WorldViews: Connecting the world through easy sharing of views

Don Kimber
FXPAL & WorldViews
3174 Porter Drive,
Palo Alto, CA, 94304
kimber@fxpal.com

Enock Glidden
WorldViews
3174 Porter Drive,
Palo Alto, CA, 94304
enockglidden@hotmail.com

Jennifer Marlow
FXPAL
3174 Porter Drive,
Palo Alto, CA, 94304
marlow@fxpal.com

ABSTRACT
The confluence of technologies such as telepresence, immersive imaging, model based virtual mirror worlds, mobile live streaming, etc. give rise to a capability for people anywhere to view and connect with present or past events nearly anywhere on earth. This capability properly belongs to a public commons, available as a birthright of all humans, and can been seen as part of an evolutionary transition supporting a global collective mind. We describe examples and elements of this capability, and suggest how they can be better integrated through a tool we call TeleViewer and a framework called WorldViews, which supports easy sharing of views as well as connecting of providers and consumers of views all around the world.

Author Keywords
Telepresence; interactive media; GIS systems; multimedia browsing; social media

ACM Classification Keywords
H.5.1 Multimedia information systems, H5.2 user interfaces, H.m Miscellaneous.

INTRODUCTION
This paper examines trends in digital media related to how people see the world, starting from a global perspective of human evolution. Pierre Teilhard de Chardin has described evolution in broad terms as having given rise not only to the geologic structure of our planet and its atmosphere, and then of a living biosphere, but ultimately to a noosphere - the space of mind and thought. [1] Human intelligence is not primarily an intelligence of the individual, but a collective intelligence. Everything we know and think as individuals, lays within a cultural matrix tying us together, and all our achievements, whether as grandiose as walking on the moon, or as mundane as meeting our basic daily needs, are part of a collective process.

4th International Workshop on Interactive Content Consumption at TVX'16,
June 22, 2016, Chicago, IL, USA

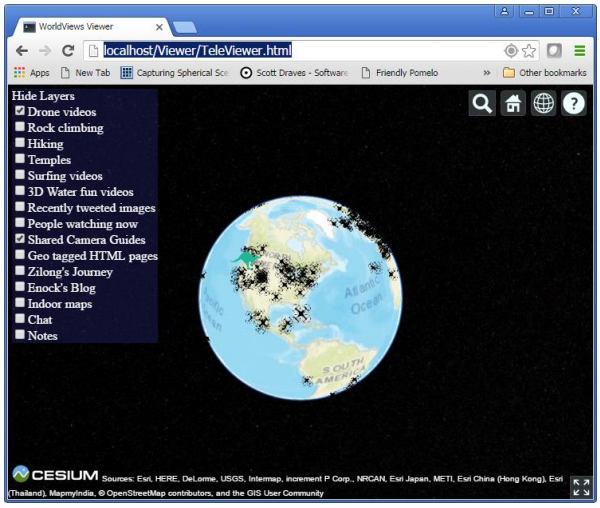

Figure 1. TeleViewer: a virtual earth based viewer portal providing easy views of any place on earth.

Teilhard de Chardin's noosphere, is now more commonly termed cyberspace. It entails not only the online computational and informational world, such as all data and knowledge resident in the cloud, but the association of human minds and consciousness with and through that space. John Barlow Perry writes "We will create a civilization of the Mind in Cyberspace." [2] The development of language, writing, printing, have all been important steps in the evolution of noosphere. But the development of electronic media, and ultimately of the internet, are to the evolution of human collective intelligence, as the transition from purely chemical signaling between cells to electrically based neural signaling was in the development of intelligent organisms. In the same sense as the development of nervous systems, and then ever more elaborate brains gave rise to intelligence, the deployment of information technology enables an awakening and blossoming of human collective intelligence.

