) Last Clock. *Computer Graphics and Applications, IEEE* 25 (1): 20–23.
- Cruz, P., and Machado, P. (2011) Visualizing the Circulatory Problems of Lisbon. In *ACM SIGGRAPH 2011 Posters*, 92. ACM.
- Finsterwald, J-M., Grefenstette, G., Law-To, J., Bouchard, H., and Mezaour, A. D. (2012) The Movie Mashup Application MoMa: Geolocating and Finding Movies. In *Proc. of the ACM GeoMM'12*, 15–18. ACM.
- Gil, N., Silva, N., Dias, E., Martins, P., Langlois, T., and Chambel, T. (2012) Going through the Clouds: Search Overviews and Browsing of Movies. In *Proc. of Academic MindTrek'12, ACM*, 158–65.
- Hao, J., Wang, G., Seo, B., and Zimmermann, R. (2011). Keyframe Presentation for Browsing of User-Generated Videos on Map Interfaces. In *Proc. ACM Multimedia*, 1013–16. ACM.
- Harley, J. B., and Woodward, D. (1987) *The History of Cartography*. Vol. 1. Univ. of Chicago Press.
- Jorge, A., and Chambel, T. (2013) Exploring Movies through Interactive Visualizations. In *Proc. of the 27th Int. BCS Human Computer Interaction Conf.*, 6 pgs, Sep 2013, London, UK.
- Jorge, A., Gil, N., and Chambel, T. (2012) Time for a New Look at the Movies through Visualization. In *Proc. of Artech'2012, the 6th Int. Conf. on Digital Arts*, 269–78. Faro, Portugal.
- Kraak, M. J., and Ormeling, F. (2010) *Cartography: Visualization of Geospatial Data*. Harlow; New York: Prentice Hall.
- Levin, G. (2005) An Informal Catalogue of Slit-Scan Video Artworks and Research. *FLONG*. http://www.flong.com/texts/lists/slit_scan/
- Noronha, G., Álvares, C., and Chambel, T. (2012) Sharing and Navigating 360° Videos and Maps in Sight Surfers. In *Proc. of Academic MindTrek'12*, 255–62. ACM.
- Nunes, M., Greenberg, S., Carpendale, S., and Gutwin, C. (2007) What Did I Miss? Visualizing the Past through Video Traces. In *ECSCW 2007*, 1–20. Springer.
- Peterson, M.P. (1997) Trends in Internet Map Use. In *Proc. of the 18th Int. Cartographic Conf., Stockholm, Sweden*, 1635–42.
- Peuquet, D. J. (1994) It's about Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. *Annals of the Association of American Geographers* 84 (3): 441–61.
- Peuquet, D. J. (2001) Making Space for Time: Issues in Space-Time Data Representation. *Geoinformatica* 5 (1): 11–32.
- Peuquet, D. J. (2002) *Representations of Space and Time*. New York: The Guilford Press.
- Ramalho, J., and Chambel, T. (2013) Immersive 360° Mobile Video with an Emotional Perspective. *Proc. ImmersiveMe'2013*, ACM Press, pp.35-40.
- Rego, A., Baptista, C., Silva, E., Schiel, U., and Figueirêdo, H. (2007) VideoLib: A Video Digital Library with Support to Spatial and Temporal Dimensions. In *Proc. of ACM SAC'07*, 1074–78.
- Serra, S., Jorge, A., and Chambel, T. (2014) Multimodal Access to Georeferenced Mobile Video through Shape, Speed and Time. In *Proc. of the 28th Int BCS Human Computer Interaction Conf.*, 6 pgs, Sep 2014, Southport, UK.
- Viégas, F., Boyd, D., Nguyen, D. H., Potter, J., and Donath, J. (2004) Digital Artifacts for Remembering and Storytelling: Posthistory and Social Network Fragments. In *Proc. of HICSS'04*, 10 pgs.
- Viégas, F., and Wattenberg, M. (2009) "Flickr Flow." *HINT.FM*. <http://hint.fm/projects/fli>