Gigapixel Television

David J. Brady

Duke Imaging and Spectroscopy Program, Duke University, Durham, North Carolina, USA
e-mail: dbrady@duke.edu

Abstract

We suggest that digitally zoomable media will emerge from the integration of broadcast television and interactive networks. We review progress in multiscale cameras, consisting of parallel arrays of microcameras behind a common spherical objective, and physical layer compressive measurement. Each of these technologies is essential to “zoomcast” media in which each viewer will be able to analyze events at the fundamental physical limits of spatial and temporal resolution.

1. Introduction

Four score and seven years ago, Kenjiro Takayanagi began research in television. An anniversary of four score and seven is famously remembered in Abraham Lincoln’s speech at Gettysburg as the time between the American Revolution and Civil War. In comparing these two anniversaries, one is struck by how much more dramatic transformations in life and technology were in the twentieth as opposed to the nineteenth century. When Takayanagi began his work, who would have imagined that one would today be able to instantaneously access images and information from anywhere in the world from a tablet or cell phone?

Developments in television since Takayanagi’s day, on the other hand, have been relatively modest. In the half century following Takayanagi’s initial demonstration TV evolved through broadcast standards and the introduction of color. Live broadcast of sporting events and news transformed culture and connected the world in many ways. Only recently, however, have HDTV standards replaced NTSC and PAL systems that stood for many years. High definition standards such as 1080p are not much more than double the resolution of standards deployed 70 years ago and even the best available digital video cameras, operating under the 4K standard, are only 8x the resolution of NTSC.

At the same time, progress in media delivery has been truly revolutionary. Where 70 years ago real-time media was dominated by large TV and radio networks, the cellular networks and internet technologies today enable any individual to establish their own broadcast channel on YouTube and similar sites. Interactivity is the defining feature of emerging media. Users expect the ability to search for their own perspective and control their own content. At the same time, interest in the broadcast of major sporting and news events remains high. In recent years such broadcasts have been augmented with digital technologies and denser camera networks to improve the viewer experience, but the basic nature of the broadcast image has remained unchanged.

This paper considers strategies to radically increase the information content of broadcast media. Current image formats have are designed to approximately match human perception. They operate at 30-60 frames per second, capture three colors and focus on megapixel scale images over the central angular range of human perception. While these standards may be appropriate when millions or even billions of viewers are satisfied to view identical images, they do not allow interactive personalized experiences. The broadcaster may zoom in time and space to focus on interesting aspects of an event, but viewers are not allowed to explore on their own.

As we enter a second century of broadcast media, our goal should be to capture and broadcast images, sound and data at the limits of physical space-time resolution rather than at the limits of human resolution. The angular resolution

Presented at the 14th Takayanagi Kenjiro Memorial Symposium, Shizuoka University, Hamamatsu, Japan, 27 November 2012

of imaging systems is limited by atmospheric effects, but may exceed 30 to 50x human acuity at sporting events. Physically achievable temporal resolution is limited by photon flux, but often exceeds human perception by several orders of magnitude. Images broadcast with such resolution would allow not just television or telepresence, they would allow viewers to independently and/or collaboratively explore vast hyperspaces. Using mobile networks, even viewers attending live events will augment their experience using tablets or head-mounted displays.

Of course, it will not be possible or even desirable to transmit the entire broadcast data cube to every viewer. Each viewer will receive a data stream at bandwidths matched to the capacity of their display device. But every viewer will be able to search the entire data cube and digitally zoom in and out in space and time. With this capacity in mind, we refer to this new paradigm as “zoomcasting.”

Zoomcasting will require sophisticated data networks to transmit, cache and distribute media. Current streaming media platforms, such as YouTube and Netflix, offer hints of this potential and static gigapixel image distribution platforms, such as gigapan.org, illustrate the potential for real-time interactive zoom.

Compact low-power gigapixel cameras are an essential enabling technology for zoomcasting. The Duke Imaging and Spectroscopy Program is developing two critical technologies for such cameras. The first, multiscale cameras, combines a new approach to wide field high resolution lens systems with a parallel electronic architecture to enable single objective aperture capture of wide field images with instantaneous field of view at 10-50 microradians. The second, compressive video, introduces physical layer coding strategies to increase effective frame rate and to manage depth of field and dynamic range while reducing operating power. This paper reviews these technologies and their implications for zoomcasting.

2. Multiscale Cameras

The number of pixels a camera can resolve is limited by wave diffraction to the space-bandwidth product, which is proportional to the ratio of the aperture area and the square of the wavelength. For systems operating in the visible spectrum, the space bandwidth product is in the range of 1 megapixel for 1 mm aperture cameras, 100 megapixels for 1 cm aperture cameras and 10 gigapixels for 10 cm aperture cameras.

Conventional lens systems do not achieve diffraction-limited information capacity for apertures larger than 1 cm because geometric and chromatic abberations, rather than diffraction, dominate lens performance on these scales. To address this problem, my group proposed a hierarchical approach to lens design in which secondary microcameras locally correct field curvature, defocus, geometric and chromatic aberrations arising in a larger objective lens [1]. In subsequent work, we developed a scalable architecture for cameras resolving 1-50 gigapixels combining spherical monocentric objective lenses with parallel microcamera arrays [2-4].

