Shift-encoded optically multiplexed imaging

Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Shift-encoded optically multiplexed imaging

Vinay Shah
Yaron Rachlin
R. Hamilton Shepard
Tina Shih

SPIE.

Vinay Shah, Yaron Rachlin, R. Hamilton Shepard, Tina Shih, "Shift-encoded optically multiplexed imaging," *Opt. Eng.* **56**(4), 041314 (2017), doi: 10.1117/1.OE.56.4.041314.

Optical Engineering 56(4), 041314 (April 2017)

Shift-encoded optically multiplexed imaging

Vinay Shah, Yaron Rachlin,* R. Hamilton Shepard, and Tina Shih
Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, Massachusetts, United States

Abstract. In a multiplexed image, multiple fields-of-view (FoVs) are superimposed onto a common focal plane. The attendant gain in sensor FoV provides a new degree of freedom in the design of an imaging system, allowing for performance tradeoffs not available in traditional optical designs. We explore design choices relating to a shift-encoded optically multiplexed imaging system and discuss their performance implications. Unlike in a traditional imaging system, a single multiplexed image has a fundamental ambiguity regarding the location of objects in the image. We present a system that can shift each FoV independently to break this ambiguity and compare it to other potential disambiguation techniques. We then discuss the optical, mechanical, and encoding design choices of a shift-encoding midwave infrared imaging system that multiplexes six 15×15 deg FoVs onto a single one megapixel focal plane. Using this sensor, we demonstrate a computationally demultiplexed wide FoV video. [DOI: 10.1117/1.OE.56.4.041314]

Keywords: multiplexed imaging; computational imaging; infrared; noise equivalent differential temperature; midwave infrared.

Paper 170059SSP received Jan. 9, 2017; accepted for publication Mar. 7, 2017; published online Mar. 20, 2017.

1 Introduction

The vast majority of optical sensors used today are designed so that each pixel views only a single contiguous region in object space in each frame. In these established approaches, increasing the field-of-view (FoV) while maintaining spatial resolution requires either more pixels or more time. Methods have been developed to create wide FoV systems without sacrificing fine image detail by stitching together images from multiple narrow FoV sensors, by scanning a single narrow FoV sensor across the scene, and through super-resolution techniques that combine a series of images with subpixel shifts. These are well-established approaches useful in many applications. In situations in which focal planes are prohibitively expensive and/or require extra cooling infrastructure, such as advanced digital focal plane arrays (FPAs), sensitive infrared imagers, or fast framing imagers, stitching multiple images may be a cost or resource prohibitive option. A scanning system relies on mechanical systems that can limit scene revisit. Achieving resolution over a wide FoV through super-resolution techniques relies on wide FoV optics, which may result in distortions that limit resolution gains, as in a fisheye lens, and requires a static scene over multiple frames.

The image formed by an optically multiplexed system is the superposition of multiple images formed by discrete imaging channels. This has been investigated in designs that use multiple lenses to form images on a single FPA,[1] a cascade of beam splitting elements to divert multiple FoVs into a single lens,[2–5] and by placing an interleaved array of subaperture microprisms in front of a single lens.[6,7] In a recent paper,[8] we presented a new optical design architecture based on a division of aperture technique that divides the pupil area of a single lens into a number of independent imaging channels. As discussed in that paper, this method offers advantages over prior approaches through its flexibility to individually direct and encode the optical channels, and it yields a significant volume advantage in systems with a high degree of multiplexing at a cost of an optical resolution loss due to the divided aperture.

A single multiplexed image undersamples the scene, and without additional measurements or prior knowledge about the scene, contains inherent ambiguities. While it is feasible to detect objects in a multiplexed image, the angular position of a detected object is uncertain since it could have appeared in any of the image channels. Static encoding schemes, such as changing the point spread function of each channel,[8] combining measurements from two or more multiplexing sensors with different multiplexing channel parameters,[9] or relying on differential channel overlap and rotation,[3] enable disambiguation and tracking of localized objects. Dynamic time-encoded schemes, in which the encoding varies across frames, enable imaging of scenes without any *a priori* knowledge of the spatial structure. Image reconstruction comes at the cost of multiple exposures, and thus requires the scene to be relatively stable across this set of exposures. Reconstruction via dynamic encoding has been previously achieved by using shutters to attenuate individual imaging channels[4] or by using a slow moving element to continuously shift a single channel's image between samples for a two channel multiplexed imager.[2]

In prior work, we described hardware design architectures for optically multiplexed imaging,[8] methods of image encoding, and computational image reconstruction,[10] and the design and calibration of a six channel shift-encoded mid-wave infrared (MWIR) multiplexed imager.[11] A high-level flow diagram of the optically multiplexed imaging process with image shifting as a means of encoding is shown in Fig. 1. Multiple sections of the scene, or FoVs, are multiplexed onto the imaging sensor through the multiplexing assembly. The multiplexing assembly splits the aperture of the objective lens into equal area subapertures corresponding to each of the FoVs. Additionally, the FoV of the subapertures can be independently shifted, which serves as an encoding function. The ability to change the

*Address all correspondence to: Yaron Rachlin, E-mail: yaron.rachlin@ll.mit.edu

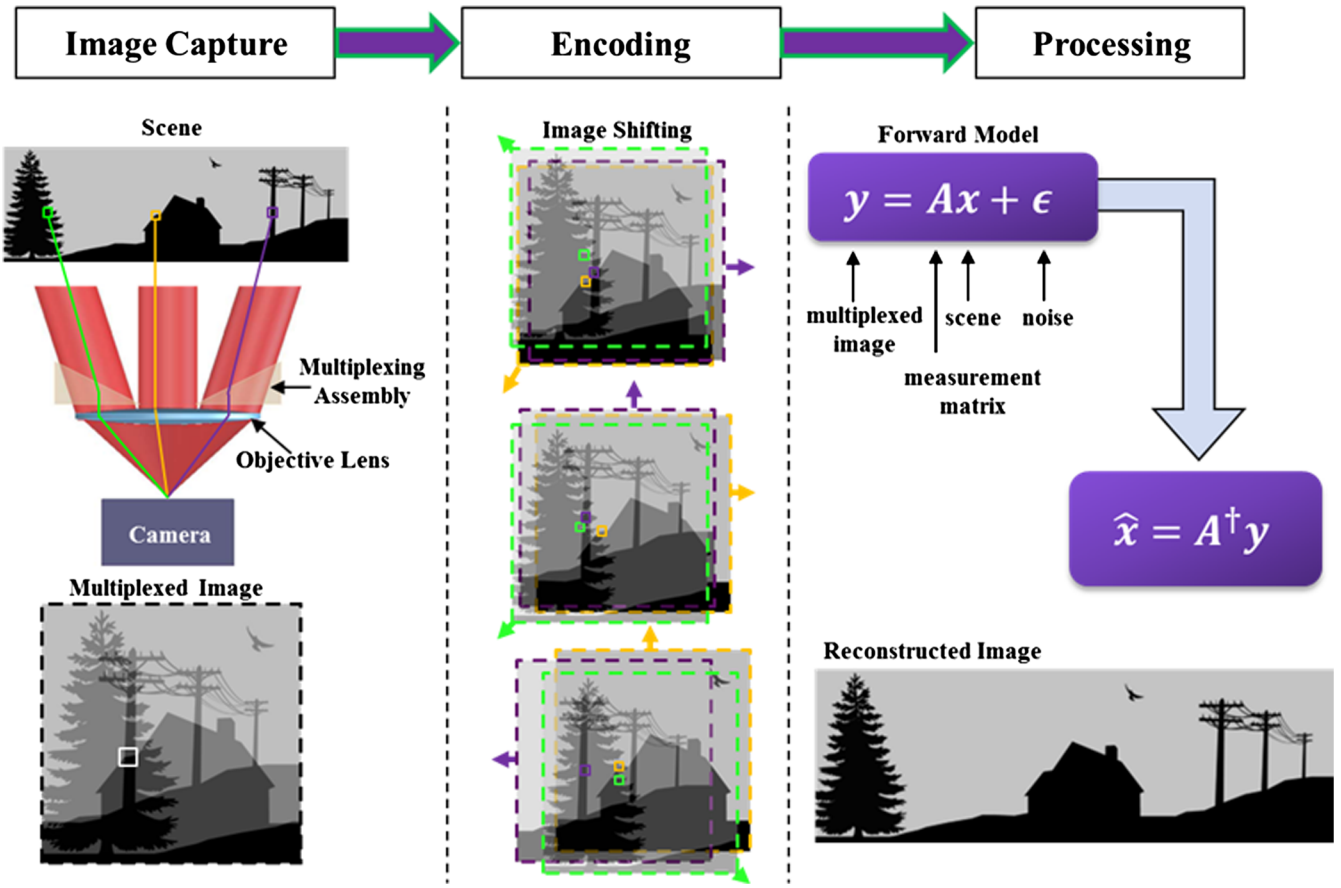

Fig. 1 Flow diagram of the optically multiplexed imaging, shift encoding, and computational decoding processes realized in the MWIR prototype discussed in this paper.

encoding per frame is what we refer to as dynamic encoding. A collection of shifted multiplexed measurements can result in an overdetermined, underdetermined, or full rank measurement matrix depending on the number of multiplexed measurements and shift selection. These measurements can be processed to extract information about the underlying scene. Broadly applicable examples of this include detecting objects of interest in the multiplexed images using standard detection methods and recovering an estimate of the scene by applying a pseudoinverse to the linear measurements.

