

Parallel cameras

DAVID J. BRADY,^{1,2,3,*}  WUBIN PANG,^{1,2}  HAN LI,⁴ ZHAN MA,⁴ YUE TAO,⁴ AND XUN CAO⁴

¹Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina 27708, USA

²Duke Kunshan University, Kunshan, Jiangsu, China

³Aqueti, Inc., Shanghai, China

⁴School of Electronic Science and Engineering, Nanjing University, Nanjing, Jiangsu, China

*Corresponding author: dbrady@duke.edu

Received 16 October 2017; revised 21 December 2017; accepted 21 December 2017 (Doc. ID 309322); published 29 January 2018

Parallel lens systems and parallel image signal processing enable cost efficient and compact cameras to capture gigapixel scale images. This paper reviews the context of such cameras in the developing field of computational imaging and discusses how parallel architectures impact optical and electronic processing design. Using an array camera operating system initially developed under the Defense Advanced Research Projects Agency Advanced Wide FOV Architectures for Image Reconstruction and Exploitation program, we illustrate the state of parallel camera development with example 100 megapixel videos. © 2018 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

OCIS codes: (110.1758) Computational imaging; (120.3620) Lens system design; (150.6044) Smart cameras.

<https://doi.org/10.1364/OPTICA.5.000127>

1. INTRODUCTION

A parallel, or array, camera is an imaging system utilizing multiple optical axes and multiple disjoint focal planes to produce images or video. Using “parallel” as in “parallel computer,” a parallel camera is an array of optical and electronic processors designed to function as an integrated image acquisition and processing system. While array cameras have long been used to capture 3D, stereo, and high-speed images, we limit our attention here to arrays designed to function as conventional cameras, meaning that the output of the system is nominally an image observed from a single viewpoint. The motivation for such arrays is the same as the motivation for parallel computer: both optical and electronic processing can be simplified and improved using parallel components. Since image processing is particularly amenable to parallel processing, parallel focal planes and image processing electronics are particularly useful in reducing the system cost and complexity. Array camera and parallel computer design both address the same design challenges in selecting processor granularity, communications architecture, and memory configuration.

Just as parallel computers have developed the terminology of CPUs, microprocessors, graphical processing units (GPUs), and processing cores to describe the system design, terms are emerging to describe array camera design. To a large extent, array cameras are identical to parallel computers. They use arrays of CPUs, GPUs, and image signal processing chips (ISPs) to process parallel data streams. In addition to these components, array cameras include image sensor and lens arrays. We refer to the modular component consisting of one image sensor and its associated lens and focus mechanism as a “microcamera” and the whole array camera as a “macrocamera” or simply a camera. As discussed below, some

current designs use discrete microcameras with essentially conventional lenses and some use microcameras that share a common objective lens. We call the second category “multiscale systems” [1].

The transition from monocomputers to multicomputers has substantially improved computing capacity and the rate of capacity improvement [2]. Similarly, multicamera designs have already demonstrated 10–100× improvements in pixel processing capacity relative to conventional designs. More significantly, as arrays become increasingly mainstream the rate of improvement in pixel processing capacity is expected to substantially increase. While dynamic range, focus, sensitivity, and other metrics are also critical, given adequate image quality, spatial and temporal pixel sampling, and processing rates are the most fundamental measures of camera performance. Parallel architectures have already driven a transition from megapixel (MP) to gigapixel (GP) scale spatial sampling [3]. Parallel cameras have also excelled in temporal processing, with systems capable of 1–10 GP/second currently readily available. However, while supercomputer capacity has continuously improved for over half a century, it is not clear how far supercameras can be developed beyond the gigapixel scale. Atmospheric considerations are likely to limit the aperture size of reasonable cameras to 10 cm, and the flux of natural light limits the frame rate. At the diffraction limit, a 10 cm aperture exceeds a 10 GP resolution, operating at kilohertz frequencies with 10–100 spectral channels, so one can imagine supercameras reaching pixel processing rates in excess of 10^{15} pixels per second. While this limit is 2–3 orders of magnitude beyond current limits, one expects that it may reasonably be achieved in the next decade. On the other hand, the size, weight, and power of current supercameras is also several orders of magnitude greater than physical

limits. Making gigapixel-scale cameras increasingly compact and energy efficient may be a project that can span the next half century.

Photographic array cameras have a long history, dating from Muybridge's studies of animal motion [4], Lippmann's integral photography [5], and stereo photography [6]. Muybridge's work begins a long tradition of using arrays to improve temporal sampling rate; recent examples of this approach are presented in [7–10]. Lippmann was motivated in part by the parallel optical systems found in insects and other animals, and the “bug eye” analogy has remained a common theme in array camera development over the intervening century. Recent versions of bug eye systems include TOMBO [11] and related systems [12,13]. While the advantages of array architectures in biology derive from the simplicity of neural processing, recent biologically inspired imagers have focused on digital super-resolution [14–17] and sensor diversity [18–20]. The computer vision community also has a long history of multi-aperture imaging, mostly focusing on “light field imaging,” which allows camera arrays to reconstruct diverse viewpoints [7,21] or focal ranges [22,23]. This work is also reflected in the many companies and universities that have constructed 360° panoramic cameras looking out [24] or in [25] on a scene.

While there are fewer examples of array cameras designed to nominally produce a single viewpoint image, the 16-camera array developed by Light Incorporated is a recent example of such a camera [26]. On a larger scale, projects such as LSST [27], Pan-STARRS [28], and ARGUS [29] have created large scale staring arrays for astronomy and high-altitude surveillance.

In the context of these diverse examples, our review focuses on parallel cameras with real-time video processing to produce integrated images. We use parallel optical and electronic design to continue the natural evolution of pixel count from megapixels to gigapixels. In pursuit of this goal, basic concepts relating to camera function and utility must be updated. We consider these concepts in the next section of this paper before describing current strategies for optical and electronic design.

2. COMPUTATIONAL IMAGING

We define a “camera” as a machine that captures and displays the optical information available at a particular viewpoint. The basic design of a camera has been stable for the past two hundred years; a lens forms an image and a chemical or electronic sensor captures the image. With the development of digital sampling and processing systems over the past quarter-century, however, this approach is no longer ideal, and many alternative designs have been considered. The fundamental design question is “what set of optics and electronics should be used to most effectively capture and display the optical information at a particular viewpoint?” Parallel cameras are one approach to this “computational imaging” design challenge. We define a “computational imaging system” as a camera in which physical layer sampling has been deliberately co-designed with digital processing to improve some system metric. Under this definition, a camera designed with a good modulation transfer function (MTF) for high-quality focal plane sampling is not a computational imaging system, even if substantial post-capture image processing is applied to improve color balance, reduce noise, or improve sharpness. On the other hand, the use of a color filter array, such as the Bayer RGB filter, is a form of computational imaging.

To our knowledge, the first paper explicitly proposing a camera design to optimize post-capture digital processing appeared in 1984 [30]. The first widely discussed example of a non-obvious computational imaging camera was the extended depth of field system proposed by Dowski and Cathey [31]. The Cathey and Dowski system used “pupil coding,” consisting of the deliberate introduction of lens aberrations to improve the depth of field. The many subsequent computational sampling strategies may be categorized into (1) pupil coding systems, which modulate the lens aperture; (2) image coding systems, which modulate the field at or near the image plane; (3) lensless systems, which use interferometric or diffractive elements to code multiplex measurements; (4) multi-aperture systems; and (5) temporal systems, which vary sampling parameters from frame to frame.

