# Multi-aperture imaging for flat high-resolution cameras

GUILLEM CARLES[1] AND ANDREW R. HARVEY[1,*]

[1] School of Physics and Astronomy, University of Glasgow, Glasgow, G12 8QQ, UK
[*] Corresponding author: Andy.Harvey@glasgow.ac.uk

Compiled January 5, 2021

**Imaging with high angular resolution requires large apertures and long focal lengths. This has prevented the integration of telephoto lenses into thin devices such as modern mobile phones. We report a camera module employing multiple rotated rectangular apertures and folding of the optical system into the plane of the camera, enabling an order-of-magnitude reduction in depth compared to traditional telephoto lenses. Multiple images are fused in the frequency domain to yield a single high-resolution image equivalent to that yielded by a single circular aperture. The diameter of this equivalent aperture may be several times wider than the depth of the camera module. We propose two architectures and present illustrative optical designs to demonstrate the concept. Simulations of ray-traced image acquisition and computational image reconstruction demonstrates the potential for high-quality, high-resolution imaging from thin, flat lens modules.** 

http://dx.doi.org/10.1364/ao.XX.XXXXXX

The phenomenal development of camera technology within mobile phones has enabled a ten-fold enhancement in pixel count over the past decade – even as the thickness of the cameras has reduced to fit into thinner mobile phones. The recent introduction of multiple cameras within a single phone has enabled multiple focal lengths and fields-of-view (FoV) to be employed simultaneously. Image processing can then be used to provide users with a continuous variation of the FoV between the fields of the fixed focal lengths, emulating the optical zoom provided in traditional cameras. However, the incorporation of a longer focal length lens with a narrow FoV into a thin camera module has proved elusive.

The track length and hence depth of a narrow-FoV, high-resolution camera module is associated with the need for a wide optical aperture. A typical 10-megapixel miniature camera module with a FoV of 40° will have an aperture of about 1.5 mm, a diffraction-limited angular resolution of about 0.45 mrad and a focal length of about 3 mm. A reduction in the FoV requires a proportionate increase in the diameter and focal length of the lens. A ten-fold increase in magnification consequently yields a camera that is significantly greater than the thickness of the device. Because mobile-phone camera modules are required to fit within a thickness, their focal length and therefore their angular resolution are limited[1]. Telephoto optical designs aim at fitting longer focal lengths within a fixed total track length to maximize magnification, but the severe constraint on thickness means that optical zooms greater than about three are simply not possible. Optical zooming requires changing the focal length of the camera (and its aperture, in order to preserve F-number). This is typically achieved by precise mechanical movement of lens elements. However, porting traditional techniques onto mobile phone logistics requires 'camera bumps' with volumes that are not accepted by the thickness-driven market (an example attempt is the *Samsung Galaxy K Zoom*, with 10× zoom). As an alternative, folded strategies enable optical zooms of up to 5×, by folding the optics train along the device plane, which enables accommodating larger focal lengths (the *Hoya Asus Zenfone Zoom* and the *Light* camera are early examples). A further step to fit even longer focal lengths is to implement multiple foldings (one of the cameras of the recent *Huawei P40 Pro+*, with five foldings, is a remarkable example). These approaches allow for longer focal lengths because the optical track length is no longer limited by the device thickness. However, the diameter of the aperture becomes limited by the device thickness, setting a limit above which further focal length increase implies a larger F-number, resulting in increased oversampling (and reduced light sensitivity) but not in higher angular resolution.

We describe a technique that enables long focal length lenses to be incorporated into a module that can not only be more than an order-of-magnitude thinner than the focal length, but also thinner than the effective aperture of the system – enabling 10× or higher zoom factors, and an angular resolution not possible with conventional or conventionally-folded approaches.

