Petapixel Photography and the Limits of Camera Information Capacity

David J. Brady*[a], Daniel L. Marks[a], Steven Feller[a], Michael Gehm[b],
Dathon Golish [b], Esteban Vera [b], David Kittle[a]
[a]Department of Electrical and Computer Engineering, Duke University, Box 90291, Durham, NC 27708,
[b]Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ

ABSTRACT

The monochromatic single frame pixel count of a camera is limited by diffraction to the space-bandwidth product, roughly the aperture area divided by the square of the wavelength. We have recently shown that it is possible to approach this limit using multiscale lenses for cameras with space bandwidth product between 1 and 100 gigapixels. When color, polarization, coherence and time are included in the image data cube, camera information capacity may exceed 1 petapixel/second. This talk reviews progress in the construction of DARPA AWARE gigapixel cameras and describes compressive measurement strategies that may be used in combination with multiscale systems to push camera capacity to near physical limits.

Keywords: Multiscale cameras, gigapixel imaging

1. INTRODUCTION

Over their brief history [1], digital cameras have reduced the cost, in both time and money, of image capture, storage and communication by many orders of magnitude. Despite this progress, however, camera capacity remains far below fundamental limits. Imagers have often been designed to match the limits of human acuity at 300 milliradian instantaneous field of view (ifov), 3 color channels and 30-60 frames per second. While this represents an apparently formidable 1 gigapixel/second of image data, cameras that greatly exceed human acuity are both desirable and feasible.

Growth in image data rate is determined by technical, economic and political constraints. Technical constraints include optical resolution, focal plane pixel technology and electronic read-out, communication, processing and storage technologies. Economic constraints include both the cost of constructing and operating camera systems and the market demand and economic value of images. Politics plays a role in the definition and implementation of imaging and broadcast standards. As illustrated in Fig. 1, electronic image pixel count has been relatively static since the first introduction of electronic imaging for broadcast television in 1940. Indeed, the first consumer digital cameras, introduced in the 1980's, were "still video cameras" operating with pixel count matched to the original NTSC standard [1]. Modern "HD video" exceeds the pixel count of the original standard by a factor of 5. This increase in image quality over ¾ century pales in comparison to the 10 order of magnitude increase in single channel communication bandwidth over the century and half from Morse code to wavelength multiplexed optical fiber and the similar increase in computing capacity in half of century of Moore's law. Since its launch at video resolution, the pixel capacity of digital still imaging has thus far grown along an exponential path characteristic of modern information technologies.

Primary factors limiting video pixel count have been, first, the assumption that image data should be limited by the 6-20 MHz channel capacity of radio broadcasts and, second, the assumption that little advantage would accrue from broadcasting at rates much in excess of the single channel human information capacity. These assumptions have limited growth in both audio and video broadcast and communications standards. For example, the G.722 telephony standard merely doubles ancient 3.5 kHz bandlimits and, as noted above, HD video covers a mere 2-3x improvement in resolution relative to standard definition. While political and technical barriers make more substantial improvements challenging, it is important to note that substantially different economic and market forces bear on video and audio. Audio data processing, while technically feasible and of occasional market interest, is much less common than video processing. Market interest in spatial and temporal zoom is much higher for image data as people study scenes to find otherwise unnoticed events and features.

*david.brady@duke.edu; phone 919 660 5394; www.disp.duke.edu

Computational Imaging XI, edited by Charles A. Bouman, Ilya Pollak, Patrick J. Wolfe,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8657, 86570B • © 2013 SPIE-IS&T
CCC code: 0277-786X/13/$18 • doi: 10.1117/12.2014274