In this paper, we discuss the performance implications and processing of a shift-encoded optically multiplexed imaging system. Section 2 provides a model of the formation of a multiplexed image. The shift encoding of our system is accomplished via a division of aperture multiplexing approach, whose implications we discuss in Sec. 3. The optical and mechanical design choices for a six channel MWIR multiplexed imager are described in Sec. 4. Section 5 discusses image reconstruction using image shift encoding and examines the effect of shift selection on image reconstruction error. In Sec. 6, we describe encoding and localizing features of interest without generating demultiplexed images, and compare shift-encoding to other encoding approaches. Lab measurements confirming the effect of multiplexing on noise scaling and imagery from field testing are presented in Sec. 7. The imagery includes frames from a wide FoV video generated using our MWIR multiplexed imaging prototype.

2 Sensor Model

We model the imaging process as a linear transformation from object to image space with additive noise. This can be represented as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^{l}_{\geq 0}$ is the measured image observed on the focal plane, $\mathbf{A} \in \mathbb{R}^{l \times m}_{\geq 0}$ is the imaging transformation matrix, $\mathbf{x} \in \mathbb{R}^{m}_{\geq 0}$ is a discretized m-pixel representation of the scene, and $\boldsymbol{\epsilon} \in \mathbb{R}^{l}$ represents the noise corrupting each pixel measurement. A multiplexed imager has a transformation matrix comprised of an encoding, a selection, and a multiplexing transformation. Thus, the multiplexing imaging process can be written as follows:

$$\mathbf{y} = \mathbf{A}_{\mathbf{multiplex}}\mathbf{A}_{\mathbf{selection}}\mathbf{A}_{\mathbf{encoding}}\mathbf{x} + \boldsymbol{\epsilon}. \tag{2}$$

The encoding operation, $\mathbf{A}_{\mathbf{encoding}} \in \mathbb{R}^{nm \times m}_{\geq 0}$, produces an encoded version of the underlying scene for each of the n channels. Various encoding schemes are described in Sec. 6. The selection matrix, $\mathbf{A}_{\mathbf{selection}} \in \{0,1\}^{\ln \times nm}$, maps the shifted scene coordinates to focal plane coordinates. Finally, the multiplexing operation, $\mathbf{A}_{\mathbf{multiplex}} \in \mathbb{R}^{l \times \ln}_{\geq 0}$, sums over the n channels, which are physically superimposed on the l pixel focal plane. In general, solving for x given y is an ill-posed problem as $l < m$ due to the multiplexing

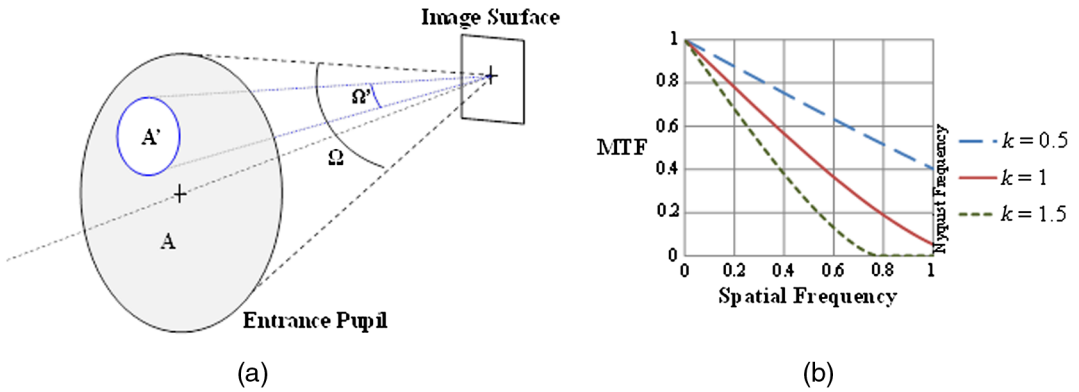

Fig. 2 (a) Aperture area and solid angle for a division of aperture imaging system. (b) Optical system MTF for undersampled, critically sampled, and oversampled systems.

operations. However, taking p images of a static scene with a varying encoding can be written as

$$\tilde{\mathbf{y}} = \tilde{\mathbf{A}}\mathbf{x} + \tilde{\boldsymbol{\epsilon}}, \tag{3}$$

where $\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \\ \mathbf{y}_p \end{bmatrix} \in \mathbb{R}^{pl}_{\geq 0}$, $\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \\ \vdots \\ \mathbf{A}_p \end{bmatrix} \in \mathbb{R}^{pl \times m}_{\geq 0}$, and $\tilde{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \boldsymbol{\epsilon}_3 \\ \vdots \\ \boldsymbol{\epsilon}_4 \end{bmatrix} \in \mathbb{R}^{pl}$. In Sec. 5, we will show that if $pl \geq m$, then $\tilde{\mathbf{A}}$ can be full rank or overdetermined for appropriately chosen encoding matrices. Using the pseudoinverse of $\tilde{\mathbf{A}}$, one can solve for an estimate of the scene, which minimizes the mean squared errors (MSEs) between the measurements predicted by the estimate and the observed measurements:

$$\hat{\mathbf{x}} = \tilde{\mathbf{A}}^{\dagger}\tilde{\mathbf{y}}. \tag{4}$$

Other formulations of signal recovery are also feasible, for example, that incorporate prior assumptions about the scene as explored in the compressive sensing literature[12,13] or interframe scene dependencies,[2] though in this paper, we focus on the performance of linear least squares estimators.

3 Sensor Implications of Shift-Encoded, Division of Aperture Multiplexed Architecture

The realization of the imaging model described in the prior section can be achieved using a number of imaging architectures. An architecture based on aperture division and shift encoding possesses a number of advantages, including volume, the size of the encoding elements, and performance, as discussed in a recent paper (Ref. 8). Primary disadvantages include the diffraction losses when using contiguous subapertures.

In a division of aperture, shift-encoded architecture, the entrance pupil is subdivided into sections of equal area. An array of mirrors placed near the entrance pupil directs a number of FoVs into an objective lens to form a multiplexed image. Image encoding is performed by rapidly shifting individual layers of the multiplexed image via tilting the mirrors. A concrete example of such a design is described in Sec. 4. In this section, we discuss the implications of this architecture on image resolution and signal-to-noise ratio (SNR).

If each of the n channels is designed to have an equal aperture area, A', then its area is equal to A/n, where A is the area of the entrance pupil of the objective lens, as shown in Fig. 2(a). Approximating the apertures as circular, the relative entrance pupil diameter for each channel scales by $1/\sqrt{n}$ relative to the objective lens. It follows that the channel's effective $F/\#$ scales by $\sqrt{n}$.

As a consequence of dividing the aperture, the diffraction-limited resolution of the individual channels decreases relative to the objective lens. If the system resolution is limited by the optical resolution of the full aperture, a resolution loss is incurred due to the division of aperture. However, in optical systems, in which the resolution is limited by the pixel size, for example, in many infrared imaging systems, in which focal planes have a large pixel pitch, division of aperture can yield a FoV gain without a system resolution loss. Consider a critically sampled system in which the Nyquist sampling frequency is equal to the diffraction-limited optical resolution. With coarser sampling (i.e., undersamping), the image resolution is digitally limited, and with finer sampling (i.e., oversampling), the image resolution is limited by diffraction. Equation (5) expresses this criterion in terms of system $F/\#$, wavelength, and pixel pitch, d_{pixel}. Figure 2(b) shows modulation transfer function (MTF) curves for oversampled, undersampled, and critically sampled systems. If the multiplexed channels are undersampled or critically sampled, a FoV gain of n can be achieved without imaging resolution loss. The MWIR prototype described in Sec. 4 was designed to operate at a central wavelength of 4.2 μm. An $F/2$ objective lens was used with a FPA of 25 μm pixels. This produced a highly undersampled image with $k = 0.2$ for the objective lens and $k = 0.5$ for each of the multiplexed channels. For this particular design, FoV can be expanded by a factor of 23 without resolution loss, though with SNR implications discussed below:

$$\frac{0.61\lambda F\#}{d_{\text{pixel}}} = k, \begin{cases} k < 1 \text{undersampled} \\ k \approx 1 \text{critically sampled} \\ k > 1 \text{oversampled} \end{cases}. \tag{5}$$

Table 1 MWIR prototype optical design parameters.

