Benefits of optical system diversity for multiplexed image reconstruction

Hseuh-Ban Lan, Sally L. Wood, Marc P. Christensen, and Dinesh Rajan

Algorithms that use optical system diversity to improve multiplexed image reconstruction from multiple low-resolution images are analyzed and demonstrated. Compared with systems using identical imagers, systems using additional lower-resolution imagers can have improved accuracy and computation. The diverse system is not sensitive to boundary conditions and can take full advantage of improvements that decrease noise and allow an increased number of bits per pixel to represent spatial information in a scene.

OCIS codes: 100.3010, 100.6640, 100.3190, 110.0110.

1. Introduction

Computing high-resolution images from multiple low-resolution images is a well-known, ill-posed, and computationally intensive problem[1–3] that has recently become of interest in two application areas. A high-resolution camera may be considerably more expensive than a lower-resolution camera with only 10% to 25% as many pixels. If there is adequate time to capture multiple slightly offset images of the same stationary objects, it is possible that a high-resolution image could be obtained from a low-cost camera such as a web camera or television camera.[4–6] In this application the relative offsets caused by camera motion are typically computed by using image correlation rather than by relying on camera position and orientation information.

A second application area is the design of a camera with a flat form factor, achieved by using smaller optical elements that must be positioned closer to the image detectors.[7–11] In this case the low-resolution images are a result of moving the imaging elements closer to the sensor array, resulting in a physically smaller image at the detector for the same field of view (FOV). The individual detectors in the array cannot be proportionately reduced in size because of manufacturing constraints and increased noise levels as the detector size is reduced, so computational imaging methods must be used to create a higher-resolution image. The camera design may have fixed[7,8] or steerable[9–11] imaging resources that will determine the relative offsets of the low-resolution images.

Radio astronomy (e.g., Ref. 12) and medical imaging (e.g., Ref. 13) have a long and successful history of computing high-resolution reconstructed images from sampled data measurements with lower-resolution in some dimensions. In these cases the resulting image is often used as input for further computation or detection algorithms as well as for viewing, so the number of bits per pixel in the computed image will often be limited by the noise environment rather than by the observer viewing requirements. Methods of error analysis, system design, and performance specifications used in these fields can be applied to current flat-camera design problems of image reconstruction accuracy and efficient computational methods.

In both the flat-camera application and the single movable low-resolution camera application, results have been demonstrated with resolution improved by factors in the range of 3 to 5. For example, the thin observable module by bound optics (TOMBO) system[8] used a conventional 240 × 240 pixel sensor array divided into a 6 × 6 array of 40 × 40 pixel subarrays, each of which had a 500 μm aperture microlens with $f = 1.3$ mm. With the object distance fixed at 26 cm, a resolution improvement of a factor of 4 was measured, although the design parameters suggest that a factor of 5 or 6 is possible. How-

H.-B. Lan and S. L. Wood (swood@scu.edu) are with the Department of Electrical Engineering, Santa Clara University, 500 El Camino Real, Santa Clara, California 95053, M. P. Christensen and D. Rajan are with the Department of Electrical Engineering, Southern Methodist University, 6251 Airline Road, Dallas, Texas 75275.

Received 24 August 2005; accepted 25 October 2005; posted 11 January 2006 (Doc. ID 64324).