Each of these strategies has been applied in many different systems. We mention a few representative studies here, with apologies for the many interesting studies that we neglect for lack of space. In addition to Cathey and Dowski, pupil coding was developed in earlier pioneering studies by Ojeda-Castaneda *et al.* [32]. Among many alternative studies of aperture coding, the work of Raskar *et al.* has been particularly influential [33]. Image coding includes color filter arrays, such as the Bayer filter mentioned above [34], as well as the Lytro light field camera [22] and various pixelated spectral imaging systems [35,36]. Lensless systems have a very long history, dating back to the *camera obscura*. Recent lensless optical imaging systems focus on coded aperture and interferometric designs [37]. FlatCam [38] is a recent example of a coded aperture design; rotational shear interferometry [39] is an example of an interferometric lensless camera. Various multiple aperture systems are listed above; we conclude this very brief overview by mentioning a couple of examples of temporal coding. The canonical example is high dynamic range (HDR) imaging, which uses multiple frames to synthesize high dynamic range images [40]. HDR coding has already been widely implemented in mobile camera applications [41], making it perhaps the second widely adopted form of computational imaging (following color filter array processing). Alternative forms of multiframe processing, such as focal stacking for extended depth of field [42] and 3D imaging, have also been implemented in phone cameras. HDR and focal stacking are examples of multiframe temporal coding for computational imaging, and recent studies have also explored dynamic modulation of capture parameters for single frame computational imaging [43]. In particular, we note in [44] that sensor translation during exposure has the same capacity to be encoded for extended depth of field and 3D imaging with the advantage that the coded point spread function can be dynamically tuned and switched off to achieve a high MTF.

Based on these many studies, it is important to recognize computational imaging strategies that have been more and less successful. Future development will come from building on success and abandoning failure. In our view, pupil coding and lensless imaging research has not revealed useful strategies for visible light computational imaging. The challenges for pupil and lensless coding are (1) these techniques inevitably lead to substantial reductions in signal-to-noise ratio (SNR) for given optical flux, and (2) they are therefore not competitive with alternative sampling strategies to achieve the same objectives. Pupil coding and lensless sensors are examples of “multiplex sensors” in which multiple object points are combined in a single measurement. While multiplexing is inherent to many measurement systems, particularly in

tomography [45], its impact on optical imaging systems is universally problematic. Our group worked on various interferometric and coded aperture lensless imaging systems in the late 1990's and early 2000's, but our interest in lensless optical imaging ended with a study finding no scenario under which such systems surpass the performance of focal systems [46]. Challenges arise both from the ill-conditioned nature of the forward model for multiplexed systems with nonnegative weights and with the impossibility of arbitrarily multiplexing optical information in physical systems [47].

The challenge of optical multiplexing may most simply be explained by noting that a lens does a magical thing by bringing all the light from a single object point into focus at a single image point, despite the fact that this light spans many different spectral modes. A typical visible camera field spanning 400–700 nm captured in 10 ms has a time bandwidth product of 10^{12} . The number of temporal modes detected is approximately equal to the time bandwidth product. The number of photons detected during this span is typically 10^4 – 10^6 , less than one millionth of a photon per mode. Absent a lens, it is impossible to combine information from these different modes with high SNR. Of course, for three dimensional objects, there is no mechanism for simultaneously bringing all object points into focus. As noted above, however, temporal coding through focal sweeping and multiple aperture solutions is as effective as pupil coding in scanning 3D objects, but has the advantage that they can be dynamically and adaptively coded to maximize the SNR. We therefore suggest that it is extremely difficult to find an operating scenario where deliberate multiplexing using pupil coding or lensless imaging makes sense for visible light. In contrast with pupil coding, image coding in the form of color filter arrays and temporal processing remains a key component of commercial computational imaging systems. Multi-aperture and temporal coding, on the other hand, have demonstrated clear and novel utility but are only beginning to emerge in commercial cameras.

The lesson learned is that a well-focused high MTF image has enormous advantages. However, in contrast with conventional cameras, modern computational cameras only require that the image be locally of high quality. In fact, by breaking an image into sub-images, focus may be more effectively mapped onto 3D scenes. Image discontinuities and distortions can be removed in software. Imaging is naturally a highly parallel information processing, and imaging systems can be designed to optimize sampling in parallel for local regions with the idea that the full image is recovered in software. As we discuss below, however, for multiple parallel images captured with different exposures, sampling times, and focal states, definition of the “full image” may present challenges. For now, we are ready to move to the next section, which discusses lens design for parallel cameras.

3. OPTICS

The general goal of camera design is to capture as much visual information as possible subject to constraints on size, weight, power, and cost. These constraints weigh on both the optical and electronic camera components. In most modern parallel cameras, the size, weight, power, and cost of the electronic capture and processing components are dominant factors. However, in conventional high-performance cameras, such as single-lens reflex cameras using zoom lenses, lens size, weight, and cost are often dominant. This difference arises because the conventional lens

volume and complexity grows nonlinearly as the information capacity increases. Parallel design reduces the lens complexity by removing the need for a mechanical zoom and by reducing the sub-image field of view (FoV) as the camera scale increases.

Two different lens design strategies may be considered. The first uses discrete arrays of conventional lenses, with each microcamera having independent optics. The second strategy uses multiscale lenses in which microcameras share a common objective lens. Discrete arrays have been commonly used in very wide FoV systems, such as 360° cameras. Multiscale arrays were used to reduce the lens volume and cost in the Defense Advanced Research Projects Agency (DARPA) Advanced Wide FOV Architectures for Image Reconstruction and Exploitation (AWARE) program [3]. Emerging designs include hybrid systems consisting of discrete arrays of multiscale cameras. Here, we discuss basic design requirements driving FoV granularity and when to use discrete and multiscale designs.

A. Multi-aperture Optics

For reasons discussed in Section 2, the lens is and will remain the basic workhorse of optical imaging. A “lens” consists of one or more blocks of transparent material, each with spherical or aspherical surfaces that modulate the light path in a desired fashion. The lens designer's goal is to find a lens system that meets the functional requirements with minimal cost. “Cost” here refers to a function of lens parameters that may include the actual material cost, but in modern design more commonly refers to the system volume and complexity. The central design question is “what are the limits of cost and how do we achieve these limits?” We can also phrase this question in another way. “Given a fixed cost budget, what is the best way to design and manufacture a lens system that maximizes the camera performance?”

To answer to this question, we draw our inspiration from the divide and conquer strategy of parallel computing. Dividing the task into parallel portions being solved individually may produce a great reduction in complexity. Parallel lens arrays accomplish the imaging task by segmenting the full FoV into small units denoted as FoV_s . Each lens in the array processes the field only from its assigned subfield. The designer selects FoV_s to minimize the lens cost.

The lens cost can be evaluated according to the number of elements, volume, weight, and materials as well as the manufacturing cost. We use the function C to denote this cost. C is a function of the system FoV, focal length f , aperture size $F/\#$, wavelength range, pixel number (information capacity), and other image specifications such as distortion, uniformity of luminance, mapping relationship ($f, f \sin \theta, f \tan \theta$), and the lens configuration. In another words, it is a multivariable function depending on numerous factors. Among all these principal factors, FoV is the distinguishing factor between a monolithic lens and a parallel lens array. For this reason, we would like to examine the relationship between the cost function C and argument FoV while keeping the other variables constant, which is equivalent to examining the cross section of the cost function along the FoV axis. An expression of the cost function under the scheme of a parallel lens array is

$$C_A(\text{FoV}, \text{FoV}_s) = \left(\frac{\text{FoV}}{\text{FoV}_s} \right)^2 C(\text{FoV}_s), \quad (1)$$

where C_A denotes the cost function of the lens array, and C without a subscript denotes the cost function of a monolithic lens.

$(\text{FoV}/\text{FoV}_s)^2$ is the number of FoV_s lenses needed to fully sample the FoV .