Parameter	Objective lens	Multiplexed system
Focal length	96 mm	96 mm
$F/\#$	2	4.9
FoV	15.2×15.2 deg	90×15.2 deg
Instantaneous field of view	260 μrad	260 μrad
Distortion	<1%	<1%

The SNR scaling of multiplexing as a function of the number of images channels depends on the dominant noise sources of the system and the nature of the signals being multiplexed. In an infrared system, the division of aperture design reduces the image irradiance of each channel by n; however, the background irradiance remains constant due to the superposition of the images. Thus, the background shot noise is set by the full solid angle of the objective lens while the channel signal levels scale by $1/n$. This has the effect of scaling the single-frame signal transfer function (SiTF) and noise equivalent differential temperature (NEDT) of the multiplexed channels by $1/n$ and n, respectively, compared to the full aperture objective lens.[14] This scaling relationship assumes the same integration time (T_{int}) for both systems, which is reasonable for cooled systems in which T_{int} is limited by the well depth and for uncooled systems in which T_{int} is limited by the thermal response time of the pixel. In the case in which noise is dominated by the shot noise of the signal photons rather than the background shot noise (e.g., a hot object against a cold background), the single frame NEDT would instead scale by $1/\sqrt{n}$. As in a conventional system, the SNR can be increased through additional integration time, combining multiple frames if not limited by other factors such as scene or platform motion, or increasing the aperture area of the multiplexed sensor. The SNR implications of combining multiple multiplexed images to form a conventional image are discussed in Sec. 5.

4 Prototype Design

4.1 *Optical Design*

The prototype design process began with camera selection. Given the resolution and SNR implications of optical multiplexing outlined in the prior section, a camera with large undersampled pixels with a high well capacity was sought, leading to the selection of the IRC910 from IRCameras.[15] This camera provides a 1024×1024 image format with 25 μm pixels and a 10 Me^- well depth. The dewar contained a custom $F/2$ cold shield with a 3.4 to 5 μm CO_2 notched cold filter. A 96-mm focal length requirement followed such that the lens would provide a 90×15.2 deg panoramic multiplexed FoV. The first order optical design parameters of the prototype are shown in Table 1.

A distortion requirement of <1% was selected to balance the overall FoV requirement with the image processing complexity. Constraining the distortion to a small value prevented the horizontal FoV from falling below 90 deg. It also provided an approximately rectilinear mapping of the scene such that additional FoV overlap was not required to avoid blind spots between the channels. And finally, high levels of distortion introduce a variation in apparent scene motion across the FoV as the images are shifted for encoding, which would require additional correction in the image processing algorithms. In principle, an F-theta lens would be desired for a uniform image shift, but this would conflict with the second objective (rectilinear mapping). For the small image shift magnitudes (~10 pixels) and the relatively narrow semi-FoV angle (10.7 deg), the difference between F-theta and F-Tan (theta) distortion was negligible.

The geometry for the multiplexing optical assembly was selected to divide the entrance pupil into six sections of equal area. This pupil division arrangement is shown in Fig. 3(a). The design intent was to produce the same aperture area and SNR at the center of each channel; however, variations in the image irradiance within the FoV of each channel were caused by vignetting on the multiplexing mirror structure, a variation between the planes of the mirrors and the plane of the entrance pupil, and pupil aberrations. The resulting per channel SiTF and single-frame NEDT are shown in Fig. 11 and Table 2.

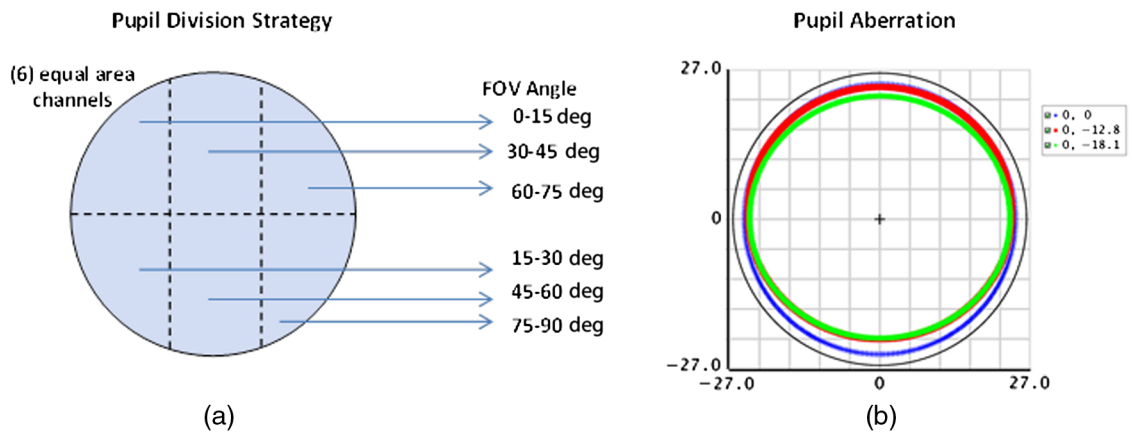

Fig. 3 (a) Pupil division geometry and the associated FoV coverage for each channel. (b) Pupil aberration illustrated with a footprint diagram at the entrance pupil. Relative field angles of 0, 0.707, and 1 are shown.

Table 2 NEDT noise in units of Kelvin.

Parameter	Full aperture	Channel 1	Channel 2	Channel 3	Channel 4	Channel 5	Channel 6
NEDT	0.018	0.122	0.106	0.089	0.094	0.109	0.123

The optical design for this prototype unit includes a pupil relay that reimages the cold stop to the multiplexing optical assembly, which is located in the front of the system. A threshold pupil aberration requirement was derived to constrain the maximum relative marginal ray error at the entrance pupil to <10% for any field angle. This limited the maximum variation in image irradiance due to pupil aberration to roughly 25%. Additional image irradiance variations resulted from vignetting. The requirement to correct pupil aberrations became a design driver and required the use of more lens elements than would otherwise be needed.

Pupil aberration was controlled in the optimization merit function by constraining ray heights at the entrance pupil location. The chief ray height for a number of field angles was minimized to reduce spherical aberration of the pupil, which causes a lateral shift of the beam footprint[16] with field angle. Magnification, anamorphism, and distortion of the pupil were constrained by limiting the variation in marginal ray heights at the pupil plane for multiple field angles. Simultaneous correction of image distortion and pupil coma proved particularly challenging. The image distortion requirement converged toward a balancing of third and fifth order distortion; yet, simultaneous correction of image and pupil aberrations places restrictions on the ability to balance third and fifth order aberrations.[17] Uncorrected image distortion translates into coma of the pupil, which is observed by the increasing anamorphic pupil magnification with field height in Fig. 3(b). Ultimately, an approach of minimizing aberrations by splitting lenses in the rear group was required to satisfy the requirement set.

A layout of the final optical design is shown in Fig. 4(a). The prescription is a reimaging design that projects the entrance pupil ~100 mm in front of the lens. Ten lens elements were used with aspheres on the surfaces closest to the intermediate focus and the aperture stop. Color correction over the 3.4 to 5 μm waveband was achieved with only two materials, silicon and germanium. The spacing between the intermediate field lens and rear group was stretched to accommodate a fold mirror for packaging, as shown in Fig. 4(b).