0003-6935/06/132859-12$15.00/0

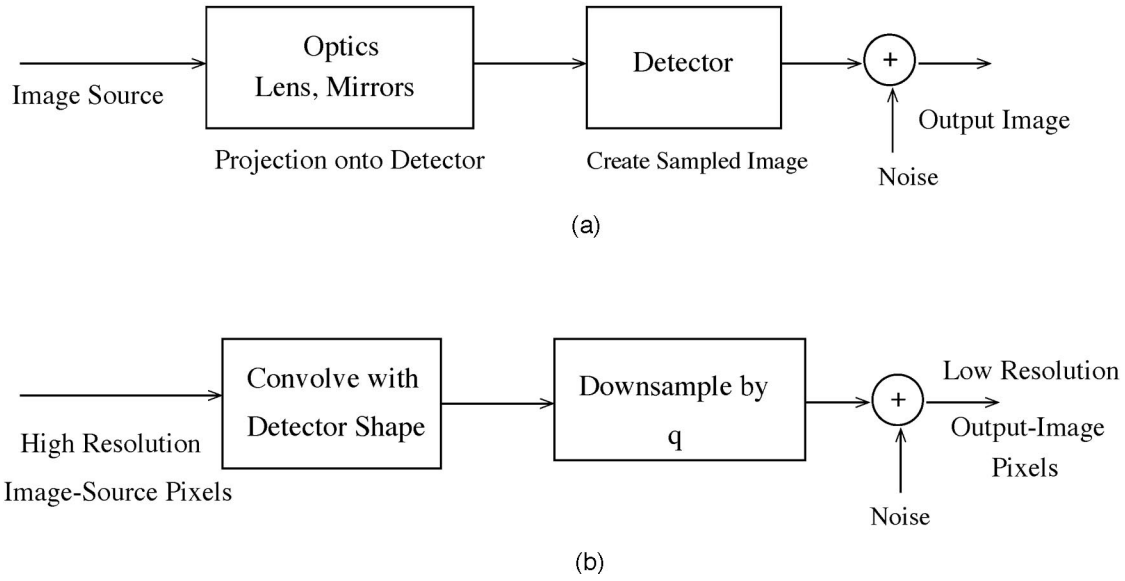

Fig. 1. Basic block diagrams for (a) the physical optical system and (b) the corresponding mathematical model.

ever, the pseudoinverse-based reconstruction method caused undesirable noise amplification in the reconstructed images.

The design of manufacturable camera geometries and associated reconstruction algorithms must be concerned with the accuracy of the reconstructed image and the computational burden of creating it. The ill-posed nature of the inverse problem requires methods that use almost all observed image data for the computation of each reconstructed pixel. Usually the image quality is measured in terms of the actual or expected mean-square error (MSE), which is a more suitable measure for image information available for postprocessing than for viewing quality. The visual effect of a particular MSE level can vary dramatically, depending on the local contrast of the desired image and the spatial-frequency content of the error. However, for applications in which computational methods for feature detection or segmentation will be used, the MSE will be more closely related to overall performance. In these cases increasing the number of bits per pixel above 8, as is typical in medical imaging, may be important.

This paper describes a computational imaging approach to reconstructing high-resolution images by using magnification diversity in the low-resolution image-acquisition architecture. This results in better expected and actual performances when measurements are made in terms of MSE. Diversity also results in computational benefits because good-quality local reconstructions of small tiles from the full image can be computed independently, leading to a reduction in computation and natural parallelization opportunities. In addition, noise propagation is greatly reduced by using a minimum variance estimator (MVE) with magnification diversity. When the noise associated with the acquisition of low-resolution images can be reduced, methods that use diversity can provide more bits per pixel in the computed high-resolution image. The mathematical notation and analysis of the expected error performance for arrays of low-resolution images are derived in Sections 2 and 3, respectively. The improvements expected from diversity are analyzed and demonstrated in Section 4, and the performance of an optical system with mild distortion is explored in Section 5.

2. Image Data Model

Figure 1 shows block diagrams of a basic physical optical model of digital image formation and the corresponding mathematical model used in this paper. The input to the first block of the physical model is a spatially continuous image source. This single block represents the projection of the image source by using combinations of lenses and mirrors, such as those described in Refs. 7–10 onto a detector array of discrete sensor elements. In the second block each pixel in the sensor array integrates light intensity spatially over the finite physical size of the sensor element and temporally over the image-acquisition time. Measurement noise from various sources is added to the spatially and temporally sampled output image read from the detector array.

The corresponding abstracted mathematical block diagram shown in Fig. 1(b) also uses two blocks that are both part of the physical detector block above it. Here the input image is modeled as a continuous image source that is a pixel array at the resolution of the desired output. Although the actual desired pixel size or angular resolution of the image source will be measured in milliradians, in this mathematical model the pixel size is measured after projection onto a detector array in the same units used for the detector pixels. The first block represents the convolution of the high-resolution input image with an averaging filter whose width and shape are determined by the detector pixel size. If the point-spread function of the optical system is significant compared with the de-

