VR Webcams: Time Artifacts as Positive Features

Michael Naimark

www.naimark.net

A spatially contiguous triptych of three different times of day in Timbuktu

Abstract

"Virtual Reality" and "webcams" are currently incompatible suppositions, placing sensory richness in opposition to liveness. Large immersive images, sent through a "narrow pipe" such as today's Internet, must "accumulate" over time. Time artifacts result, since not everything can be transmitted at the same time.

Such time artifacts were explored using visual material from a previous art installation, filmed with a custom-built camera system, where such factors as frame rate, lens angles, and panning speed were known. Though the footage was pre-recorded, it approximated what a live "VR webcam" could be.

Scenes of the same places at different times of day were combined in various ways to simulate "narrow pipe" time artifacts. Studies produced from this footage suggest that time artifacts, while reducing the verisimilitude of the imagery, can increase its density or activity. In such "hyper-real" images, "more" can "happen." A "VR Webcam" is proposed.

Introduction

In 1560, Flemish painter Pieter Bruegel the Elder painted "Children's Games," depicting, like much of Bruegel's work, everyday life [1]. In it, over a hundred children can be seen actively playing dozens of games in a village square. Though the scene may have actually taken place, we know it didn't. Too much is happening at once, and all the action is perfectly composed. Not even Cecil B. DeMille could have created such a scene with a set and live actors. We assume that "Children's Games" is a realistic representation of an unrealistic event, an aggregate composition based on a "accumulation" of instances in Bruegel's memory or in his imagination.

In 1979, American cartoonist Robert Crumb drew "A Short History of America," depicting, in 12 frames, the progression of a single place from an untouched meadowland to a frontier village to an American street corner complete with convenience store and a clutter of power lines [2]. Even as a cartoon, the details are comprehensive. If a camera existed in the early days of colonial American history, and was positioned motionless in the same place for 200 years, "A Short History of America" could have been a time-lapse film.

Bruegel's painting and Crumb's cartoon are both place-based works depicting "accumulated" views, Crumb, over time in a progression of frames, and Bruegel, all at once. The elements that are accumulated, in theory, can be stored as separate data, and added or deleted interactively. This class of "hyper-real" imagery may be a model for cameras on the Internet.

The VR/Webcam Dilemma

The dream of "virtual reality" and the reality of "webcams" could not be further distanced. We associate VR with multi-sensory, high-bandwidth, immersive, interactive experiences, while webcams are associated with postage-stamp size images that rarely update faster than once per second. While the attraction of VR is sensory richness, the attraction of webcams is liveness.

This dilemma exists for several reasons, such as the need for rich, immersive source material and the need for immersive display technology, but the most prominent reason is due to the narrow pipe of the Internet. Consider that a good Internet connection speed for the home (e.g., DSL or cable modem) is rarely higher than 1 megabit per second. Uncompressed high definition television is one thousand times higher, 1 gigabit per second, and Imax is approximately ten times higher than HDTV. The bottleneck for an immersive "VR-like" webcam experience is the narrow pipe of the Internet.

Even with a narrow pipe, it is possible to use a great deal of inexpensive computational horsepower and digital memory at both ends. For example, one could build an immersive camera system (e.g., high definition, stereoscopic, panoramic) with a local host computer which stores short sequences and transmits them slowly to remote destinations, where they are restored.

Since such a system can not operate in real time, decisions about what to transmit will be necessary. These decisions can be of a smaller granularity than that of the actual frame. Consider, for example, having the ability to only transmit "interesting" elements from a common street scene - lovers walking hand-in-hand, a dog jumping in the air, a bird in flight - even if these events are not simultaneous.

Now imagine having a library of such events. One could under-populate or over-populate the scene as one desires. (Imagine an interactive Bruegel!) But the scene will never look perfect, in the sense of credible verisimilitude, because of time artifacts. Events occurring even a few minutes apart will often exhibit time artifacts due to the change of sunlight. Such artifacts are not of the sort easily fixable in PhotoShop. Semantic knowledge of the scene and events are required. Indeed, transforming an element recorded at night to appear during the day may never be truly possible.

Studies

What would such time artifacts look like? Will an image retain its wholeness as a "hyper-representation?" Will the place represented retain its "place-ness?"