The recent and dominant trend in the field is to employ and integrate several apertures with different focal lengths and computationally fuse the images to create an effective optical zoom. Of course, this approach is limited to the angular resolution of the aperture with highest focal length (highest angular resolution), and so there is a strong need for systems capable of long focal length imaging with acceptable F-numbers.

Besides efforts in the optical design process[1], previous works to achieve imaging with compact devices include the use of planar optics[2, 3], multi-aperture imaging[4–6] and annular folded optics[7–9]. Imaging with planar optics such as using metasurfaces is currently attracting significant interest and promises interesting further developments, enabling flat optical components for imaging. In its current state however, developments focus on flattening the optical surfaces and so the length of the imaging system is still comparable to the lens focal

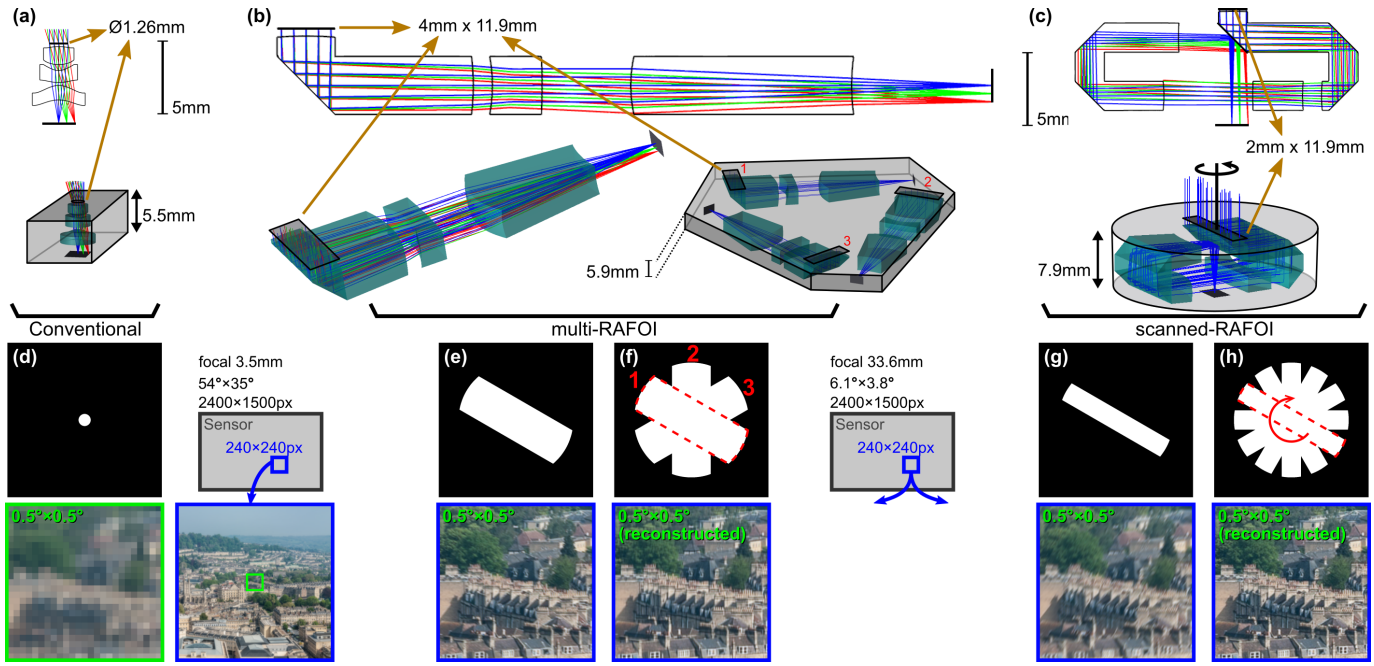


**Fig. 1.** Comparison of a conventional camera module (a) and RAFOI camera modules (b,c) (upper, ray-traced plots are drawn to scale; scale bar is 5 mm). RAFOI modules have a ten-fold longer focal length, but a similar axial thickness. Pupil apertures are depicted for conventional (d), single RAFOI (e,g) and multi- or scanned-RAFOI (f,h); simulation of image formation in the lower images of (d-h) employed rigorous ray-tracing, and (f,h) show reconstructed high-resolution images.