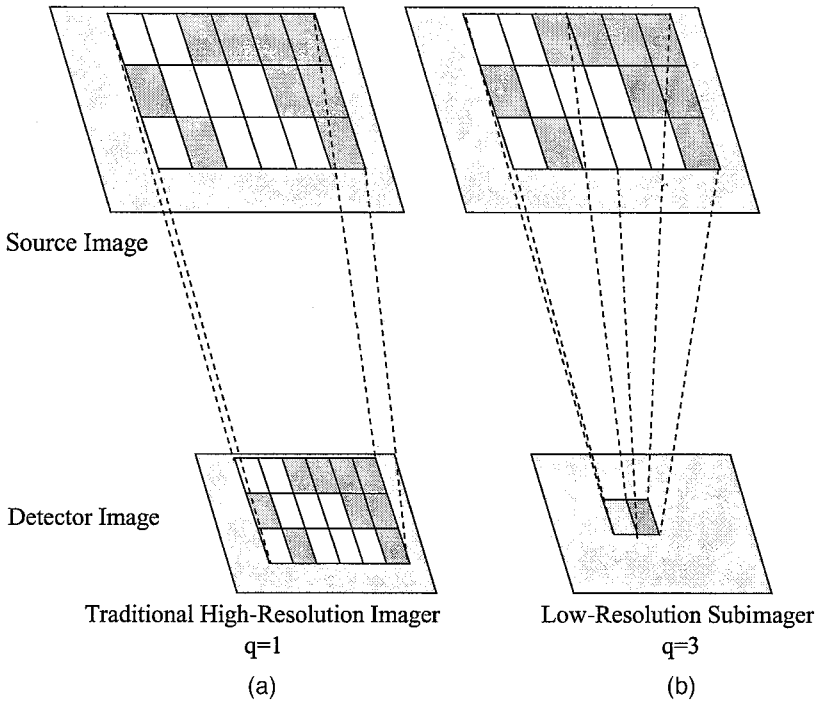

Fig. 2. Projections of source-image pixels onto detector arrays.

tector size, that convolution would also be included here. However, it is assumed that the diffraction-limited spot size of the optical system is ~5 times smaller than the detector size[9,10] and the blur discussed in this paper is caused by the finite detector size, not the resolution limitations of the optical system. The convolution result is the input to the second block, which implements a sampling operation to select values at the detector pixel centers. Figure 2(a) schematically shows a projection in which the detector pixels have the same resolution as the source pixels, while in Fig. 2(b) nine source pixels are averaged by each detector pixel.

The three main objectives of computational imaging analysis are to recover the best estimate of the source image from the image or images captured from the detectors to define performance limits and to improve the optical system design by making it more efficient in producing information useful for the image reconstruction. Let F represent the two-dimensional desired source image modeled by a pixel array at the desired resolution with N_y rows and N_x columns. Let $N = N_yN_x$ be the total number of source pixels. For computation and analysis, the source-image array F is stored by rows in an $N \times 1$ column vector f, with source pixel value $F(n_y, n_x)$ stored in $f(n_yN_x + n_x)$ for $0 \le n_y \le N_y - 1$ and $0 \le n_x \le N_x - 1$. Similarly, the $M_y \times M_x$ detector image, G, will be stored by rows in an $M \times 1$ column vector, g, where $M = M_yM_x$.

The characteristics of the image-acquisition geometry, the behavior of the optical elements, the integration and sampling effects of the detector pixels, and the relative physical positions and sizes of the detector pixels and projected source pixels are all used to compute a set of coefficients that define the relative contribution of each source-image pixel to each detector pixel. These coefficients form the observation partial matrix H. A linear mathematical model of the measurement system appropriate for both space-invariant and space-variant imaging systems is given as

$$g = Hf + v. \tag{1}$$

Although shift invariance is not assumed in this superposition sum model, intuitive understanding is often obtained from considering the more restricted case in which H represents a separable convolution sum. An $M \times 1$ noise vector v may include uncorrelated random noise from a variety of sources.

A MVE $\hat{f}$, can be computed by using Eq. (2) (e.g, Ref. 14) if the measurement noise and the class of possible image sources can be reasonably modeled as Gaussian random variables:

$$\hat{f} = f_0 + K(g - Hf_0). \tag{2}$$

The initial estimate of f is $\hat{f}_0 = f_0$, where $f_0 = \mathbb{E}(f)$ is the expected value f. This initial estimate is updated based on the measurement prediction error in Eq. (3) by using the reconstruction matrix K in Eq. (4) to produce a MVE estimate. The matrix K is computed by using estimates of the noise covariance, $R_v = \mathbb{E}(vv^T)$, and the covariances of the image class, $P_0 = \mathbb{E}[(f - f_0)(f - f_0)^T]$. These estimates are indicated

by $\hat{R}_v$ and $\hat{P}_0$:

$$\tilde{g}_0 = g - \hat{g}_0 = g - Hf_0, \tag{3}$$

$$K = \hat{P}_0 H^T \left(H\hat{P}_0 H^T + \hat{R}_v\right)^{-1}. \tag{4}$$

The derivation of Eq. (4) assumes that the noise has a zero-mean value and that there is no correlation between the noise and the image data so that $\mathbb{E}(fv^T) = 0$.