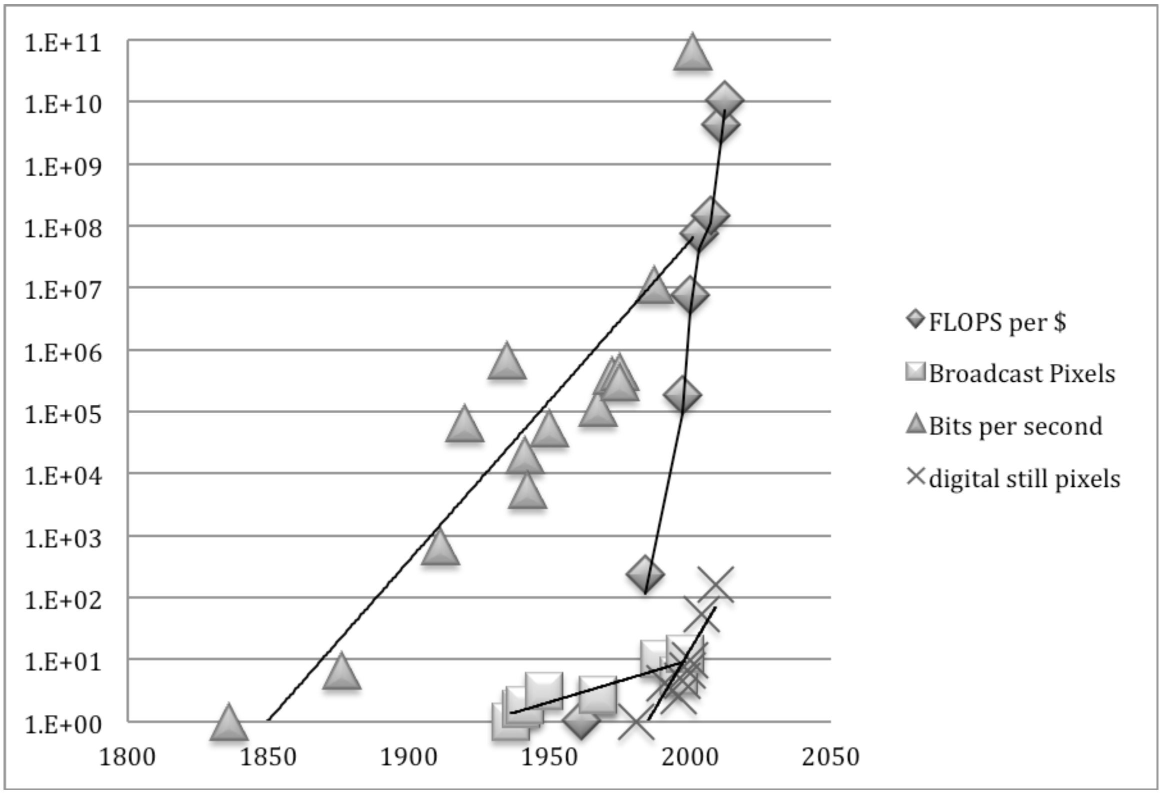

Figure 1. Measures of single channel communications capacity, computing power, broadcast image quality and still image quality relative to first commercial demonstrations vs. year. Communications is normalized to human transmission of Morse code, broadcast to the original NTSC standard and digital stills to the first "still video" cameras.

Assuming that technical challenges could be overcome, one may easily imagine digitally zoomable broadcast systems in which viewers could digitally zoom magnification in space and time over several orders of magnitude. Where current broadcast systems allow only a narrow field porthole view, future systems may offer substantially more than human acuity, making telepresence more than a match for actually being there. One may imagine systems that autonomously search for features of interest and display multiple details on wide field scenes.

A transition from serial image read-out to parallel optical and electronic processing is essential to achieving this vision. Building on the multiscale approach described previously [2, 3], this paper considers the physical limits of camera capacity and discusses continuing efforts to use multiscale design to allow camera capacity to leverage exponential improvements in processing and communications.

2. INFORMATION CAPACITY AND MULTISCALE DESIGN

The information capacity of a camera, i.e. the maximum number of nominally independent pixels it can resolve per unit time, is limited by optics, electronics and the physics of the scene being imaged. The space bandwidth product, which determines the number of independent degrees of freedom one may capture per spectral band per unit time, is the most basic measure of information capacity. This number is proportional to the aperture area in units of square wavelengths [4]. To achieve this limit, however, one must construct a lens that eliminates geometric aberration in analog image formation. This challenge is historically challenging due to the restriction of digital cameras to planar focal surfaces and linear scaling in geometric aberration as a function of lens scale. However, in our recent work we have shown that the combination of monocentric spherical optical designs and local field correction in multiscale systems enables diffraction limited image formation over a wide range of lens apertures [5, 6]. In practice, aperture size is limited by fabrication challenges rather than lens design.