4.2 *Mechanical Design*

In order to achieve the image encoding of each channel via rapid and precise movement of each FoV, a multiplexing reflective mirror assembly was designed and mounted at 45 deg at the entrance pupil of the final optical design, as shown in Fig. 4(b). The multiplexing mirror assembly consisted of six diamond-turned aluminum mirror facets, each mounted to an invar two-axis piezo stage from nPoint (RXY3-276). The mirror facets are sized such that beam projections normal to the optical axis divided the full aperture of the optical system into equal sections, resulting in the pupil geometry described in Fig. 3(a). In order to eliminate stray sources of background light during the piezo actuation of each FoV, every side of each mirror facet terminated in knife edges, each mirror facet was designed to be slightly larger than its projected section, and the mirrors were assembled in such a way that the facets overlapped slightly with each other such that there was never a gap in the mirror assembly during sampling. Figure 5(a) shows four of six mirror facets mounted on piezo actuators being assembled. The multiplexing mirror assembly was nominally aligned to create one contiguous image spanning 90 deg×15 deg in the nominal, nonactuated configuration by shimming each piezo-mounted mirror. Care was taken to ensure that there

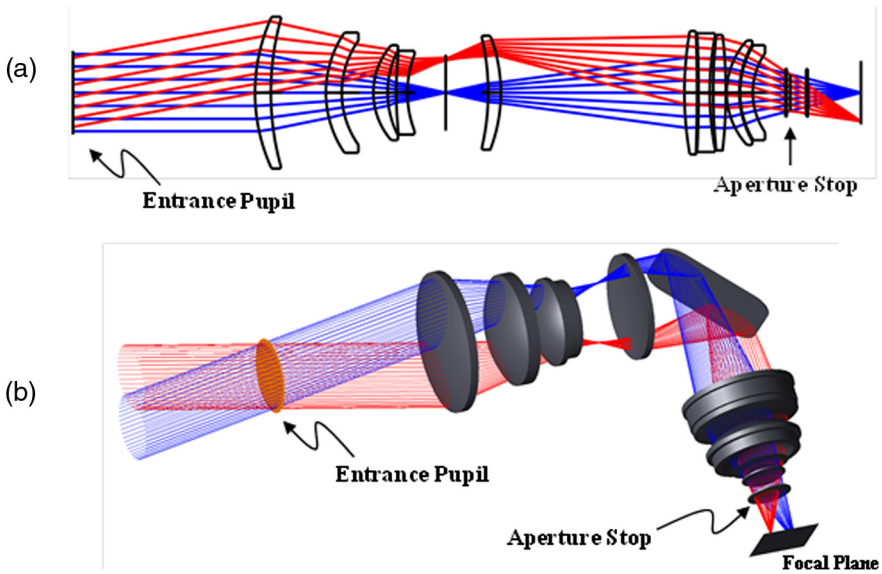

Fig. 4 (a) Unfolded prescription. (b) Isometric view showing the pupil relay and fold mirror.

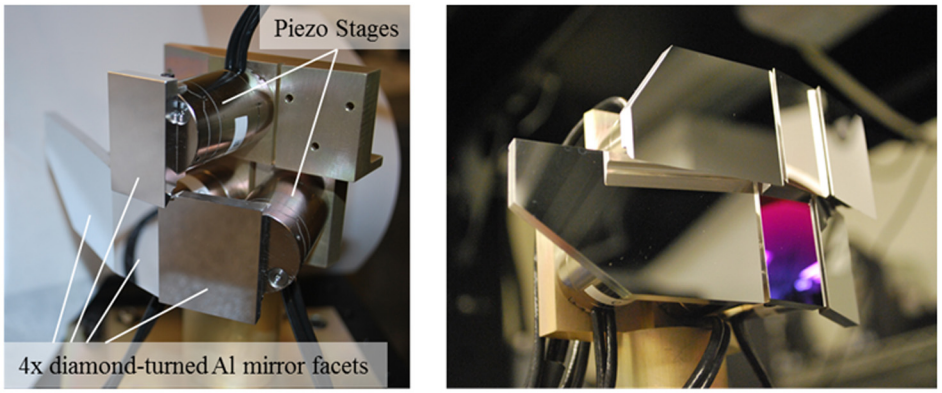

Fig. 5 (a) Multiplexing mirror assembly mid-construction; only four of six mirror facets are mounted to piezo stages and secured to the assembly base. (b) Completed six-facet multiplexing mirror assembly.

were a few pixels of overlap between the FoVs of each channel so that there is no loss of information. For thermal considerations, each of the six diamond-turned mirror facets is imbued with an interface collar in the rear that mates athermally to the invar piezo stages by way of a 0.015″ thick 566-RTV radial bond line. The completed multiplexing mirror assembly can be seen in Fig. 5(b).

The mass of each mirror facet affects the performance of the piezo actuators and its associated resonance. Effort was made to tune each piezo's control loop parameters to drive the additional mirror mass for a rapid step and settle time. These stages also provide precise positioning relative to the camera's 260 microradian I FoV with a RMS position error of 50 nanoradians. Calibration was performed for each stage to translate between the piezo stage's local coordinates to focal plane pixel coordinates. This calibration is discussed in additional detail in Ref. 11.

Additionally, tests were run to ensure that the mirror facets will not come into physical contact with each other across the full range of motion of the piezo actuators. Since image quality depends on the flatness of each mirror facet, structural, modal, and thermal analyses were performed across the entire operational range and showed that the optical performance of the multiplexing assembly is sufficiently ruggedized for stability across a temperature range of 0°C to 50°C. This enabled us to take the system to various outdoor test sites to collect pertinent data.

5 Imaging

As described in Sec. 2, by combining multiple, shift-encoded, multiplexed frames, a demultiplexed image of the scene can be created. We have developed two imaging methods, one that produces a conventional image and one that generates gradient images. Conventional image formation requires the inversion of a large sparse matrix and amplifies noise as a function of the conditioning of that matrix. Gradient image formation is computationally inexpensive since it operates in linear time, suffers from a noise increase, and can provide sufficient information for tasks, such as object detection, tracking, and recognition. Both methods produce artifacts from saturated pixels, unknown camera motion error, uncertainty in the FoV shifting, and scenes that vary over the course of the measurements. A thorough study of the impact of these various factors on image quality is a topic for future work.

5.1 *Conventional Image Formation*

In order to provide an example of the reconstruction of a conventional image, we consider the linear formulation of the imaging process, as described in Sec. 2. Figure 6 illustrates the multiplexing image transformation. In this example, integer pixel image shifts are used as the means of encoding for a two channel ($n = 2$), three pixels per FoV ($l = 3$), six pixels in the demultiplexed image ($m = 6$) multiplexed imaging system. Each of the shaded boxes in the encoding matrices

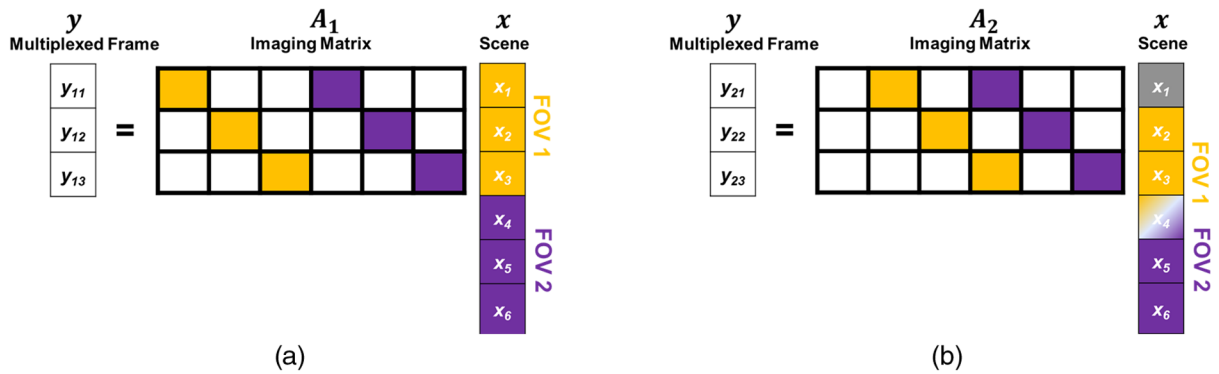

Fig. 6 Matrix formulation of multiplex measurements of two continuous FoVs with three pixels per FoV. (a) The FoVs are unshifted and (b) the first FoV is shifted to the right by one pixel.

