# From Photos to Memories: A User-Centric Authoring Tool for Telling Stories with Your Photos

Fons Kuijk, Rodrigo Laiola Guimarães, Pablo Cesar, and Dick C.A. Bulterman

Centrum Wiskunde & Informatica,
Science Park 123, 1098 XG Amsterdam, Netherlands
{fons.kuijk,rlaiola,p.s.cesar,dick.bulterman}@cwi.nl

**Abstract.** Over the last years we have witnessed a rapid transformation on how people use digital media. Thanks to innovative interfaces, non-professional users are becoming active nodes in the content production chain by uploading, commenting, and sharing their media. As a result, people now use media for communication purposes, for sharing experiences, and for staying in touch. This paper introduces a user-centric authoring tool that enables common users to transform a static photo into a temporal presentation, or story, which can be shared with close friends and relatives. The most relevant characteristics of our approach is the use of a format-independent data model that can be easily imported and exported, the possibility of creating different storylines intended for different people, and the support of interactivity. As part of the activities carried out in the TA2 project, the system presented in this paper is a tool for end-users to nurture relationships.

**Keywords:** Animation, Content Enrichment, Multimedia Document, Pan and Zoom, Photo Sharing, SMIL, Storytelling, Togetherness.

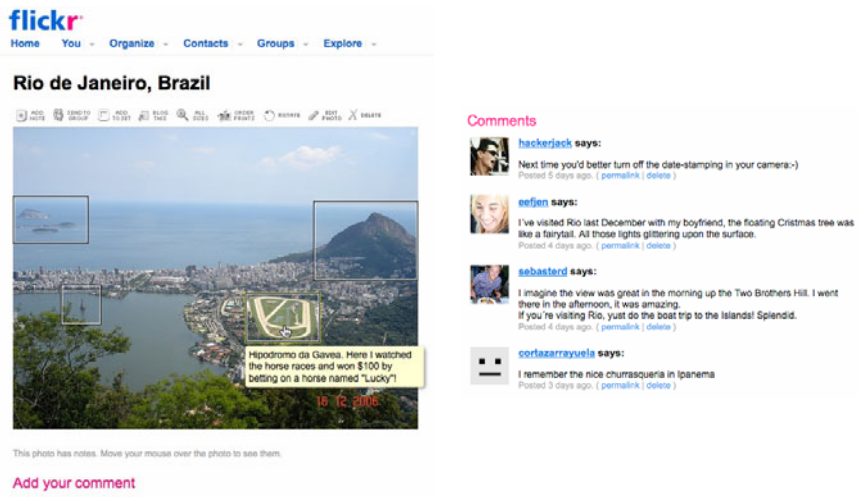
## 1 Introduction

The Internet has been designed to facilitate exchange of information in an effortless way. Lately, it has turned into a social environment that facilitates communication, services, and interaction. Technology innovation has led to equipment and mechanisms to produce and distribute multimedia content efficiently and at a low cost. The World Wide Web enables easy access to this multimedia content. Web interfaces such as blogs, podcasts, video casts, image sharing, and instant messages have emerged, and their fast rising popularity indicates the need for facilities so that the ‘common’ user can become an active node in the content production chain. This trend, the user becoming a media producer and distributor, and the use of his media, is referred to as User Centric Media. In a current report from the EC [13], scientific challenges have been identified for research in the area of Future Media Internet. One of these challenges is the capability to actively support the common user in creating new content and sharing it with others. Professional artists have found their way to use ‘technological’ tools of the earliest kind to create media; the challenge now is to invent tools for the common user that focus on creation, rather than on technology.

P. Daras and O. Mayora (Eds.): UCMedia 2009, LNICST 40, pp. 13–20, 2010.

## 2 Motivation and Contribution

In this paper we focus on sharing old memories – photos –with friends and family that do not live in the same household. Currently, the Web offers many interfaces (e.g., Flickr, Picasa, MySpace, and Facebook) that allow people to share photos. In some cases, users can turn their photos into movie-like presentations and slideshows, converting photo sets into an interesting class of temporal documents (Animoto, Stupeflix). Users may include metadata describing a photo, or parts of a photo, and add comments. *Fig. 1* shows a photo of Rio de Janeiro as present in Flickr, apparently published by a tourist that wants to share his experiences with others. Our tourist added notes: rectangular-shaped regions – indicating locations he visited and objects he observed – with associated descriptions. In addition, some of his friends have commented on the photo. This example represents one instance on how user-centric media can leverage togetherness between people living in different locations by providing users with easy-to-use interfaces to describe, comment, and tag their media.