length. Conversely, multi-aperture imaging has been successful at reducing optical track-length of imaging systems, but only through mitigation of the aliasing imposed by the pixel pitch of available detector arrays, and so it also does not directly tackle the issue of achieving high focal length to system depth ratios. Annular folded lenses, on the other hand, are an example of thin lenses with potential for high telephoto imaging, but their small aperture area reduces their étendue and they pose significant manufacturing and tolerancing challenges (including chromatic aberrations) that prevent high-quality imaging.

Here we propose a computational-imaging concept that achieves, for the first time, an order-of-magnitude optical zoom increment whilst satisfying the thickness constraint and providing high-quality imaging performance. The concept is based on rectangular-aperture, folded-optics imaging (RAFOI). It exploits the fact that for a rectangular aperture, the optics can be folded transversely (one or multiple times) in the direction perpendicular to the long side of the aperture to fit within a thin space envelope. Of course, the rectangular aperture does not transmit the full spatial-frequency bandwidth of an equivalent full circular aperture. To solve this, RAFOI uses multiple images acquired at different orientations to provide a diversity of high-angular-resolution directions. Post-detection image reconstruction can then be used to combine all images and reconstruct a single high-resolution image comparable to that produced by a high-magnification, full-aperture system.

The concept is illustrated in Fig. 1, where cameras with complementary fields of view are depicted: a traditional short focal-length lens ($f = 3.5$ mm) with a small aperture in (a) and two long focal-length RAFOI designs ($f = 33.6$ mm) in (b) and (c). The corresponding apertures and simulated image acquisitions are depicted in (d), (e,f) and (g,h) respectively. The traditional short focal-length and long-focal length, RAFOI lenses may be used as the basis for emulation of optical zooming. The angular resolution of the former camera is fundamentally limited to 0.031° by the diffraction limit of the 1.26 mm diameter aperture, while the 2400×1500 pixel, 1.5 µm pitch sensors provides a FoV of 54°×35°. In (d) we show a 240 × 240 region of the simulated recorded image (right), and a 0.5°×0.5° area of this image (left). The RAFOI cameras depicted in (b) and (c), provide a complementary narrow FoV and higher diffraction-limited angular resolution of 0.0032° due to the larger apertures and longer focal length. The rectangular apertures of the RAFOI designs enable folding of the optics. Fusion of images from multiple orientations enables an image to be synthesized, which has the near-isotropic high-resolution of near-circular apertures as shown in (f,h). A 120×120 pixel crop of the simulated recorded images and reconstructed images are shown in the lower images of (e,g) and (f,h) respectively.

To synthesize full-aperture, high-resolution imaging, several rectangular-aperture images are required with different orientations. One has the design choice of using thinner rectangular apertures at the expense of requiring more orientations to cover all the full frequency space equivalent to a full circular aperture, or to use fewer but wider apertures. In this Letter, two exemplar architectures are proposed, and for each, an illustrative optical design is used to perform simulation of image formation (using rigorous ray-tracing of realistic lens modules) and image processing is employed to reconstruct a high-quality image.

The first example, multi-RAFOI, uses replication of $n$ identical systems. An $n = 3$ system is illustrated in Fig. 1(e,f) and ray-traced and schematics are shown in (b). It is capable of reconstructing an image of equivalent resolution of a F/2.8 lens with 33.6 mm focal length, with a camera thickness less than 6 mm and an effective 11.9 mm aperture. This ten-fold increase in focal length, provides a 10× optical zoom. Such a focal length to system depth ratio, enabled by the proposed computational approach, has not been possible before. Although this approach

