

Journal of Electronic Imaging

SPIDigitalLibrary.org/jei

Resolution and sensitivity of wafer-level multi-aperture cameras

Alexander Oberdörster
Hendrik P. A. Lensch



Resolution and sensitivity of wafer-level multi-aperture cameras

Alexander Oberdörster

Fraunhofer Institute for Applied Optics and Precision Engineering
Albert-Einstein-Straße 7, D-07749 Jena, Germany
E-mail: alexander.oberdoerster@iof.fraunhofer.de

Hendrik P. A. Lensch

Eberhard Karls University Tübingen
Computer Graphics Group Sand 14, D-72076 Tübingen, Germany

Abstract. *The scaling limits of multi-aperture systems have been widely discussed from an information-theoretical standpoint. While these arguments are valid as an upper limit, the real-world performance of systems for mobile devices remains restricted by optical aberrations. We argue that aberrations can be more easily controlled with certain architectures of multi-aperture systems, especially those manufactured on wafer scale (wafer-level optics, WLO). We complement our analysis with measurements of one single- and one multi-aperture WLO camera. We examine both sharpness and sensitivity, giving measurements of modulation transfer function and temporal noise, and showing that multi-aperture systems can indeed reduce size without compromising performance. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JEI.22.1.011001](https://doi.org/10.1117/1.JEI.22.1.011001)]*

1 Introduction

In multi-aperture optics, a single optical system is replaced by an array of optical channels side by side. In a single-aperture system, its focal length and pixel pitch determine the rate it samples object space. In a multi-aperture system, this relation can be broken up by interlacing the views of adjacent channels so that they supersample object space. After capture, the channel microimages are assembled digitally to obtain a continuous image. Using this principle, the system thickness can be reduced while keeping sampling of object space constant. When the optics is considered to be diffraction-limited, however, either sensitivity or effective resolution has to be sacrificed.¹

On the other hand, when system thickness is reduced, lens dimensions are reduced along with it. multi-aperture systems are often realized with micromanufacturing techniques, which are more accurate for lenses with small diameters and sags, leading to better optical performance.

We examine the balance of these two effects using the electronic cluster eye (eCLEY)² as one example of a multi-aperture system. The eCLEY uses supersampling to

reduce system thickness and lens dimensions. Additionally, the total field of view of the system is divided; each channel only images a small field of view.

After reviewing related work in this area (Sec. 2), we discuss performance scaling and manufacturing issues in Sec. 3. Next, we treat the effects of image reconstruction on sharpness and noise (Sec. 4). We then compare the theoretical results to the actual performance of the eCLEY using measurements of the modulation transfer function (MTF) and the temporal noise in Sec. 5. Finally, we compare the MTF with a state-of-the-art single-aperture camera manufactured with wafer-level optics.

2 Related Work

An early small multi-aperture system was TOMBO,³ which uses a low number of identical channels with the same viewing direction. The same principle has also been applied to macroscopic infrared focal plane arrays for remote sensing applications.⁴ Flexible laboratory setups such as the Stanford large camera array have been valuable to investigate possible applications and configurations of multi-aperture systems, as well as yielding practical insights on how to calibrate these systems.⁵ The eCLEY, in contrast, is specifically designed for precise and cost-effective manufacturing with microfabrication techniques and contains unique channels with different viewing directions.

Supersampling with multi-aperture systems is a natural extension of super-resolution from video sequences. Park et al.⁶ have conducted a comprehensive review of existing methods. Registration techniques as well as reconstruction algorithms have been adapted to multi-aperture systems, for example by Nitta et al.⁷ and Kanaev et al.^{8,9} However, for images from real-world systems, simple shift-and-add schemes preceded by calibration with sub-pixel accuracy have remained popular, for example as reported by Kitamura et al.¹⁰ An extended version of this type of scheme is also used for reconstructing images from the eCLEY.¹¹

Independent of the applied reconstruction algorithm, the theoretical performance limits of thin optical systems were comprehensively investigated by Haney.¹ He concludes that multi-aperture systems with reduced length can only match the performance—sensitivity and resolution—of single-aperture systems at a significant increase in footprint.

Paper 12337SSP received Sep. 3, 2012; revised manuscript received Nov. 21, 2012; accepted for publication Dec. 4, 2012; published online Jan. 31, 2013.

Measurements of both sensitivity and resolution from experiments are rare. Figures for peak signal-to-noise ratio comparing ground truth with a simulation are stated most frequently, along with example images from the actual system. Portnoy et al.⁴ give contrast measurements for a single frequency along with the signal-to-noise ratio.

We provide an analysis of the sensitivity and the resolution of multi-aperture, systems. We confirm our theoretical model with measurements of the MTF and the temporal noise of a specific system, the electronic cluster eye, which is described in Sec. 3.

3 Scaling in Multi-Aperture Systems

In this section, we discuss scaling in general multi-channel systems. As we will see with the example of the eCLEY, there are two aspects to any multi-aperture configuration that have different impacts on system volume and performance.

The eCLEY is based on the principle of interlaced tiles, as introduced in Ref. 2. Each optical channel of the eCLEY has a small field of view (FOV) and a unique viewing direction. The FOVs of adjacent channels overlap, together creating a larger FOV (Fig. 1). Their viewing directions are carefully tuned, so that pixels of one channel sample object space inbetween pixels of the adjacent channels (Fig. 2). In practice, one pixel does not have a discrete viewing direction; it integrates light over a solid angle. The implications are discussed in Sec. 3.2.

These two aspects of the concept serve specific purposes:

- *Segmenting* the system FOVs into smaller channel FOVs reduces the field each channel has to image. Aberrations can be controlled with a less complex lens system, reducing cost and making manufacturing easier and less prone to degradation because of tolerances.
- *Interleaving* the tiles achieves supersampling of object space and is responsible for reducing the effective focal length of the system, which is the lower limit to thickness.

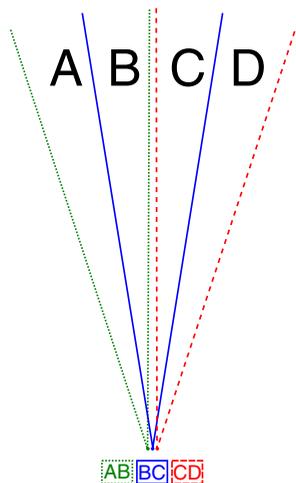


Fig. 1 The field of view of the (e)CLEY is segmented into multiple channels, each viewing a part of the total FOV. The FOVs of the channels overlap. Here, three channels are shown in different colors, in one dimension. The actual eCLEY has 17×13 channels.