In considering universal aspects of mind as they apply to the human collective mind, we begin to appreciate their connection with social and digital media. A full discussion

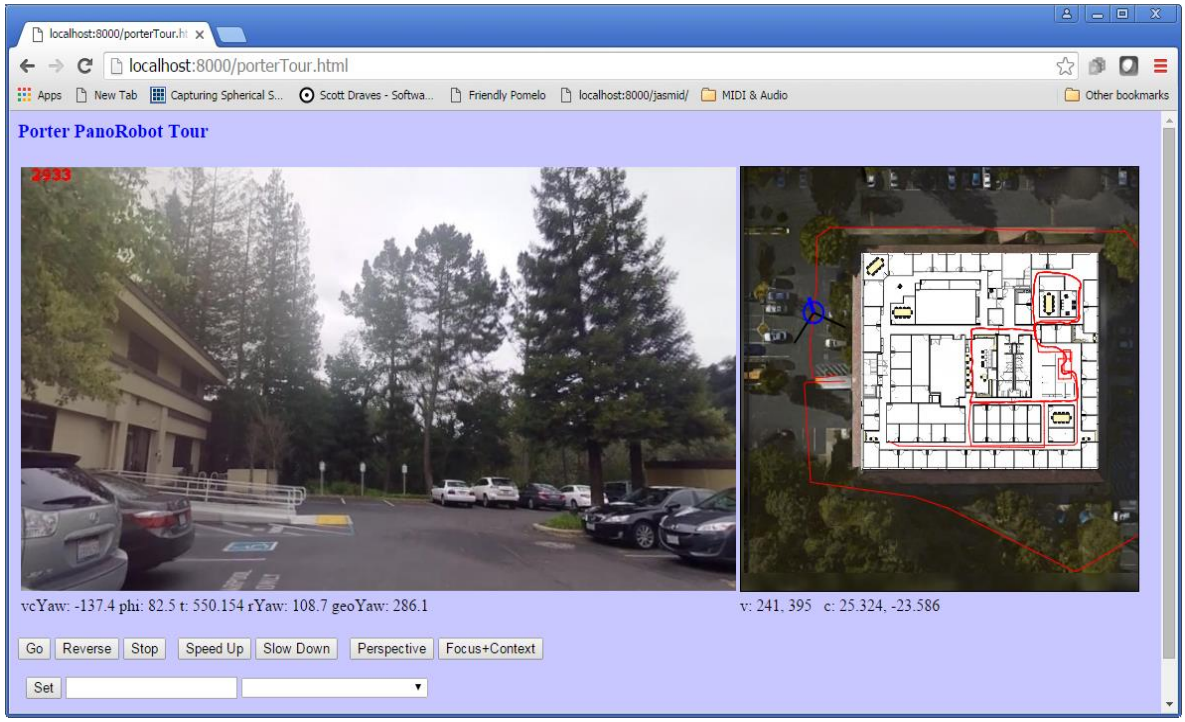

Figure 2: Panoramic Video Map

Somewhat like google street view, video is collected along paths, indoor and outdoor. The paths are shown on the map, and a user may drag anywhere along the path to see those from a given position. Tthe user can also "Go" to watch continuous motion along the path, while freely choosing the view direction. Video can be collected by robot or person carrying the special camera.

of this is off topic for this paper, but briefly, key aspects of mind include memory, perception, proprioception, emotion, cognition, intention for action, and facets of consciousness, such as mechanisms for control of attention. Digital media, and its integration through shared models of the world (e.g. digital maps and globes and our shared images of the earth) can be thought of as supporting humanity's collective cognitive and emotional systems - of how we think about and understand ourselves and each other. Streaming media are a key element of our collective perception and proprioception - how humans together perceive what is happening on Earth.

People now can increasingly view all kinds of events and places all over the planet. The ways people see remote events and places can be placed in several categories, which roughly are:

Image Based: Recorded images played back as images, video, possibly panoramic video, possibly along constrained paths, or possibly using image based rendering. This approach also includes lightfields and lumigraphs.

Model Based: Computer graphics generated "mirror worlds" that show what a place looks like - as far as the model is correct, from nearly any view.

Telerobotic: Live views from fixed or controllable cameras, or possibly on robotic devices - telerobots, drones, UAV's, etc.

Human Guided: Streamed views, in which people carry and stream from cameras, typically on their phone, or with wearable or assistive devices such as google glass, or "Polly" described below.

Rather than questioning which of these means is better than the others, we consider them as complementary and synergistic and having some overlap. For example, live video feeds can be embedded into a "mirror world" to help give the spatial context of the views, and can also be recorded as a supplement to the stored collection of images for image based viewing. Stored images can also be used to help in the generation and updating of virtual models used for. The models of are helpful in planning live events and in coordinating during those events.

We are undertaking a project called WorldViews, with the goal of better supporting the use of all these methods for

viewing the world One aspect of this is a prototype tool we call TeleViewer, which seeks to integrate these methods and providing a single portal through which to access them. A second is what we call the 'coodination layer' which is the means of using social media not only to push views, but to request or organize produced views.