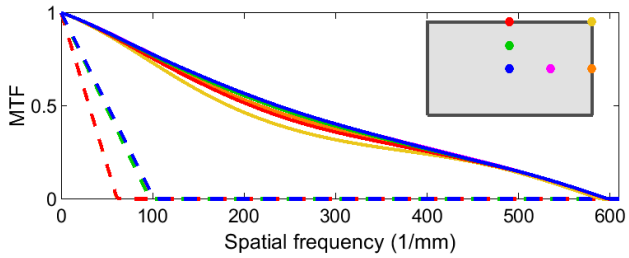


**Fig. 2.** MTF plots for various field positions (color-coded as per the top-right plot of the sensor plane) for the multi-RAFOI example (the scanned-RAFOI example has similar MTFs in the long aperture axis). Continuous (broken) lines correspond to MTFs associated with the long (short) axis of the aperture.

can implement long focal length imaging in a thin device, the requirement for system replication requires a significant volume, although shaped in a flat form.

Our RAFOI optical designs in Fig. 1(b,c) were based on the traditional optical design employing three aspheric singlets shown in Fig. 1(a). They were up-scaled to increase focal length while preserving the sensor size, modified with rectangular apertures and re-optimized for imaging performance. This illustrative design is sufficient for demonstration of the technique, although further optimization would improve system size, weight and cost, and reduce the thickness of the lens elements.

The second example, scanned-RAFOI, is based on the same concept but does not require multiple apertures and detectors, and instead uses time-sequential rotation of a single lens system with respect to the optical axis to capture $n$ images on a stationary detector. In this case, the rectangular aperture is made even thinner to enable the required additional folding of the optics. In this example, folding based on six mirrors is used as shown in Fig. 1(c). Besides the additional folding and thinner apertures, the optical prescription is however identical to the multi-RAFOI example, and so focal length is 33.6 mm. This design ensures that the aperture optical axis is coincident with the center of the sensor. The system scans an equivalent full aperture in time sequence. Importantly, rotation does not blur the image acquisition provided the rotation axis coincides with the optical axis. This scanned-RAFOI architecture has the advantage of providing a very compact system. In this example, the system fits in a cylinder of 24 mm in diameter by 8 mm height. Time scanning introduces a delay before reconstruction of the output image, but note that this example produces a high-quality image from only 6 acquisitions. Frame-rate video is possible, although delayed due to the scanning, provided that the mechanical rotation of the lens assembly is implemented at $1/n$ of the frame rate.

Interestingly, scanned-RAFOI can potentially be implemented using a single acquisition. In this mode, the rotation speed should be synchronized to one revolution per exposure time, such that the spatial-frequencies of all rotations would be integrated. The resultant averaged MTF would then be less than for a diffraction-limited full aperture. Post-detection image processing could then be used to restore the image to a near-diffraction-limited MTF, although some modest noise amplification would be expected. This has some similarity to the use of through-focus scanning for increased depth of field[10] that was reported at the dawn of computational imaging.

The simulation of image formation is as follows. Point-spread functions (PSF) were calculated using raytracing – the optical

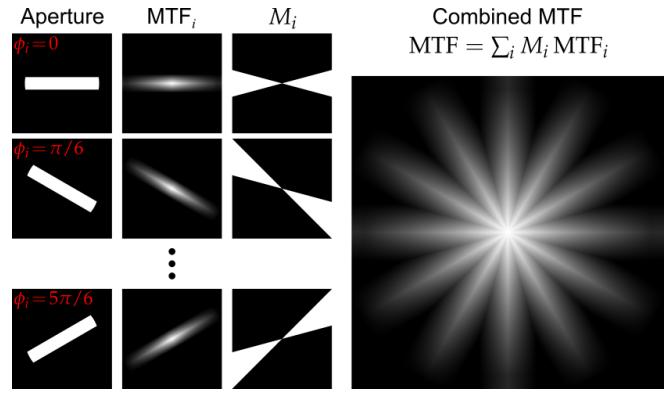


**Fig. 3.** Illustration of the recovered MTF. Images on the left show the aperture orientations, their corresponding MTF and combination mask. An effective, combined, system MTF can be calculated and is shown in the right.