Equation (2) explicitly allows f to have a nonzero-mean value. If it is assumed that the detector images are always adjusted by Hf_0, as explicitly shown in Eq. (2), then simpler versions of the equations that assume a zero-mean f can be used, and f_0 can be simply added to the computed estimate to produce $\hat{f}$.

Simulations can demonstrate how good an estimate of f is created by using the reconstruction matrix K, but it is also useful to consider the analytical expressions for the expected squared error by using the covariance of the estimate error $\xi = \mathbb{E}(\tilde{f}\tilde{f}^T)$, where $\tilde{f} = f - \hat{f}$. The variance of each individual pixel in $\tilde{f}$ can be computed by using Eq. (5). This expression is valid for any matrix K, and it explicitly allows an actual noise covariance R_v and image class covariance P_0 to be different from the estimated values $\hat{R}_v$ and $\hat{P}_0$. The diagonal of the matrix ξ contains the expected variance for each element of the error vector $\tilde{f}$:

$$\xi = \mathbb{E}\left(\tilde{f}\tilde{f}^T\right) = (I - KH)P_0(I - KH)^T + KR_vK^T. \tag{5}$$

These equations provide a method for comparing the expected and actual performances of ideal and simplified estimators as a function of the actual noise variance. For the specific case in which K is computed from Eq. (4) by using accurate models so that $\hat{R}_v = R_v$ and $\hat{P}_0 = P_0$, Eq. (5) can be simplified to the following:

$$\xi = P_0 - P_0 H^T \left(HP_0 H^T + R_v\right)^{-1} HP_0. \tag{6}$$

A. Traditional Imaging System

These performance measures can be applied to a traditional high-resolution imaging system in which the sensor pixel size is matched to the desired resolution of the projected image, and there is a one-to-one correspondence between the elements of f and the elements of g. This is illustrated in Fig. 2(a) Typically it is simply assumed that $\hat{f} = g$, which implicitly assumes that the image source is projected onto the sensor array with correct focus, that H is the $N \times N$ identity matrix I_N, and that the measurement noise is zero. With knowledge of P_0 and R_v, Eqs. (2) and (4) can be used to compute the reconstruction matrix K for the MVE as follows:

$$K = \hat{P}_0 \left(\hat{P}_0 + \hat{R}_v\right)^{-1}. \tag{7}$$

For the special case in which the individual noise samples are uncorrelated, the covariance matrix R_v can be written as a scaled identity matrix $\sigma^2 I_M$. If $P_0 = p_0 I_N$ then $K = [p_0/(p_0 + \sigma^2)]I_M$, where $M = N$. For the MVE, the values of σ^2 and p_0 will determine the best relative weighting of the observed pixel values and the mean value of f.

B. Resolution Improvement with a Single Image

The pixel resolution of the sensor image g may be lower than the desired resolution of $\hat{f}$. With only one image, resolution improvement through restoration relies on implicit or explicit *a priori* assumptions that may not always be appropriate. Interpolation can be used to make the resolution appear to be improved, but this does not restore frequencies that have been zeroed by the spatial averaging of individual detectors and does not repair aliasing artifacts.

A common model for the low-resolution sensor image assumes that each detector in a sensor array averages a $q \times q$ pixel area from the desired high-resolution source image, which is perfectly focused on the sensor arrays. For simplicity, it will also be assumed that the pixel grid boundaries of G are aligned with every qth pixel boundary of F, although allowing relative translations and rotations does not fundamentally change the results. Now the desired number of source pixels is $N = q^2 M$. The response of a single pixel of the detector is modeled by a corresponding row of the $M \times N$ rectangular matrix H. Each row will have q^2 nonzero elements, each with a value of $1/q^2$, positioned to correspond to the q^2 elements of F in the FOV of the detector pixel. Because the optics produces a small point-spread function relative to a detector pixel size, there will be no significant overlap in the FOV of any pair of detectors, and the dot product of any pair of rows of H will be zero. This imaging system with one low-resolution detector array will be called a subimager (SI).