PREVIOUS AND RELATED WORK

In this section we describe examples of each of the types of viewers described above. Although we give other examples, for reasons of familiarity, most of the examples will be of projects carried out at FXPAL.

Image Based Views

Many projects at FXPAL and elsewhere have shaped our views about live video streaming. For the FlyCam project in 2000 we built a panoramic video camera that could be used to live stream or record. [5] We investigated various usage scenarios, such as meetings, recording of events, etc. but found that one of the more interesting uses, inspired by the "Aspen Movie Map" [4], was "spatial capture" in which the camera is moved around along interesting paths such as roads or paths through a garden or park etc. Using this system, called FlyAbout, viewers could then later "virtually move" along those paths and get a sense of moving through the space. [6] This essentially provided a means by which people could make a virtual visit to a space, as well as capturing images that could be used as a backdrop for seeing an event. Figure 2 shows a web based video map of the FXPAL research center using such a map. Google street view has done something similar, on an impressively comprehensive scale, although it still does not provide video giving a smooth sense of motion through the space. Current efforts to crowdsource the collection of these kinds of comprehensive imaging from around the world include the startup Mapillary, which collects images taken by smartphone's during drives.

Panoramic cameras, indexed images from moving, or any other methods for collecting large numbers of images which are in some sense comprehensive or encyclopedic, could be viewed very generally as sampling of a sort of "God's-eye-view from everywhere" - the so called plenoptic function. [21] In its full glory, this is a seven dimensional function of position, direction, time, and wavelength, giving the intensity of illumination for any color viewed from given position in the given direction at the given time. While it is not possible to exactly determine the plenoptic function, many methods for trying to comprehensively capture, and support rich views, may be thought of as sampling of the plenoptic function. [3] Lightfields, and Lumigraphs are advanced methods for giving a richer sampling and playback of the plenoptic function. [7,8]

Another project that influenced our thinking about event streaming was FlySPEC which investigated the live streaming and recording of presentations and meetings in a room with 4 special cameras, each combining a Panoramic view, with a PTZ camera to gather close-ups. [9] The system used a hybrid human/automatic control system designed to work sensibly in the case of no human direction, direction from one person, or multiple inputs. The system used the low resolution panoramas and a model of the information value of various possible close-ups - to choose views. If one person made a view request it would be honored, by moving the camera accordingly. If many people made view requests, it would try to best satisfy the overall requests in terms of the information value model.

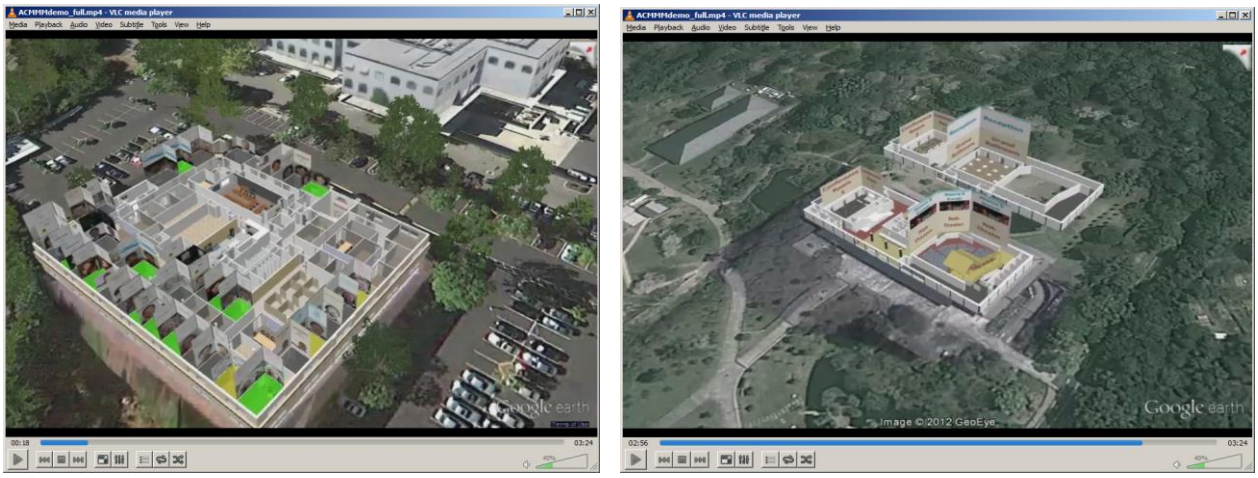

Figure 3: Mirror World showing FXPAL and ACM Multimedia conference in Nara

A Mirror World viewer running in the Google Earth web plugin. The left shows a view of the FXPAL research center, with colors indicating which rooms are occupied, billboards above rooms showing information about offices, etc. The left view is of the ACM Multimedia conference in Nara Japan, with billboards showing session information.