Until very recently, electronic pixel pitch and focal plane array size were as significant as optics in limiting camera capacity. With the recent commercialization of micron-scale pixels and the ability to mosaic small focal planes in arrays of arbitrary effective size using multiscale lenses, electronic pixel count is no longer a significant barrier. Therefore, with the optical and electronic barriers nearly overcome, one may reasonably wonder what is the fundamental limit of pixel count?

Since optics and electronics have been dominant for so long, we often forget the role that the scene plays. A camera is ultimately a transceiver for information emitted or reflected by an object and then transmitted through air. It is important to understand that the information encoded on the optical field on emission or reflection is finite and that this information is further degraded by the transmission medium. Turbulence in the air reduces the spatial coherence of wavefronts emanating from a common point, ultimately limiting the maximum coherent aperture size and achievable spatial resolution. Similarly, the limited photon flux of the object field limits temporal and spectral resolution.

We assume that the optical quality of the atmosphere, measured by the Fried parameter, limits spatial resolution to about 5 microrad. This is the limit imposed by the effects of atmospheric turbulence at most locations, and in practice for terrestrial viewing the resolution may be significantly above this [7, 8]. For temporal resolution we may consider an object illuminated by ambient Solar illumination of 1 kW/m^2 at a range of 100 meters. We resolve a 1 mm patch, corresponding to 1 milliwatt/patch. With a 10 cm aperture, we detect ~.5 nW/patch or 10^{10} photons/sec. Assuming 1000 photons per pixel, this allows about 10 million pixels split between temporal and spectral degrees of freedom. Assuming 10000 fps, 100 spectral channels and 10 focal range bins, this suggests 100 petapixels per second as a reasonable physical limit.

While this limit exceeds current camera capacity by 6-7 orders of magnitude, one must view this deficiency as an opportunity to move television, in its original meaning of as an electronic telescope allowing high-resolution observation of remote events, on to a more promising track than that shown in Fig. 1. With optics and pixel sampling eliminated as barriers, the primary remaining hurdle is that the power and communications bandwidth necessary for petapixel transmission is unsustainable. It is important to note, however, that the actual object information contained in the raw petapixel photon flux is always much less than the pixel limit in natural fields. Noting recent progress in using compressive sampling to reduce bandwidth and pixel sampling rates [9-12] in addition to coding spectral [13] and focal [14] degrees of freedom, one may reasonably expect physical layer compression to reduce the camera data load by 4-5 orders of magnitude. In combination with parallel read-out enabled by multiscale design, this suggests that full petapixel data cube sensitivity may eventually be achievable. As a first step in this direction, we consider the current status of gigapixel resolution multiscale cameras.

3. IMAGES

We consider images captured by the AWARE 2 120 degree field of view 40 microradian ifov and the AWARE 10 100 degree field of view 25 microrad ifov cameras described in [2]. These cameras consist of three element monocentric spherical objective lenses surrounded by arrays of 100-300 14 megapixel microcameras. Each microcamera has independent focus, exposure and frame rate control. Camera data is read-out through a hierarchy of control circuitry, allowing flexible provisioning of readout bandwidth.

A composite from a single microcamera of the AWARE 10 system moved into 10 different locations in the array is shown in Fig. 2. Zoomed in regions of this color composite are shown in Fig. 3. The distances to these targets varies from 54 meters to 162 meters. The low f/# optics give a shallow depth of field, requiring precise focusing.