If the function C has the form of FoV_s raised to power γ , i.e., $C(\text{FoV}_s) = c\text{FoV}_s^\gamma$, where c is a constant, then we have

$$C_A(\text{FoV}, \text{FoV}_s) = c\text{FoV} \times \text{FoV}_s^{\gamma-2}. \quad (2)$$

According to Eq. (2), if $\gamma > 2$, a parallel lens array reduces cost. For $\gamma = 2$, the cost function is same for both strategies. If $\gamma < 2$, a parallel scheme increases the overall cost of the lens system. The conclusion here is whether the parallel lens array is preferable depends on the cost function of the monolithic lens $C(\text{FoV}_s)$. We may express this function in a more general way by using the polynomial series

$$C(\text{FoV}_s) = c_1\text{FoV}_s + c_2\text{FoV}_s^2 + c_3\text{FoV}_s^3 + \dots + c_n\text{FoV}_s^n + \dots \quad (3)$$

There is no constant term because no lens is needed in the case of $\text{FoV}_s = 0$. Substituting into Eq. (1) yields

$$C_A(\text{FoV}, \text{FoV}_s) = \text{FoV}^2 \left(c_1 \frac{1}{\text{FoV}_s} + c_2 + c_3\text{FoV}_s + \dots + c_n\text{FoV}_s^{n-2} + \dots \right). \quad (4)$$

Setting the first derivative with respect to FoV_s equal to 0, we find

$$c_1 = c_3\text{FoV}_s^2 + c_4\text{FoV}_s^3 + \dots + c_n\text{FoV}_s^{n-1} + \dots \quad (5)$$

If c_1 and any higher-order terms are nonzero, then there exists a nonzero value of FoV_s at which C_A has a minimum value. To explain this intuitively, by choosing a camera array, the required number of camera units increases quadratically along with the total FoV . If this quadratically increased number of cameras overwhelms the increase of complexity in the monolithic case, the employment of the camera array loses its advantage in reducing the cost. Only if the complexity of the monolithic lens grows much faster than the number of units needed in the array we stand to profit by switching to a camera array. It can be prohibitively difficult to derive an explicit expression for the cost function. However, some properties of this function can be projected based on empirical knowledge in lens design. Here, we present two basic conjectures. (1) The cost function of a monolithic lens contains high-order (terms higher than second order) components, which means that for a given FoV coverage there is an interval of FoV_s within which an array strategy outperforms a monolithic choice. (2) This cost function also consists of a first-order term, i.e., c_1 in Eq. (5) is nonzero, which indicates that there is an optimal FoV_s value that produces a camera array minimizing the total cost. The nonzero value of the first-order term not only implies the existence of a minimum total cost but also predicts a lower threshold of FoV_s under which the array becomes an inferior choice. These two conjectures are visually illustrated in Fig. (1). The blue (lower) curve represents the cost function for one single lens with FoV corresponding to the abscissa axis. The upper black curve is the derived cost function for a lens array; here, the abscissa represents FoV of each individual channel, while the cost on the vertical axis represents the total cost of the array for achieving an 80° total coverage. From these two curves, for a total coverage less than 20° the monolithic solution is preferred. For a total FoV greater than the 20° , an array scheme is favored. If the assumption holds to be true, within a wide range of total FoV demand, a parallel lens array system outperforms a monolithic one in terms of cost. It should be clear that this diagram is

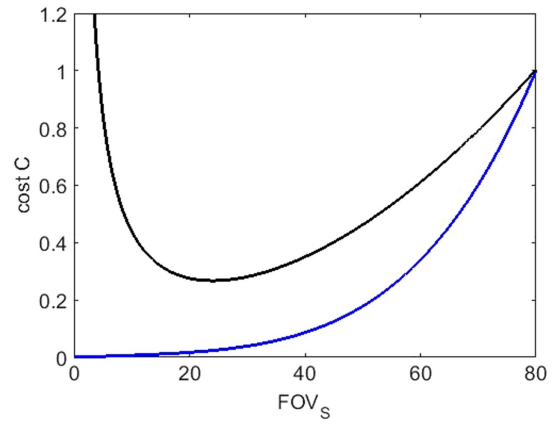


Fig. 1. Cost curve under the assumption of a polynomial function form. The blue line represents the cost of a single lens from the parallel lens array, while the black line represents the total cost of the array system for a full FoV coverage of 80° and here this curve shows a minimum cost around $\text{FoV}_s = 20^\circ$.

generated to show the general idea of our conjecture and not plotted from any calculation or simulation.

The relationship between the lens cost and FoV is very complicated, and it is prohibitively challenging to find an explicit expression. However, to test our conjecture, it is possible to use discrete experimental data to approximate the function curve. One approach would be to compare the cost of a collection of lenses from some lens catalogs with near identical specifications other than FoV . However, the lens parameters differ extensively in current camera lens categories, which renders it impossible to sort out a collection of useful samples from commercial lens categories for our purpose.

Instead, we build our own lens datasets using computer-aided design software (ZEMACS). Figure 2 shows the results from one of our datasets. Each sample lens in this example features a 35 mm focal length, 587 nm design wavelength, F/3.5 aperture size, and uses BK7 glass for all elements. In making all other specifications identical for each lens, we try our best to eliminate the effect of factors other than FoV . Nonetheless, there is a difficulty in doing so, and it is impossible to keep all the different lenses achieving an “identical” imaging quality, which can be indicated either by the MTF curves or image spot sizes. To address this issue, we demand that every design should achieve a near diffraction-limited performance. Of course, for each set of design requirements, there is an infinite number of valid design solutions. All these solutions have different costs or complexities. It is the work of the lens designer to not only find a qualified solution but also a solution with a cost as low as possible. In creating our lens datasets, each design has been optimized to trim away unnecessary expense in terms of the system volume, weight, as well as the number of elements. Therefore, these design examples represent our best effort for pursuing an approximation of the “law of cost” in lens design. For simplicity, we have designed and evaluated lenses at a single wavelength, neglecting chromatic aberration. On one hand, chromatic aberration is one of many geometric aberrations, and we assume that the trend between the FoV and cost function will not change significantly if it is also corrected. On the other hand, chromatic aberrations often demand correction through the employment of different lens materials, which would also complicate the cost

analysis substantially. While the net result would be to shift optimal FoV, to smaller angles, we assume that single-wavelength consideration captures the essential point of our analysis.

A total of 9 lenses with FoV ranging from 5° to 80° were produced for this experiment. Design details are included in a lens design dataset; the first part is in Supplement 1. In our analysis, the system volume, overall weight, and number of elements of each design were chosen separately as measurements of the cost. The results are shown in Fig. 2; each design is represented by a dot in all the graphs, and the dashed lines are used to visualize the trend of the changes. In Fig. 2(a), both the curves of the system volume and the overall weight resemble that of an exponentially growing one, while the number of elements grows in a nearly linear fashion. In a $f \tan \theta$ lens, the information throughput of a lens is proportional to the area of the image plane, which can be expressed as $\pi(f \tan(\theta))^2$, where θ represents the semi-FoV of the lens. Since the information throughput can also be described by the total pixel account resolved by the lens, we would like to examine the cost per unit information or cost per pixel, since this quantity measures the system performance in terms of the information efficiency. Dividing the system volume, the overall weight, and the number of the elements by the pixel numbers of each design, we obtain the plots shown in Fig. 2(b), in which the valley-shaped curves show up in the volume per pixel as well as the weight per pixel, with the minimum value located in place of FOV = 30°. This result implies that, for a set of fixed design targets with varying FoV, there exists a specific FoV in which the system may have the highest information efficiency in terms of the cost per pixel. Nonetheless, there is a deflection at the end of the curve, the design for FOV = 80°, rather than a rise of the information efficiency. This is because we were unable to achieve a satisfactory design at this limit. Assuming we want to achieve

a desired FoV of 80°, we must instead use an array of lenses with each lens covering only a fraction of the whole FoV, and the desired FoV target can be pieced together by the group. The question is: will this strategy reduce the overall cost? As demonstrated in Fig. 2(c), the answer is yes, at least in this experimental case. The “divide and conquer” solutions are always more optimal than using just a single-aperture lens, with the best solution corresponding to a microcamera FoV of 30°. It is worth noting that the number of elements per pixel and the number of elements under the lens array strategy decrease monotonously as the FoV increases, which is not surprising. As the FoV increases, the number of the elements does not increase significantly compared with the volume and weight; instead, the small-aperture size elements are replaced by large-aperture size elements. In other words, the pixel capacity increases much faster than the number of elements. However, large numbers of small optics do not necessarily indicate a higher cost than that of a small number of much larger optics, since the manufacturing processing is much easier in the former case than that in the latter.

By building our own lens dataset, we have investigated the relationship between the cost function and FoV. As demonstrated in our results, the cost and complexity of imaging lenses grows such that the cost per pixel plot features a V shape with a minimum position. By implementing the approach of parallel design, we can reduce the overall cost in optics while still accomplishing our design target.