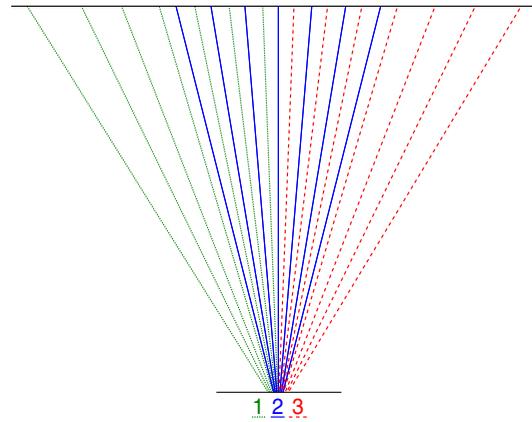


Fig. 2 The viewing directions of the eCLEY channels are adjusted carefully so that the pixels of one channel sample object space between the pixels of the adjacent channel. Three channels shown in one dimension, with seven pixels for each channel. The eCLEY has 39×39 pixels per channel.

Both aspects act in concert to decrease lens diameters. Interleaving achieves this goal directly, because at the same F -number, a smaller focal length leads to smaller lens diameters. Segmentation achieves the same goal indirectly, as less complex lens systems tend to have smaller lens diameters: The further away a lens is from the aperture stop, the larger it has to be to avoid vignetting of marginal rays. The more lenses a system has, the larger the axial extent of the system, leading to large lenses far away from the aperture.

We now investigate how multi-aperture systems compare to single-aperture systems in terms of light collection efficiency, resolution and physical size. For better clarity, we treat the effects of segmentation and supersampling separately.

3.1 Light Collection

First, we determine the light collection efficiency of a single-aperture system. Consider a setup with a scene emitting the radiance L , a lens with diameter D and effective focal length f , and an image sensor [Fig. 3(a)]. The sensor has the extent $w \times h$, divided into $n_x \times n_y$ pixels with a pitch of p_x . From the image plane, the lens subtends a solid angle of

$$\Omega = \frac{\pi(D/2)^2}{f^2}.$$

As the aperture takes on the radiance of the scene,¹² the sensor receives an irradiance of

$$I = \eta_{\text{lens}} \cdot \Omega \cdot L,$$

with the lens having an optical transmittance of η_{lens} . Each pixel integrates I over its photosensitive area, collecting a radiant flux of

$$\Phi_{\text{pix}} = \gamma_{\text{pix}} \cdot p_x^2 \cdot I,$$

where γ_{pix} is the fill factor of the pixel. The total flux collected by the sensor is

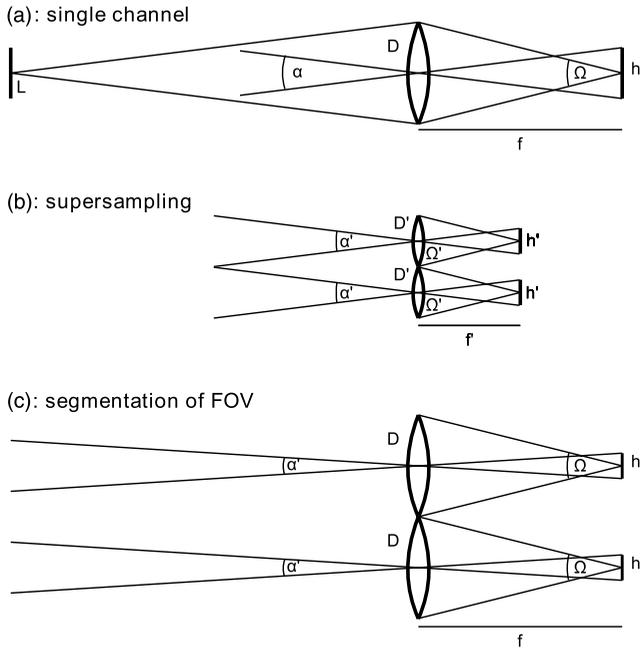


Fig. 3 Geometric properties of a single-aperture system (a) and two multi-aperture systems, supersampling (b) and segmenting (c) with lens diameter D , focal length f , image height h , and field of view α . The lens subtends an angle Ω at the image plane.

$$\Phi_{\text{tot}} = n_x n_y \cdot \Phi_{\text{pix}},$$

neglecting effects such as distortion or vignetting.

We discuss a supersampling multi-aperture system next [Fig. 3(b)]. To decouple sampling rate from pixel pitch, the single optical system is replaced by $N \times N$ channels side by side, with the supersampling factor N . In case of the eCLEY, the supersampling factor N is 2, though the number of channels is higher because the FOV is segmented.

Each channel is a scaled version of the original system (see Ref. 2), so $f' = f/N$, $D' = D/N$ and each system retains the F-number of the original camera. Therefore, in each system, $\Omega' = \Omega$. Consequently, each pixel records the same flux Φ_{pix} as in the single-aperture case. As the system samples the same FOV (solid angle) as the original system with the same sampling rate, the total amount of samples—or pixels—stays the same. Therefore,

$$\Phi'_{\text{tot}} = \Phi_{\text{tot}}.$$

Next, we segment the FOV α of the camera into $M \times M$ channels [Fig. 3(c)]. In case of the eCLEY, M is 8 horizontally and 6 vertically. The geometry of each of these channels is identical to that of the original optical system: Both f and D stay the same. The FOV of each channel is limited, however, by reducing the image size in each channel. The viewing direction of the channel is selected by introducing a lateral offset between optical system and image. Again neglecting distortion and vignetting, in each channel, partial FOV is α/M and image size is $w/M \times h/M$. As sampling rate and pixel size are kept the same, each channel now uses $n_x/M \times n_y/M$ pixels. If either distortion or vignetting are not corrected in the optical system, they affect single-aperture and multi-aperture systems in the same way.

Because the focal length is still f and the aperture diameter is still D , Ω remains the same and $\Phi'_{\text{pix}} = \Phi_{\text{pix}}$. As the total amount of pixels in the system does not change, $\Phi'_{\text{tot}} = \Phi_{\text{tot}}$.

In summary, both segmenting and supersampling multi-aperture systems collect the same amount of light as single-aperture systems, as long as the F -number and the total photodetector area are kept constant.

3.2 Sharpness

In this section, we will first examine the effects of supersampling on image sharpness. Segmentation of the FOV will be of relevance in the course of the discussion.

By using supersampling, a digital camera can be made thinner without sacrificing sampling rate in object space and without requiring a smaller pixel pitch. To retain actual optical resolution in object space along with sampling rate, however, the MTF of the channels in image space has to keep up with the sampling rate.