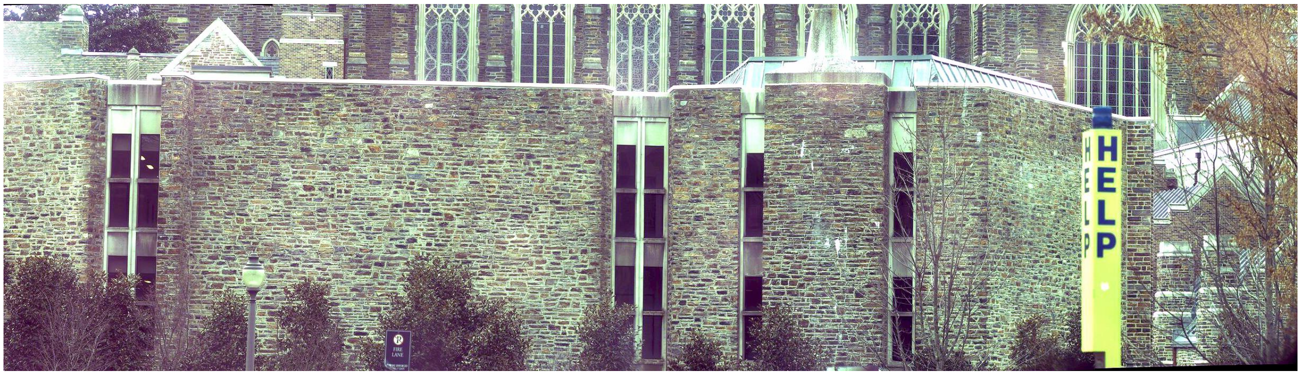

Figure 2: Color compositing is a straightforward extension of the architecture developed for the grayscale camera. Here, compositing is performed on the red, green, and blue color channels independently and fused afterword. Data in this image comes from a single microcamera in the AWARE 10 system placed at several different field locations behind the objective.

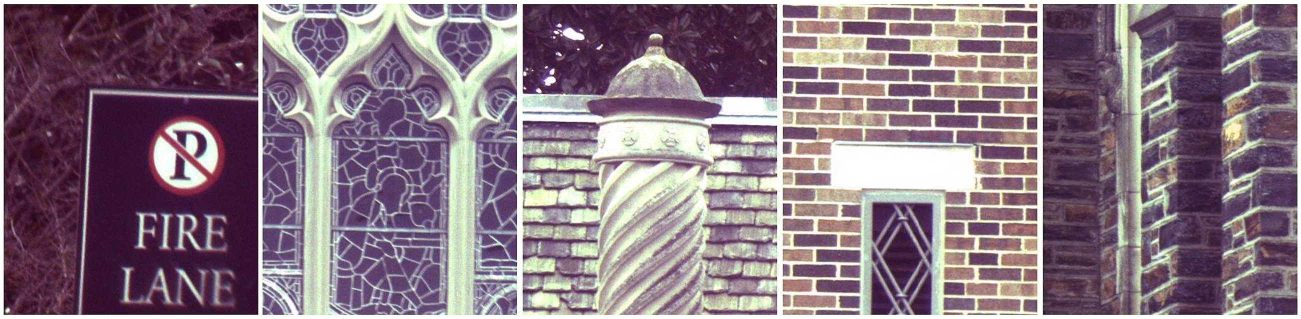

Figure 3: Zoomed in regions of Fig. 2 showing the detail of the AWARE 10 camera system. Distance to object from camera, Left to right: 54m, 162m, 123m, 117m, 162m.

The multiscale philosophy of the camera extends into the image processing architecture. Because modern displays typically have <~2 megapixels, image formation can be limited to only the data required to fill these displays. Video-rate image display becomes accessible by framing image formation as a highly parallelizable task, via the MapReduce framework [15]. Such a task is amenable to processing on current GPUs, which can have a large number (>1000) of processors on a single card. The image formation architecture treats imaging as describable via a parametric model that maps input pixels (from individual focal planes) to output pixels (for display), making the mapping process parallelizable on an input-pixel basis. The intensity of each pixel is predicted via a parametric model of the vignetting, taking into account the exposure time of the focal plane to which that pixel belongs. Image formation therefore happens in a 32-bit high dynamic range (HDR), wherein the 8-bit range from every microcamera is placed according to its exposure. Overlapping input pixels (intra- and inter-camera) are combined with knowledge of this intensity and exposure variation, making the reduction process parallelizable on an output-pixel basis. The final step converts the 32-bit luminosity data into an 8-bit (suitable for display on most devices) representation of the scene. Several methods are available, but a spatially-varying tone mapping method produces results most representative of those seen by the human eye. The resulting image contains information about both extremely bright (sunlit) and dim (shadowed) regions of the scene (see Fig. 4) [16].