Rio de Janeiro, Brazil

Hipodromo da Gavea. Here I watched the horse races and won $100 by betting on a horse named "Lucky"!

This photo has notes. Move your mouse over the photo to see them.

Comments

hackerjack says:

Next time you'd better turn off the date-stamping in your camera:-)
Posted 5 days ago. ( permalink | delete )

eefjen says:

I've visited Rio last December with my boyfriend, the floating Cristmas tree was like a fairytail. All those lights glittering upon the surface.
Posted 4 days ago. ( permalink | delete )

sebasterd says:

I imagine the view was great in the morning up the Two Brothers Hill. I went there in the afternoon, it was amazing.
If you're visiting Rio, yust do the boat trip to the Islands! Splendid.
Posted 4 days ago. ( permalink | delete )

cortazarrayuela says:

I remember the nice churrasqueria in Ipanema
Posted 3 days ago. ( permalink | delete )

**Fig. 1.** A typical Flickr annotated photo of Rio de Janeiro with its thread of comments

While digital photo sharing systems have been widely used for studying knowledge extraction [8], location mapping [6], and social networks [4][11], their restrictive rich media capabilities have not been challenged yet. Consider again the image in *Fig. 1*: regions of interest with metadata have been specified, and people have commented on the photo. The image itself, however, remains a static object, with an undefined focused area of interest. One way of improving the presentation of this image is to add a variant of the Ken Burns Effect, in which various descriptions associated with each region are presented as dynamic, synchronized pans and zooms across the image content. The presentation may start, for example, with the general description of the photo using some styling [2] and interactive options can be incorporated in the form of temporal and spatial hyperlinks to the comments. The presentation then continues with a number of sequenced pans and zooms across the content. The story becomes even more compelling if the resulting presentation can be customized for different friends and family members [7].

The contribution of this article is an authoring tool that provides users with an array of possibilities for creating meaningful and customizable stories as sketched above. The authoring tool uses an underlying declarative multimedia document format with the following characteristics:

- Regions of interest: users are able to identify regions of interest, which can be imported from and exported to other formats (eg., MPEG-7, SMIL);
- Descriptions and comments: users may associate annotations and comments to identified regions of interest;
- Path creation: users have the ability to animate transitions and zoom in on the regions of interest;
- Temporal synchronization: users can synchronize different media types (eg., audio commentary and animation within the photo);
- Base-Media integrity: users cannot alter the underlying based content. Regions, annotations, and comments are linked to media. They are not embedded into it;
- Customization: users may re-use regions/paths for the creation of customized stories using existing content control mechanisms;
- Accessibility: users are able to create accessible multimedia presentations; and
- Interactivity: users can create interactive presentations based on temporal and spatial hyperlinks.

The authoring tool for adding dynamic visual features to static images is targeted for the common user and focused on the creative aspect of storytelling. The tool generates a dynamic presentation of the author's story, in the form of a declarative multimedia document that can be shared with non-collocated peers.

An essential element of the multimedia document format is the SMIL MediaPanZoom module[1], an extension we have proposed to add to the SMIL 3.0 recommendation [3]. The extension is now standardized and supported by popular multimedia players such as the Ambulant Player[2] and the RealPlayer [3].

## 3 Related Work

Many photo management tools (e.g. iPhoto) and video production systems (such as iMovie or Photo Story) are available for creating digital artifacts from personal media to share experiences with others. Users can employ these applications to tell stories, but these applications typically package the resulting media in some encoded format (e.g. slideshow or video). This feature does not keep the image integrity or allow navigation based on timed enrichments, nor does it support selective viewing (since the annotations are hardcoded with the base photos).