With the same assumptions used for the traditional high-resolution imaging system with respect to R_v and P_0, we have $HP_0H^T = (p_0/q^2)I_M$. The resulting expression for K is given by

$$K = p_0 I_N H^T \frac{q^2}{p_0 + q^2\sigma^2} I_M = \frac{p_0 q^2}{p_0 + q^2\sigma^2} H^T. \tag{8}$$

In the noiseless zero-mean case $\hat{f} = q^2 H^T g$. Since the nonzero values in H are all $1/q^2$, the reconstruction simply backprojects the value of each detector pixel onto the q^2 source pixels in its FOV. This is sometimes called zero-order interpolation. As measurement noise increases, $\hat{f}$ is reduced, and increasing the value of q increases the noise term in the denominator. Using only image data from one SI, we cannot distinguish any structures within the FOV of a detector pixel, we cannot repair aliasing artifacts due to the lower spatial sampling rate, and we lose all information about the presence of spatial frequencies

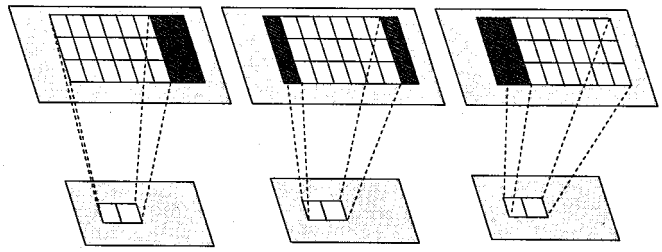

Fig. 3. Projections of source-image pixels onto three different SI detector arrays.

with an integral number of periods equal to the width of a detector.

3. Arrays of Subimagers

An array of identical SIs with lateral shifts of FOV equal to the desired resolution can provide information needed to resolve some structure at the desired resolution and avoid aliasing due to the reduced sampling rate. For simultaneous image capture, this array of SIs with lateral shifts of FOV can be accomplished in a fixed geometry such as the one used for the TOMBO system.[7,8] However, the fixed geometry requires a fixed distance between the image source and the camera for a specific resolution improvement. To avoid this limitation, one can make a SI array with a more flexible steerable imaging system such as the PANOPTES (processing arrays of Nyquist-limited observations to produce a thin electro-optic sensor) system.[9,11] For multiple images of a stationary scene, the array of low-resolution images can be obtained from a single SI if there is physical motion of the camera between acquisition of the images.[1,6,5]

The mathematical representation of a single SI in Section 2 can be extended to represent an array of SIs. Let $H_{0,0} = H$ represent a single SI in which the detector FOVs are not overlapping and that each detector averages a $q \times q$ pixel area of the desired image. A $q \times q$ array of SIs with FOVs offset from each other by the width of one source pixel will have a combined FOV that is extended by $q - 1$ pixels horizontally and vertically with $N_{xE} = N_x + q - 1$ and $N_{yE} = N_y + q - 1$. The number of pixels in this extended FOV will then be $N_E = N_{xE}N_{yE}$. This is shown in Fig. 3 for three small SIs. On the left, a 3×3 pixel region of f is projected to each of two pixels in the SI array in the same manner as shown in Fig. 2(b) except that the size of F has been extended by two pixels on the right. The two additional shifted SIs each have a FOV shifted by the width of one source pixel, and the combined FOV of all SIs is now 3×8 pixels. Note that, for the source model used here, the subimage from a FOV shifted by a fraction of a source pixel can be written as a weighted sum of two subimages with an offset of exactly the width of one source pixel, so there is no loss of generality in assuming the regular spacing of SI FOVs.

To represent these observations mathematically, we define a shift matrix Z_l as an $N \times N_E$ matrix in which $Z_l(i, j) = \delta[i - (j - l)]$, where $\delta(k) = 1$ for

