Super-resolution imaging using a camera array

Guillem Carles,[1] James Downing,[1,2] and Andrew R. Harvey[1,*]

[1]*School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK*

[2]*STMicroelectronics, Pink Hill, Edinburgh EH12 7BF, UK*

**Corresponding author: andy.harvey@glasgow.ac.uk*

Received November 6, 2013; accepted February 1, 2014;
posted February 20, 2014 (Doc. ID 200843); published March 21, 2014

The angular resolution of many commercial imaging systems is limited, not by diffraction or optical aberrations, but by pixilation effects. Multiaperture imaging has previously demonstrated the potential for super-resolution (SR) imaging using a lenslet array and single detector array. We describe the practical demonstration of SR imaging using an array of 25 independent commercial-off-the-shelf cameras. This technique demonstrates the potential for increasing the angular resolution toward the diffraction limit, but without the limit on angular resolution imposed by the use of a single detector array. © 2014 Optical Society of America

OCIS codes: (100.3020) Image reconstruction-restoration; (100.6640) Superresolution; (110.1758) Computational imaging.

http://dx.doi.org/10.1364/OL.39.001889

While the angular resolution of an imaging system is fundamentally limited by diffraction, the finite size of the detector pixels in commercial imaging systems normally yields a lower limit on angular resolution. That is, aliasing occurs in the sampling of the image by the detector array since the sampling frequency is limited by the pixel size to be less than twice the optical cutoff frequency. Larger pixels offer the advantages of higher optical sensitivity and reduced fabrication challenges. In these circumstances multiaperture imaging using lenslet arrays can enable super-resolution (SR) of recorded images [1] to yield an angular resolution that is close to the diffraction limit of the individual lenslets. If there are m lenslets across the width of the array, then a reduction in lens length by a factor of m is possible compared to a single lens of equal field-of-view, although the diffraction-limited resolution is limited to that of an aperture of width D/m, where D is the width of the detector array. Recent advances employing multiaperture imaging include increased field-of-view, implementation in the infrared, and improvements in signal processing, see [2–6] and references therein. None of these tackle the above angular-resolution limit due to use of a single-detector array. Here we report the practical demonstration of a multiaperture imaging system implemented using an array of discrete cameras so that the angular resolution is independent of array dimension D (enabling multiaperture imaging with high angular resolution), and we explore experimentally the trade between image quality and number of cameras, demonstrating that the resulting randomized sampling offsets enable SR.

From simple geometrical considerations, the number of pixels across an image is $N_{\mathrm{LR}} = 2f \tan\theta/p$, where θ is the half-angle field of view, f is the focal length, and p is the detector pitch. When m independent imagers are combined in parallel, the number of independent pixels may be increased by SR in one dimension to $N_{\mathrm{HR}} = mN_{\mathrm{LR}}$. For a given pixel count in the final image, parallel aliased imagers may alternatively enable a reduction in the focal length of the lenses by a factor m, yielding a commensurate reduction in depth of the imaging system.

For optical spatial-frequencies ν_{opt}, in the range $\nu_s/2 > \nu_{\mathrm{opt}} > \nu_o$, where $\nu_s = 1/p$, $\nu_o = 1/(\lambda f/\#)$ is the optical spatial cutoff frequency, λ is the wavelength and $f/\#$ is focal ratio; aliasing is manifest as artifacts in the detected image. In the frequency domain these artifacts are associated with frequency components in the baseband of spatial-frequencies ($\nu_{\mathrm{det}} < \nu_s/2$) introduced by heterodyning of optical spatial-frequencies with the multiple harmonics of the frequency of the sampling comb [7]. There are $n = \mathrm{int}(\nu_o/\nu_s)$ such frequency components aliased into the baseband, but provided that $n < m$, it is possible, using SR, to accurately recover the aliased spatial-frequencies in the range $\nu_s/2 > \nu_{\mathrm{opt}} > \nu_o$ to yield artifact-free images with a spatial-frequency cutoff that is limited by the diffraction limit of the low-resolution images, ν_o, rather than by Nyquist sampling frequency, $\nu_s/2 = 1/(2p)$. Ideal reconstruction using SR also requires that the multiple cameras are aligned such that detectors sample the image with offsets in integer multiples of p/m so that the image is fully sampled at mN_{LR} nonredundant points (in one dimension).