Supersampling multiplies the image plane sampling frequency f_S and the Nyquist frequency f_{Ny} by a factor of N . Therefore, the MTF should now show significant modulation up to $f'_{\text{Ny}} = N \cdot f_{\text{Ny}}$. Consequently, it has to improve considerably.

The MTF of a camera is the product of the MTF of the lens and the sensor, where the sensor MTF consists of a geometrical component and a component resulting from crosstalk between pixels:

$$\text{MTF}_S = \text{MTF}_O \cdot \text{MTF}_G \cdot \text{MTF}_C.$$

MTF_G describes spatial integration over the photodetector. For square photosites, it is the Fourier transform of the rect function with the width of the photosensitive area p_p :

$$\text{MTF}_G(f) = \text{sinc}(\pi p_p f).$$

The pixel pitch p_x stays the same. While we are still free to choose a smaller p_p , light sensitivity decreases with photosensitive area, or p_p^2 . Therefore, we assume MTF_G to be constant.

Crosstalk depends on the chief ray angle of light incident on the sensor and on sensor technology. Neither of them changes for multi-aperture systems. Therefore, MTF_C is constant as well.

The burden of improving the system MTF is therefore placed entirely on the optical MTF. As described by Lohmann,¹³ if an optical system is scaled by the factor $1/N$, the area of an image point A_p scales as

$$A_p(N) = \lambda^2 F^2 + \left(\frac{1}{N}\right)^2 \bar{\xi}^2, \quad (1)$$

for light of the wavelength λ and with aperture stop number F and the lateral aberration ξ ($\bar{\xi}^2$ is its Gaussian moment).

When diffraction is negligible, the diameter of an image point is

$$d_p(N) \propto \sqrt{A_p(N)} \approx \frac{1}{N} \bar{\xi}$$

and the resolution limit therefore scales linearly with the size of the system. Supersampling with N scales each individual

optical channel of the multi-aperture system by $1/N$. Point diameters are therefore scaled by $1/N$.

Segmenting the FOV also has beneficial effects. Many of the Seidel aberrations depend on field height h .¹² Field curvature and astigmatism, for example, increase with h^2 . Therefore, segmenting the FOV into M parts reduces aberrations accordingly.

However, quantifying the benefit exactly is not possible so easily. The well-known scaling laws for Seidel aberrations only apply to imaging with a single lens. In practice, aberrations are partially corrected with multilens systems, whose behaviour is more complex. This is true even for low-cost mass-market cameras for mobile devices. With a certain amount of correction, higher-order aberrations cannot be neglected any more; these aberrations also defy description by simple scaling laws.

In conclusion, optical MTF is indeed improved considerably by scaling. This is necessary to retain optical resolution in object space. As an example on how this works out, Fig. 4 shows the MTF of a system with aberrations ($N = 1$) and the effect of scaling down this system ($N > 1$). First, only optical MTF is plotted on an absolute frequency axis (a). Optical MTF is improved as expected for increasing N . However, when the f axis is normalized to the sampling frequency f_s , which scales with N , improvement is less apparent (b). When we include pixel MTF, system MTF is similar for all N (c). Therefore, object-space sharpness of the super-sampled systems is comparable to the original system.

Increasing N further still improves optical MTF, but pixel MTF cancels this gain.

Enhancement to the optical MTF itself is limited by diffraction, which is independent of system scaling. This is illustrated in Fig. 4(d). Here, we used the same optical system as in Fig. 4(a), but scaled it down by 4, so f is now 1 mm. Again, optical MTF is improved for $N = 2$, but improvement is limited by diffraction (dashed line). The resulting object space sharpness for $N = 2$ is lower than the sharpness of the original system.

3.3 Manufacturing Tolerances

When manufacturing a lens system, deviations in lens curvatures, distances, decenter and tilt degrade the system performance. When scaling a lens system down, deviations have to be smaller as well, or performance is compromised. As a simple example, consider a single thin lens with focal length f and diameter D positioned so that it focuses light from point P onto an image plane (Fig. 5). When the lens is moved from its correct image plane distance f by a deviation Δs , defocus leads to an image point diameter $d_p = \Delta s \cdot D/f$. The smaller the system, the smaller d_p has to be to retain sharpness. Accordingly, Δs has to be smaller as well.

The same is true for the focal length of the lens: For a plano-convex lens, f is proportional to lens radius R ,¹² so a deviation ΔR leads to a new focal length f' with $\Delta f = f - f'$. Δf effectively is a defocus shift Δs , leading to an image point diameter analogous to a lens shift.