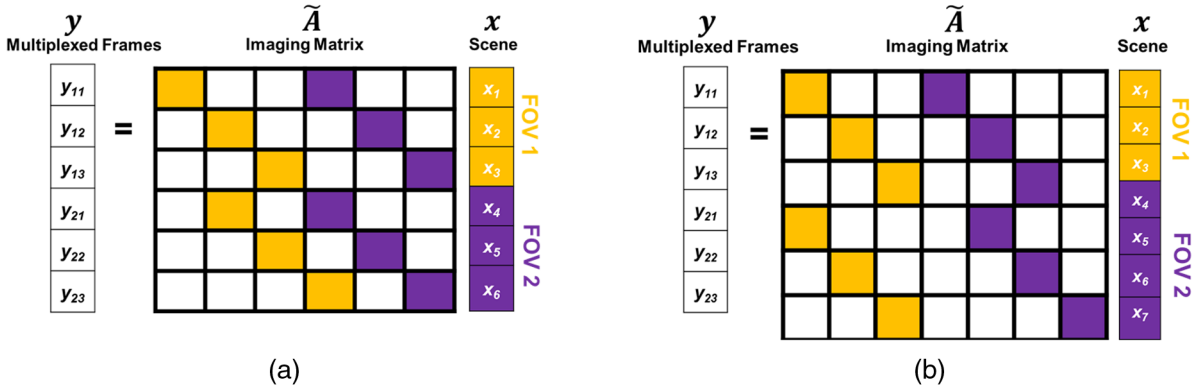

Fig. 7 Two different multiplexed imaging matrices for a two FoV three pixel/FoV. The imaging matrix in (a) is full rank where as the one shown in (b) is underdetermined.

is one and the remaining values are zeros. The first frame in Fig. 6(a) shows the linear imaging process when no shifts are applied. In Fig. 6(b), FoV 1 is shifted by one pixel to the right.

Concatenating the transformation matrices, forming $\tilde{\mathbf{A}}$, and the measurement vectors forming $\tilde{\mathbf{y}}$, as described in Sec. 2, result in a fully determined set of equations, as shown in Fig. 7(a).

It is important to note that having at least n measurements does not guarantee a fully determined transformation matrix, since some shift choices lead to a rank deficient matrix. The shift of FoV 1 to the right is considered an inward shift as it did not require increasing the size of the unknown x vector. If instead FoV 2 shifted to the right, an additional element, x_7, would have to be introduced. This would result in having fewer measurements than unknowns and the matrix would no longer be full rank, as shown in Fig. 7(b).

The $\tilde{\mathbf{A}}$ matrix is simple, composed of only ones and zeros, and is sparse. The number of nonzero elements is equal to the product of the number of pixels in a multiplexed frame, the number of FoVs, and the number of multiplexed frames. The total number of elements in $\tilde{\mathbf{A}}$ is the product of the number of pixels in a multiplexed frame, the number of multiplexed frames, and the number of pixels in the scene. Thus, the sparsity is approximately equal to 1 over the number of pixels per FoV. For the one megapixel frames of our MWIR prototype, the sparsity is 10^{-6}.

We formulate the estimation of the underlying scene as a minimum MSE problem. For small images, it may be possible to explicitly compute the solution to this problem, as shown in Eq. (4). For our six channel multiplexed imaging sensor described in Sec. 4, the $\tilde{\mathbf{A}}$ matrix contains over 10^{12} elements, since we are solving for a six megapixel scene, and thus explicitly computing the pseudoinverse is intractable due to both memory and computational constraints. However, the sparsity can be leveraged by using a sparse matrix iterative solver, such as LSQR[18] that can approximately compute minimum MSE solution. The conventional imaging results shown in Sec. 7 were obtained using the MATLAB implementation of the LSQR algorithm developed by members of Stanford's Systems Optimization Laboratory,[19] with modifications allowing it to run on a GPU. Also, for certain shifts as described in the next paragraph, portions of the scene can be solved independently from each other. Thus, the solution of a much smaller problem is required and each portion can be solved precisely and in parallel.

Selecting which set of shifts used to encode the image strongly affects image reconstruction performance. Shifts chosen at random are not guaranteed to be full rank. As described above, inward shifts are required to have a full rank matrix without requiring additional measurements. Within the subset of full rank shifts are shifts that measure regions of the image in a decoupled manner, which thus can be solved for independently of each other. One such example that we have found is performing inward shifts in only one dimension, the multiplexed dimension. For our system, since the FoVs are horizontally multiplexed, this means shifting only in the horizontal direction. This decouples the rows from each other and each row can be independently solved for in parallel. Within this reduced set of shifts, there exists a set of shifts that has the best noise performance.

5.1.1 *Imaging performance*

A demultiplexed image is a function of multiple frames, and as a result, the noise in the demultiplexed image can be less than the multiplexed per frame NEDT described in Sec. 3. In the case of viewing a uniform background scene, in which the noise is dominated by the shot noise of the background photogenerated electrons, and with a sufficiently large number of photons in each pixel, we can approximate $\tilde{\boldsymbol{\epsilon}}$ as normally distributed and uncorrelated, with identical variances σ^2. Under these conditions, we consider the expected MSE, between the true discretized version of the underlying scene and that of the reconstructed scene:

$$\text{MSE} = \frac{\sigma^2}{m} Tr(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}. \tag{6}$$

For a nonsignal amplifying $\tilde{\mathbf{A}}$ matrix (e.g., passive optical systems) whose entries are bounded in absolute value by 1, the $Tr(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}$ is ≥ 1. If $\tilde{\mathbf{A}}$ is chosen to be the Walsh–Hadamard matrix, this lower bound is achieved, minimizing the MSE.[20] The resulting MSE in the estimated image is $n\sigma^2$. In this case, the square root of the MSE is $\sqrt{n}$ larger than the NEDT in the unmultiplexed case, when comparing a full narrow FoV aperture to a multiplexed aperture divided into n regions, as discussed in Sec. 3. The two cases are not directly comparable since the reconstruction is over a wider scene and requires n times longer integration (where $p = n$ for this square matrix example). This result is in contrast to the n times larger NEDT in a single frame of the multiplexed image shown in Sec. 3 and suggests that for computational

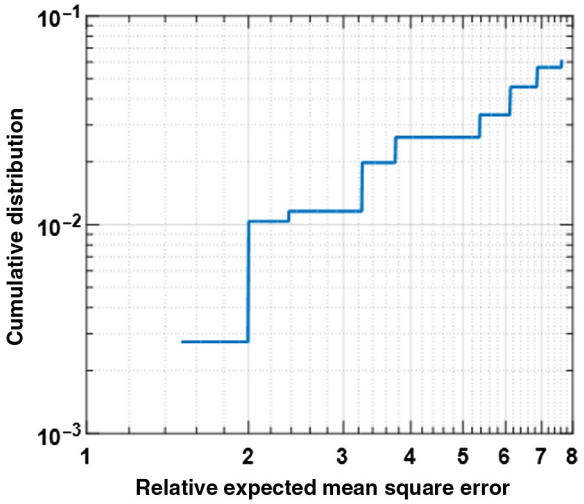

Fig. 8 For a multiplexed imaging simulation, a cumulative distribution, as a fraction of all possible shifts, of the relative expected MSE for shifts, which result in a full rank $\mathbf{A}$ matrix. The error performance is relative to that of the optimal Walsh–Hadamard matrix, and the percentage of shifts that lead to an invertible matrix is 6% of all possible shifts.

imagers, a reconstructed image NEDT would be a useful concept for discussing the scaling of system noise. This analysis changes with assumptions about scene sparsity or the dominant sources of noise. In practice, a dense Walsh–Hadamard matrix is difficult to realize due to multiple challenges, most prominently, the presence of dense negative and positive entries. Shift encoded multiplexed imagers realize a sparse multiplexing matrix. This has benefits in terms of noise for the shot-limited case in the multiplexed frame, since fewer variables are combined in each measurement. However, an optimal choice of shifts to minimize the $Tr(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}$ term in the MSE equation above is a topic that requires additional research. In the next subsection, we empirically investigate optimal shift selection for a restricted multiplexed imaging problem.

5.1.2 *Shift selection*

To empirically explore the impact of shift selection on image reconstruction error, we exhaustively computed the MSE, for all possible horizontal integer shifts for a two channel, eight pixels per multiplexed row imaging system. Since the rows are decoupled due to shift selection, this results in a 16×16 forward matrix for which there exists a Walsh–Hadamard Matrix, which we can compare performance against. The maximum shift was limited to half of the FoV, or four pixels. Thus, each channel can be shifted by d pixels, in this example, one of nine values (-4 to $+4$), in each of the two frames. The total numbers of shifts, z, are $d^{p\times n}$ or $9^{(2\times 2)} = 6561$. Of these, only 404 (6%) of them were found to be full rank. The expected MSE for each set of shifts was compared to that of a Walsh–Hadamard matrix and the cumulative distribution of the relative expected error is shown in Fig. 8. Only 18 (0.27%) of the shift sequences attain the minimum error, which is 1.5 times that of a Walsh–Hadamard Matrix. Although shift encoding did not achieve the optimal bound, the result is encouraging given the resulting matrix is restricted compared to a Walsh–Hadamard matrix as mentioned in the previous section.