B. Multiscale Optics

The nonlinear increase in the lens complexity as a function of the FoV discussed above is based on the assumption that the lens must correct for geometric aberrations. The five Seidel aberrations are the traditional starting point for considering geometric aberrations. However, one of these aberrations, field curvature, does not degrade the image quality if one allows image formation on a curved surface. Using a Luneberg lens [48], one can image without geometric aberration between two given concentric spheres. The Luneberg design is independent of the aperture scale, so the same lens design would, in principle, work at all aperture sizes and pixel capacities. Unfortunately, Luneberg lenses require graded index materials, which are difficult to manufacture. One can, however, approximate Luneberg lenses using discrete layers of spherical materials. Such “monocentric objectives” can also achieve a near diffraction-limited performance on a spherical focal surface. Spherically symmetric structure features an identical imaging property for all directions that facilitates wide-angle imaging. The primary challenges of this approach are (1) curved focal planes are not readily available, and (2) the object space focal surface of the Luneberg lens is also spherical. Focusing requires adjustment of the radius of curvature of both the image and object surfaces.

The multiscale design provides a middle ground between Luneberg and conventional designs. The multiscale method is a hybrid of the single-aperture design and the parallel multi-aperture design. Multiscale systems share a common objective lens at the front with a microcamera array at the rear. The secondary microcameras may be mounted on a curved surface to relay the intermediate focal surface formed by the objective onto conventional planar image sensors. In previous work, we constructed various multiscale systems through the DARPA AWARE program [3]. Table 1 shows characteristics of three AWARE cameras constructed from 2012–2014.

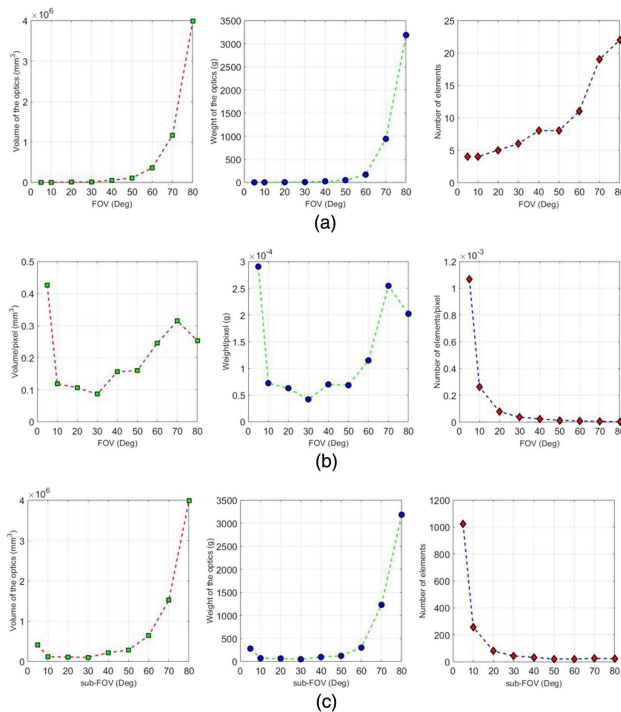


Fig. 2. Lens cost estimation in terms of system volume, weight, and number of elements. (a) The cost curves from the nine design examples. (b) The graphs of cost per pixel plots showing the information efficiency. (c) The cost curves by applying the lens array strategy.

Table 1. Characteristics of as-Constructed AWARE Cameras

System	FoV	iFoV	Resolution	Optics Volume (m^3)
			Elements (FoV/iFoV ²)	
AWARE-2	120° × 50°	2.3×10^{-3}	1.3×10^9	8×10^{-3}
AWARE-10	100° × 60°	1.4×10^{-3}	2×10^9	<0.1
AWARE-40	36° × 36°	6.17×10^{-4}	3.6×10^9	0.1

Multiscale designs correct the field curvature locally by each microcamera unit, thus leading to a low system complexity. Because of the shared objective lens, multiscale systems preserve correlation information between different sub-image units and allow more uniform brightness and color, consistent magnification, and accurate relative positions. The AWARE multiscale designs are telescopes, permitting easy access to a high angular resolution or long focal length [49]. By sharing one common objective lens, the multiscale method also tends to yield a camera volume that is more reduced than the non-multiscale parallel design for a given set of specifications.

As with parallel computers, design of secondary optics in multiscale systems begins with the problem of selecting the processor granularity. In practice, the designs of the objective lens and secondary microcameras are closely correlated, which indicates that the choice of FoV segmentation has an effect not only on the secondary optics but also on the front objective. Here, the challenge is like that what we have faced in the multi-aperture parallel lens array, which is to find the optimal sub-FoV that results in the best system solution in terms of the camera cost and functionality.

The cost of a camera lens can involve a wide variety of factors. In this investigation, we pick the system volume as a representative of the overall cost. To simplify the analysis without losing the key argument, we could discuss the effect of granularity of microcameras while keeping the objective lens fixed, which produces a highly curved intermediate image of objects from different depths of field. As we increase the granularity of the microcamera array by decreasing the sub-FoV of each microcamera unit, the size of the array is scaled down accordingly. A small FoV also indicates a simple lens characterized by fewer elements and weak surface profile. The extreme case of this scale down in volume and complexity is one pixel versus one microcamera unit, which has been reduced to an optical fiber array [50]. Unfortunately, under this approach it is not possible to locally adjust the focus. In practice, high-resolution imaging of complex scenes requires that each individual microcamera focus independently to capture objects at various distances. The focusing capacity is proportional to the aperture size of the microcameras. An optimal choice for the sub-FoV should be able to strike a balance between the lens cost and focus capacity.

By “focus capacity” we mean the ability of each microcamera to accommodate a targeted focal range. From a near point of the object to infinity, the object position observed by the microcamera varies from an infinite conjugate focal surface of the objective to a point displaced by F^2/z_N from that focal surface, where F is the focal length of the objective and z_N is the near point in the focal range. For $F = 25$ mm and $z_N = 2$ m, for example, the range of the focal surface is 300 μm . To focus the multiscale array camera, each microcamera must be capable of independently focusing over this range. If each microcamera is only a single pixel,

each pixel or fiber would need to be independently displaced over this range to focus. As the aperture of the microcamera grows larger, the nominal size of the microcamera displacement required remains constant, but since the ratio of the required displacement to the microcamera aperture falls, the difficulty in implementing the focal adjustment is reduced. This is to say that it is easier to move a 1 mm aperture by 300 μm than to move 100 100 μm apertures each by 300 μm . With this in mind, we estimate that the “focal capacity” of a microcamera improves inversely in aperture size. On the other hand, making the microcamera aperture larger groups pixels that may have different focus requirements and, more ominously, increases the microcamera cost function.

To explore this trade-off, we used ZEMAX modeling to produce 7 multiscale designs distinguished by different sub-FoVs. As shown in multiscale lens design in the second part of Supplement 1, for each design we set the focal length $f = 30$ mm, the aperture size $F/\# = 3$, and overall FoV = 120°. Figure 3(a) models the inverse relationship between the microcamera aperture and the focus capacity, while Fig. 3(b) shows the microcamera lens cost function (the same cost function as used above for discrete arrays) as a function of the microcamera FoV. Figure 4 merges the two plots of Fig. 3 together by equally weighting each factor. This approach suggests that this focal length microcamera FoV between 3° and 6° optimizes the lens cost and focus capacity. We have incorporated the imaging

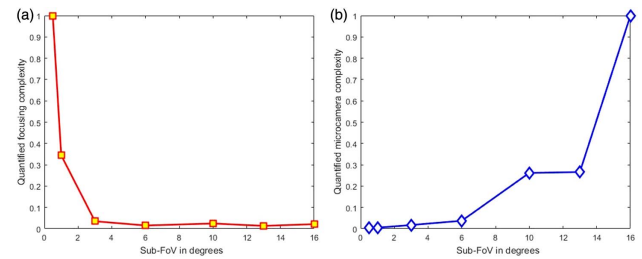


Fig. 3. Lens complexity and focusing complexity versus sub-FoV. (a) The focusing complexity skyrockets when the sub-FoV moves to the left side of the axis. (b) The lens complexity grows rapidly as the sub-FoV increases.