performance is close to the diffraction limit, as can be seen in the modulation-transfer function (MTF) plots in Fig. 2. The PSFs are approximately spatially invariant and are elliptical in accordance with the diffraction limits of the rectangular apertures. The image formed at the sensor plane was simply calculated by convolution of the PSFs for each color channel with the simulated scene. The rectangular apertures produce some vignetting of the off-axis fields in the transverse direction of the apertures. Using a sensor of 3.6mm×2.25mm, these systems have a full FoV of 6.1°×3.8°. Raytraced simulations of the multi-RAFOI example predict this effect to be almost negligible, only present at fields above 1.5° (i.e. the outer 10% of the FoV) and increasing to a vignetting factor of only 95% at the edge of the vertical FoV at 1.9°. For the scanned-RAFOI example, because the apertures are thinner, vignetting starts affecting at a smaller FoV of 0.75° and increases to a vignetting factor of 66% at 1.9° (this can be appreciated in the ray-traced plot in Fig. 1(c) as the off-axis fields are cropped by intermediate apertures).

Image recovery involves simply constructing an image with a spatial-frequency spectrum that contains the union of the spectra of all the rotated images. To do this we define the combination masks for sub-system $i$ in frequency space as,

$$M_i(u,v) = \begin{cases} 1 & \text{if } D_i(u,v) < \frac{\pi}{2n} \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(1)}$$

where $n$ is the number of rectangular apertures, and $D_i(u,v)$ is the angular distance between the direction $\theta$ of the spatial-frequency component $(u,v)$ and the direction of the long axis of the rectangular aperture $\phi_i$, computed as,

$$D_i(u,v) = \left| \text{mod}\left(\theta(u,v) - \phi_i + \frac{\pi}{4}, \frac{\pi}{2}\right) - \frac{\pi}{4} \right| \quad \textbf{(2)}$$

where mod is the modulo operator (remainder after division).

The reconstructed image is then calculated by,

$$r = \mathcal{F}^{-1}\left\{ \sum_i M_i \mathcal{F}\{g_i\} \right\} \quad \textbf{(3)}$$

where $\mathcal{F}$ is the Fourier-transform operator and $g_i$ are the individual images acquired. Note that taking the weighted sum of all spatial frequencies, rather than the union, would enable some enhancement in signal-to-noise ratio in reconstructed images but this is not important for these simulations.

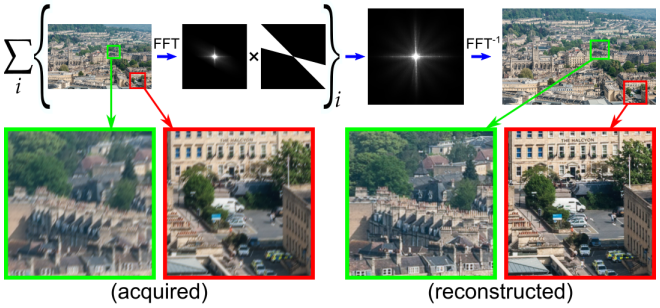


**Fig. 4.** Illustration of image recovery. The top row sequence illustrates how the acquired images of are combined in the Fourier domain using the combination masks, as Eq. (3). Lower images are close-up views to show the increase in resolution through the recovery process.

In Fig. 3 we depict the combined MTF based on six rotated rectangular apertures, their corresponding MTFs and masking functions. The construction of a high-resolution image in the spatial-frequency domain from the spectra of the individual images is illustrated in Fig. 4 using an example image.