For a perfectly aligned system, the varying geometrical projection of the image onto the individual detector arrays varies the sampling offset with range and hence the degree of redundancy and the efficacy of SR also vary. Small, random variations in sampling offset, as occur with typical manufacturing tolerances for assembly of the camera array, tend to decohere the effects of varying redundancy across the array yielding an image quality that is approximately invariant with range and yields an image quality that is almost as high as a perfectly aligned array with zero redundancy [8].

We describe here a multiaperture imaging system consisting of 25 commercial-of-the-shelf (COTS), low-cost cameras assembled into a plane array as shown in Fig. 1. Each individual camera employs an $f/2.8$ two-element compound lens with a nominal focal length of 2.96 mm and a full-angle field-of-view of 39×51 degrees with a detector array of 1616×1208 pixels. Camera centers are separated by 5.2 mm, and $p = 3.5$ μm for each color channel of the Bayer matrix, which has a pitch of 1.75 μm. Therefore the optical cutoff-frequency is $\nu_o = 650\ \mathrm{mm}^{-1}$ for the green channel and the Nyquist frequency is $\nu_s/2 = 1/(2p) = 143\ \mathrm{mm}^{-1}$. The system modulation transfer function (MTF) for a single camera is therefore limited by the detector. Ideal sampling and SR would increase the effective Nyquist frequency by

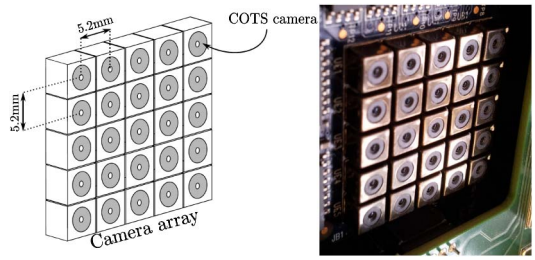

Fig. 1. Assembled device comprising a 5×5 camera array.

five to 715 mm^{-1}, potentially enabling recovery at spatial frequencies up to ν_o: a factor of 4.5 increase in image bandwidth. In practice, suppression of the MTF by the pixel response and optical aberrations may prevent high-quality SR recovery at higher spatial frequencies. Cameras are assembled onto a printed circuit board with the relatively low precision typical of electronic assembly. Because of this there are random pointing variations between cameras, yielding variations in sampling offsets typically greater than p so that offsets are essentially randomly distributed in the range 0 to p modulo p.

In this prototype, the multiple camera outputs are recorded using time-sequential multiplexing so real-time recording is not yet realized. Future developments will enable a parallel readout of all cameras to enable high-resolution, video-rate snapshot imaging to be recorded.

Image recovery entails two steps: system calibration using image registration and image reconstruction. System calibration can be accomplished for an arbitrary range and arbitrary plane. This includes short-range imaging for which the perspective varies significantly across the camera array. Multiview image registration is implemented using feature-based registration, where points of interest are identified using the scale-invariant feature transform (SIFT) detector [9]. This provides point localization with subpixel accuracy. Image registration is accomplished in two stages: (1) interest points are matched using the SIFT descriptors and unmatched points are discarded; (2) the geometrical image transformation is estimated by fitting the variable parameters of a geometrical transformation model to the transformation determined between points matched using the SIFT algorithm. A Random Sample Consensus (RANSAC) [10] procedure is used to discard outliers. This process allows the use of arbitrary perspective transformations that are convenient for general, natural scenes.