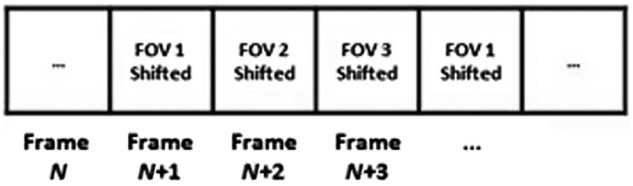

Fig. 9 Frame by frame shift sequence for gradient image generation.

This exhaustive search approach does not scale well for a larger number of multiplexed FoVs due to the exponential scaling of the size of the search space. These empirical results indicate the strong dependence of reconstruction error on shift selection and suggest that a random shift selection approach is unlikely to approach optimal performance. However, this empirical investigation was restricted to the case in which the number of measurements is equal to that of the number of multiplexed FoVs and the shift is along the multiplexing direction. We expect that not restricting the shifts to the multiplexed dimension should enable more favorable noise properties, however, the rows would no longer be independent and thus, the image formation process could not be decoupled into multiple subproblems. Understanding what types of shifts result in favorable noise properties is a focus of ongoing work.

5.2 *Gradient Images*

To generate gradient images, it is necessary to shift only one FoV per frame, as shown in Fig. 9. By looking at the difference between consecutive frames, the signals from the non-shifted FoVs are common and cancel out and only an edge, or gradient, from the shifted FoV remains. A gradient of a FoV is generated per frame and so to generate a gradient image for all n channels, n frames are required. Gradient image formation is computationally inexpensive as it consists primarily of a linear time frame difference. Gradient images convey much of the useful information in the image. They are readily human interpretable but could also be used as an input for automated processing. For example, one could compute stable features for vision-based navigation, or perform object detection, classification, and tracking. Alternatively, some of these automated processing tasks can be carried out directly in the multiplexed frame, as discussed in Sec. 6 below.

6 Detection and Localization

Demultiplexing by solving for the underlying scene may produce information in excess of the requirements of tasks, such as object detection and tracking. Furthermore, image formation can introduce additional computational complexity and noise. An alternative approach is to detect objects in the multiplexed frame and then localize only those objects. Uttam et al.[3] explore this concept for a multiplexed imaging sensor, though that work does not include the division of aperture multiplexing architecture we discuss in this paper.

Due to the linearity and local geometry preservation of the multiplexing operation, many traditional detection algorithms, such as background subtraction, spatial–temporal filtering, image segmentation, and hotspot detection, can be applied to multiplexed frames without modification. By performing detection directly in the multiplexed frame, one does not incur the additional noise penalty introduced in the

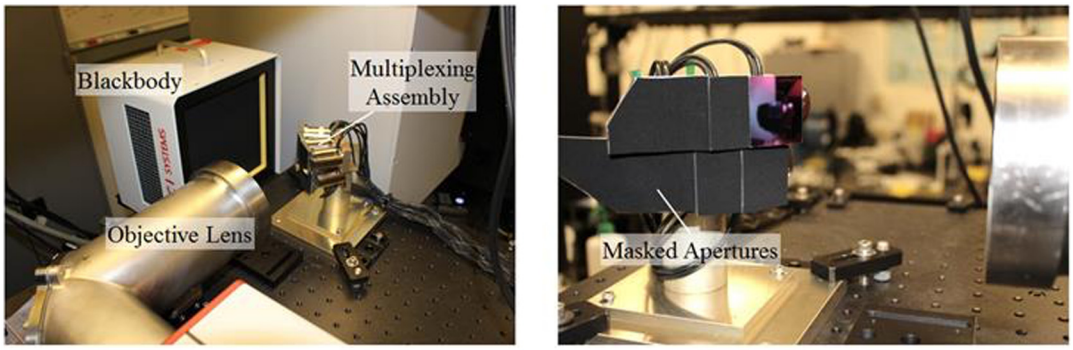

Fig. 10 (a) Flood illumination of the system with an extended area blackbody. (b) Mirror masking to measure individual channels.

image reconstruction process. After detection, there is a residual uncertainty in the object location due to the FoV ambiguity inherent to the multiplexing design. Thus, an additional localization step must occur.

In order to localize, the measurements must uniquely encode each channel. Many encoding schemes exist, each balancing various performance metrics. For example, in prior work, point spread function encoding, rotation encoding, and shutter encoding were used as a means to detect and localize a star field.[1] The subsequent discussion compares these techniques to the shift-based encoding discussed in this paper and used in the MWIR multiplexed imager prototype described in Sec. 4. Unlike shuttered encoding, FoV shifting allows for persistent observation. Also, no *a priori* knowledge of the object motion is required as is the case with rotation encoding. A disadvantage of the time-encoded FoV shifting approach as opposed to the point spread function encoding approach is the inability to simultaneously detect and localize within a single frame.

A shutter or FoV-shift encoding multiplexed imager has the capability to decouple the detection and localization operations. As a result, detection performance does not suffer from the SNR reduction that would occur when encoding the signal for localization. For example, the shift-encoding system could operate in a static nonshifting mode to detect, and similarly, a shuttering system could operate with all shutters open. By contrast, a point spread function encoding scheme that spreads light over multiple pixels suffers from additional background and detector noise. For an IR sensor, like the prototype described in Sec. 4, the design of shutters that do not add background noise themselves is a significant engineering effort.

In a shift-encoded multiplexed imaging system, localization is informed by sequential measurements, in which each FoV has shifted by a controlled amount. Though the initial postdetection distribution over the object location is multimodal, object motion that is correlated with a known channel motion can be used to estimate the true object location. Similarly, for object attenuation that is correlated with shuttering, a particular channel can be used to identify true object location. In a point spread function encoded system, a matched filter-based approach, with one filter tuned for each channel, can be used to localize.[8]

7 Results

7.1 *Optical Performance and Calibration Procedure*

Testing was performed to verify performance predictions and collect calibration data required for the image processing algorithms. The following three categories of tests were performed: (1) image quality (MTF), (2) SNR and relative image irradiance (SiTF, NEDT, two-point gain correction). Test equipment consisted of a differential black body source, a 12″ diameter collimator, and a collection of pinhole and knife-edge targets.

Radiometric data were collected by flood illuminating the system with an extended area black body shown in Fig. 10. Each channel was measured individually by masking the

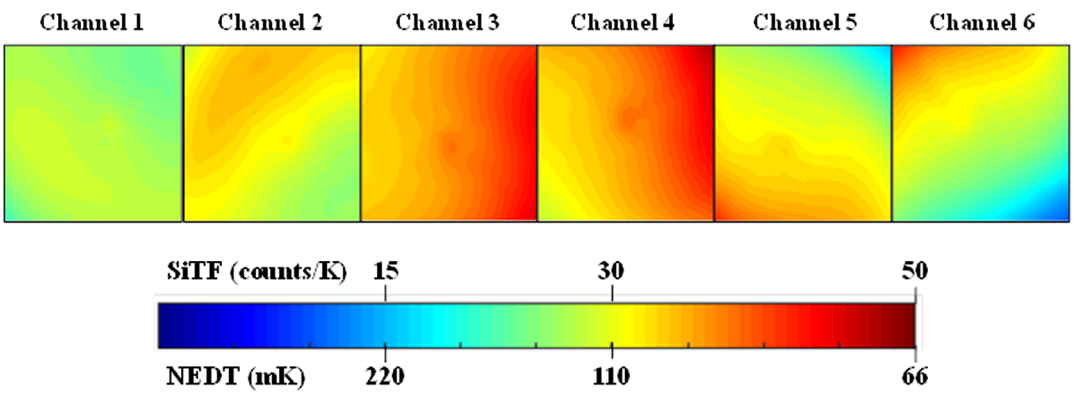

Fig. 11 SiTF and single-frame NEDT for each imaging channel with a 2-ms integration time.