Model Based Views – Mirror Worlds

We worked for several years at FXPAL on a system with many cameras around our lab, that could track movement of people, but also provide views of locations where most activity was happening. [10] Although this was conceived partly as a possible surveillance system, we also considered how it might be used at venues such as conference centers. We combined the system with 3D modeling of the location, to create viewers that helped users understand the spatial layout and camera placement. [11] But also we built viewers we called 'Magic Mirrors' that showed the 3D model of the location, with the camera views, on large displays. These displays could be locally at locations around our space (e.g. we had one in the lobby, one in a main conference room, and one in our kitchen) that gave people awareness of the other locations, but also could be placed at remote locations such as our sister lab in Japan. Having the local displays of the views provided by the cameras made the space feel less "creepy" but did not fully address reciprocity when remote people were watching. That issue is addressed by in socially mediated systems discussed below.

The culmination of our work combining multiple cameras with 3D models, was our "Mirror World" project, in which we had detailed models of several locations, including our own lab, our sister lab in Yokohama, and a conference venue in Nara, Japan. (Figure 3.) These models could all be viewed in a web browser using the Google Earth plugin, with Javascript API. Using this system, someone could see what was happening in our lab, including seeing live views from several cameras, views of presentation screen content updated live during presentations, and images on some whiteboards updated when pictures of them were taken with a special smartphone app. The user could then "fly over to Yokohama" and get a similar view there, or to the conference venue model to see how that looked. The tool could be used by prospective attendees and conference planners beforehand for planning. For example they could see the models of the rooms presentations would take place in, and also could see in Google Earth nearby temples. Although the Nara venue did not have advanced camera, screen capture, and display facilities, we anticipate that many venues in the future will, as well as having telerobots for remote participants.

Telerobotic Views:

Increasingly telepresence robots are used at many locations to make those locations available to remote viewers. Some of these are private for enterprise use, but many are starting to be available at public locations like museums. [16] Also, drones are increasingly used to provide views of many locations. For example the sites travelbydrone.com and travelwithdrone.com each provide thousands of curated videos showing scenery from all over the world, with scatter maps showing the locations where videos are available. Makers of popular consumer drones, like DJI and Parrot provide sites for sharing videos taken by customers, and periscope recently announced they will