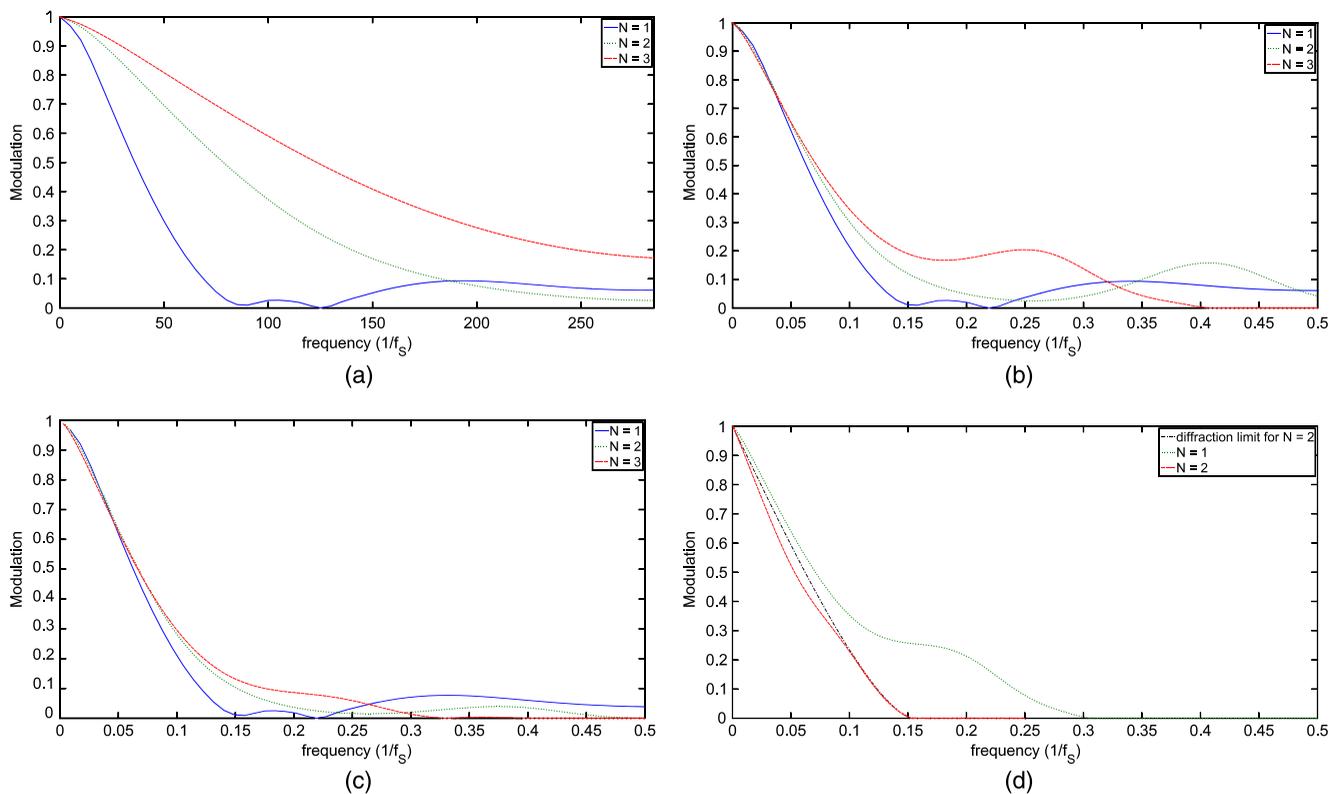


Fig. 4 Scaling effects for a system with two optical surfaces and $f = 4$ mm, F2.4, simulated on-axis MTF curves. (a) Optical MTF improves for scaling the optical system down. Supersampling factors $N = 1$ to 3; $N = 1$ is the original system. Plotted on absolute frequency axis (cycles/mm). (b) The same curves plotted on a frequency axis relative to the image-space sampling frequency f_s , which scales with N . (c) Optical MTF multiplied with geometrical pixel MTF. (d) Improvement in optical MTF is limited by diffraction effects, shown by scaling the system down by a factor of 4.

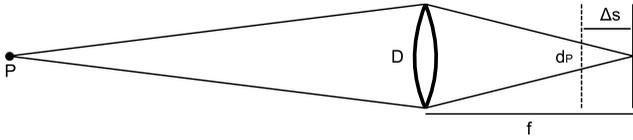


Fig. 5 Image point diameter d_p of a defocused optical system with focal length f and lens diameter D , with the image plane moved by Δs .

For perspective, with current pixel technology, $d_p < 2 \mu\text{m}$ is desirable. This requires a focus shift of less than $2d_p = 4 \mu\text{m}$.

The ability to meet the required tolerances depends on the technology that is used to manufacture and assemble the lens components. A suitable technology for multi-aperture systems is wafer-level optics (WLO), as multiple lenses side by side are manufactured and aligned in parallel. Assembly of the lens components can be achieved with the required micron precision.¹⁴

Critical, however, is precision during replication of the lens components. Lenses are manufactured from certain polymers by molding and ultraviolet curing. During hardening, these materials shrink significantly. The amount of shrink is proportional to the lens volume. Lens volume grows with the square of the lens radius and linearly with lens sag. Therefore, small lenses with low sags are preferable. Molding tools are adjusted to anticipate shrink; however, shrink has a certain spread that is proportional to shrink itself. The hardened lenses therefore still have form deviations that scale with N^3 .

Using a multi-aperture architecture—either supersampling or segmenting—decreases lens diameters. For a supersampling factor of N , lens diameter decreases by N , lens sag also decreases by N and we can expect deviations to decrease with the cube of N . Segmenting the FOV also reduces lens diameters, having a similar effect on deviations.

In conclusion, while tolerances have to be tighter for scaled-down lens systems, the fact that small lenses can be manufactured with less shrink makes it easier to meet these tolerances. Therefore, sharpness of actual, mass-manufactured camera systems can benefit significantly from a multi-aperture architecture. This result contradicts the theoretical analysis in Sec. 3.2, which suggested that multi-aperture systems can at best reach a performance comparable to single-aperture systems.

3.4 Volume

We already established that system thickness is reduced by supersampling. In some applications, however, total system volume is more relevant than thickness. Therefore, we now examine how multi-aperture system volume V' compares to that of a single-aperture system V . We again treat the two different architectures (supersampling and segmented FOV) separately. In both cases, we first derive the footprint of the system. It is given by either sensor footprint A_{sens} or total aperture area A_{tot} , depending on which one is larger. In the single-aperture case, A_{tot} is simply the area of the single system aperture. The values for the multi-aperture system are A'_{sens} and A'_{tot} , which is now the sum of all individual aperture areas A' . Next, we derive system height. In both cases, system height scales with effective focal length f . To f , a part of the optical system thickness h_{opt} is

added, depending on system complexity and placement of the principal planes. We disregard the thickness of the image sensor, sensor carrier and casing, as these values are small compared to the focal length and are not affected by the system architecture.

Supersampling: As noted in Sec. 3.1, neither the pixel pitch nor the total number of pixels on the sensor change. Therefore, $A'_{\text{sens}} = A_{\text{sens}}$. This is also true for A'_{tot} :

$$A'_{\text{tot}} = N^2 A' = N^2 \pi \left(\frac{D'}{2} \right)^2 = \pi \left(\frac{D}{2} \right)^2 = A_{\text{tot}},$$

assuming circular apertures with diameter D . Therefore, system footprint stays the same. f , in contrast, is reduced by a factor of N . As the lens dimensions all scale with N , $h'_{\text{opt}} = h_{\text{opt}}/N$. Therefore, $h'_{\text{tot}} = h_{\text{tot}}/N$ and $V' = V/N$. In conclusion, a supersampling system is not only thinner, but also has less volume than a single-aperture system.

Segmentation of FOV: Again, pixel size and number stay the same, so $A'_{\text{sens}} = A_{\text{sens}}$. However, the single-aperture with area A_{tot} is now replaced with M copies of the original aperture. Aperture area therefore is increased:

$$A'_{\text{tot}} = M \cdot A_{\text{tot}}.$$

The proportion of A_{tot} to A_{sens} in a camera is approximately the proportion of the corresponding lengths:

$$\frac{D_{\text{sens}}}{D} = \frac{2f \tan \frac{\alpha}{2}}{\frac{f}{N}} = 2N \tan \frac{\alpha}{2}.$$

Miniaturized cameras tend to have a large FOV. If we assume $N = 2.8$ and $\alpha = 70^\circ$, $D_{\text{sens}}/D \approx 4$. Therefore, the sensor width is larger than the lens diameter and system footprint is given by A_{sens} for $M \leq 4$.