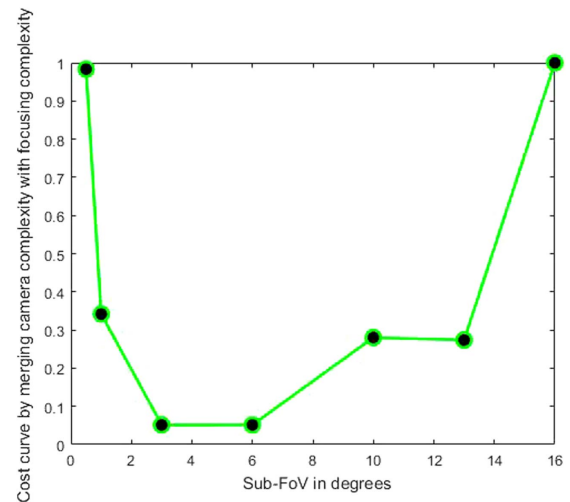


Fig. 4. By merging the two plots together, the optimal sub-FoV falls into a region between 3° and 6° in our specific case. The green solid line is an equally weighted addition of the two plots in Fig. 3.

quality of different designs into the result by applying the cost per pixel instead of the total cost.

This result is anticipated and can be easily explained. As illustrated in Fig. 3, when the sub-FoV decreases, the number of microcameras grows quadratically. The cost of the focusing mechanism for individual microcameras increases and the number of focusing units also increases, which rapidly leads to an impossible task. On the other hand, when the sub-FoV shifts toward the other end of the opposite direction, each microcamera subtends a highly curved intermediate image that requires a complex secondary optics to correct the field curvature. The resulting lens would increase in its longitudinal track, causing the total volume to grow in a cubic fashion. Consequently, the choice of the granularity of the microcamera array really needs to strike a balance between these two aforementioned factors.

Both simulations in this section require families of “identical” lens designs varying only by FoV. “Identical” means that image-related specifications and metrics, such as $F/\#$, focal length, MTF, and distortion, should be the same except for FoV. However, it is impossible to really keep these quantities identical. The $F/\#$ and focal length can be controlled very precisely by the design software, but MTF and distortion cannot be pointwise identical. To have a valid simulation result, we try our best to grind each design sample to achieve as near diffraction-limited MTF as possible under minimum complexity. In each design, the image distortion is constrained to under 4% by applying the operand “DIMX” in the ZEMAX merit function in hope of reducing its interference to as minimum as possible.

By combining the benefits of the approximation of a Luneberg lens with a microcamera array, the multiscale method overcomes the traditional scaling constraints of a large aperture and FoV. The natural remaining question is how to choose between the discrete arrays with which we began our discussion and the multiscale arrays with which we have concluded. It is also important to note that hybrid designs using arrays of multiscale systems are also possible and attractive for cameras with FoV exceeding 120° .

At smaller FoVs, the choice between conventional and multiscale arrays is not presently an optical issue. As we have seen, increasing the FoV with conventional flat focal surface sensors leads to nonlinear increases in the lens complexity. Luneberg-style multiscale systems, in contrast, support FoVs up to 120° with relatively simple microcameras. As illustrated by the range of systems constructing in the AWARE program, multiscale systems can be built with aperture sizes of several centimeters without substantially increasing the microcamera complexity. From a purely optical perspective, Luneberg-style multiscale lenses enable wide FoV imaging with a smaller optical volume per resolved pixel at essentially all aperture sizes. However, with current technology, the optical volume and lens cost is not a large driver of the overall camera cost for aperture sizes less than 5 mm. Currently, 5 mm aperture microcameras operating at $f/2.5$ over a 70° FoV are produced in mass quantities at extremely low cost for mobile phone modules. A larger FoV is most economically produced using arrays of such lenses. On the other end of the spectrum, the AWARE 40 optics volume is approximately $100\times$ smaller than the volume of an array of discrete cameras with an equivalent pixel capacity. At the 160 mm focal length of the AWARE 40, the lens cost dominates the microcamera cost and a multiscale design is highly advantageous.

The most interesting question in modern camera design is how to design systems with apertures between 5 mm and 5 cm. Even

for the AWARE 40 system, the cost and volume of the electronics were much greater than those of the optics. While recent designs suggest that it is possible to further reduce the optics volume of AWARE-style designs by an order of magnitude or more [51], the most pressing current problem is how to manage the size, weight, power, and cost of electronics in high-pixel-count cameras. We expect that multiscale designs will eventually be attractive at aperture sizes spanning 1–5 cm, but at present the cost and volume of 1–2 cm $f/2.5$ lenses are so small compared to the electronics needed to operate them that multiscale integration may be premature. Keeping this in mind, we turn to a discussion of electronic components in the next section. As discussed below, the electronic volume for AWARE-based cameras has been reduced by more than $100\times$ over the past five years. A similar volume reduction in the next five years will impact the choice between discrete and multiscale arrays.

4. ELECTRONICS

The first 100 years of photography relied solely on optical and chemical technologies. In the third half-century, from 1925–1975, film photography and vacuum tube videography co-existed. The first digital camera was built in 1975, with 0.01 MPs [52]. In the fourth half-century, from 1975 to the present, electronics have become increasingly integral components of digital cameras. Indeed, where the original camera consisted of just two parts, the lens and the focal plane sensor, the modern camera consists of three parts—the lens, the sensor, and the computer. With this evolution, the lines between photography and videography are increasingly blurred as interactive features are added to still photographs and photographs are estimated from multiple frames rather than just one. Video is also changing. Conventional video assumes that the capture resolution (e.g., SD, HD, 4K, or 8K) and the display resolution are matched. The array cameras, however, can capture video at a much higher resolution than any display can support, and, therefore, require a cloud layer where video streams can be buffered and transcoded to match the display requirements.

The array camera electronics include the image sensor, initial image signal processing components, as well as memory and communications. This section reviews each of these components in turn and discusses the current state of the art in array camera implementation.

A. Imaging Sensor

Since the electronic camera [52,53] appeared, the basic trend has been for the image sensor resolution to increase. There was a $100\times$ improvement of the first prototype in 1975 to the MP performance of commercial cameras in 1990. Since cameras reached the 10 MP scale, however, improvements have been more gradual. While 100 and 150 MP image sensors are commercially available (for example, Sony IMX211 and IMX411 [54]), a higher resolution does not automatically translate to a higher image quality. Instead, for a given sensor area, the sensor with more pixels has a smaller pixel size, and thus has a lower SNR in practice. High-quality large sensors are expensive and cannot run at high frame rates. Of even greater significance, for reasons discussed above, 5–10 mm aperture-size lenses with smaller format sensors are more attractive from an optical design perspective.

As in the optics section, granularity is a fundamental question in sensor design. Currently, 4 K sensors with a video rate or faster frame rates are readily available and image processing pipelines are

optimized for 4 K systems. In considering moving to larger or smaller sensors, one must analyze what metrics may be improved. The mechanical overhead and electronic interfaces suggest that very small sensors will have a higher cost per pixel than 4 K sensors. But it is far from clear that 100 MP sensors have a lower cost per pixel or better noise performance than 4K. As with optics, there is some optimal array size at which the system cost per pixel will be minimized.

In this regard, it is important to note that the actual sensor contributes relatively little to the cost or volume of current digital camera systems. For the AWARE cameras the sensor cost was less than 2% of the overall system cost, and the size, weight, and power of image processing, communications, and storage systems were vastly larger than that of the sensor itself. These subsystems are naturally parallelizable.

B. Image Signal Processor

Digital cameras process images after acquisition. Processing tasks include analog to digital conversion, sensor bias correction (e.g., pixel non-uniformity, stuck pixels, dark corner effect, and geometric distortion), demosaicing, automatic configuration (e.g., auto-focus, auto-exposure, and auto-white-balance), random noise removal, and image/video coding. As with lenses and sensors, parallel arrays of image signal processors can handle much larger pixel rates than single-vector processors.

In fact, parallel processing is highly desired for images and videos with spatial resolutions beyond 4 K or 8 K formats. Typically, a single high-resolution image or video frame will be sliced into tiles spatially, and each tile could be processed independently. In codec design, because of the neighbors involved for prediction, it usually uses the expensive on-chip buffer (such as SRAM) to cache the pixels from the upper line for fast retrieval without on-/off-chip buffer transfer. 16 K video requires $16\text{ K} \times 1.5 = 24\text{ KB}$ to only host neighboring pixels for the upper line. This is unbearable for a consumer level codec with only tens of KB of SRAM, which is also loaded by motion estimation, logic, etc. On the other hand, content adaptive binary arithmetic coding (CABAC) is utilized in the advanced video coding standards (e.g., H.264/AVC and H.265/HEVC) to improve the coding efficiency. However, CABAC operates at a sequential behavior. The overall throughput is highly dependent on the pixel numbers. For a 16 K video, the encoder frame rate could be just 15 fps if we assume the encoder could offer 240 fps at 4 K resolution. But with parallel tiles, 16 K or even higher-resolution videos can be split into multiple videos at a lower spatial resolution, where off-the-shelf chips could handle the real-time encoding and processing easily.