An example points-of-interest detection and matching, following filtering by the RANSAC procedure, is shown in Fig 2. The scene includes two dominant components: a far-field for which one geometrical transformation model is appropriate and a near-field scene for which a different model is appropriate. Registration in this first instance has been accomplished based on a transformation model only for the far-field scene. It is clear that the dominant far-field perspective plane is selected by the transform-estimation algorithm. This is regarded as the plane of interest. Calibration results in a geometrical relationship between each image and the reference image, which may be written in homogeneous coordinates as

$$[u \quad v \quad w]^{\mathrm{T}} = \mathbf{T}[x \quad y \quad 1]^{\mathrm{T}}, \tag{1}$$

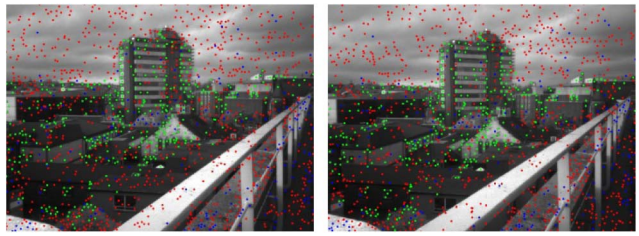

Fig. 2. SIFT point detections from two green-channel images. Red points are discarded because their SIFT descriptor was not matched, blue points are discarded by RANSAC deemed to be outliers, green points are the matched points used to calculate the geometrical transformation.

where $[u \quad v]^{\mathrm{T}}/w$ and $[x \quad y]^{\mathrm{T}}$ are the image coordinates of the two images and $\mathbf{T}$ is the 3×3 transformation matrix known as a homography [11].

The reconstruction step utilizes the registration information to construct the output high-resolution image. Many SR techniques have been reported, of which non-uniform interpolation, error-reduction energy, maximum-likelihood estimation, maximum *a priori* estimation, and projection into convex sets are prominent [12–14]. Results reported here follow maximum-likelihood estimation, as it makes no assumption of any *a priori* information regarding the scene. For this, a forward model is constructed as the basis of image retrieval:

$$\mathbf{y}_i = \mathbf{D}\mathbf{W}_i\mathbf{x} + \mathbf{e}_i, \tag{2}$$

where $\mathbf{y}_i$ is a lexicographical ordered vector of the acquired image of camera i, $\mathbf{D}$ is a rational decimation operator, which down-samples by averaging blocks of high-resolution subpixel samples into detector pixels; $\mathbf{W}_i$ is the warping matrix defined by the geometrical transform relating the image recorded by camera i with the reference image; $\mathbf{x}$ is the high-resolution scene to be estimated; and $\mathbf{e}_i$ is the noise that corrupts image i.

The warping matrices $\mathbf{W}$ are constructed from the transformation matrices $\mathbf{T}$ obtained from the registration procedure. Each high-resolution pixel coordinate is projected into the high-resolution reference image-space (according to the estimated $\mathbf{T}$), and $\mathbf{W}$ is constructed by interpolation between the reference points obtained. This construction of $\mathbf{W}_i$ does not introduce image-reconstruction artifacts.

SR is independently performed on each of the color planes of the Bayer-encoded images (prior to demosaicing of the color planes) and have dimensions of 808×604 pixels. Constructing an image from the 5×5 array super-resolves these images by a factor five in each direction to yield a 4040×3020 pixel image, per color plane. The warping matrices, $\mathbf{W}_i$, thus have dimensions of $12,200,800 \times 12,200,800$ and the decimation matrix, $\mathbf{D}$, has dimensions of $488,032 \times 12,200,800$. Although large, these matrices are very sparse and so computation remains tractable.

We define the system matrix as

$$\mathbf{M} = \begin{bmatrix} \mathbf{D}\mathbf{W}_1 \\ \mathbf{D}\mathbf{W}_2 \\ \vdots \\ \mathbf{D}\mathbf{W}_{25} \end{bmatrix}, \tag{3}$$

and the image retrieval model becomes a classic restoration model:

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{e}, \tag{4}$$

where $\mathbf{y}$ and $\mathbf{e}$ are the concatenation of all $\mathbf{y}_i$ and $\mathbf{e}_i$. Due to the large size of $\mathbf{y}$ and $\mathbf{e}$, it is not possible in practice to invert Eq. (4), but iterative methods are available to obtain a good estimate of the object $\mathbf{x}$. Maximum-likelihood estimation attempts to estimate $\mathbf{x}$ using Bayesian estimation without any prior. This resembles Lucy–Richardson iterative deconvolution

$$\mathbf{x}_{n+1} = \mathrm{diag}(\mathbf{x}_n)\mathbf{M}^{\mathrm{T}}(\mathrm{diag}(\mathbf{M}\mathbf{x}_n))^{-1}\mathbf{y}, \tag{5}$$

where $\mathbf{x}_n$ denotes the nth iterative approximation to the latent high-resolution scene and $\mathrm{diag}(\mathbf{x})$ denotes diagonal matrix with elements of vector $\mathbf{x}$. The number of iterations used to obtain the image estimate trades data fidelity (better fit of reconstructed image to observations) against noise amplification.

Reconstruction results are shown in Fig. 3 for a United States Air-Force (USAF) target, a central area of an ISO12233 chart and the image of the outdoor scene shown in Fig. 2. The increase in resolution and image quality is evident. A comparison of the Michelson contrast of the various bar targets in the images of the USAF target indicates an enhanced angular resolution of a factor of 2.25, as determined by the ratio of spatial frequencies yielding similar contrast. For this image, full-field, single-plane, Fourier-based correlation registration was used to increase robustness as finite, uniform range is assured and the edge-based content of the USAF resolution target provides insufficient interest points for the SIFT algorithm. Clearly, targets with suitable content for robust registration can also be used to calibrate the system for a particular plane. Calibration and reconstruction have been applied independently for each color channel using the green channel as the reference to produce the color super-resolved image. Color artifacts due to both Bayer-matrix sampling and lateral chromatic aberration are thus avoided, as can be observed in Fig. 3. Note that no image deconvolution has been included in the SR reconstruction to sharpen the reconstructed images. For the outdoor scene, the handrail, being displaced from a distant plane used for reconstruction is severely blurred. In principle three-dimensional scene reconstruction and segmentation would enable sharp accurate SR reconstruction throughout the image volume; however, this has not yet been attempted. Motion blurring of clouds occurs in the reconstruction due to the sequential recording of the data—this will be eradicated in a future snapshot implementation.

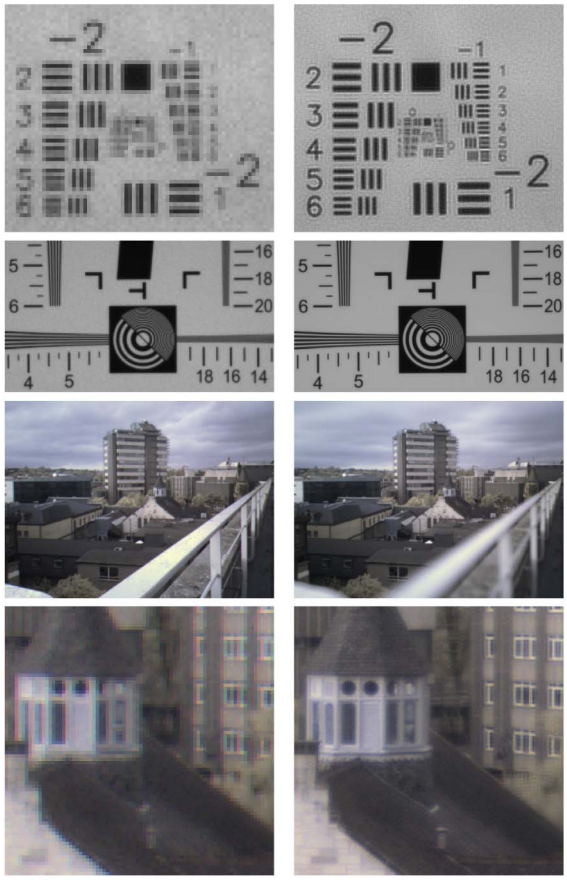

Fig. 3. From top to bottom: image of USAF target; portion of ISO12233 chart with edges; outdoor color scene; and detail of its central region. Images in the left are from a single camera and images in the right are super-resolved.