[1] http://www.w3.org/TR/SMIL3/smil-extended-media-object.html

[2] http://www.ambulantplayer.org

[3] http://www.realnetworks.com

StillMotion[4] and Amara Photo Animation[5] are slideshow-authoring tools that enable users to create Flash presentations with sound, navigation, transitions, and pan and zoom functionality. These tools pack all media in a self-contained media file, so media integrity is not retained. Even though these tools produce Flash, they do not support navigation based on timed annotations or selective viewing. MemoryNet Viewer [12] is a Peer-to-Peer system for sharing and enriching photos (by adding voice and text comments) among people in personal networks. MemoryNet does not enable end-users to add timed enrichments, such as pan and zoom, nor to export to an open document suitable for other systems. StoryTrack [1] supports a touchscreen device for creating personal stories based on digital photos. Stories and metadata are stored in XML, which allows for translation to be shared with others who do not have a StoryTrack system. This tool does not allow users to zoom in to regions of interest and does not allow an audio track to go beyond the presentation of a single image. Flipper System [5] is a photo sharing application both for mobile and desktop environments. Users may add text comments to any image. It does not offer timed enrichments, and it is not possible to create a story based on a set of images. iTell [10] is a narrative composition tool that leads storytellers stepwise through the writing and media production processes. Forced to follow this stepwise approach, a user can specify sequencing of the imagery and voiceover. iTell departs from the typical timeline metaphor to the notion of associations in order to indicate relationships between narration script, voiceover and images. It exports stories to SMIL, but does not support timed annotations or hyperlinks to extra information or related stories. Web-based photo sharing services (e.g. Flickr) and community-sharing environments (e.g. Orkut) do allow users to share pictures with the ability to add notes and comments. End-users even can add spatial notes to third-party photos if they have been given editing rights to do so. Yet, annotations are a temporal and site-specific.

## 4 Authoring Process

To illustrate the creative process we envision for storytelling based on images, picture a family member that went on holiday to Rio de Janeiro, Brazil. He decides to tell the story of his trip based on the photo of *Fig. 1*, on which memorable locations are visible. Using our authoring tool, he identifies regions of interest within the photo (like Flickr's notes). He orders these regions on a timeline that is part of the user interface (see *Fig. 2*). Ordering regions defines temporal transitions: the default transition, inserted by the system automatically, is panning along a straight path from one region to the next. In this way, we obtain a presentation that starts by slowly zooming in on the first region of interest. Then by panning, following the connecting paths, a transition to the next region of interest is made, and so on, concluding with zooming out to reveal the entire image. A timeline representation of this presentation is shown in *Fig. 3*. The author can link recorded audio files (e.g., saying "On this mountain I saw….") and other information (textual descriptions and annotations) to regions as well as to transitions. Note that regions, transitions, and associated linked media are not embedded, thus assuring customizable and accessible capabilities.

[4] http://www.imagematics.com

[5] http://www.amarasoftware.com

We experimented with the system set-up and concluded that we could recognize two preferred modes of operation: operating on the basis of the visual aspects (the regions of interest), or operating on the basis of the audio (the narratives). Both modes of operation are supported by the authoring system shown in *Fig. 2*.

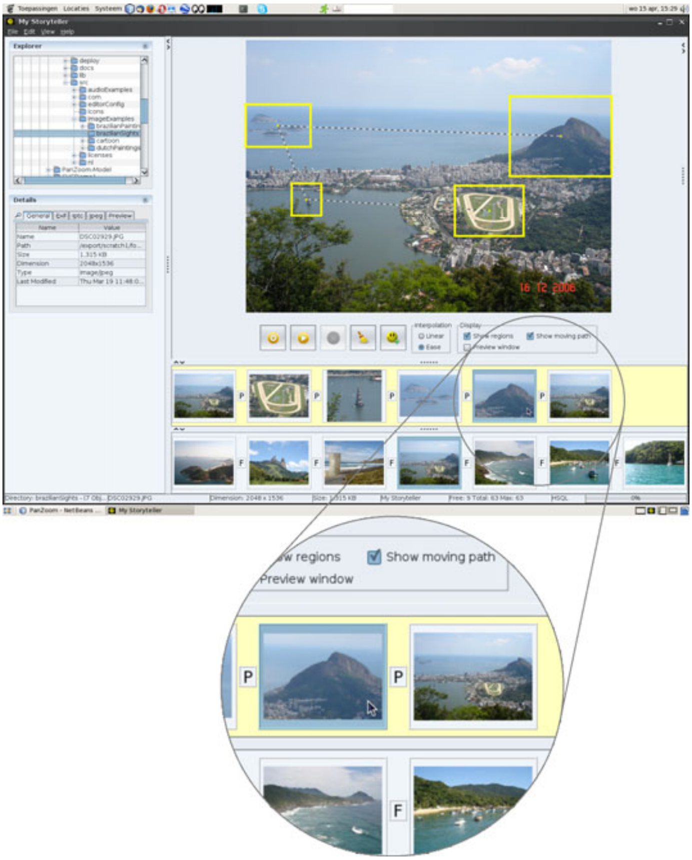


**Fig. 2.** User interface. On the left we see a *browser* for selecting images and an area to display *properties*. The large *work area* on the right of that serves for editing regions of interest. Below we see two timelines: the *image-level timeline* (bottom row) for ordering images, and the *region-level timeline* (upper row) for ordering regions.