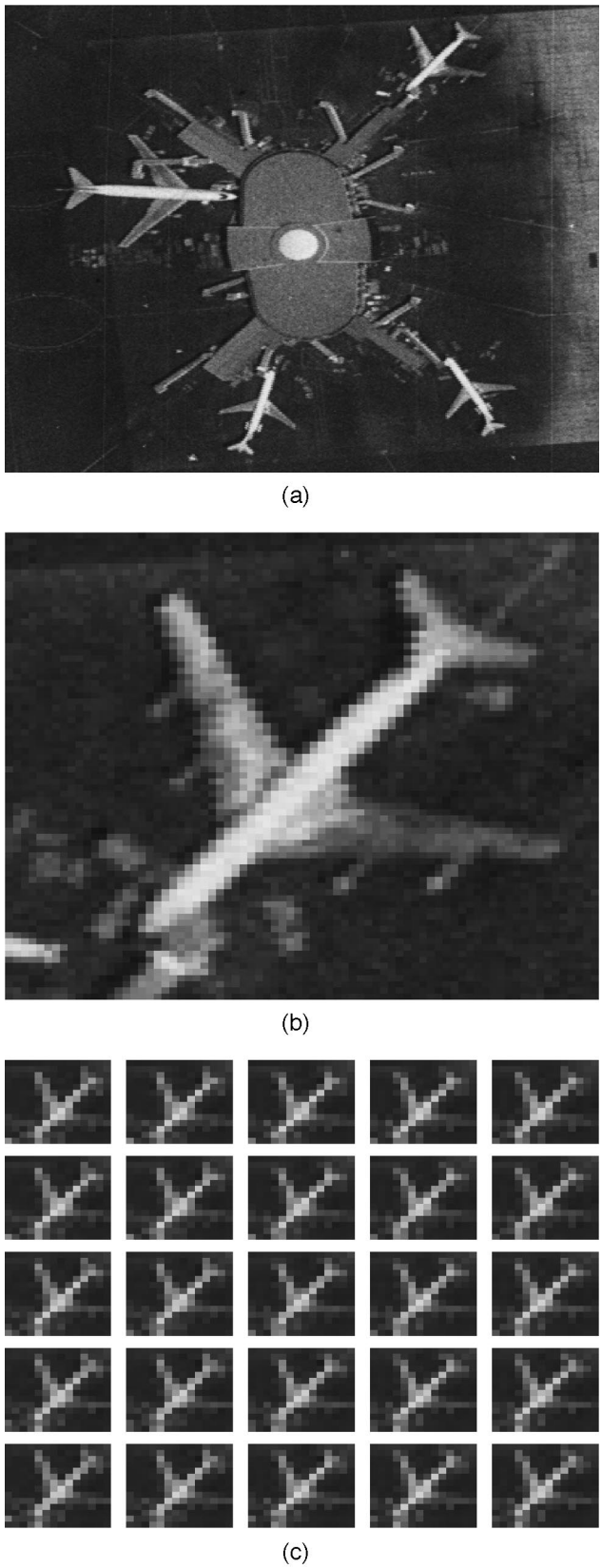

Fig. 4. High-resolution 400×400 pixel image of (a) an airport (from the University of Southern California Signal and Image Processing Institute Image Database), (b) the image tile 76×76 pixels, (c) the SI outputs for $q = 5$.

$k = 0$ and has a value of zero for all other values of k. If the H matrix of a SI is postmultiplied by Z_l, where $0 \le l \le q - 1$, all columns of H will be shifted by l positions to the right, and vectors of zeros will fill the first l columns. This postmultiplication will effectively shift the FOV of the SI by l pixel positions in the f vector and will place the SI's FOV correctly in the extended composite FOV of all SIs. Row shifts of k rows will require a multiplication by $Z_{kN_{xE}}$ for the two-dimensional image data stored in f in vector form.

Let the vector $g_{k,l}$ represent the detector image from the (k, l)th SI. Using the measurement model of Eq. (1), we can write the individual equations for each SI:

$$\begin{aligned} g_{0,0} &= HZ_0 f + v_{00} = H_{0,0} f + v_{0,0}, \\ g_{0,1} &= HZ_1 f + v_{01} = H_{0,1} f + v_{0,1}, \\ g_{1,0} &= HZ_{N_{xE}} f + v_{10} = H_{1,0} f + v_{1,0}, \\ &\vdots \\ g_{q-1,q-1} &= HZ_{(q-1)(N_{xE}+1)} f + v_{q-1,q-1} \\ &= H_{q-1,q-1} f + v_{q-1,q-1}. \end{aligned} \tag{9}$$

The total number of pixels from the $q \times q$ array of M-pixel SIs is equal to q^2M, which is the same as N. So the number of observed pixels is now the same as the number of pixels in a traditional high-resolution imager. However, for the array of SIs, the total FOV is extended because no assumptions are made about boundary conditions. Figure 4(b) shows a 74×74 pixel high-resolution image tile of an airplane taken from the larger source image shown in Fig. 4(a). Figure 4(c) shows the 25 reduced-resolution 14×14 pixel images obtained from a SI array by using $q = 5$. The loss of significant image information in the individual low-resolution images is readily apparent. A reconstruction algorithm is needed to compute an estimate of the high-resolution image from the array of low-resolution images.