Therefore, inspired by the great success of parallel computer system, the parallel image signal processor [3,7], which senses and processes the large images with huge number of pixels by a set of sub-processors, are proposed. Integrating dozens local signal processors together, the parallel camera works synergistically to acquire more information, e.g., higher spatial resolution, light field with more angle information or high-speed video with finer temporal resolution. The electronic system structure diagrams of both conventional cameras and parallel cameras are presented in Fig. 5. As shown in Fig. 5(a), the electronic part of a conventional camera is simply composed of an image sensor and an image signal processor, which are usually integrated on a single chip for a compact camera design. For the parallel system shown in Fig. 5(b), to make

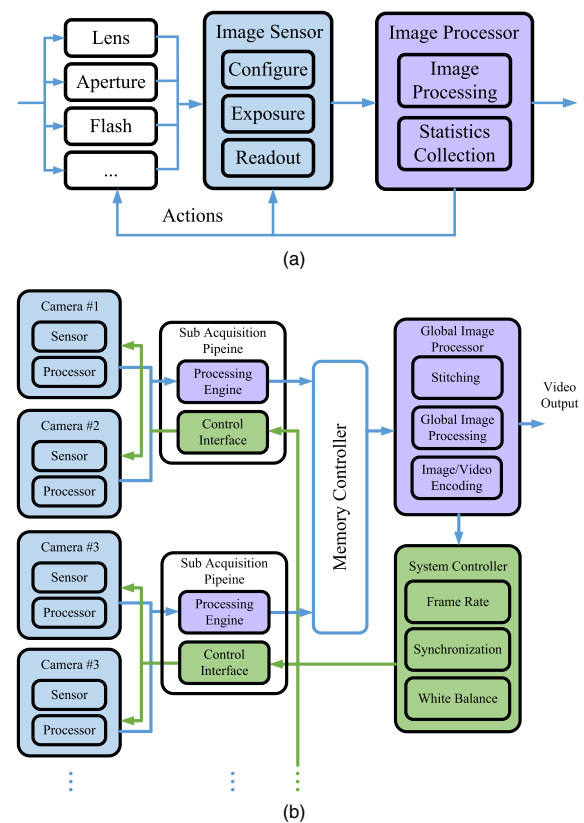


Fig. 5. System structure for (a) the conventional camera and (b) the parallel camera.

the sub-cameras work together in a proper way, a complex hierarchical electronic structure is required.

C. Hierarchical Structure for a Parallel Electronic System

It is difficult to design an electronic system to sense and process the entire image data all at once for cameras with large data throughput, like gigapixel cameras [3,7]. Therefore, it is natural to use the parallel framework to handle large-scale data. Figure 5(b) illustrates a hierarchical structure for parallel cameras. In such systems, the sub-cameras are divided into several groups, and each of these groups is an independent acquisition module. For instance, Brady *et al.* [3] use a FPGA-based camera control module to provide an interface for local processing and data management. Wilburn *et al.* [7] handle 100 cameras with four groups, and four PCs are used to control the groups accordingly and record the video streams to a striped disk array.

It is worth noting that except for the complete parallel cameras, which are composed of a set of individual cameras, there are also two kinds of hybrid structures, i.e., the cameras with a parallel optical system+single electronics [11,22] and the cameras with a single optical lens+parallel electronics [27]. As for the electronic part, the former one is just like the single camera, but the latter one is very similar to parallel cameras. As a typical example, LSST [27] has a single optical lens, but uses 189 scientific sensors to capture an image with 3.2 GPs. To handle the data at such a huge scale, the hierarchical structure is also applied, i.e., each of the nine sensors are assembled into a raft, and each raft has its own dedicated electronics.

As an example of the scale of electronic processing required, Aqueti, Inc. developed a software platform to allow video operation of AWARE cameras. AWARE cameras used field programmable gate arrays (FPGAs) to collect data from the microcameras. The FPGA's required water cooling to process 6 frame per second images with 3 W per sensor capture power. Data compression and storage were implemented in a remote computer cluster, requiring nearly 1 Gb/sensor/second of transmission bandwidth between the camera head and the server [55]. Real-time stitching and interactive video from this system used a CPU and network attached storage array requiring more than 30 W per sensor.

More recently, Aqueti has extended this software platform in the "Mantis" series of discrete lens array cameras. Mantis cameras use a NVidia Tegra TX1 "system on module" microcamera controllers. Each Tegra supports two 4 K sensors with 10 W power such that the system runs at 30 fps, with image processing and compression implemented in the camera head with 5 W power per sensor. The Mantis cameras produce 100 MP images coded in H.265 format with 10–25 MBs bandwidth to a remote render machine. While Mantis does not require camera head water cooling, as used in AWARE, the Mantis head dissipates 100 W power. While the overall image processing and compression volume is decreased by $>100\times$ relative to AWARE, the electronic system remains larger and more expensive than the optics.

5. IMAGE COMPOSITION AND DISPLAY

The astute reader will have noted by now that we have not accounted for any of the potential disadvantages of parallelizing camera image capture and processing. The primary such disadvantage is that, while a conventional camera captures a continuous image of a relatively uniform quality, the image captured by an array is only piece-wise continuous. At the "seams" between images captured by adjacent cameras, "stitching defects," which have no direct analog in conventional cameras, may appear. We have heretofore neglected this problem because it is, in fact, relatively difficult to objectively evaluate. From a raw information capacity perspective, image discontinuities have little impact on camera performance. However, such discontinuities are naturally disconcerting to human viewers expecting the camera to truthfully render the scene.

In the AWARE camera systems, each section of a scene was assigned to an independent camera. The microcamera FoV overlapped by 10–20% to allow feature matching for control points and image stitching. A fully stitched image was estimated using image-based registration. The control points found in one frame could then be used to compose fully stitched images of subsequent frames. Very occasionally, fully stitched images were printed on paper for wall hangings. At 300 dpi, printed AWARE 2 images are 1.8 m high and 4.4 m long. While there are certainly applications for video tile displays on this scale, in almost all common uses of gigapixel-scale cameras the video display resolution is much less than the raw camera resolution. In such cases the stitched image must be decomposed into "tile" components for interactive display. In the case of the AWARE cameras, a model-based architecture was created to use the pull data from microcamera streams at the video rate to compose the real-time display [56,57] without forming a completely stitched image. More recently, the Aqueti Mantis array cameras included a wide FoV camera along with the narrow field array. At low resolution, images from the wide-field camera are presented to the

viewer, and as the viewer zooms in the display switches to data from high-resolution cameras. While the camera uses a discrete array of 18 narrow-field cameras, the current display window can always be estimated from, at most, four cameras.

The use of inhomogeneous sensor arrays has a long history prior to Mantis. In some scenarios different types of image sensors or different configurations (e.g., exposure time and frame rate) are required to achieve multimode sensing. Some examples follow: Shanakr *et al.* [18] studied microcamera diversity for computational image composition; Wang *et al.* [58] combined a DSLR camera and low-budget cameras to capture high-quality light field images with low-cost devices; Kinect [59] captured the RGB and depth images by using two types of sensors; and Wang *et al.* [60] achieved high-quality spectral imaging by combining the high-resolution RGB sensor and a coded aperture-based spectral camera together. The Large Synoptic Survey Telescope (LSST) [27] uses three types of sensors, i.e., common image sensors, wavefront sensors, and guide sensors to correct the aberration effect caused by the atmospheric disturbance from images.

As these examples illustrate, there are many possible uses and configurations for array cameras, with many configurations yet to be explored. Traditional photography and videography implement a one-to-one mapping between focal-plane pixels and display pixels. For example, standard definition television, HD television, and 4 K television all operate under the assumption that the image captured by the camera is the image seen on the display. With array cameras, however, such a one-to-one mapping is no longer possible or even desirable. Instead, high-resolution array camera data streams require context-sensitive and interactive display mappings similar to those under development for virtual reality broadcasting [61]. In previous work, we have described the real-time interface developed for AWARE cameras [55]. This interface allows a single user connected to an array camera to digitally pan, tilt, and zoom over high-resolution images while also controlling the flow of time. The AWARE architecture is a network structure consisting of "microcamera controllers," memory, and "render agents." Multiple render agents may be connected in parallel to a camera, but in practice fewer than five such systems have been connected.