Effective focal length is not affected. Reduced optical system complexity in each channel, however, decreases h'_{opt} slightly. Therefore, system volume V' is smaller than V for moderate segmentation of FOV, but increases with M^2 for large M .

This analysis does not consider additional volume consumed by the system casing, structures for suppressing stray light or walls separating channels. The latter are needed to prevent crosstalk between channels. In current systems such as the eCLEY, structures for crosstalk suppression do consume a considerable amount of space between channels. They therefore increase the total volume of the system and lead to unused areas on the image sensor. For reducing this waste of space and sensor area, very thin vertical or slanted walls have to be manufactured. Techniques for cheaply fabricating these structures are currently being developed.

4 Reconstruction

In the last section, the theoretical and practical scaling characteristics of multi-aperture systems were discussed. In the next section, these characteristics are verified with measurements. To compare the analysis with the measurements, we have to consider that in a multi-aperture system, a multitude of images have to be combined into a continuous image. This image reconstruction step has effects on image sharpness and alters the noise characteristics of the system. In principle,

neither can be improved without negatively affecting the other. As the focus of this publication lies on the scaling characteristics of multi-aperture systems per se, we do not attempt an exhaustive analysis of this topic. Instead, we quantify the effects of a single, simple reconstruction scheme, a shift-and-add algorithm with Gaussian interpolation. In this case, the effect is a decrease in noise and a loss in sharpness.

4.1 Algorithm

We treat each recorded pixel as a measurement of the light incident on the camera from a specific direction. The pixel viewing directions are derived from the model of the optical system; it includes effects such as geometric distortion. We intersect each of these pixel viewing rays with a virtual focal plane (Fig. 6). The intersection points of viewing rays and focal plane form a two-dimensional cloud of measurements, an irregular sampling of the scene (irregular because of parallax and geometric distortion of the channels; Fig. 7). To render an image from this point cloud, we create a regular sampling of the scene by interpolation. For each pixel of the target image, contributions from the nearest measurement points available are added, weighted with the distance from the measurement coordinate to the target pixel (Fig. 8).

From the distance r , the weight $W_{x,y,i,j}$ of the neighbor j , contributing to the target pixel at coordinates x and y , is calculated as

$$W_{x,y,i,j} = e^{-wr^2}, \quad (2)$$

where w is an adjustable filter width. The weights $W_{x,y,i,j}$ are normalized so that $\sum W_{x,y,i,j} = 1$.

The algorithm is presented in full in Ref. 11.

4.2 Sharpness

Interpolation can be treated as a spatial filter. Calculating the Fourier transform of the filter kernel yields the MTF of the interpolation operation. The interpolation kernel is the continuous version of Eq. (2), the Gaussian

$$K(r) = e^{-wr^2},$$

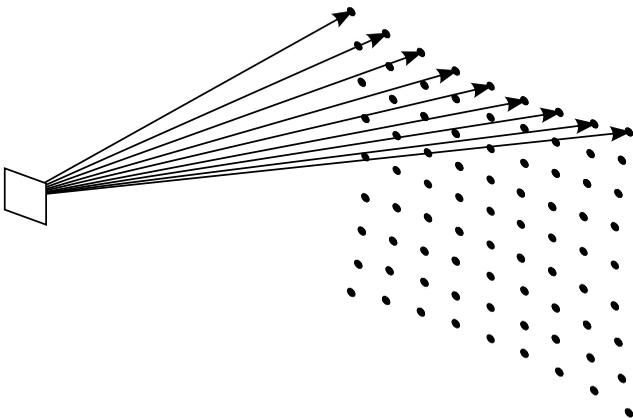


Fig. 6 Viewing directions of the pixels of one channel of a multi-aperture system (arrows), intersected with a virtual focal plane (points). The points show how this channel samples object space on a specific focal plane.

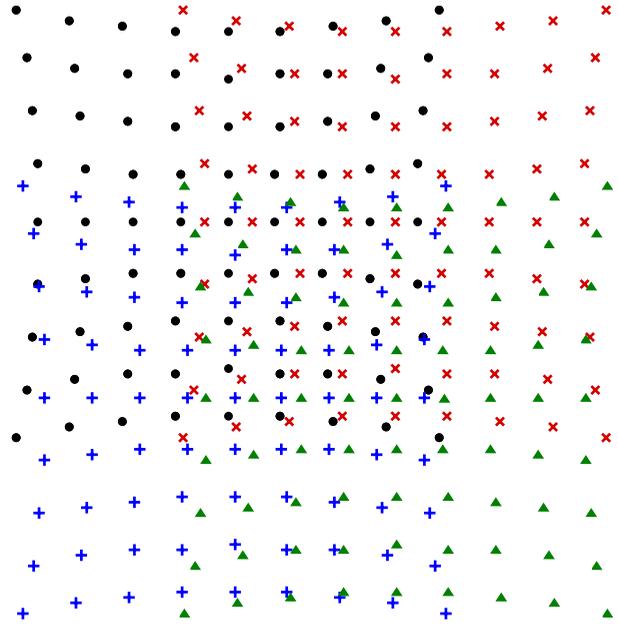


Fig. 7 Placing the measurement of multiple channels (four in this case, shown in different colors) on a common focal plane according to the distances of channels and focal plane creates an irregular point cloud.

again with the filter width w and the distance from target pixel to measurement coordinate r , in units of pixels. The effect is a loss in modulation at higher frequencies.

4.3 Noise

Each target pixel is calculated from the weighted mean of ν measurements. If a value V is calculated as the weighted sum of measurements $m_{x,y,i,j}$ with equal uncertainties σ_m ,

$$\sigma_V = \sigma_m \cdot \frac{1}{\sqrt{\sum W_{x,y,i,j}^2}},$$

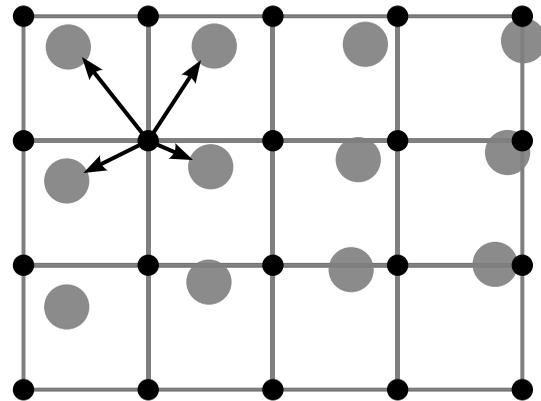


Fig. 8 To render an image from an irregular point cloud, a regular grid is overlaid on the point cloud. Each of the grid intersections represents a target pixel in the image. Each of the target pixels (black) is calculated by interpolating the nearest measurements from the point cloud (grey).

with the weights $W_{x,y,i,j}$. The weights are different for each target pixel. They depend on the distance of the measurement to the target pixel r and on the filter width w .