The column vectors of image outputs from all the SIs in Eq. (9) can be combined into a single array of data with Eq. (10), in which column structures are written in a more compact transposed row form. As shown here, the vector of SI data is sorted first by SI and then by rows within one subimage. A subscript q has been added to the combined H matrix to indicate matrices for a $q \times q$ array of SIs with resolution reduced by a factor of q:

$$\begin{aligned} g &= H_q f + v, \\ g^T &= \left[g_{0,0}{}^T,\ g_{0,1}{}^T,\ \ldots,\ g_{q-1,q-1}{}^T \right]^T, \\ H_q{}^T &= \left[H_{0,0}{}^T,\ H_{0,1}{}^T,\ \ldots,\ H_{q-1,q-1}{}^T \right]^T, \\ v^T &= \left[v_{0,0}{}^T,\ v_{0,1}{}^T,\ \ldots,\ v_{q-1,q-1}{}^T \right]^T. \end{aligned} \tag{10}$$

The expected squared error for the MVE reconstruction can be estimated if the singular values of H_q are known. Then H_q can be written in terms of its singular-value decomposition or spectral representation[15] $H_q = USV^T$, where U and V are $N \times N$ and $N_E \times N_E$ unitary matrices and S is an $N \times N_E$ matrix with singular values along the diagonal. Then Eq. (5) can be rewritten by using that form. For simplicity, assume as before that $R_v = \sigma^2 I_N$ and $P_0 = p_0 I_{N_E}$. Then $H_q P_0 H_q{}^T = p_0 USS^T U^T$ and, using Eq. (6), $\xi = V \Lambda V^T$, where the elements of the $N_E \times N_E$ diagonal matrix Λ are given by Eq. (11). The elements of Λ will determine the average expected squared error when $\hat{f}$ is computed by using the MVE with the detector images in Eq. 10:

$$\lambda_i = \begin{cases} p_0 \sigma^2 / \left(p_0 s_i{}^2 + \sigma^2 \right) & \text{for } 1 \le i \le N \\ p_0 & \text{for } N + 1 \le i \le N_E \end{cases}. \tag{11}$$

Analysis of the expected errors for the array of subimages is described more intuitively if the combined data vector g is reordered with the observed pixels for a row of SIs interleaved so that the overlapping individual FOVs of adjacent detector pixels are shifted by the width of one source pixel. Thus the first three elements of the reordered g would be $[g_{0,0}(0, 0), g_{0,1}(0, 0), g_{0,2}(0, 0), \ldots]^T$ rather than $[g_{0,0}(0, 0),\ g_{0,0}(0, 1), g_{0,0}(0, 2), \ldots]^T$. With this ordering, g can be interpreted as the output of a single array of detector pixels ideally sampled at intervals equal to the desired source pixel size. However the samples are taken from a source image blurred by convolution with a rectangular function of the same size as the actual detector prior to projection onto the detector.

In this form a one-dimensional convolution would create a Toeplitz matrix H and the two-dimensional separable convolution would create a Toeplitz–block–Toeplitx matrix H.[15] As the size of a Toeplitz matrix increases, its eigenstructure approaches the eigenstructure of a circulant matrix[16] in which the eigenvalues are the discrete Fourier transform (DFT) of the first row and the eigenvectors are DFT vectors. The eigenvalues for the H matrix for a two-dimensional convolution in this case will be the product pairs of the eigenvalues for an H matrix for one-dimensional convolution.[15] This provides a simple method for estimating the performance of an array of SIs based on the size of the reconstructed image tile, the value of q, and the amount of measurement noise. Exact computations of the expected squared error without the circulant structure assumption are given in Section 4, and these results are consistent with the analytical approximation given here. Note that the DFT values will be zero for spatial frequencies with an integral number of periods in one detector width and that, when boundary conditions are not known, any amount of these spatial frequencies can be added to $\hat{f}$ without changing the predicted values of g[11,17]. This will limit the performance of an array of identical SIs as described above,

and reducing measurement noise, σ^2, below the point at which most of the error is due to zero-valued eigenvalues will not continue to improve the reconstructed images.

4. Performance Improvements with Diversity

In Section 3 we analyzed the expected performance of arrays of identical shift-invariant SIs when the FOV of adjacent SIs is shifted by one pixel width of the desired high-resolution image. Adding more identical SIs with subpixel shifts of FOV can provide more data for noise averaging, but since this does not change the spatial frequencies that will produce a response of zero, the additional images do not improve the observability of the desired image. Even if the simple uniform averaging model for the detector pixel response may be modified somewhat, the detector pixel will still perform an averaging function over its surface and will have low sensitivity to some spatial frequencies, which will limit reconstruction accuracy.