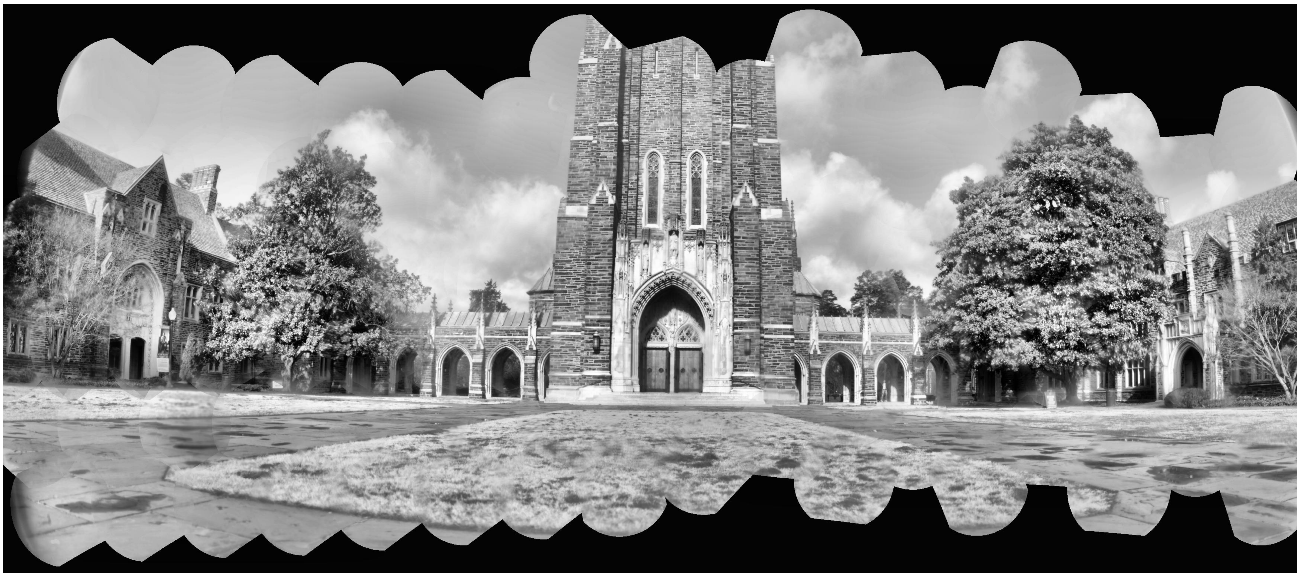

Figure 4. A high dynamic range image of the chapel on Duke University campus. 98 microcamera images are composited into a one gigapixel 32-bit image, which here has been downsampled and converted to 8-bits for display.

Importantly, this architecture allows us to consider a gigapixel camera as a fundamentally multi-user tool. As the electronics are architected to allow requests from multiple sources, many independent processing workstations can form a live stream of images for a group of independent users. These users may be looking at portions of the full field of view that are disjoint, overlap, or are at dramatically different scales (see Fig. 5).

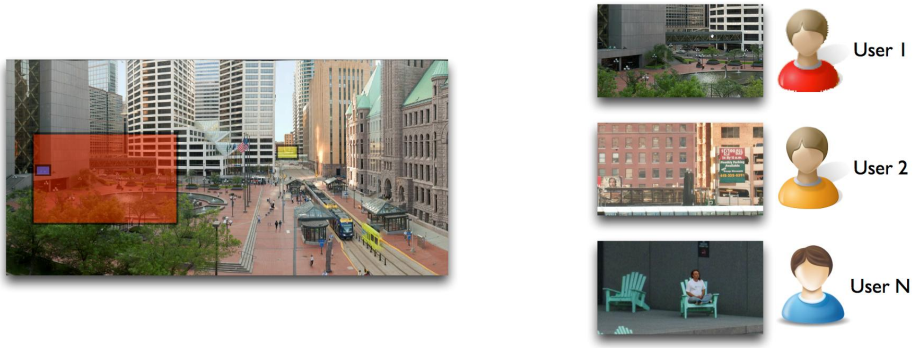

Figure 5. Multiple users can interact with the camera array simultaneously. Each user can have a window of any scale or position without interfering with other users. (Image courtesy of Tom Nelson [17]).