In the *region driven* mode, the author selects images and annotates them with regions of interest. The author specifies the order in which images and regions are to be shown by ordering thumbnail representations on the image timeline or region-timeline. Clicking and dragging in the work area specify new regions of interest. To begin with, default duration is assigned to images and regions and default transitions are inserted automatically. The user can record narratives or specify audio files that go with an image, a region or a transition. Selecting an image, a region, or a transition, enables the record button. The duration of the audio fragment or recording defines the duration assigned to the selected item.

In the *audio driven* mode, the author starts recording the narratives and in the meantime makes up the visuals by selecting images and regions that relate to what is being told. As the recording and selection occurs in realtime, the time between selections defines the temporal characteristics of the presentation: the time images and regions are displayed (anticipating for transitions).

In both modes, default transitions are inserted to complete the presentation in a professional and visually attractive way. For this, the author does not need to have any expertise on animation. The default transition for image changes is a simple fade-in fade-out. The default transition between two regions of the same image is enhanced panning (a combination of panning and some zooming out and in that helps the viewer not to lose the context).

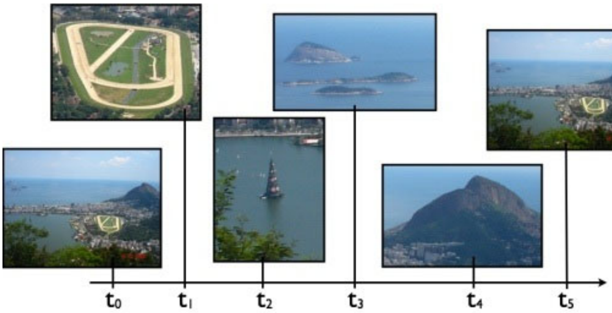


**Fig. 3.** Timeline of the presentation. The regions of interest and the order in which show up.

The author can export the presentation as a structured multimedia document. Currently the encoding is based on SMIL that – being an open format – offers navigation and customization to the end-user. The images are referred to via URI's, maintaining the integrity of the sources. The timed annotations for transitions between the regions of interest include functionality of the SMIL MediaPanZoom module[6]. The visual component of a storyline is in effect a sequence of static and dynamic `panZoom` components [9]. Support for interaction is obtained by using temporal and spatial hyperlinks. Specific storylines can be targeted to individual users, so that watching the presentation may become an interactive, personalized experience. Although there is a common ground, our tourist may want to convey a story to his family that differs from the one he will tell his close friends.

## 5 Implementation

The authoring system is implemented in Java. We use a format-agnostic data model (see *Fig. 4*), designed to accommodate the authoring process rather than the process of

[6] http://www.w3.org/TR/SMIL3/smil-extended-media-object.html

generating the structured multimedia document. An XML representation of the model is used for persistence and to cut, copy and paste a storyline or parts there of. This organization allows for import and export of functionality to manifold formats, such as MPEG-7 and SMIL.

A `DataManager` handles all access to the data model. A storyline is represented by `Path`, being a collection of `PathElements`. Each `PathElement` has an associated `TimeInterval` that specifies duration; it can be coupled with an audio fragment. We distinguish two types of `PathElements`: `Transition` and `Hold`. A `Path` of a typical storyline is a series of alternating `Holds` and `Transitions`. A `Hold` is linked with one `Region`. It represents the part of a storyline when a region of interest is highlighted (e.g. by zooming in on that region and playing the audio fragment that may be coupled to `TimeInterval`). A `Transition` is linked with two `Regions`. It represents the animated transition from one region of interest to another (e.g. by panning). Its `TimeInterval` may also be coupled with an audio fragment.

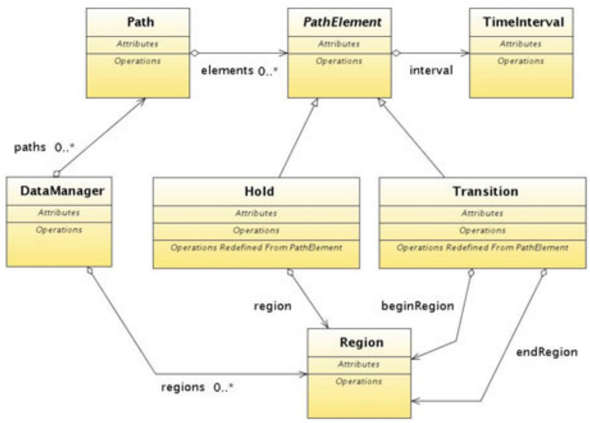


**Fig. 4.** Data model of the authoring system

## 6 Discussion

The EU-funded project TA2[7] studies how technology can help to nurture family-to-family relationships by sharing experiences, holidays, celebrations and moments of fun and laughter. TA2 media and communication support can be characterized by naturalness; clear, relaxed voice communication and intelligently edited video.

