Using Film-Making Techniques to Synthesize Compelling Video Shows from Consumer Photographs

Eamonn O'Brien-Strain, Jim Rowson, Simon Widdowson, Abbe Don,
Prakash Reddy, Kurt MacDonald[1]
HP Laboratories Palo Alto
HPL-2005-140
August 1, 2005*

photograph, video, synthesis, storytelling, cinematography, consumer

We describe a software architecture that combines storytelling and cinematographic techniques to guide consumers in creating compelling video shows from their photographs. The software implementation of this architecture could run in cameras, consumer appliances, PCs, or web services.

* Internal Accession Date Only
[1]USC School of Cinema-Television
Approved for External Publication

Using Film-Making Techniques to Synthesize Compelling Video Shows from Consumer Photographs

Eamonn O'Brien-Strain, Jim Rowson, Simon Widdowson, Abbe Don, Prakash Reddy, [†]Kurt MacDonald

OST/HP Labs, [†]USC School of Cinema-Television

{eob, jim.rowson, simon.widdowson, abbe.don, prakash.reddy}@hp.com, kurt.zero@gmail.com

Abstract

We describe a software architecture that combines storytelling and cinematographic techniques to guide consumers in creating compelling video shows from their photographs. The software implementation of this architecture could run in cameras, consumer appliances, PCs, or web services.

Problem Statement

Personal photographs are undervalued when they languish unwatched and unshared.

We hypothesize that repeated viewing and sharing are not more widespread because easy ways of presenting photographs are boring to the viewer. Digital photo albums often have more in common with the family slideshows that used the slide projector in the living room than they do with new media.

As desktop publishing democratized publishing, we want to democratize video creation. We want to enable average people to create compelling video shows from their photographs. This will help to unlock their value and personal meaning, and in turn increase the perceived value of HP imaging products and web services.

Our Solution

The television and film industry is very successful at creating compelling media. Writers and directors are skilled in storytelling techniques and conventions and, as Bruce Block shows [1] directors make movies particularly captivating by manipulating visual components in sophisticated ways that evoke in the viewers an intensity that matches the structure of the story. We want to allow untrained consumers to take advantage of these film techniques.

Story Telling

Personal photographs and personal stories can be thought of as the folktales of the 21st Century. As such, we can apply the standard storytelling frameworks of Joseph Campbell [2] and Vladimir Propp [3] to the common genres of personal photographs, such as: "Vacation Trip", "Wedding", "Birthday Party", "Quinceañera", "First Communion", "Bar Mitzvah", or "House Tour". Based on these principles, professional storytellers can create a library of well-crafted *storyboard templates* that enables the average consumer to tell compelling stories with their photographs. They simply choose a template from the library that most closely describes the content of their particular set of photographs.

The storyboard encodes the dramatic arc of a story as a sequence of story "intensity" numbers.

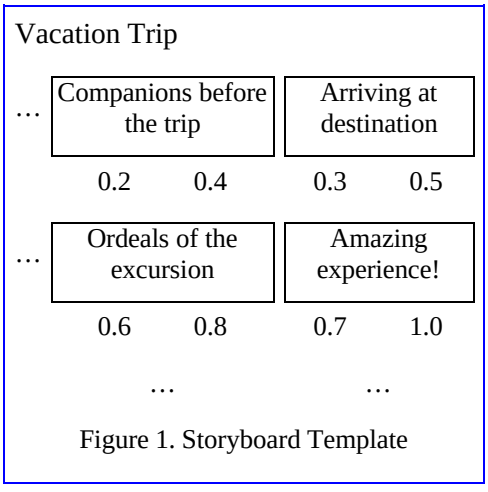

Figure 1. Storyboard Template

Visual Style

Bruce Block's approach to film is amenable to automation in part because it is more reductionist than other approaches. In his theory, film directors use seven categories of visual components to help tell a story: colour, line, movement, rhythm, shape, space, and tone.

A *look* is a particular set of values of these visual components. For example `{saturation=[0.6,0.8], tone=[0.8,0.9], hue=[67,212], line=[92,89], space=[0.2,0.3]}` describes an analyzed photograph as being fairly saturated, quite bright, greenish-yellow and cyan-blue, with vertical lines, and fairly flat space. An example desired look `{saturation=[1], hue[+-180], movement[+-0], space=[1]}` requests highly saturated contrasting hues with uniform movement and lots of depth cues.