Microcamera arrays create virtual focal plane arrays scalable to arbitrary size. The disadvantage of this approach relative to monolithic or mosaicked integration of large electronic focal planes is that the effective pixel size is increased, which means that the objective lens must ideally be designed to operate around f/3, and quantum efficiency is reduced by vignetting. Advantages of the multiscale approach are that the yield and cost of large effective focal planes is linear in the yield and cost of the much smaller microcamera focal planes, that independent focus and exposure control can be implemented in each subfield and that parallel electronic read-out and processing is enabled. The ability to operate microcameras at diverse frame

Figure 1. Baseball game captured using AWARE 2. An interactive version of this image is online at http://www.gigapan.com/gigapans/111587

rates and resolutions is also important for system power management.

As described in [5], our team has developed optical manufacturing and sensor packaging technologies for multiscale camera systems. We have also created a largescale camera operating system that enables gigapixel snapshot capture in 10 milliseconds. With current hardware retrieving images from the camera requires 12 seconds per gigapixel, but networking technologies in future cameras will reduce this time.

Our first multiscale camera series, AWARE 2, was constructed in 2011. AWARE 2 achieves 40 microradian ifov over a 120 degree field of view. The current system is asymmetrically populated with microcameras to capture a 120 by 60 degree field.

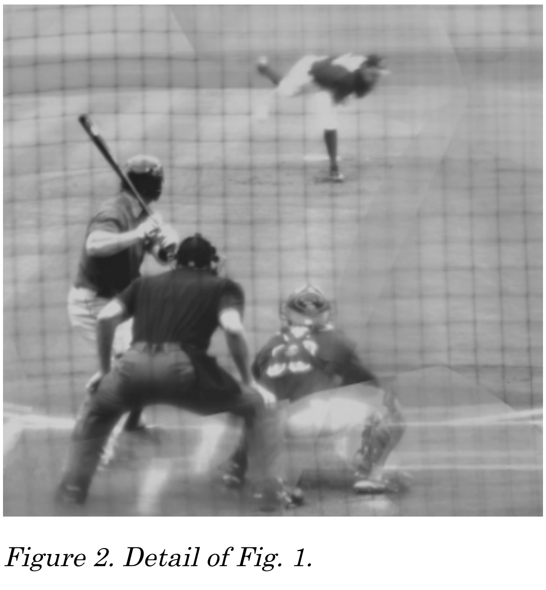

Figure 2. Detail of Fig. 1.

Figure 1 is an AWARE 2 image captured at the Durham Bulls vs. Columbus Clippers baseball game on August 7, 2012. Oval sections in the image correspond to the 98 microcamera frames stitched to form the full panorama. As illustrated by the ball in hanging between the pitcher and the batter in Fig. 2, this image fully captures an instant in the game. While the first generation of AWARE 2 cameras has been successful as a demonstration of the capacity for gigapixel image capture and manipulation, image artifacts in visible in Fig. 2 illustrate imperfections in the microcamera fabrication process that currently keep AWARE 2 from achieving diffraction or pixel limited performance. AWARE 2's size, weight and power (SWaP) are also larger than desireable. The camera is 0.75 by 0.75 by .5 meters in size, weighs ~75 kg and requires 500W of power during image capture.

Building on manufacturing and operational lessons learned from AWARE 2, our team has constructed second generation lens and electronics platforms in 2012. These platforms are being used to construct the AWARE 10 camera, which resolves 25 microradian

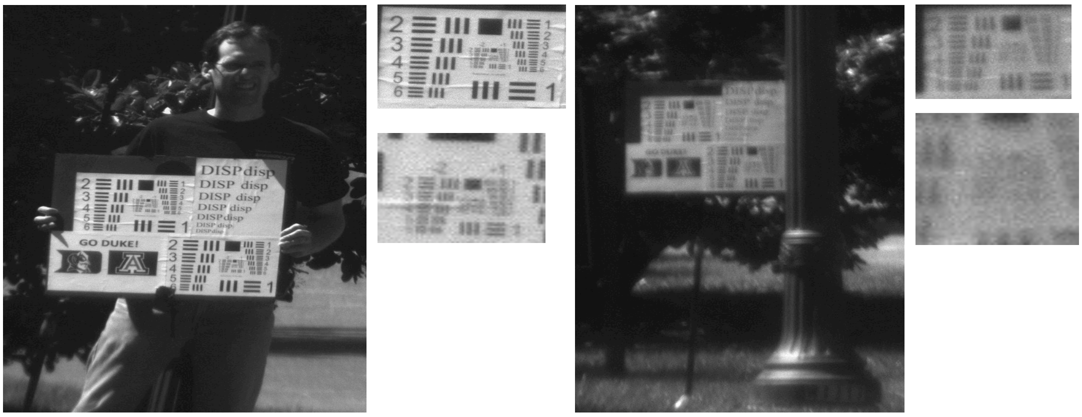

Figure 3. Resolution target captured at 66 meters by AWARE 10 optics (left) and AWARE 2 (right), with zoomed details for each.

ifov over a 100 by 60 degree field of view using approximately 400 microcameras to capture a 5 gigapixel image. As illustrated in the comparison shown in Fig. 3, AWARE 10 avoids the optical artifacts observed in the first generation AWARE 2 design and achieves near diffraction-limited performance.