The achievable increase in resolution is ultimately limited by either the diffraction limit of individual cameras or by the increased effective sampling frequency achieved using multiple detector arrays, which in this case is a factor of five larger. In practice, the resolution increase is also determined by registration accuracy, optical aberrations, and degree of nonredundancy [8]. We consider now how experimentally determined image quality varies with increasing numbers of cameras. We have assessed image quality using the circular patterns shown in Fig. 3 since they give rise to clear aliasing artifacts. As a measure for attained quality we employ the root-mean-square error, $\varepsilon_{\mathrm{rms}}$, between the image restored using a subset of the 25 cameras and the image obtained using all 25 cameras (which is expected to be the highest quality image obtained and is our best estimate of the ultimate image quality). Note that pixel count is increased five times in each direction even though the system matrix in Eq. (3) uses less than 25 cameras. The variation of $\varepsilon_{\mathrm{rms}}$ with number of cameras, m, for up to 100 different randomly selected camera combinations for each value of m, is shown in Fig. 4. The best- and worst-case images

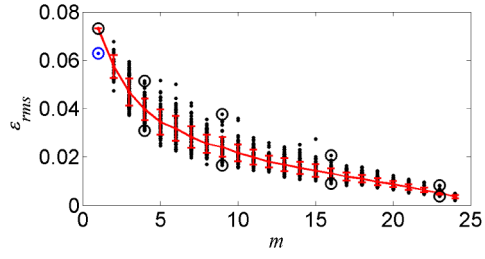

Fig. 4. Metric of aliasing against number of cameras, m, used in reconstruction. Mean and standard deviation (in red) for 100 randomly selected camera combinations are plotted. The blue point indicates bi-cubic interpolation. Reconstructed images corresponding to encircled points are shown in Fig. 5.

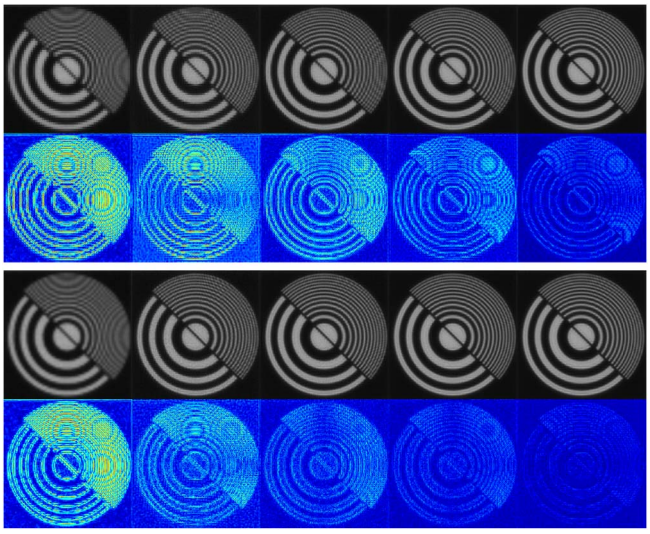

Fig. 5. Reconstructed images and their pseudo-colored error map compared to the 25 cameras reconstruction. Each column shows worst-case (upper two) and best-case (lower two) images out of 100 cases for 1, 4, 9, 16 and 23 cameras.

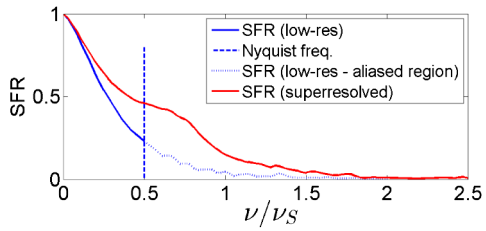

Fig. 6. SFR for low-resolution and super-resolved images.

corresponding to the encircled data points in Fig. 4 are shown in Fig. 5. The magnitude of aliasing artifacts is clearly reduced with increasing m and also by selection of the cameras statistically yielding best-case imaging. The variation between best-case and worst-case is associated with a better or worse set of sampling offsets for the selected cameras (all cameras were measured to be essentially identical). The general improvement in image quality with m is due mainly to the improved sampling afforded by a larger number of cameras. Noise reduction, evident in uniform parts of the images in the test targets of Fig. 3, however, is due to averaging.