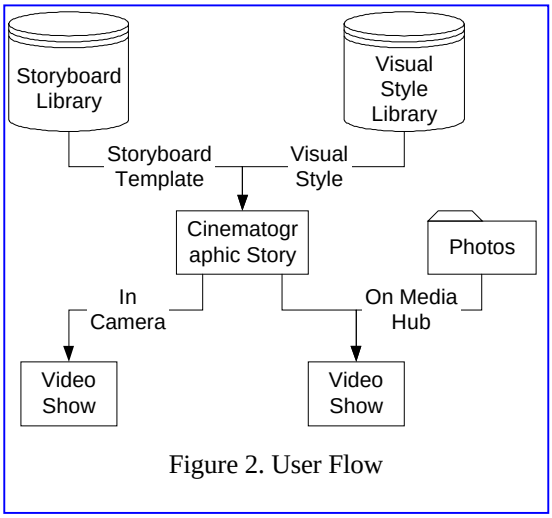

Figure 2. User Flow

A library of pre-built *visual styles* enables cinematography for everyone. Users control the look of their video show by selecting from visual styles such as "Hitchcock", "Fellini", "Ingmar Bergman", "Film Noir", "CSI Miami", "MTV", or "Six Feet Under". Skilled visual filmmakers create the visual styles. Perhaps in future visual styles can be extracted automatically from existing movies.

A visual style is simply a mapping from story intensity values to desired looks. Typically a visual style will keep some of the visual components fixed while varying other components with the intensity of the story. An example visual style might be `{hue=vary([+-0],[+-180]), saturation=fix([0.5]), space=fix([0]), line=vary([90],[45,135])}` This specifies medium saturation (pastel) colours with few three-dimensional depth cues and, as the story intensity increases, uniform hues changing to contrasting hues and the dominant line orientation changing from mostly vertical to mostly diagonal.

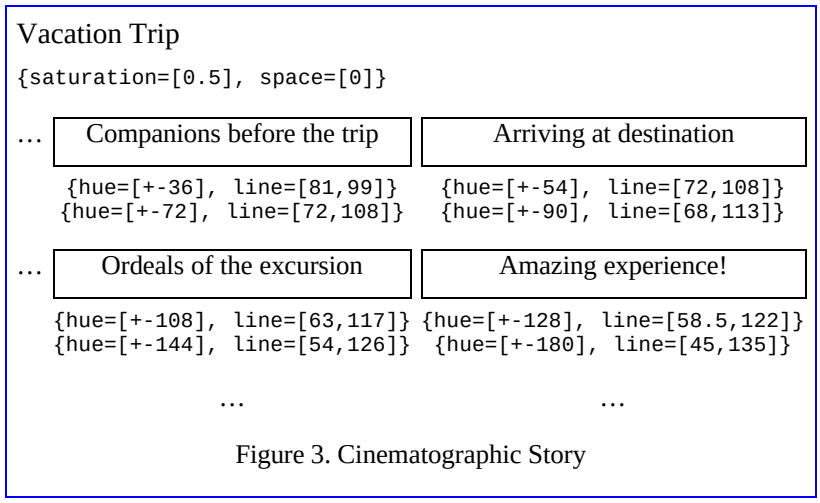

Figure 3. Cinematographic Story

Cinematographic Story

A storyboard template plus a visual style yields a *cinematographic story*, which can be used with some photographs to synthesize a video show. The visual look of each shot in the video show is as close as possible to the desired look at the corresponding place in the cinematographic story.

One way of achieving a desired look is by selection from analysed photographs. Visual components such as tone, colour, and spatial rhythm extracted from a photograph guide the selection of the most suitable photographs.

Another way is by modification of photographs. For example, tone or colour saturation can be modified automatically within certain ranges to achieve a desired value without making the images appear to be "fake".

Finally, a suitable video shot generator is chosen to synthesize the desired look for each point in the story. For example, a rostrum camera (Ken Burns) pan-and-zoom shot generator can create movement, rhythm, and line. Another shot generator moves and manipulates multiple photographs on the screen to synthesize shape.

User Interface

We assume the limited user interface of a consumer appliance or digital camera. The screen is of limited resolutions and there are simple arrow and selector buttons. The user creates a cinematographic story by picking a storyboard template, which describes their photographs generically, and a visual style they would like to apply to the storyboard.