This framework was designed with a limited number of users and processing capacity in mind. It can, however, be extended to processing the full resolution image, which becomes acutely necessary as the number of users is increased. For example, at a public event with 10,000 users able to interact with the camera via mobile devices, processing individual fields of view would require an order of magnitude more processing than is strictly necessary for the amount of input data in AWARE-2. More generally, consider that a given user receives P_{user} output pixels for their display (~2 MP for most modern displays). The processing resources required for a number of users, N_{users}, can easily exceed the total number of pixels in the system, P_{total}. Therefore when $N_{users} * P_{user} > P_{total}$, the camera is better served by providing enough resources to form the full resolution image at the desired framerate. The parallelized image formation architecture allows for this by dividing the full field of view into many smaller regions, each of which can be processed independently. The host of smaller composites can be served to users individually, in groups, and at different scales as required.

The extension of this framework from grayscale to color is straightforward, as the mapping and reduce steps can be done independently on the red, green, and blue color channels. Alternatively, processing efficiency can be improved by transmitting and processing color image data in the YCbCr or raw color spaces. Regardless of the color space, the described image formation architecture can be applied to the task. Initial composites, done with a single microcamera, confirm the efficacy of the architecture for color processing, as shown in Figure 2.

As discussed above, the electronics maintain a time-sequential buffer of recent images at a constant system framerate, allowing a user to step backward and composite images from a previous frame. However, image acquisition could be structured to fill the buffer with heterogeneous data, as shown schematically in Figure 6. This data could include varied frame rates, exposures, focal positions, resolutions, or spectral bands. The buffer history then becomes an N-dimensional data cube that greatly expands the type of information about the scene that be explored with the camera. Moreover, this data store can be viewed as being undersampled, allowing computational sensing approaches to generate various types of imagery. The desired imagery would guide the design and capabilities of the microcameras in an array (e.g. imaging rate or spectral sensitivity). Imaging with such cameras is then a combination of what data to measure and what algorithms to apply to synthesize a desired result.

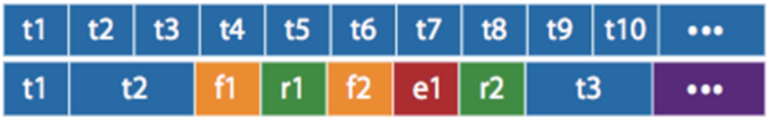

Figure 6: Currently, the frame buffers on the electronics are filled with time-sequential data at a constant framerate (top). Alternatively, the buffers could be filled with heterogeneous data (variable framerate, focal position, exposure, resolution, etc) to build a more complex data cube that can be mined to synthesize additional imagery (bottom).

4. DISCUSSION

The AWARE series of multiscale cameras, constructed under the DARPA AWARE Wide Field of View Program, demonstrate that optics and electronic sampling provide no barrier to camera information capacity. Rather, capacity is ultimately limited by photon flux and atmospheric turbulence. In the near term, however, capacity is limited by communications and processing. In exploring real-time gigapixel image capture and streaming, we begin a process, common in the history of information technologies, of moving over successive generations toward fundamental limits, even as we explore and question what those limits may be.

REFERENCES

[1] Toyoda, K., *Digital still cameras at a glance*, in *Image sensors and signal processing for digital still cameras* 2005, CRC Press. p. 1-19.

[2] 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.

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

[4] Brady, D.J. *Optical imaging and spectroscopy* 2009, Hoboken, N.J.; [Washington, D.C.]: Wiley; Optical Society of America.

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

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

[7] Hardy, J.W., *Adaptive optics for astronomical telescopes.* 1998; New York: Oxford University Press.

[8] Tyson, R.K. *Principles of adaptive optics*. 2011; CRC Press

[9] 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.

[10] 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.

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

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

[13] Gehm, M.E., R. John, D.J. Brady, R.M. Willett, and T.J. Schulz, *Single-shot compressive spectral imaging with a dual-disperser architecture.* Opt. Express, 2007. **15**(21): p. 14013-14027.

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

[15] Dean, J. and S. Ghemawat, *Map Reduce: Simplified data processing on large clusters.* Commun. ACM, 2008. **51**(1): p. 107-113.

[16] Golish, D.R., E.M. Vera, K.J. Kelly, Q. Gong, P.A. Jansen, J.M. Hughes, D.S. Kittle, D.J. Brady, and M.E. Gehm, *Development of a scalable image formation pipeline for multiscale gigapixel photography.* Opt. Express, 2012. **20**(20): p. 22048-22062.

[17] Nelson, T. *Governmnet plaza on April first.* Available from: http://gigapan.com/gigapans/46074/.