Finally, to restore those spatial-frequency components that are more attenuated (see recovered MTF in Fig. 3), the reconstruction may include a filter $H$ defined as,

$$H(u,v) = \frac{\mathrm{MTF}_{\mathrm{dl}}(u,v)}{\sum_i M_i \,\mathrm{MTF}_i(u,v) + k} \tag{4}$$

where $\mathrm{MTF}_i(u,v)$ and $\mathrm{MTF}_{\mathrm{dl}}(u,v)$ are the modulation transfer functions of sub-system $i$ and of a full aperture system (diffraction limited) respectively, and $k$ is a small constant to avoid division by zero and noise amplification.

It is worth noting that for multi-RAFOI, the FoV captured by each camera will be rotated with respect to each other, and therefore there is smaller FoV that is common to all cameras and can be fused to achieve a full, isotropic high-resolution. The peripheral regions imaged by a subset of cameras will be imaged with a reduced high-frequency content (using a sensor with 1.6 aspect ratio and $n = 3$, the 54% of the central FoV is imaged by 3 cameras and 89% by 2, and the overall FoV is 40% larger, in total solid angle). Furthermore, in the case of finite-conjugate imaging with multi-RAFOI, registration will be required to correct for parallax effects. This adds some complication, especially if the scene has components at variable depths, but also adds the felicitous benefit of enabling stereopsis-based 3D ranging.

The principle of combining multiple rotated non-isotropic PSFs to achieve a near-isotropic PSF has parallels in other computational-imaging modalities. In computed tomography, near-isotropic volume resolution is achieved by combining multiple angularly-diverse tomographic measurements with high transverse resolution and low axial resolution[11]. Similarly, in radio-frequency aperture synthesis, such as is used in radio astronomy[12] and terrestrial remote sensing[13], the use of non-isotropic arrays of non-isotropic antennas enables more practical architectures, while yielding a near-isotropic system PSF. There may be some scope for adapting image-recovery algorithms from these cognate fields for improved recovery with RAFOI.

As a generalization, the proposed RAFOI concept can be understood as imaging using multiple slab-shaped lens systems. We have described the folding of these slab lenses to enable imaging from thin camera modules, but imaging using non-folded slab cameras, arranged in other geometries is also of great interest. RAFOI can be implemented with multiple slab-lens cameras configured into geometries that are conformal to a wide range of platforms that do not easily facilitate integration of high-resolution cameras. Example applications include imaging through aerodynamic surfaces or domes of airborne vehicles, hollow cylindrical (i.e. polygonal) high-magnification surveillance systems, or wearable imaging systems such as might be integrated into helmets. Furthermore, the concept also includes distribution of slab cameras through independent networked platforms such as autonomous robots, vehicles, or wearables.

In summary, we have presented a new computational-imaging technique that is able to provide high-angular-resolution imaging from thin camera modules. It is based on the fusion of images obtained from multiple slab cameras, which allow the lens systems to be folded onto flat, thin volumes. The concept is particularly attractive for mobile-phone cameras, as it provides the possibility of synthesizing a lens aperture that is much larger than the depth of the phone. We report two illustrative designs: a multi-camera system with a focal length of 33.6 mm and a depth of only 5.9 mm and a system that employs a rotating folded slab system to enable high-resolution imaging using a single detector within an envelope that is only 7.9 mm deep. The additional costs of multiple cameras and/or moving parts (for scanned RAFOI) may be a significant disadvantage for some cost-sensitive applications, however it is noteworthy that these issues have been successfully addressed in the introduction of multi-lens cameras and mechanical autofocus in modern phone cameras. Similarly, although the time-sequential imaging employed for scanned RAFOI prevents snapshot image acquisition, this issue is common to several widely-adopted modes, such as are used for generating panoramic and 3D images.

The image reconstruction process is simple and computationally inexpensive, which is particularly attractive for low-power devices. This technique enables high-resolution imaging and large zoom ratios from mobile-phone cameras for the first time. More generally the technique has excellent potential for high-resolution imaging using multiple slab cameras integrated conformally within a wide range of platforms.