For the following analysis, we first assume uniform density of measurements. In one extreme case, the target pixel is exactly on top of a single source pixel. Choosing a filter width of $w = 2$ and setting $r = 0$ in Eq. (2), $W'_i = 1$ (not normalized yet). Four other pixels are at the distance of $r = 1$ pixel, yielding $W'_i = 0.13$. Four further pixels are at a distance of $r = \sqrt{2} = 1.4$, yielding $W'_i = 0.02$. Normalizing yields contributions of $W_i = 0.63, 0.08$ and 0.01 , respectively. Noise is consequently reduced by a factor of $1/\sqrt{0.63^2 + 4 \cdot 0.08^2 + 4 \cdot 0.01^2} = 1.54$. In the other extreme, the target pixel is exactly between four source pixels, each contributing equally. No other pixels contribute significantly due to their large distance. Noise is decreased by $\sqrt{4} = 2$.

In conclusion, noise is decreased by a factor of about 1.54 to 2.

5 Results

In this section, to support the conclusions of the last section, we compare one state-of-the-art single-aperture WLO camera, the OmniVision CameraCube, with a WLO multi-aperture system, the eCLEY. To verify sharpness, we directly compare the MTF of these systems. Direct comparison of the sensitivity of the two cameras is not useful, as they employ different sensors with different pixel pitches ($1.75 \mu\text{m}$ versus $3.2 \mu\text{m}$). The route taken is described next.

5.1 Sensitivity

According to theory, the eCLEY should have the same sensitivity as a single-aperture camera with the same aperture F3.7. For verification, we took an image of a uniformly lit target with the image sensor used in the eCLEY, with a single-aperture 16-mm lens (Schneider Cinegon) attached and set to F3.7. The same target was also recorded with an eCLEY. To avoid linearity issues, the exposure time t_{exp} was adjusted so that both cameras recorded roughly the same mean value (DN) on the target area. The values recorded and the corresponding exposure times t_{exp} were:

	Cinegon	eCLEY
Value	146	144
t_{exp}	3.3 ms	4.2 ms

The longer exposure time for the eCLEY suggests a lower sensitivity (by a factor of 0.77). We suspected that this discrepancy is caused by the way the eCLEY objective is attached to the sensor. The clear epoxy filling the gap between objective and substrate has a refractive index close to that of the per-pixel microlenses on the sensor, thereby rendering them ineffective.

We validated our suspicion by attaching a plane glass to one half of the sensor, again filling the air gap with epoxy. We then recorded the same target area with the treated sensor, again imaging the target area with the Cinegon lens set to F3.7. We measured values of 140 on the sensor half without plane glass and 110 on the other half, yielding

a relative sensitivity $\gamma = 0.71$. This figure also gives us an estimate on the relative area of the photosensor on each pixel. We assume a perfect efficiency of the pixel microlenses and set the fill factor of the sensor pixels to $\eta_{\text{pix}} = 0.71$.

Sensitivity of the eCLEY consequently has to be adjusted by a factor of 1.40, yielding an adjusted relative sensitivity of about $\gamma = 1.1$, higher than the single-aperture lens. The new discrepancy is most likely caused by losses due to internal reflections in the Cinegon lens, which has more air-glass surfaces than the eCLEY objective.

Note that the loss in sensitivity due to the loss of the per-pixel microlenses is not inherent to multi-aperture systems or WLO. The attenuation can be avoided by replacing the bottom substrate with a spacer layer that introduces an air gap between optics module and sensor.

5.2 Noise

As illustrated in Sec. 4.3, the reconstruction scheme that we use interpolates measurements, which should reduce noise. To verify this claim, we first established the image noise of the sensor used in the eCLEY.

To this end, we recorded 100 images of a scene with a wide dynamic range, using the eCLEY. The recorded images contain microlens images with all values in the dynamic range of the camera, from 0 to 255. As we are interested in temporal noise, we evaluated the temporal behaviour of each pixel. For each of them, the mean and the standard deviation were calculated. Pixels were then distributed into bins of integer values according to their mean. The resulting distribution of standard deviation over image signal is plotted on a log-log scale in Fig. 9.

We proceeded to process each of the recorded images with our reconstruction algorithm, creating continuous images from the raw images. Filter width w was set to 2.0. These processed frames were characterised pixel by pixel as before, yielding another distribution of standard deviations, this time including reconstruction. This distribution is also plotted in Fig. 9.

Comparing the plots shows that noise is attenuated by a factor of 2.0, being at the top end of our prediction from Sec. 4.3 and validating our model of the reconstruction algorithm.

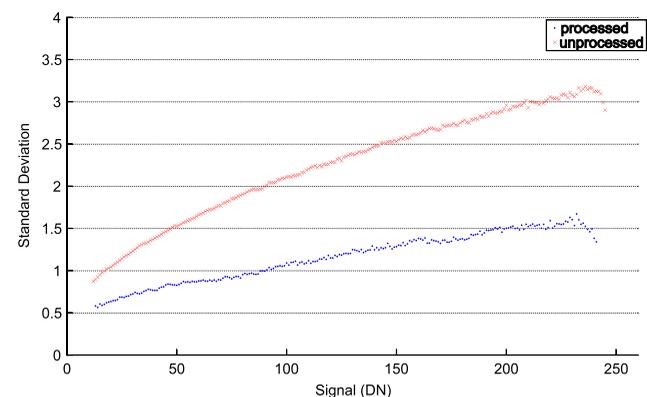


Fig. 9 Temporal pixel noise (standard deviation) of the sensor in the eCLEY, unprocessed microlens image and reconstructed image (processing filter width $w = 2.0$).