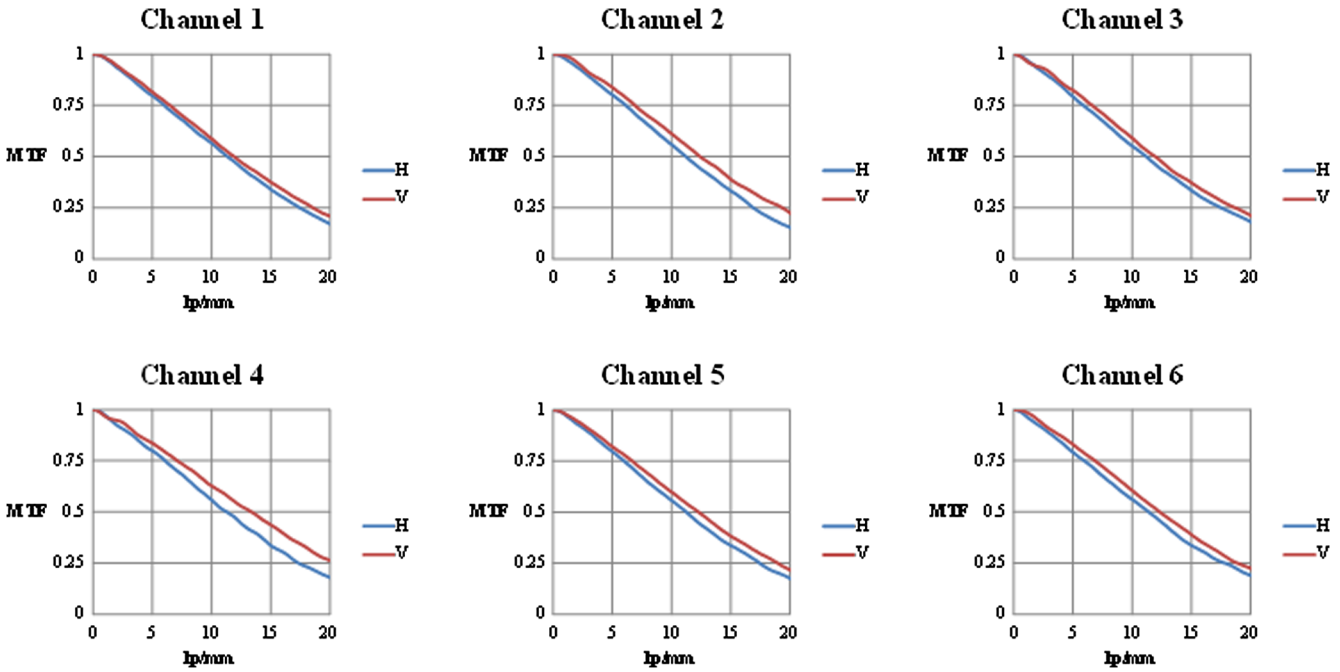

Fig. 12 MTF test results plotted to the Nyquist frequency of the focal plane (20 lp/mm).

other channels with a diffuse high emissivity material, such that they contributed only uniform ambient background radiation. The integration time was set to 2 ms to fill the pixel well to ~50%.

Two-dimensional (2-D) SiTF maps were measured by collecting 128 frames at black body temperatures of 20°C and 40°C. The frame average was computed at each temperature and SiTF was computed by taking the per-pixel difference and scaling by the 20-deg differential. These data are shown in Fig. 11. Average SiTF values for each channel ranged between 26.7 and 36.7 counts/°C. The sum of the per-channel SiTF values (188.1 counts/°C) was roughly equivalent to an SiTF measurement of the full aperture objective lens (186.3 counts/°C) as predicted.

Nonuniformity correction of the demultiplexed image was performed by using the 2-D SiTF maps as gain correction terms in the image reconstruction algorithm. Specifically, the SiTF map was normalized and inserted as a scaling matrix in the forward model. This provided a supplemental per-channel gain correction on top of a conventional full aperture two-point NUC applied at the camera hardware level. One-point offset correction was performed prior to data collection by "externally NUCing," the raw multiplexed image with a uniform diffuse surface placed outside of the lens.

NEDT was computed from an additional set of 128 images collected with a blackbody temperature of 30°C. Noise statistics in units of digital counts were computed from the image cube and then scaled by the SiTF to obtain noise in temperature units (mK). The NEDT of the full aperture objective lens was measured to be 18 mK. The theory described in Sec. 3 predicts that the single-frame NEDT of each channel should be approximately six times higher than the objective lens. Per-channel NEDT measurements ranged between 89 and 123 mK with an average value across the full FoV of 106 mK. This average value was 2% lower than the prediction and may be partially explained by the measurement process. In the multiplexed system, measurement noise statistics were computed by observing a 30°C blackbody in only one channel (one-sixth of the aperture area), but the remaining five-sixths of the aperture area observed a ~20°C blackbody via the masked channels. This would produce a slightly lower background shot noise than in the 30 deg measurement of the full aperture objective lens. NEDT is shown in Table 2.

Image quality was determined by MTF testing. Images of vertical and horizontal back-illuminated edge targets were captured at the center of the FoV of each channel. The black body was set to a 50 K temperature differential to place the span of the edge ~25 dB above the background noise floor.

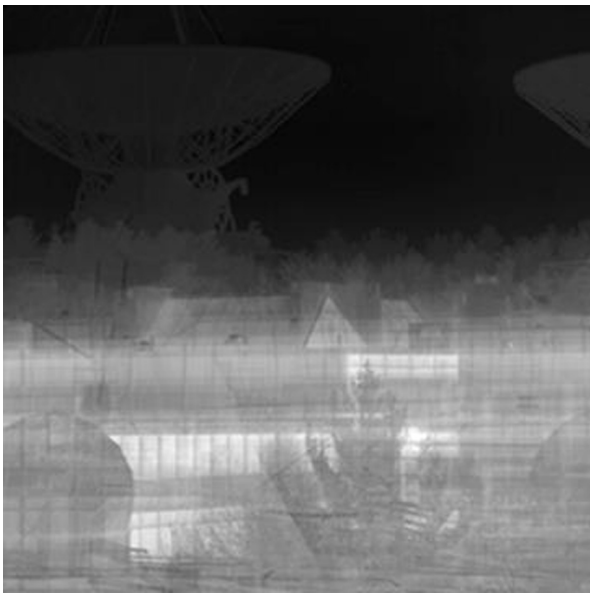

Fig. 13 A one megapixel multiplexed image with a FoV of 15.2 × 15.2 deg.

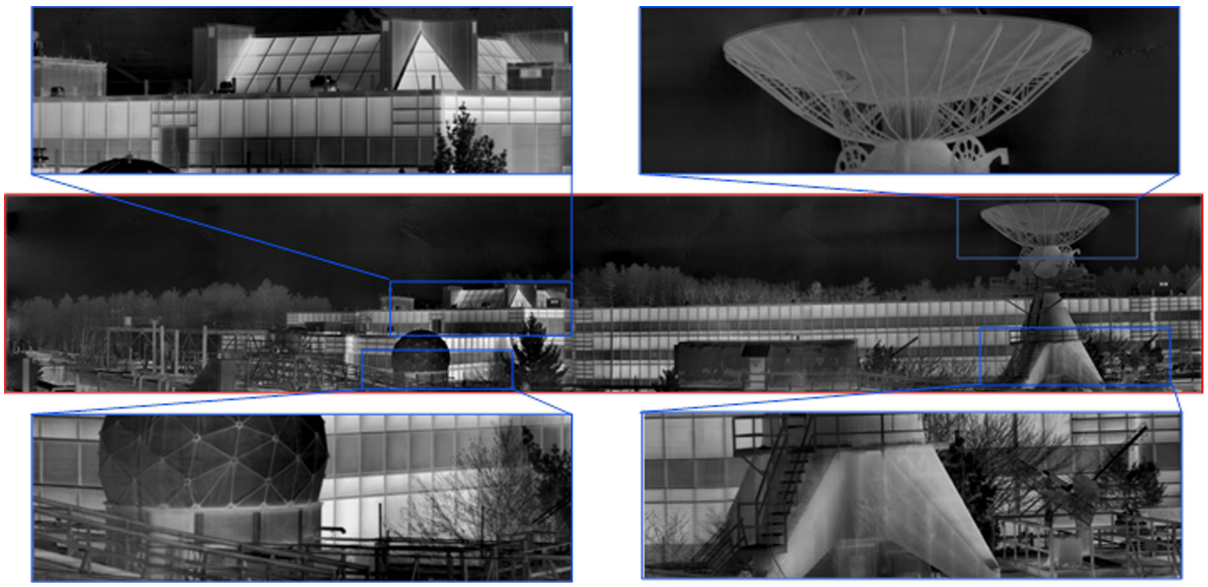

Fig. 14 A six megapixel reconstructed image with a 90×15.2 deg FoV.

MTF was computed using the ISO12233 tilted edge method. Results are plotted in Fig. 12. The system was shown to retain positive MTF contrast out to beyond the Nyquist sampling frequency of the detector, which implies that the channel images are still undersampled after pupil division and will therefore allow for a full factor of six FoV gain without resolution loss (a sixfold pixel gain in the digital image).