To further assess the improvement in image quality, the spatial-frequency response (SFR) was calculated following the procedure described in ISO12233 [15] for single-camera and camera-array super-resolved images, and is shown in Fig. 6. Although in both cases the SFR is limited by diffraction, SR mitigates aliasing by an effective increase of the sampling frequency. Increasing numbers of iterations result in increased SFR enhancement with an associated increase in noise amplification.

The number of iterations and the number of cameras also affect the computational load. Implementation in Matlab using a current PC required 0.85 s per image for registration, 7 s per camera for system-matrix construction and 13 s for a 10-iteration image reconstruction.

In summary, our experimental results demonstrate SR from 25 independent co-aligned COTS cameras. Although nominally identical, camera-to-camera variations in magnification and pointing exist due to normal optical and electronic manufacturing tolerances. These effects are incorporated and mitigated by the calibration and image recovery processes. Algorithms for image registration have been implemented for system calibration and image reconstruction algorithms super-resolve images to yield an effective increase in resolution.

We found that image quality and consistency of image quality improved asymptotically with the number of cameras used in image reconstruction. Although in principle, a diffraction-limited enhancement in resolution by a factor of 4.5 is possible, the recovered image quality is comparable to a system with a factor of 2.25 enhancement. This appears to be limited by suppression of the MTF by the pixel response and optical aberrations of the low-cost COTS cameras, which in turn reduce registration accuracy. Improved SR, approaching the diffraction limit of the individual cameras [8], will require detectors with higher signal-to-noise ratios and lower aberration optics enabling also an associated improvement in image co-registration. This SR factor provides scope for a commensurate reduction in lens track length. The computational load for image reconstruction scales approximately linearly with pixels per camera and so much larger pixel counts, through combining high-resolution cameras, is feasible and attractive. Importantly, we have shown that with independent cameras, the angular resolution is not limited by the size of the detector array.

We acknowledge support from Qioptiq, St Asaph, Wales, the KTP, STMicroelectronics, Edinburgh, Scotland and EPSRC IDC in Optics and Photonics Technologies.

References

1. J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka, Appl. Opt. **40**, 1806 (2001).
2. L. Li and A. Yi, Appl. Opt. **51**, 1843 (2012).
3. A. Portnoy, N. Pitsianis, X. Sun, D. Brady, R. Gibbons, A. Silver, R. Te Kolste, C. Chen, T. Dillon, and D. Prather, Appl. Opt. **48**, 2115 (2009).
4. M. Shankar, R. Willett, N. Pitsianis, T. Schulz, R. Gibbons, R. Te Kolste, J. Carriere, C. Chen, D. Prather, and D. Brady, Appl. Opt. **47**, B1 (2008).
5. K. Choi and T. Schulz, Appl. Opt. **47**, B104 (2008).
6. A. Kanaev, D. Scribner, J. Ackerman, and E. Fleet, Appl. Opt. **46**, 4320 (2007).
7. G. Muyo and A. R. Harvey, J. Opt. A **11**, 54002 (2009).
8. J. Downing, E. Findlay, G. Muyo, and A. R. Harvey, J. Opt. Soc. Am. A **29**, 921 (2012).
9. D. G. Lowe, Int. J. Comput. Vis. **60**, 91 (2004).
10. M. A. Fisher and R. C. Bolles, Commun. ACM **24**, 381 (1981).
11. R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd edition (Cambridge University, 2004).
12. S. C. Park, M. K. Park, and M. G. Kang, IEEE Signal Process. Mag. **20**(3), 21 (2003).
13. S. S. Young and R. G. Driggers, Appl. Opt. **45**, 5073 (2006).
14. M. Elad and A. Feuer, IEEE Trans. Image Process. **6**, 1646 (1997).
15. International Standard ISO 12233:2000(E).