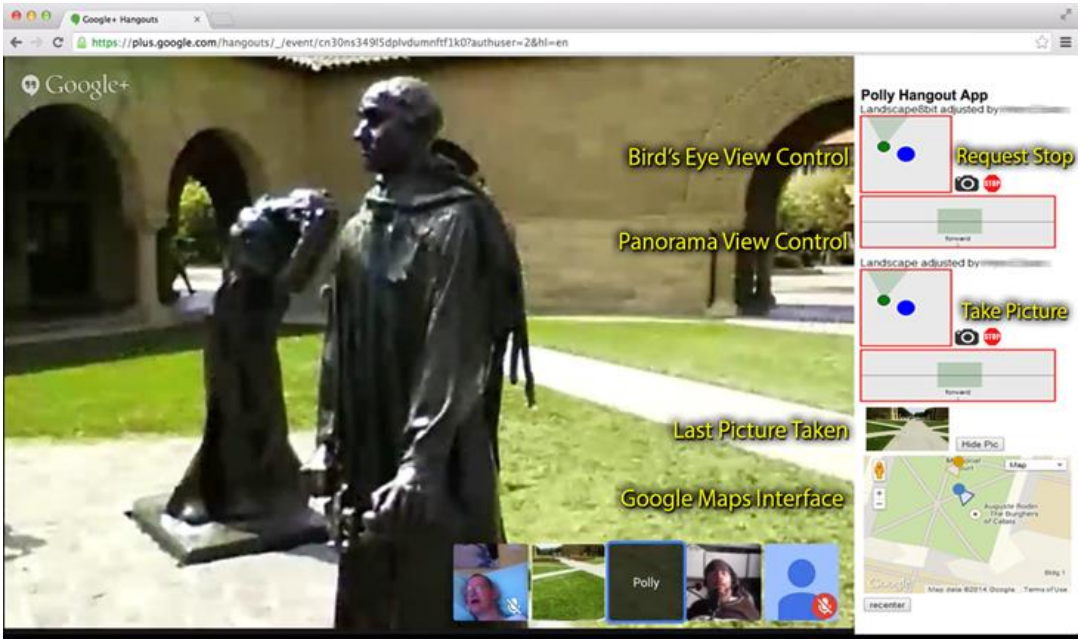

Figure 4: Polly Hangout interface for remote participants

Remote users controlled Polly through a google hangout and app. Here there were 3 remote participants, and two Polly devices, which could be controlled using a GUI on the right. A map show positions of the guide is also shown on the right.

partner with DJI to allow streaming of drone video through periscope.

Human Provided and Socially Mediated Views

The projects describe above mostly address the three aspects of rich image capture, use of models, controllable live streams, but except for some of the telepresence robots, don't address some of the more important social issues. These include the human involvement by local participants as guides for providing views to remote participants, or reciprocity of view, so that local people would not feel 'spied on' and remote people would not feel 'disembodied'.

The Polly project takes a different approach to providing remote views of places or events, following ideas from the tele-actor project [13] of letting a person with a camera act on a behalf of remote participants. A `guide' carries a device, which remote viewers may connect to and control. The metaphor is of a "parrot on the shoulder" serving as an avatar for the remote participant. The device includes a remotely controllable camera, with a stabilization gimbal that reduces annoyance of shaking, but also provides a measure of view independence for the remote participant. [14]

We implemented various Polly prototypes, but most of our investigation was with versions that used a smartphone for the camera, and with the phone display providing a view of the remote participant. (Figure 5.) This allows social interaction with the guide, or with other people present locally. We mostly looked at scenarios with a single guide and single remote viewer, but also tried a few events, such as a tour of the Stanford University campus, with multiple

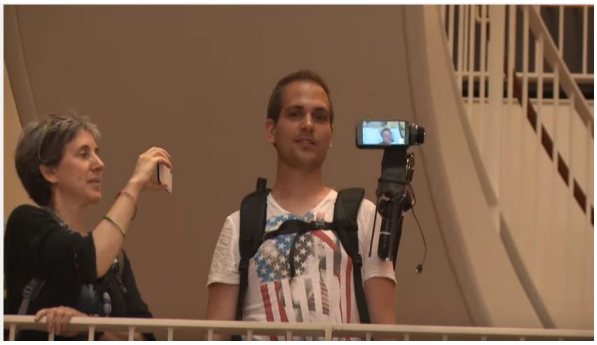

Figure 5: Polly telepresence device. Shoulder worn metaphorical "parrot on shoulder" provides telepresence to remote person.

remote participants joined together in a google hangout, which also could have more than one Polly available. A hangout app allowed anyone in the hangout to control any Polly. (Figure 4.)

Contention was controlled in a simple manner and did not appear to be a problem with say 5 people in the hangout. When a Polly device was available for control, any user could perform UI actions to remotely control it. The interface would then indicate to everyone that Polly was under that viewer's control, and only that viewer would be able to control it. After a timeout of a few seconds without any control actions, that Polly became available again to anyone wishing to control it. Meanwhile, everyone could