On a consumer appliance, the user selects from previously shot photographs, automatically guided towards the best visual match for each part of the storyboard. The available photographs are sorted so that the best matching ones appear first.

On a camera, the software gives the user guidance in what photographs to shoot and how to compose them. This can be done by some text on the LCD screen and crosshairs showing where to place objects in the frame.

Competitive Approaches

Story development software guides screenwriters and other authors to plan stories. They use concepts such as Vogler's [2] for structuring the plot and standard character types such at the protagonist, antagonist, and sidekick,

With little or no user effort, simple rostrum camera panning and zooming (à la Ken Burns) is provided as a standard screen saver on Macintosh computers. As an improvement over random movement, Maurizio Pilu uses image saliency analysis to pick the start and end of the pan-and-zoom.[4] General-purpose video editing software can be used to create video shows from a photograph collection, but it requires a considerable investment of time for an average person.

There are several languages similar to our intermediate form for describing video. These include SMIL, HTML+TIME, XTL, and XMT. We found these languages do not have convenient support for describing motion.

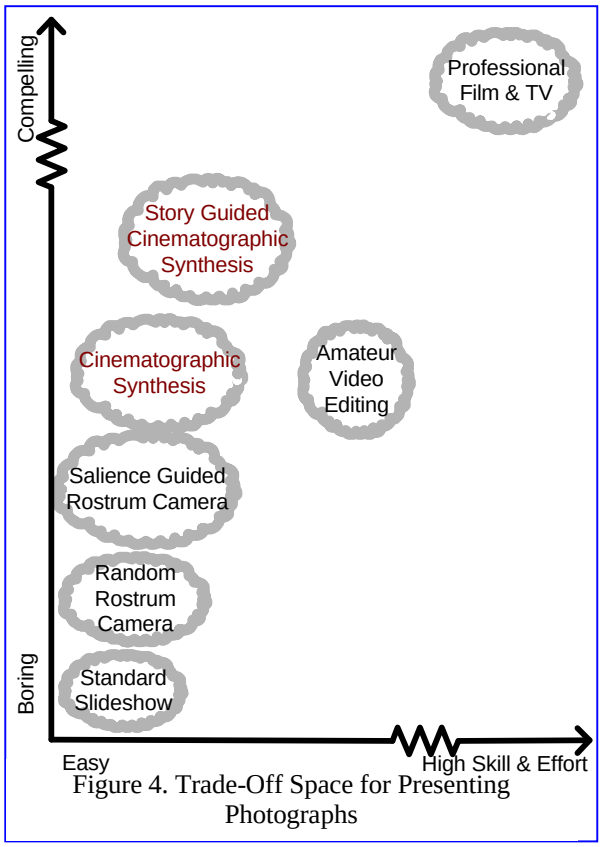

Figure 4. Trade-Off Space for Presenting Photographs

Current Status, Evidence the Solution Works, and Next Steps

We have created manually crafted video shows and categorized photographs to use as models in developing our software. Our visual component analysis modules can already extract colour, tone, and line from photographs. We have several video shot generation modules including a saliency-guided rostrum camera pan-and-zoom generator and a rhythm-modulating generator. We have a prototype video player of our video description language.

We need to plan user studies to test our two hypotheses: that our method is easy enough for anyone to use and that it produces video shows that engage the viewer's interest like professionally created video.

Our video synthesis generates a compact intermediate form that allows specification of time-based events, motion, and interaction of multiple simultaneous images on the screen. A computationally and memory-efficient player on the consumer appliance will directly play this. Alternatively, we can encode this into MPEG for playing on any video playing device.

Acknowledgements

We would like to thank our colleagues Rich Gossweiler, Rakhi Rajani, and David Frohlich. And a special thanks to Bruce Block whose seminar at HP Labs was the inspiration for this paper.

References

[1] Block, B. 2001, The Visual Story: Seeing the Structure of Film, TV, and New Media. Focal Press
[2] Vogler, C. 1993, The Writer's Journey: Mythic Structure for Writers. Michael Wiese Productions
[3] Propp, V. 1968, The Morphology of the Folktale. University of Texas Press.
[4] Pilu, M. 2004. A method for automatic panning and zooming of photographs. HP Labs Report