These questions were addressed in a series of studies made from pre-existing footage from one of my earlier installations, "Be Now Here" [3]. Be Now Here was filmed in four "endangered" cities on the UNESCO World Heritage list using a custom camera configuration. It consisted of two synchronized motion picture cameras side by side (for stereopsis), 60-degree (horizontal) wide-angle lenses (for immersion), and a precision motorized tripod that rotated once per minute (for panoramics). In the final installation, visitors wore inexpensive polarized glasses for 3D and stood on a slowly rotating floor, rotating in sync with the image, resulting in the illusion that the movie was rotating around the visitors. Four-channel sound was composed from asynchronous recordings made at each location. (It is noteworthy that artificial accumulation of sound elements into a single composition often has no loss of credibility.) Five times of day were recorded at each of the four locations, plus in San Francisco.

Three studies were produced from the Be Now Here material to explore time artifacts [4].

The first study involved "match-cutting" three different times of day as the camera panned, starting with one cut per second and increasing to faster rates. The results are ambiguous, depending on what the viewer fixates. When one fixates on transient elements, such as people walking, the results are jarring. But when one fixates on the non-transient elements, such as buildings, whose color and shadows transform but remain stationary, the results appear smooth.

The second study required only two frames from the same location, with the camera pointing in the same direction, at different times of day. A small circular mask was made in PhotoShop that allowed a portion of one image to appear through the other. The mask could be moved in real time. The result was like an interactive "hole in time," with all non-transient elements (trees, buildings, etc.) staying perfectly registered. This simple effect appeared magic to many viewers, who thought much more was occurring. Anyone can replicate this effect with a camera, a tripod, and PhotoShop.

The third study was produced by projecting three images side by side as a triptych. Given the properties of the footage, several experiments were made. The most obvious was to simply offset the same footage by ten seconds between each screen, recreating a spatially seamless 180-degree scene of the same place at almost the same time. With no transient elements, the scene looked virtually perfect, since the sun and clouds did not change enough in ten seconds to be noticeable.

With transient elements in the scene, things become more complex. When the scene contains slow, prominent moving elements (such as a camel caravan in Timbuktu), the ten-second offset was enough to create mis-alignment of the moving elements between screens. When the scene contains fast, prominent moving elements (such as a security truck in Jerusalem), the repeated motion of the same element on all three screens is obvious. When motion occurs at the edge of the frame (such as a little boy standing still, then walking away at the instant his image exits the frame), this too is very obvious. But when many non-prominent transient elements appear in the frame (such as a crowd of people), the repetition on all three screens, offset by ten seconds, is difficult to detect.

Another three-screen experiment displayed the same place, spatially synchronized, but at very different times of day, e.g., dawn, mid-day, and early evening. In both a rural example (Angkor Wat) and an urban one (Dubrovnik), time artifacts were obvious: shadows fall in different directions, the sky and clouds change, and the color temperature shifts. Yet it's obvious that the triptych still represents the same place: "place-ness" appeared retained.

A final three-screen experiment displayed the same time, temporally synchronized, but in different places. Sunrise sequences were synchronized such that the sun appeared to move smoothly across the frame in Jerusalem, then continued moving across the next frame in Dubrovnik, then again in the next frame in Timbuktu. This sort of continuity is difficult to describe. The triptych clearly represented a noticeable continuity, but one that was more abstract than simple spatial continuity. Some observers noted the existence of continuity but couldn't detect what it was.

The Grounded VR Webcam

What makes such spatially coherent, accumulated images possible, in the end, is a grounded camera. Being physically anchored to a particular location, it enables perfect spatial registration on different image elements. This camera can be big, with immersive optics and robotic movement, and it can employ powerful computing. It can also be connected to the Internet via a very broadband connection (allowing "wide pipe" alternatives for destinations that also have such connectivity). It could also serve as a local head-end for smaller wireless cameras. Such an integrated system would be ideal not only for accumulated "hyper-images," but possibly for accumulated environmental data as well.

While much of the high-tech community is focussing on wireless, it is also accepting the compromises such low-bandwidth access entails, such as loss of large-scale immersion. Such a loss too often strips imagery of "sense of place." Grounded "VR webcams" offer an alternative and complimentary way of making and experiencing images, particularly place-based ones.

References

[1]
http://www.artchive.com/artchive/B/bruegel/bruegel_childrens_games.jpg.html
[2] http://www.crumbmuseum.com/history2.html
[3] http://www.naimark.net/projects/benowhere.html
[4] a web-based version of this publication, including video clips of the studies, can be found at:
http://www.naimark.net/writing/vrwebcam.html