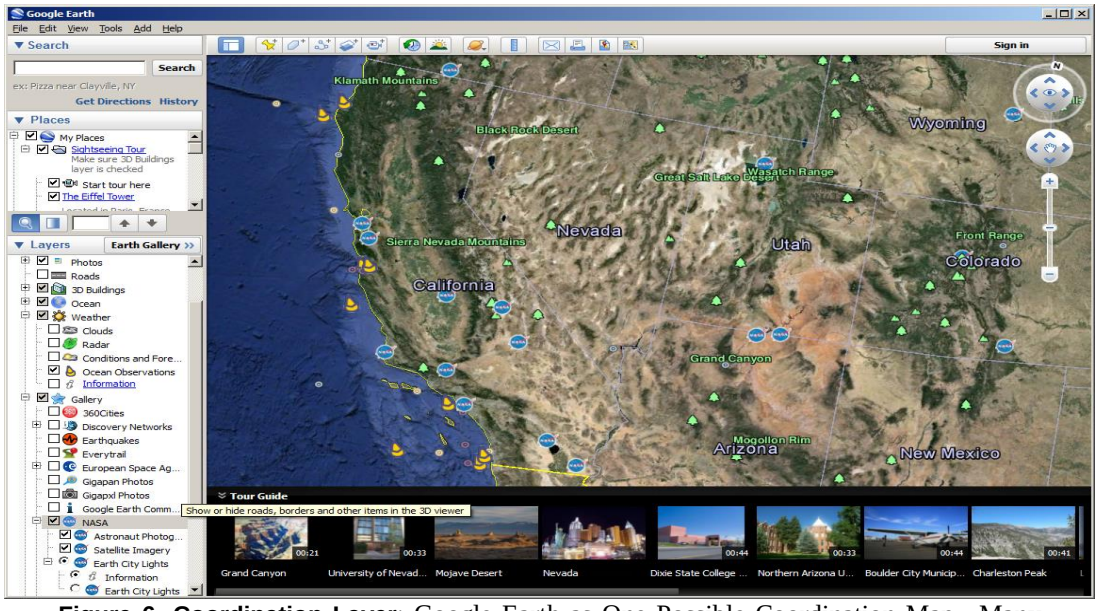

Figure 6: Coordination Layer: Google Earth as One Possible Coordination Map. Many layers can be listed showing different kinds of source views (e.g. drone, robot, human carried, etc.) as well as positions of guides, events, questions, etc.

see the views from the camera or whichever Polly they chose to watch. A map also showed the position on a map of the Polly devices and guides. We consider this a prototype of what could be called a "coordination map" for much larger events, or access to guides and views.

The Polly prototype was interesting in that it gave remote users some autonomy, and some local presence through the display, but was bulky, unwieldy and fragile. We expect that over the next few years, much better Polly like devices will be build, with varying capabilities. The Parrot Bebop drone for example, could form the basis for a lightweight Polly that pushes the parrot metaphor further by not only sitting on the shoulder - but by flying under remote control. That drone has a large field of view 14 megapixel camera with the capability for digital image stabilization in real time, eliminating the weight and fragility of a mechanical gimbal.

However, we felt the greatest opportunity for short term contribution to the advancement of 'guided telepresence' scenarios is not in building the more sophisticated kinds of devices we anticipate, but actually in choosing simpler scenarios using devices widely available. A good example of this kind of approach is the Virtual Photo Walk project [15], which pairs up photographers on photo shoots at interesting locations or events, with remote participants, in many cases people with limited mobility (such as through disabilities) not able to visit many places. The rigs used are simply a DSLR camera, together with an iPhone running a hangout app. The remote participants communicate with the participants about what they would like to see, which views for pictures they would like, etc. They cannot directly control views or take pictures, but communicate these intentions through the hangout.

This inspired another project, SharedCam, which lets a local guide with a smartphone pair up with a remote participant, in taking high resolutions pictures on the smartphone. Either the guide or remote viewer may 'click the shutter' to take the picture, which is then automatically uploaded and made available at full precision. The system uses a version of the jumpch.at WebRTC app for streaming, modified to allow full resolution pictures to be taken at the same time.

ISSUES, OPPORTUNITIES & RESEARCH QUESTIONS:

Rise of a New Role: The Virtual Guide

Just as blogging provided a new and greatly expanded channel for writers and other creative individuals to share their content, cultivate a following, and take on the new role of "blogger", the new streaming technologies will enable a new role of "virtual guide." People with the right mix of personality, knowledge, access to interesting locations or events, may thrive as "local experts" or "guides." They may choose their own schedule, streaming at their whim, or at high value locations and events. But others will be "on call" by running Apps that notify them when remote people have questions or want a quick view "Who can show me the Golden Gate bridge now, and chat with me a bit about it?"

An interesting question is how the economics for guides will work out. We anticipate that there will be cases of people making money as virtual guides - indeed it may even become a full time profession for some. Others will do it for fun or to build up reputation, or motivated by a sense of service, such as for cultural sharing or to provide experiences to disadvantaged individuals who would not otherwise not have access. Mechanisms like earning hearts in periscope support the non-monetary and reputation aspects of motivation. We anticipate tourism scenarios whereby people make money as independent guides for visiting tourists, but serve as virtual guides as a way of getting business.

Figure 7: Coordination Layer: This shows a coordination map as it might be seen zoomed in to a small area.