As an example, the Aqueti Mantis 70 camera is an array of 18 narrow-field microcameras, each with a 25 mm focal length lens and a 1.6 μm pixel pitch. Each uses a Sony IMX 274 color CMOS sensor. Sensor readout, ISP, and data compression are implemented using an array of NVIDIA Tegra TX1 modules with two sensors per TX1. Custom software is used to stream sensor data to a render machine, which produces real-time interactive video with <100 ms latency. The sensors are arrayed to cover a 73° horizontal FoV and a 21° vertical FoV. The instantaneous FoV is 65 μrad , and the fully stitched image has a native resolution of 107 MPs. The camera operates at 30 frames per second. [Visualization 1](#) and [Visualization 2](#) are example video clips captured from the render interface. When zooming into the video, stitching boundaries between microcameras are sometimes visible; zooming out the interface switches to the wide field camera. These visible boundaries are mostly due to difficulty in stitching multiple images globally as well as stitching scenes from different field depths simultaneously.

Broadcasting from high-resolution array cameras to allow millions of viewers to simultaneously explore live events is a logical next step for this technology. As a simple example of interactive broadcasting, Fig. 6 shows the full-view and user-view video

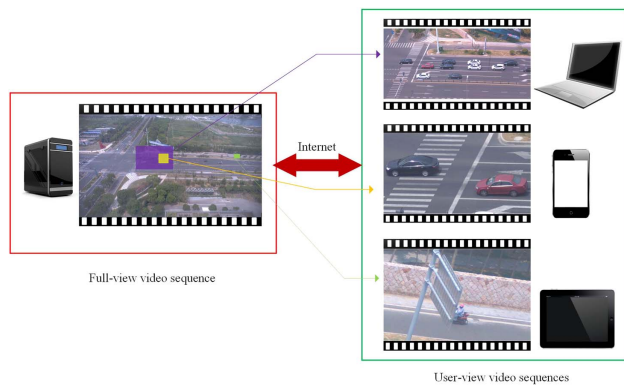


Fig. 6. In interactive video broadcasting, the user can explore the video on a window of any scale both on the spatial axis and the temporal axis.

sequences. The low-resolution wide range video sequence will be provided, and then different users can zoom in on random high-resolution regions of interest on their own mobile devices. Interactive web servers for gigapixel-scale images are as old as the World Wide Web [62], but protocols for interactive gigapixel video require further development. In addition to full video service, one imagines that novel image formats allowing exploration of short clips may emerge. A relatively simple javascript-based example is accessible at [63, 64]. This demo presents a short video clip taken with the Mantis 70 camera. The panorama in the web is created by stitching images captured by Mantis 70. For interactive display, we use the multiscale display by decomposing high-resolution images to tiles of JPEG or PNG images at different resolutions that make up an image pyramid. It enables users to zoom in or out of random parts of a large map within several seconds, because only those tiles for the user's view of the image on the screen are required to load. To experience the view in the time domain, we add a time slider to control the display of the video sequence. For example, if users zoom in on a certain region, then the view will show the video sequence of the specific area where the slider is dragged. In the web-based example, there are some different scenarios captured by the Mantis 70 camera, including Hospital, School, and Road. The corresponding distances between the scene and camera are about 30, 100 and 200 m. Even in the remote scene, users can see some details clearly such as road signs.

Stored as tiled JPEG images, 3 s of Mantis 70 video requires 2 GB residing in over 20,000 files. The online quality of service depends on the server and client quality of service and bandwidth. Here, we include servers in the United States and China to allow users to link to a close site. Large-scale deployment must rely on commercial content delivery networks with forward provisioning. Novel architectures allowing similar service using h.264 or h.265 compression may greatly reduce the bandwidth and storage requirements.

6. CONCLUSION

After several decades of development, array cameras using computational image composition are increasingly attractive. Recent introduction of commercial cameras for VR and high-resolution imaging suggest that this approach is increasingly competitive with conventional design. At the same time, lessons learned from computational imaging research allow systematic lens, electronic

hardware, and software design for array systems. One expects that this platform will allow camera information capacity per unit cost and volume to improve at rates comparable to other information technologies.

Funding. Intel Corporation; National Natural Science Foundation of China (NSFC) (61627804, 61571215).

See Supplement 1 for supporting content.

REFERENCES

1. D. J. Brady and N. Hagen, "Multiscale lens design," *Opt. Express* **17**, 10659–10674 (2009).
2. G. Bell, *Supercomputers: The Amazing Race (A History of Supercomputing, 1960–2020)* (2014).
3. D. J. Brady, M. Gehm, R. Stack, D. Marks, D. Kittle, D. R. Golish, E. Vera, and S. D. Feller, "Multiscale gigapixel photography," *Nature* **486**, 386–389 (2012).
4. J. Muybridge, "The horse in motion," *Nature* **25**, 605 (1882).
5. G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys. Theor. Appl.* **7**, 821–825 (1908).
6. A. Young, "Stereocamera," U.S. patent 2,090,017 (August 17, 1937).
7. B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," in *ACM Transactions on Graphics (Proc SIGGRAPH)* (ACM, 2005), Vol. **24**, pp. 765–776.
8. M. Shankar, N. P. Pitsianis, and D. J. Brady, "Compressive video sensors using multichannel imagers," *Appl. Opt.* **49**, B9–B17 (2010).
9. F. Mochizuki, K. Kagawa, S.-I. Okihara, M.-W. Seo, B. Zhang, T. Takasawa, K. Yasutomi, and S. Kawahito, "Single-event transient imaging with an ultra-high-speed temporally compressive multi-aperture CMOS image sensor," *Opt. Express* **24**, 4155–4176 (2016).
10. R. Horisaki, K. Choi, J. Hahn, J. Tanida, and D. J. Brady, "Generalized sampling using a compound-eye imaging system for multi-dimensional object acquisition," *Opt. Express* **18**, 19367–19378 (2010).
11. J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka, "Thin observation module by Bound Optics (Tombo): concept and experimental verification," *Appl. Opt.* **40**, 1806–1813 (2001).
12. J. Duparré, P. Dannberg, P. Schreiber, A. Bräuer, and A. Tünnermann, "Thin compound-eye camera," *Appl. Opt.* **44**, 2949–2956 (2005).
13. G. Druart, N. Guérineau, R. Hadar, S. Théas, J. Taboury, S. Rommeluère, J. Primot, and M. Fendler, "Demonstration of an infrared microcamera inspired by Xenos Peckii Vision," *Appl. Opt.* **48**, 3368–3374 (2009).
14. K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "Picam: an ultra-thin high performance monolithic camera array," *ACM Transactions on Graphics (Proc SIGGRAPH)* **32**, 166 (2013).
15. A. Portnoy, N. Pitsianis, X. Sun, D. Brady, R. Gibbons, A. Silver, R. Te Kolste, C. Chen, T. Dillon, and D. Prather, "Design and characterization of thin multiple aperture infrared cameras," *Appl. Opt.* **48**, 2115–2126 (2009).
16. M. Shankar, R. Willett, N. Pitsianis, T. Schulz, R. Gibbons, R. Te Kolste, J. Carriere, C. Chen, D. Prather, and D. Brady, "Thin infrared imaging systems through multichannel sampling," *Appl. Opt.* **47**, B1–B10 (2008).
17. G. Carles, J. Downing, and A. R. Harvey, "Super-resolution imaging using a camera array," *Opt. Lett.* **39**, 1889–1892 (2014).
18. P. M. Shankar, W. C. Hasenplaugh, R. L. Morrison, R. A. Stack, and M. A. Neifeld, "Multiaperture imaging," *Appl. Opt.* **45**, 2871–2883 (2006).
19. R. Shogenji, Y. Kitamura, K. Yamada, S. Miyatake, and J. Tanida, "Multispectral imaging using compact compound optics," *Opt. Express* **12**, 1643–1655 (2004).
20. V. R. Bhakta, M. Somayaji, S. C. Douglas, and M. P. Christensen, "Experimentally validated computational imaging with adaptive multiaperture folded architecture," *Appl. Opt.* **49**, B51–B58 (2010).
21. R. T. Collins, A. J. Lipton, H. Fujiiyoshi, and T. Kanade, "Algorithms for cooperative multisensor surveillance," *Proc. IEEE* **89**, 1456–1477 (2001).

22. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Computer Science Technical Report CSTR 2, 1–11 (2005).
23. K. Fife, A. El Gamal, and H.-S. P. Wong, "A 3D multi-aperture image sensor architecture," in *IEEE Custom Integrated Circuits Conference* (IEEE, 2006), pp. 281–284.
24. R. Anderson, D. Gallup, J. T. Barron, J. Kontkanen, N. Snavely, C. Hernández, S. Agarwal, and S. M. Seitz, "Jump: virtual reality video," *ACM Transactions on Graphics (Proc SIGGRAPH)* **35**, 198 (2016).
25. E. C. Cull, D. P. Kowalski, J. B. Burchett, S. D. Feller, and D. J. Brady, "Three dimensional imaging with the Argus sensor array," *Proc. SPIE* **4873**, 211–222 (2002).
26. S. Tallon, "A pocket camera with many eyes," *IEEE Spectrum* **53**, 34–40 (2016).
27. Z. Ivezić, J. Tyson, B. Abel, E. Acosta, R. Allsman, Y. AlSayyad, S. Anderson, J. Andrew, R. Angel, and G. Angeli, "LSST: from science drivers to reference design and anticipated data products," arXiv: 0805.2366 (2008).
28. N. Kaiser, H. Aussel, B. E. Burke, H. Boesgaard, K. Chambers, M. R. Chun, J. N. Heasley, K.-W. Hodapp, B. Hunt, and R. Jedicke, "Pan-starrs: a large synoptic survey telescope array," *Proc. SPIE* **4836**, 154–164 (2002).
29. B. Leininger, J. Edwards, J. Antoniadis, D. Chester, D. Haas, E. Liu, M. Stevens, C. Gershfield, M. Braun, and J. D. Targove, "Autonomous real-time ground ubiquitous surveillance-imaging system (ARGUS-IS)," *Proc. SPIE* **6981**, 69810H (2008).
30. W. T. Cathey, B. R. Frieden, W. T. Rhodes, and C. K. Rushforth, "Image gathering and processing for enhanced resolution," *J. Opt. Soc. Am. A* **1**, 241–250 (1984).
31. E. R. Dowski and W. T. Cathey, "Extended depth of field through wave-front coding," *Appl. Opt.* **34**, 1859–1866 (1995).
32. J. Ojeda-Castaneda, R. Ramos, and A. Noyola-Isgleas, "High focal depth by apodization and digital restoration," *Appl. Opt.* **27**, 2583–2586 (1988).
33. R. Raskar, *Computational Photography: Epsilon to Coded Photography* (Springer, 2009), pp. 238–253.
34. B. E. Bayer, "Color imaging array," U.S. patent 3,971,065 (July 20, 1976).
35. X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, and D. J. Brady, "Computational snapshot multispectral cameras: toward dynamic capture of the spectral world," *IEEE Signal Process. Mag.* **33**, 95–108 (2016).
36. G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: an introduction," *IEEE Signal Process. Mag.* **31**, 105–115 (2014).
37. D. J. Brady, *Optical Imaging and Spectroscopy* (Wiley, 2009).
38. M. S. Asif, A. Ayremlou, A. Veeraraghavan, R. Baraniuk, and A. Sankaranarayanan, "Flatcam: replacing lenses with masks and computation," in *IEEE International Conference on Computer Vision Workshop (ICCVW)* (IEEE, 2015), pp. 663–666.
39. D. L. Marks, R. A. Stack, D. J. Brady, D. C. Munson, and R. B. Brady, "Visible cone-beam tomography with a lensless interferometric camera," *Science* **284**, 2164–2166 (1999).
40. J. J. McCann and A. Rizzi, *The Art and Science of HDR Imaging* (Wiley, 2011), Vol. 26.
41. S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Transactions on Graphics (Proc SIGGRAPH)* **35**, 192 (2016).
42. C. Zhang, J. Bastian, C. Shen, A. van den Hengel, and T. Shen, "Extended depth-of-field via focus stacking and graph cuts," in *20th IEEE International Conference on Image Processing (ICIP)* (IEEE, 2013), pp. 1272–1276.
43. P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Coded aperture compressive temporal imaging," *Opt. Express* **21**, 10526–10545 (2013).
44. P. Llull, X. Yuan, L. Carin, and D. J. Brady, "Image translation for single-shot focal tomography," *Optica* **2**, 822–825 (2015).
45. D. J. Brady, A. Mrozack, K. MacCabe, and P. Llull, "Compressive tomography," *Adv. Opt. Photon.* **7**, 756–813 (2015).
46. S. Bastý, M. Neifeld, D. Brady, and S. Kraut, "Nonlinear estimation for interferometric imaging," *Opt. Commun.* **228**, 249–261 (2003).
47. D. J. Brady, "Multiplex sensors and the constant radiance theorem," *Opt. Lett.* **27**, 16–18 (2002).
48. R. K. Luneburg and M. Herzberger, *Mathematical Theory of Optics* (University of California, 1964).
49. E. J. Tremblay, D. L. Marks, D. J. Brady, and J. E. Ford, "Design and scaling of monocentric multiscale imagers," *Appl. Opt.* **51**, 4691–4702 (2012).
50. I. Stamenov, A. Arianpour, S. J. Olivas, I. P. Agurok, A. R. Johnson, R. A. Stack, R. L. Morrison, and J. E. Ford, "Panoramic monocentric imaging using fiber-coupled focal planes," *Opt. Express* **22**, 31708–31721 (2014).
51. W. Pang and D. J. Brady, "Galilean monocentric multiscale optical systems," *Opt. Express* **25**, 20332–20339 (2017).
52. G. A. Lloyd and S. J. Sasson, "Electronic still camera," U.S. patent 4,131,919 (December 26, 1978).
53. D. Praker, *The Visual Dictionary of Photography* (AVA Publishing, 2009).
54. "Sony Product Announcement," http://www.sony-semicon.co.jp/products_en/news/detail/170301.pdf (2017).
55. J. Nichols, K. Judd, C. Olson, K. Novak, J. Waterman, S. Feller, S. McCain, J. Anderson, and D. Brady, "Range performance of the DARPA AWARE wide field-of-view visible imager," *Appl. Opt.* **55**, 4478–4484 (2016).
56. D. Golish, E. Vera, K. Kelly, Q. Gong, P. Jansen, J. Hughes, D. Kittle, D. Brady, and M. Gehm, "Development of a scalable image formation pipeline for multiscale gigapixel photography," *Opt. Express* **20**, 22048–22062 (2012).
57. Q. Gong, E. Vera, D. R. Golish, S. D. Feller, D. J. Brady, and M. E. Gehm, "Model-based multiscale gigapixel image formation pipeline on GPU," *IEEE Trans. Comput. Imaging* **3**, 493–502 (2017).
58. Y. Wang, Y. Liu, W. Heidrich, and Q. Dai, "The light field attachment: turning a DSLR into a light field camera using a low budget camera ring," *IEEE Trans. Vis. Comput. Graphics* **23**, 2357–2364 (2017).
59. "Kinect," 2017, <https://www.microsoftstore.com/accessories/xbox-kinect/p/1708-00000>.
60. L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu, "High-speed hyperspectral video acquisition with a dual-camera architecture," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 4942–4950.
61. F. Duanmu, E. Kurdoglu, S. A. Hosseini, Y. Liu, and Y. Wang, "Prioritized buffer control in two-tier 360 video streaming," in *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network* (ACM, 2017), pp. 13–18.
62. K. Martinez, J. Cupitt, and S. Perry, "High resolution colorimetric image browsing on the web," *Computer Networks and ISDN Systems* **30**, 399–405 (1998).
63. "MantisDemo," 2017, <http://vision.nju.edu.cn/demo/>.
64. "MantisDemoUS," 2017, <http://aqueti.tv/demos/xiaolinRoad/>.