7.2 *Imaging Results*

Imagery and motion videos were collected using the prototype sensor. Figure 13 shows a single frame of the six-layered multiplexed image. To demultiplex the image, a sequence of eight images was collected in which each image layer was shifted uniquely by a magnitude of ~10 pixels. The shifts chosen for these results were 2-D, with the shift direction in each FoV constrained to ensure distinct shift directions across FoVs for every frame. The decoupled shifts described in Sec. 5.1 were not utilized in part due to FoV misalignment in the prototype. Experimental verification of shift optimality is a topic for future work. Using 2-D shifts, multiplexed frames were processed to reconstruct the image shown in Fig. 14.

To collect motion imagery, the camera was windowed to a 1024×512 region of interest. This was done solely to allow the camera to read out at a faster frame rate (120 Hz) and was not a general limitation. By increasing the frame rate, the total time required to collect the eight samples was reduced,

Fig. 15 A single frame from a 90×7.6 deg demultiplexed conventional video. Only the center 1024×512 region of the camera was read out in order to increase the sampling rate.

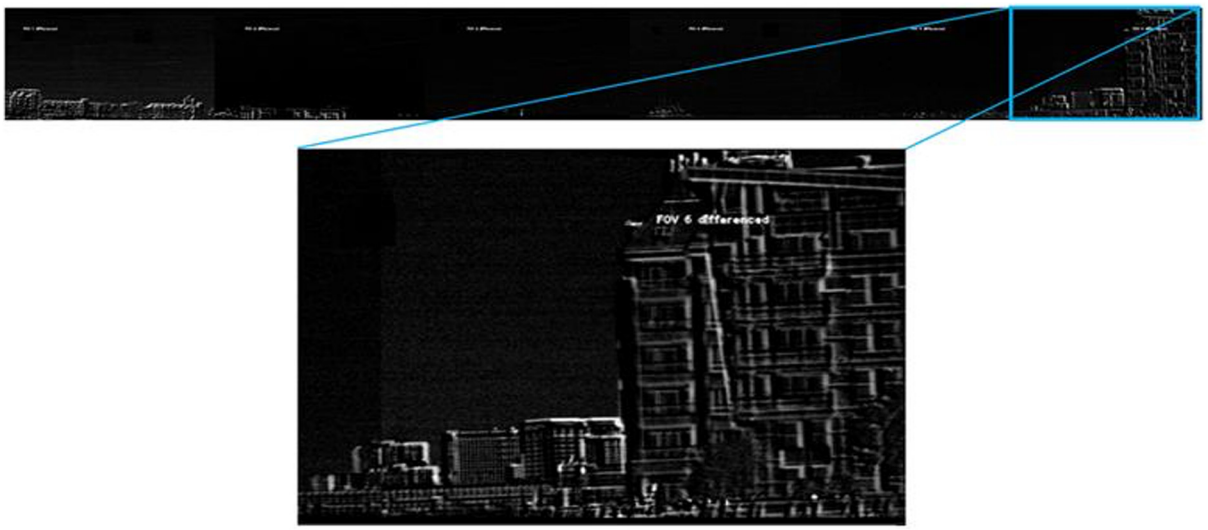

Fig. 16 A single frame from a 90×7.6 deg demultiplexed gradient video. Only the center 1024×512 region of the camera was read out in order to increase the sampling rate.

which helped to suppress motion-related image artifacts. Demultiplexed images were reconstructed on a rolling basis at 120 Hz with each reconstructed image containing data from the prior eight frames. Figures 15 and 16 show a single frame of a conventional image video and demultiplexed gradient images, respectively. The conventional video shows details, such as the reflections off the water, flags waving, and people walking on the dock. The gradient video demonstrates that much of the contextual information is retained such as the shape of the buildings and windows as well as the fence above the dock.

8 Conclusion

We have demonstrated a MWIR optically multiplexed imaging sensor and discussed some of the optical, mechanical, and system design choices for that sensor. We have found that the selection of shifts used for encoding is critical in terms of the noise in the reconstructed demultiplexed image. The optimal selection of such encoding shifts is the focus of ongoing research. Additionally, we discussed the computationally inexpensive technique of forming gradient images, as well as localization techniques, which recover only objects of interest in the scene. This work provides a foundation for future application-specific optical system designs that tradeoff performance for FoV in a previously inaccessible regime.

Acknowledgments

This work is sponsored by the Department of the Air Force under Air Force Contract #FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government.

References

1. M. D. Stenner, P. Shankar, and M. A. Neifeld, "Wide-field feature-specific imaging," in *Frontiers in Optics 2007, OSA Technical Digest*, Optical Society of America (2007), paper FMJ2.
2. R. F. Marcia et al., "Superimposed video disambiguation for increased field of view," *Opt. Express* **16**, 16352 (2008).
3. S. Uttam et al., "Optically multiplexed imaging with superposition space tracking," *Opt. Express* **17**, 1691 (2009).
4. V. Treeaporn, A. Ashok, and M. A. Neifeld, "Increased field of view through optical multiplexing," *Opt. Express* **18**, 22432 (2010).
5. R. Horisaki and J. Tanida, "Multi-channel data acquisition using multiplexed imaging with spatial encoding," *Opt. Express* **18**, 23041 (2010).
6. C. Y. Chen, T. T. Yang, and W. S. Sun, "Optics system design applying a micro-prism array of a single lens stereo image pair," *Opt. Express* **16**, 15495 (2008).
7. A. Mahalanobis et al., "Off-axis sparse aperture imaging using phase optimization techniques for application in wide-area imaging systems," *Appl. Opt.* **48**, 5212 (2009).
8. R. H. Shepard et al., "Design architectures for optically multiplexed imaging," *Opt. Express* **23**, 31419–31436 (2015).
9. R. Gupta et al., "Compressive sensing with local geometric features," in *Proc. of the 27th Annual ACM Symp. on Computational Geometry*, pp. 87–98, ACM (2011).
10. Y. Rachlin et al., "Dynamic optically multiplexed imaging," *Proc. SPIE* **9600**, 960003 (2015).
11. R. H. Shepard et al., "Design and calibration of a wide field of view MWIR optically multiplexed imaging system," *Proc. SPIE* **9948**, 994480T (2016).
12. E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?," *IEEE Trans. Inf. Theory* **52**, 5406 (2006).
13. D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory* **52**, 1289 (2006).
14. G. C. Holst, *Testing and Evaluation of Infrared Imaging Systems*, JDC Publishing, SPIE Press, Bellingham, Washington (2008).
15. "IRC910 - IR Cameras," IR Cameras, http://www.ircameras.com (1 March 2016).
16. J. Sasian, "Interpretation of pupil aberrations in imaging systems," *Proc. SPIE* **6342**, 634208 (2006).
17. D. Shafer, "Aberration theory and the meaning of life," *Proc. SPIE* **0554**, 25 (1985).
18. C. C. Paige and M. A. Saunders, "LSQR: an algorithm for sparse linear equations and sparse least squares," *ACM Trans. Math. Software* **8**(1), 43–71 (1982).
19. C. C. Paige and M. A. Saunders, "LSQR: sparse equations and least squares," http://web.stanford.edu/group/SOL/software/lsqr/ (1 July 2015).
20. M. Harwit and N. J. A. Sloane, *Hadamard Transform Optics* Academic Press, New York (1979).

Vinay Shah is a staff member at MIT Lincoln Laboratory. He received his BS in electrical engineering from Rensselaer Polytechnic Institute and his MS in electrical engineering from Syracuse University in 2005 and 2008, respectively. His research interests include computational imaging, radar signal processing, and interferometric imaging.

Yaron Rachlin received his PhD in electrical and computer engineering from Carnegie Mellon University. He is a member of the Technical Staff in the Advanced Capabilities and Technologies Group at MIT Lincoln Laboratory, where he leads the development of innovative

algorithms and sensors for infrared imaging, satellite-based Earth observation, and spectral imaging systems. His research interests include computational imaging, information theory, and machine learning.

R. Hamilton Shepard received his PhD in optical sciences from the University of Arizona. His research interests include innovative optical system design, computational imaging, and infrared systems. He is currently at Waymo serving as the optical engineering lead for lidar systems on self-driving cars.

Tina Shih is an assistant group leader of the Rapid Prototyping Group at MIT Lincoln Laboratory. She received her BS in mechanical engineering from MIT, studied abroad at Cambridge University, and received her PhD in applied physics from Harvard University. Her optics research has led to wonderful adventures: volume holography with gummy bears, summers of spectroscopy collaborations in Germany, and gained countless friends (and a thesis) through femtosecond science.