As people embark on personal adventures they will be able to use this technology to bring people along with them. They will have the opportunity to monetize their own content by using the videos on crowdfunding sites to show people what they are doing and why they need help with funding. For example, the OpenROV project set up a site called OpenExplorer to let people list planned explorations and seek followers and support. [19] In conjunction with blogging and other social media, adventurers will be able to allow people to have an immersive experience almost as if they are on the adventure with them. This will allow the adventurer to engage followers in an exciting way and may encourage their followers to help with funding, so they can see more of this content. An adventurer may even be able to have their own site where they allow a pay per view like experience where the follower pays for access to be on the adventure in a virtual way. Professional adventure guides may also take advantage of this technology by enabling them to sell hybrid guided expeditions with in person guiding and virtual guiding. When people get more funding more content will become available and more people may be inspired to go on their own adventures.

WorldViews TeleViewer

The state of the ecosystem for accessing views of remote parts of the world, is reminiscent of the state of the internet before the development of the web. A variety of information sources were accessible by various tools such as ftp, gopher, etc., but a single simple to use access tool was lacking, as was an integrative framework for easily interconnecting sources in useful ways. The advent of HTML, together with open source browsers and HTTP servers, quickly changed that. The situation today seems similar. Digital earths such as google earth, provide a "geo browser" which is a good start. But what is needed are good open source equivalents, and standards for integration of the sources into those tools, and a convergence of methods for accumulating and sharing layers of GIS data.

WorldViews is dedicated to a more unified access to all the methods described above for seeing the world. We are developing a prototype browser called TeleViewer built upon Cesium, based on WebGL. [20] Information layers corresponding all kinds of views may be added or subscribed to, and appear in the browser, or as placed links. Figure 1, shows a view of the whole world, with a set of layer choices on the left. These layers correspond to sources such as geotagged videos associated with various topics, drone videos, travel blogs, live sources such as periscope & meerkat, access to ShareCam apps, etc. There is also a layer corresponding to dynamic queries for geo tagged videos matching given kewords.

WorldViews Coordination Layer

A key component of a good event streaming system, or of the overall ecosystem of various tools and people involved in live streaming, is a coordination mechanism for seeing what events are happening, which views are available, what views or information are requested, understanding where views are taken from, etc. This could be presented in a variety of ways, but one natural way is as layers and markings on a 2D maps or 3D virtual globes. These can show were prospective guides are located, automatic or remotely controllable views are available, and what people want to see or learn about. It also supports scheduling, and a market, so that people can make arrangements with guides. Furthermore the statistics at this level, which may include "likes", "hearts", number of followers of a stream, etc. contribute to the attention focus mechanism by which large numbers of remote participants may be directed to the most interesting or relevant content. This layer coordination layer can also be used to show which video devices, particularly remotely controllable devices such as telerobots, drones or UAV's are available, or to help people schedule access for such devices. Overall this layer can act as a sort of "Uber for views" where views are provided by humans, robots or some hybrid systems such as Polly.

The Production Value Chain

As we have said, the production values for video available live during an event is typically lower than what is available after a period of post-production. One reason is bandwidth or connection limitations during the event. It is generally easy to capture the event with many high resolution cameras that record locally, such as onto SD cards in the cameras, even in remote areas with little or no connectivity. More typically some level of connection is possible, but not enough for HD video, or for more immersive media such as panoramic 4K video. Another issue is that during a large event, with many video channels, only a small fraction of the overall video may be interesting, and editing possible after the fact can produce a much higher value summary.

Of course, one typically makes best efforts to provide highest value at each stage, but an interesting question is whether the remote participants can provide input before the event, or during the event, which have impact on the final higher value video, and which give the participants a greater sense of involvement or ownership of the event. For example, could some involvement at low fidelity during the event, in which input is solicited about what to see and where to go, combined with a later viewing at higher quality or greater immersion (e.g. on a head mounted display), give an overall stronger sense of having been at the event?

CONCLUSION

We believe it is time for a convergence and integration of the methods available for people to "virtually explore" the earth, see other places and connect with other people. This

integration is a further step in the evolution of human collective intelligence, particularly of our "perceptual" and "proprioception" whereby we see and share our views of what is happening in our world. A step in that direction is a provided by open source Viewers, such as TeleViewer, that integrate sources through one tool, and that allow for discussion and bidirectional requests for views. We believe it is essential that the core technologies for this new capability exist in open source form, and that a core level of content belong to the creative commons.