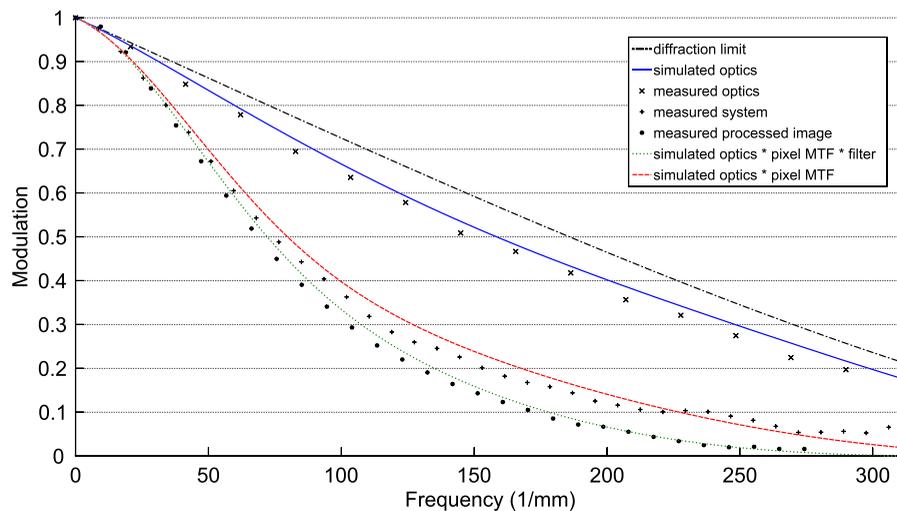


Fig. 10 MTF of the eCLEY. From the diffraction limit downwards, one additional component is added to the simulation for each curve. The complete system MTF is plotted at the bottom, with measurements confirming the model at several steps.

5.3 Sharpness

With the results from the previous Secs. 5.1 and 5.2, we have a complete model of the eCLEY transfer function. Figure 10 shows simulations of all components. On top, the diffraction limit for F3.7 is plotted. The optical MTF of the eCLEY central channel is quite close to this limit. It is calculated from a ZEMAX model of the eCLEY objective lens.

Next, the contribution of the sensor is multiplied with the optical MTF. From Sec. 5.1, we assume square photodiodes with a width of $\sqrt{\gamma} \cdot 3.2 \mu\text{m} = 2.7 \mu\text{m}$. The crosstalk was modeled as a Gaussian and fitted to results from Refs. 15 and 16.

Finally, the reconstruction step is considered by multiplying the filter kernel K to optical and sensor MTF, with filter width $w = 2.0$.

To validate this model, we measured three steps of the image formation process: The optical MTF of a single eCLEY channel, the MTF of a single channel including

sensor and the MTF of the complete system, including reconstruction. Each measurement was carried out with the slanted-edge method.¹⁷

The measurements (also plotted in Fig. 10) match the predicted MTFs quite closely, validating our model of the eCLEY. In summary, we have shown that

- Microlens arrays can be manufactured with low tolerances, so that they closely match the simulated performance;
- the image sensor plays a significant part in the total MTF of a supersampling multi-aperture system, because the photodiodes are larger than the virtual pixel pitch; and
- the reconstruction algorithm reduces noise at the expense of reduced sharpness.

Note that no calibration was necessary to align the microlens images in the reconstruction step. The distributions of

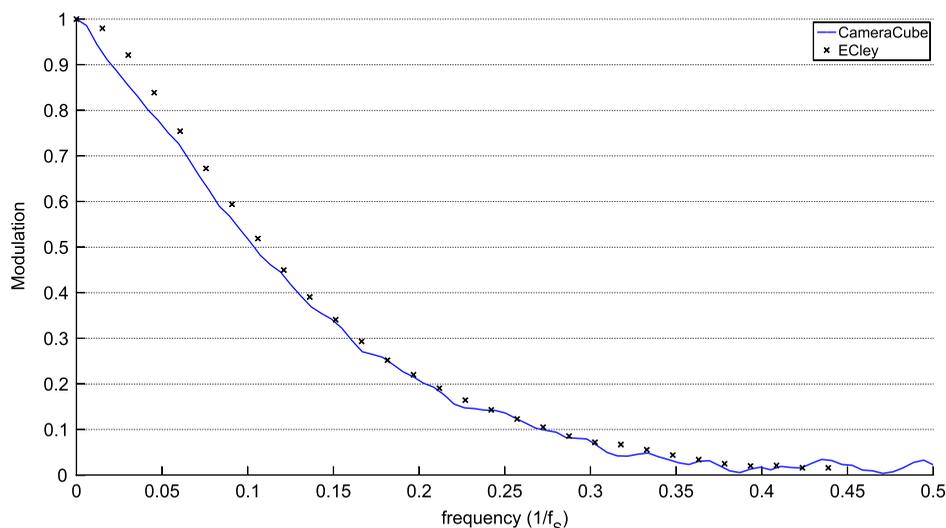


Fig. 11 MTF of OmniVision CameraCube and eCLEY. System MTFs plotted relative to the sampling frequency of each system.

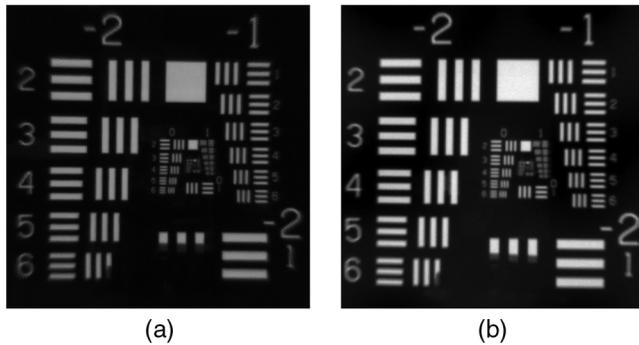


Fig. 12 Photographs of a backlit USAF test target with the eCLEY (a) and the OmniVision CameraCube (b). For comparable results, all postprocessing, sharpening and noise reduction was disabled in both cases.

the pixels on the virtual focal plane were taken directly from the ZEMAX model. This fact demonstrates the manufacturing and alignment precision of the microlens array.

Finally, we compared the MTFs of the eCLEY and an OmniVision CameraCube. Figure 11 shows the complete system MTF of both systems. We normalized the frequency axis on the image-space sampling frequency of each camera, which is $1/1.75 \mu\text{m} = 571$ cycles/mm for the CameraCube and $2 \cdot 1/3.2 \mu\text{m} = 625$ cycles/mm for the eCLEY. The eCLEY exhibits comparable sharpness at reduced total track length.

Figure 12 compares two shots of an USAF test target recorded with the eCLEY and the CameraCube. These photographs also demonstrate similar sharpness for both systems.

5.4 Volume

Despite having a larger pixel pitch ($3.2 \mu\text{m}$ instead of $1.75 \mu\text{m}$), the eCLEY has a shorter track length than the CameraCube (1.4 mm instead of approximately 2.2 mm). This is the result of $2\times$ supersampling in the eCLEY ($N = 2 \times 2$, in x and y), which cuts total track length in half. Additionally, the eCLEY has only one optical surface per channel instead of the two surfaces of the CameraCube,¹⁸ which also reduces thickness.