**Funding.** The Leverhulme Trust (ECF-2016-757).

**Disclosures.** The authors declare no conflicts of interest.

## REFERENCES

1. T. Steinich and V. Blahnik, Adv. Opt. Technol. **1**, 51 (2012).
2. P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad, and R. Devlin, Optica. **4**, 139 (2017).
3. S. Banerji, M. Meem, A. Majumder, F. G. Vasquez, B. Sensale-Rodriguez, and R. Menon, Optica. **6**, 805 (2019).
4. J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka, Appl. Opt. **40**, 1806 (2001).
5. G. Carles, G. Muyo, N. Bustin, A. Wood, and A. R. Harvey, J. Opt. Soc. Am. A **32**, 411 (2015).
6. G. Carles, J. Downing, and A. R. Harvey, Opt. Lett. **39**, 1889 (2014).
7. E. J. Tremblay, R. A. Stack, R. L. Morrison, and J. E. Ford, Appl. Opt. **46**, 463 (2007).
8. E. J. Tremblay, R. A. Stack, R. L. Morrison, J. H. Karp, and J. E. Ford, Appl. Opt. **48**, 343 (2009).
9. B. Zhang, M. Piao, and Q. Cui, Opt. Express **27**, 32337 (2019).
10. G. Häusler, Opt. Commun. **6**, 38 (1972).
11. L. W. Goldman, J. Nucl. Medicine Technol. **35**, 115 (2007).
12. A. R. Thompson, J. Moran, and G. W. S. Jr., *Interferometry and Synthesis in Radio Astronomy* (Springer, 2017).
13. A. R. Harvey and R. Appleby, The Aeronaut. J. **107**, 87 (2003).

## FULL REFERENCES


1. T. Steinich and V. Blahnik, “Optical design of camera optics for mobile phones,” Adv. Opt. Technol. **1**, 51–58 (2012).
2. P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad, and R. Devlin, “Recent advances in planar optics: from plasmonic to dielectric metasurfaces,” Optica. **4**, 139–152 (2017).
3. S. Banerji, M. Meem, A. Majumder, F. G. Vasquez, B. Sensale-Rodriguez, and R. Menon, “Imaging with flat optics: metalenses or diffractive lenses?” Optica. **6**, 805–810 (2019).
4. 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).
5. G. Carles, G. Muyo, N. Bustin, A. Wood, and A. R. Harvey, “Compact multi-aperture imaging with high angular resolution,” J. Opt. Soc. Am. A **32**, 411–419 (2015).
6. G. Carles, J. Downing, and A. R. Harvey, “Super-resolution imaging using a camera array,” Opt. Lett. **39**, 1889–1892 (2014).
7. E. J. Tremblay, R. A. Stack, R. L. Morrison, and J. E. Ford, “Ultrathin cameras using annular folded optics,” Appl. Opt. **46**, 463–471 (2007).
8. E. J. Tremblay, R. A. Stack, R. L. Morrison, J. H. Karp, and J. E. Ford, “Ultrathin four-reflection imager,” Appl. Opt. **48**, 343–354 (2009).
9. B. Zhang, M. Piao, and Q. Cui, “Achromatic annular folded lens with reflective-diffractive optics,” Opt. Express **27**, 32337–32348 (2019).
10. G. Häusler, “A method to increase the depth of focus by two step image processing,” Opt. Commun. **6**, 38–42 (1972).
11. L. W. Goldman, “Principles of CT and CT Technology,” J. Nucl. Medicine Technol. **35**, 115–128 (2007).
12. A. R. Thompson, J. Moran, and G. W. S. Jr., *Interferometry and Synthesis in Radio Astronomy* (Springer, 2017).
13. A. R. Harvey and R. Appleby, “Passive mm-wave imaging from uavs using aperture synthesis,” The Aeronaut. J. **107**, 87–98 (2003).