After working for many years on the kinds of technologies described in this paper - mostly guided by a technology perspective - we have now shifted towards a more humanistic perspective. Is it really helpful to humanity to create this kind of system, or is it just the next level of distraction to come along? We believe it is of value to humanity not only as a crucial element of an emerging global collective mind, but also as a means to connect people around the world, and to promote shared understanding and solidarity as global citizens. We anticipate ever greater ease with which people will be able to see and understand other cultures, other parts of the world, and connect through what might be called virtual tourism. We believe the connections formed in this way, which in many cases will become "real life" friendships, will play a crucial role in creating a more beautiful world.

REFERENCES

1. Teilhard de Chardin, P. (1959). The phenomenon of man (Vol. 19). London: Collins.
2. Barlow, J. P. (1996). A Declaration of the Independence of Cyberspace.
3. Zhang, C. (2004) On sampling of image based rendering data. Ph.D. Thesis, Carnegie Mellon University, Pittsburg, PA.
4. Lippman, A. (1980). Movie-maps: An application of the optical videodisc to computer graphics. ACM SIGGRAPH Computer Graphics. ACM.
5. Foote, J., & Kimber, D. (2000). Flycam: Practical panoramic video and automatic camera control. Multimedia and Expo, 2000. ICME 2000.
6. Kimber, D., Foote, J., & Lertsithichai, S. (2001). Flyabout: spatially indexed panoramic video. Proceedings of the ninth ACM international conference on Multimedia. ACM.
7. Gortler, S. J., Grzeszczuk, R., Szeliski, R., & Cohen, M. F. (1996). The lumigraph. Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. ACM.
8. Levoy, M. (2006). Light fields and computational imaging. Computer, (8), 46-55.
9. Liu, Q., Kimber, D., Foote, J., Wilcox, L., & Boreczky, J. (2002). FLYSPEC: A multi-user video camera system with hybrid human and automatic control. Proceedings of the tenth ACM international conference on Multimedia. ACM.
10. Girgensohn, A., Kimber, D., Vaughan, J., Yang, T., Shipman, F., Turner, T., et al. (2007). DOTS: support for effective video surveillance. Proceedings of the 15th international conference on Multimedia. ACM.
11. Rieffel, E. G., Girgensohn, A., Kimber, D., Chen, T., & Liu, Q. (2007). Geometric tools for multicamera surveillance systems, ICDSC'07, ACM/IEEE International Conference on Distributed Smart Cameras.
12. Kimber, D., Shingu J., Vaughan, J., Arendash, D., Lee, D., Back, M., Uchihashi, S. (2012) Through the Looking-Glass: Mirror Worlds for Augmented Awareness & Capability, Proceedings of the twentieth ACM international conference on Multimedia. ACM.
13. Goldberg, K. Y., Song, D., Khor, Y.-N., Pescovitz, D., Levandowski, A., Himmelstein, J. C., Shih, J., Ho, A., Paulos, E., and Donath, J. S. (2002) Collaborative online teleoperation with spatial dynamic voting and a human "tele-actor". In ICRA, IEEE (2002), 1179–1184.
14. Kimber, D., Proppe, P., Kratz, S., Vaughan. J., Liew. B., Severns. D., Su, W., (2014) Polly: Telepresence from a Guide's Shoulder, Computer Vision - ECCV 2014 Workshops, Volume 8927 of the series Lecture Notes in Computer Science pp 509-523
15. Virtual Photo Walks. Organization founded by John Butterill. http://www.virtualphotowalks.org/
16. Virtual Museum tours
    http://www.nma.gov.au/engage-learn/robot-tours
    http://futureofmuseums.blogspot.com/2014/05/exploring-robots-for-accessibility-in.html
17. Drone Touring sites
    http://travelbydrone.com/
    http://travelwithdrone.com/
    http://www.droneworldvideo.com/
    http://www.airpano.com/google_map.php
18. Pierce, Dennis. "Google Seeks Teachers to Pilot 3D Virtual Field Trips", The Journal, 28 Sept. 2015. Web. 22 May 2016
    https://thejournal.com/articles/2015/09/28/google-seeks-teachers-to-pilot-3d-virtual-field-trips.aspx
19. OpenExplorer, https://openexplorer.com/home
20. Cesium, "An open-source JavaScript library for world-class 3D globes and maps" https://cesiumjs.org
21. Adelson, E. and Bergen. J. The plenoptic function and the elements of early vision. in "Computational Models of Visual Processing" Landy and Movshon (eds), (pp. 3-20). Cambridge, MA: MIT Press (1991).