Footprint, on the other hand, is larger for the eCLEY, being 6.8×5.2 mm compared to 3.2×2.8 mm. This $4\times$ increase in footprint is partly due to the larger pixel pitch, partly a result of the segmentation of the FOV ($M = 7 \times 4$, in x and y), as deduced in Sec. 3.4.

6 Conclusion

We provided an analysis of the sensitivity, resolution and volume of two types of multi-aperture systems. Compared to single-aperture cameras, systems which supersample object space significantly reduce volume at constant sensitivity. Matching the resolution is challenging, but possible for low supersampling factors N in cases where the optical system is not diffraction limited. Systems that segment the FOV increase footprint and volume, but simplify the optical system, which helps reducing track length. Both principles can be used in tandem to design cameras with lower track length and sufficient sharpness, as demonstrated with our measurements of the eCLEY.

In this analysis, we assumed monochrome sensors without color filter arrays (CFAs). For a sensor with CFA, the

color channels are traditionally undersampled, potentially leading to aliasing. This is a favorable premise for a super-sampling multi-aperture system: with $N = 2$, aliasing can be avoided and, at the same time, track length can be halved. Extending the discussion of this publication to color systems therefore is a promising direction.

Finally, plenoptic cameras are in essence also multi-aperture systems. In the focused plenoptic camera, multiple channels view overlapping parts of an intermediate, demagnified image of the subject. Each channel has a limited field of view; the sampling patterns of the channels are interleaved so that the intermediate image is supersampled. This translates into increased resolution, however, only when the combined MTF of objective lens, microlens array and sensor is sufficiently large.

In multi-aperture and plenoptic cameras, filtering can regain sharpness at the price of increased noise. This is traditionally the subject of superresolution algorithms. Work in this area has focused on aligning the multiple views of the subject accurately and robustly, with the required sub-pixel resolution. When the optical system is manufactured with sub-micron precision, good alignment can be already be achieved from the geometry of the design. Similarly, the transfer function can be simulated with useful precision. To examine whether the available data is sufficient to increase sharpness without introducing artifacts would be another interesting topic.

References

1. M. W. Haney, "Performance scaling in flat imagers," *Appl. Opt.* **45**(13), 2901–2910 (2006).
2. A. Brückner et al., "Thin wafer-level camera lenses inspired by insect compound eyes," *Opt. Express* **18**(24), 24379–24394 (2010).
3. J. Tanida et al., "Thin observation module by bound optics (TOMBO): concept and experimental verification," *Appl. Opt.* **40**(11), 1806–1813 (2001).
4. A. Portnoy et al., "Design and characterization of thin multiple aperture infrared cameras," *Appl. Opt.* **48**(11), 2115–2126 (2009).
5. V. Vaish et al., "Using plane + parallax for calibrating dense camera arrays," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Vol. 1, pp. I-2–I-9, IEEE (2004).
6. S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Process. Mag.* **20**(3), 21–36 (2003).
7. K. Nitta et al., "Image reconstruction for thin observation module by bound optics by using the iterative backprojection method," *Appl. Opt.* **45**(13), 2893–2900 (2006).
8. A. V. Kanaev et al., "TOMBO sensor with scene-independent super-resolution processing," *Opt. Lett.* **32**(19), 2855–2857 (2007).
9. A. V. Kanaev et al., "Analysis and application of multiframe super-resolution processing for conventional imaging systems and lenslet arrays," *Appl. Opt.* **46**(20), 4320–4328 (2007).
10. Y. Kitamura et al., "Reconstruction of a high-resolution image on a compound-eye image-capturing system," *Appl. Opt.* **43**(8), 1719–1727 (2004).
11. A. Oberdörster et al., "Correcting distortion and braiding of micro-images from multi-aperture imaging systems," *Proc. SPIE* **7875**, 78750B (2011).
12. W. J. Smith, *Modern Optical Engineering: The Design of Optical Systems*, McGraw-Hill, New York (1990).
13. A. W. Lohmann, "Scaling laws for lens systems," *Appl. Opt.* **28**(23), 4996–4998 (1989).
14. R. Völkel et al., "Technology trends of microlens imprint lithography and wafer level cameras (WLC)," in *Conf. Micro-Optics*, Brussels, Belgium (2008).
15. J. Chen et al., "Imaging sensor modulation transfer function estimation," in *Proc. 17th IEEE Int. Conf. Image Process.*, pp. 533–536, IEEE, Hong Kong (2010).
16. X. Jin et al., "Sensitivity and crosstalk study of the zero gap microlens used in $3.2 \mu\text{m}$ active pixel image sensors," *Microelectron. Eng.* **87**(4), 631–634 (2010).
17. S. E. Reichenbach, S. K. Park, and R. Narayanswamy, "Characterizing digital image acquisition devices," *Opt. Eng.* **30**(2), 170–177 (1991).
18. R. Fraux, "Omnivision's VGA wafer-level camera," *3D Packaging* **22**, 26–27 (2012).

Alexander Oberdörster is a researcher at the Fraunhofer Institute for optics and precision engineering in Jena, Germany. He graduated in physics from the University of Düsseldorf and the Fraunhofer ISE in Freiburg. At the ISE and at Spheron VR AG, he studied and designed devices for measuring optical scattering properties of surfaces (BRDFs). Moving on to the Fraunhofer IIS in Erlangen, he developed cameras for digital cinematography and related technologies. Currently, he is working on image processing algorithms for multi-aperture imaging systems. His research interests include computational photography, non-uniform sampling and reconstruction and the measurement of light fields.

Hendrik P. A. Lensch holds the chair for computer graphics at Tübingen University. He received his diploma in computer science from the University of Erlangen, in 1999. He worked as a research

associate at the computer graphics group at the Max-Planck-Institut für Informatik in Saarbrücken, Germany, and received his PhD from Saarland University, in 2003. He spent two years (2004 to 2006) as a visiting assistant professor at Stanford University, USA, followed by a stay at the MPI Informatik as the head of an independent research group. From 2009 to 2011, he was a full professor at the Institute for Media Informatics at Ulm University, Germany. In his career, he received the Eurographics Young Researcher Award 2005, was awarded an Emmy-Noether-Fellowship by the German Research Foundation (DFG) in 2007 and received an NVIDIA Professor Partnership Award in 2010. His research interests include 3-D appearance acquisition, computational photography, global illumination and image-based rendering, and massively parallel programming.