AWARE 10 also achieves substantial reductions in electronics volume per pixel. A nominal 4x reduction in volume is achieved by operating 8 sensors per microcamera control processor rather than 2 sensors per processor in AWARE 2.0. We anticipate that 10-100 AWARE 10 and updated AWARE 2 systems will be constructed in 2013 and 2014. These systems may be used to zoomcast gigapixel frames at up to 6 frames per minute. While this is far from the dream of video rate or faster zoomcasting, it represents a significant first step.

3. Compressive Video

AWARE 10 requires liquid cooling of the microcamera control array and the frame rate is limited by maximum heat loads on the sensor heads. Even with conventional cameras, video systems at 4K resolution lag far behind 20-100 megapixel data captured with snapshot systems.

Camera operating power is linear in bandwidth. In zoomcast systems, one will generally be satisfied to reduce camera head bandwidth even if such reductions lead to increased downstream power demands from image restoration algorithms. The challenge is to avoid “hot spots” with extreme power demands s. Hot spots, in particular high power at the sensor head, determine the rate of information capture.

Over the past several years, our group has explored physical layer coding strategies to compress image data prior to digitization and thus reduce sensor bandwidth and power [6-8]. We have been particularly successful in demonstrating real-time hyperspectral imaging using image plane modulation [9]. Image plane modulation holds further promise for compressively coding focus [10] and dynamic range. More recently, several groups have explored image plane modulation for video compression [11], which may be directly effective in reducing bandwidth and power in high pixel count cameras.

4. Conclusions

Multiscale cameras and compressive measurement strategies enable revolutionary improvements in image data capture rates. Using these technologies, I suggest that it possible to broadcast individually zoomable data to millions of users, to update television to new “zoomcast” format. While many technical challenges must be resolved to enable zoomcasting, they are not greater than the challenges that Kenjiro Takayanagi overcame in his career in television.

Presented at the 14th Takayanagi Kenjiro Memorial Symposium, Shizuoka University, Hamamatsu, Japan, 27 November 2012

Acknowledgments

This work was United States Defense Advanced Projects Research Agency through the AWARE Wide Field of View program and the Knowledge Enhanced Compressive Measurement program.

References

[1] Brady, D.J. and N. Hagen, *Multiscale lens design.* Optics Express, 2009. **17**(13): p. 10659-10674.

[2] Marks, D.L., E.J. Tremblay, J.E. Ford, and D.J. Brady, *Microcamera aperture scale in monocentric gigapixel cameras.* Applied Optics, 2011. **50**(30): p. 5824-5833.

[3] Son, H.S., D.L. Marks, J. Hahn, J. Kim, and D.J. Brady, *Design of a spherical focal surface using close-packed relay optics.* Optics Express, 2011. **19**(17): p. 16132-16138.

[4] Tremblay, E.J., D.L. Marks, D.J. Brady, and J.E. Ford, *Design and scaling of monocentric multiscale imagers.* Applied Optics, 2012. **51**(20): p. 4691-4702.

[5] Brady, D.J., M.E. Gehm, R.A. Stack, D.L. Marks, D.S. Kittle, D.R. Golish, E.M. Vera, and S.D. Feller, *Multiscale gigapixel photography.* Nature, 2012. **486**(7403): p. 386-389.

[6] Pitsianis, N.P., D.J. Brady, and X.B. Sun, *Sensor-layer image compression based on the quantized cosine transform*, in *Visual information processing xiv*, Z.U. Rahman, R.A. Schowengerdt, and S.E. Reichenbach, Editors. 2005. p. 250-257.

[7] Portnoy, A.D., N.P. Pitsianis, X. Sun, and D.J. Brady, *Multichannel sampling schemes for optical imaging systems.* Applied Optics, 2008. **47**(10): p. B76-B85.

[8] Shankar, M., N.P. Pitsianis, and D.J. Brady, *Compressive video sensors using multichannel imagers.* Applied Optics, 2010. **49**(10): p. B9-B17.

[9] Wagadarikar, A.A., N.P. Pitsianis, X.B. Sun, and D.J. Brady, *Video rate spectral imaging using a coded aperture snapshot spectral imager.* Optics Express, 2009. **17**(8): p. 6368-6388.

[10] Brady, D.J. and D.L. Marks, *Coding for compressive focal tomography.* Applied Optics, 2011. **50**(22): p. 4436-4449.

[11] Hitomi, Y., J. Gu, M. Gupta, T. Mitsunaga, and S.K. Nayar. *Video from a single coded exposure photograph using a learned over-complete dictionary.* in *IEEE International Conference on Computer Vision (ICCV)*. 2011.