The authoring system we presented in this paper is part of this overall TA2 effort. Our family member can define a storyline in a natural way, simply by talking and clicking on images, much like browsing through a family photo album. Simplicity led to a set of requirements (cf. Section 2), on which we designed the structure of the underlying data model. Simplicity is key, especially to encourage elderly people (such as our grandparents) to produce and augment their stories – to help them to document

[7] http://www.ta2-project.eu/

their historic memories and experiences in general. The storyline is transferred into a SMIL document: a compact structured multimedia document that can be shared in an efficient manner with non-collocated peers, making use of URI's instead of transferring full image data wherever possible. It is important to highlight that SMIL allows the addition of metadata to any of the elements contained in the document specification. Moreover, authors can synchronize pan and zoom effects with audio or captions, and even generate customized presentations for distinct audiences.

## Acknowledgments

This work was supported by the EU FP7-ICT project TA2. The NLnet Foundation supported the development of the Ambulant Player.

## References

1. Balabanović, M., Chu, L.L., Wolff, G.J.: Storytelling with Digital Photographs. In: Proceedings of the SIGCHI Conference on Human factors in Computing Systems, CHI 2000, pp. 564–571. ACM Press, New York (2000)
2. Bulterman, D.C.A., Jansen, A.J., Cesar, P., Cruz-Lara, S.: An efficient, streamable text format for multimedia captions and subtitles. In: Proceedings of ACM DocEng 2007, pp. 101–110 (2007)
3. Bulterman, D.C.A., Rutledge, L.W.: SMIL 3.0: Interactive Multimedia for Web, Mobile Devices and Daisy Talking Books. Springer, New York (2008)
4. Cha, M., Mislove, A., Adams, B., Gummadi, K.P.: Characterizing social cascades in flickr. In: Proc. Workshop on online Social Networks, pp. 13–18 (2008)
5. Counts, S., Fellheimer, E.: Supporting Social Presence through Lightweight Photo Sharing On and Off the Desktop. In: Proc. of the SIGCHI Conference on Human factors in Computing Systems, CHI 2004, Vienna, Austria. ACM Press, New York (2004)
6. Crandall, D.J., Backstrom, L., Huttenlocher, D., Kleinberg, J.: Mapping the world's photos. In: Proc. of the 18th int. Conference on World Wide Web, pp. 761–770 (2009)
7. Jansen, J., Bulterman, D.C.: Enabling adaptive time-based web applications with SMIL state. In: Proceedings of ACM DocEng 2008, pp. 18–27 (2008)
8. Kennedy, L., Naaman, M., Ahern, S., Nair, R., Rattenbury, T.: How flickr helps us make sense of the world: context and content in community-contributed media collections. In: Proceedings of the 15th international Conference on Multimedia, pp. 631–640 (2007)
9. Kuijk, F., Guimarães, R.L., Cesar, P., Bulterman, D.C.A.: Adding Dynamic Visual Manipulations to Declarative Multimedia Documents. In: Proceedings of ACM DocEng 2009, München, Germany (2009)
10. Landry, B.M., Guzdial, M.: iTell: Supporting Retrospective Storytelling with Digital Photos. In: Proceedings of the 6th Conference on Designing Interactive Systems, DIS 2006, University Park, PA, USA, pp. 160–168. ACM Press, New York (2006)
11. Negoescu, R.A., Gatica-Perez, D.: Analyzing Flickr groups. In: Proceedings of the 2008 international Conference on Content-Based Image and Video Retrieval, pp. 417–426 (2008)
12. Rajani, R., Vorbau, A.: Viewing and Annotating Media with MemoryNet. Extended Abstracts on Human Factors in Computing Systems, CHI 2004, Vienna, Austria, pp. 1517–1520. ACM Press, New York (2004)
13. User Centric Media Cluster of FP6 projects: User Centric Media, Future and Challenges in European Research Luxembourg: Office for Official Publications of the European Communities, p.76 (2007) ISBN 978-92-79-06865-2