In contrast, adding new SIs with different ratios between the detector pixel size and the projected desired pixel size can provide new information about image components with spatial frequencies that have a weak response at the detector pixels of the first set of SIs. Consider a one-dimensional system in which a uniform average is taken over w, the width of the one-dimensional detector pixels. The spatial-frequency response due to this averaging will be a sinc function with zeros at integer multiples of radian frequency $\Omega_0 = 2\pi/\omega$. If a second set of wider detectors is added with a length that is not an integer multiple of the first length, the sinc function defining its response will have no zeros in common with the first detectors. Using subimages from both detectors will provide information about all spatial frequencies, and the accuracy of the reconstruction will be limited only by the measurement noise level. For a two-dimensional imaging system, a third set of detectors must be added to avoid loss of two-dimensional spatial-frequency components.

This optical system diversity can come from the diversity in magnification factors of SI optical systems or the diversity in the physical size of the SI sensor pixels using the same optics. In the former case the distance between the new optical components and the detector arrays would be less than the distance for the original SIs, which would cause the size of the projected image to decrease. In the latter case the increased size of the detector pixels would result in lower values of measurement noise, which would help offset the performance disadvantage of the increased value of q. For both cases, this diversity will be referred to as magnification diversity in the context of the ratio of the physical detector width and the width of the projected pixels at the desired resolution. The added SIs alone will have poorer performance than the SIs at the original magnification, but they also add information that was not observable by the original array.

The additional information from SIs with magnification diversity also affords significant computational efficiency. Reconstructions involving deconvolutions of shift-invariant averaging functions usually improve in performance as the width of the image is increased, so all observations are used in the computation of each value of $\hat{f}$. With magnification diversity, reconstructions can be computed from local measurements and there is much less improvement gained by including a larger FOV than that found with a single magnification. These smaller subsets of the desired image that are independently reconstructed will be called tiles. Reconstructing the smaller tiles reduces the amount of computation needed because of the smaller image size and also allows straightforward use of parallel computation.[17,11]

In the case of the single-magnification array, frequencies close to the nulled frequency gain some visibility as the FOV widens. For a multiple-magnification system, the visibility of those frequencies comes from the strong response of one array of SIs to a frequency that is nulled by the other array of SIs, and increasing the FOV is not as necessary.

The performance improvements due to magnification diversity can be analyzed by using an approach similar to that used for the SI array in Section 3. An H_q matrix for each magnification is generated, and then the observed images for arrays of subimages at the three magnification levels are used together as indicated in Eq. (12).

$$g = \begin{bmatrix} g_{q1} \\ g_{q2} \\ g_{q3} \end{bmatrix} = \begin{bmatrix} H_{q1} \\ H_{q2} \\ H_{q3} \end{bmatrix} f + v = Hf + v. \tag{12}$$

The performance of this new combined system can be analyzed by using Eq. (13), which is an alternate form of the estimator in Eqs. (3) and (4).[14] If $H^T = [H_{q1}{}^T H_{q2}{}^T H_{q3}{}^T]$, then $H^T R_v{}^{-1} H$ can be written as shown in Eq. (14):

$$\hat{f} = \left(H^T \hat{R}_v{}^{-1} H + \hat{P}_0{}^{-1}\right)^{-1} \times \left(H^T \hat{R}_v{}^{-1} g + \hat{P}_0{}^{-1} f_0\right), \tag{13}$$

$$H^T \hat{R}_v{}^{-1} H = H_{q1}{}^T \hat{R}_{v1}{}^{-1} H_{q1} + H_{q2}{}^T \hat{R}_{v2}{}^{-1} H_{q2} + H_{q3}{}^T \hat{R}_{v3}{}^{-1} H_{q3}. \tag{14}$$

Using the same approximation as in Section 3,[15] it can be argued that, for a tile large enough for use of the DFT approximation, the eigenvectors of all three $H_{qi}{}^T H_{qi}$ matrices will be the same. This would allow the eigenvalues of the sum of the matrices in Eq. (14) to be computed by simply adding the eigenvalues of each of the $H_{qi}{}^T H_{qi}$ matrices scaled by the appropriate inverse noise variance. This supports the intuitive concept of adding new types of SIs to capture all spatial frequencies. For the small tiles where the circulant approximation is less appropriate, computation of the exact average expected squared error shows the same behavior predicted by the mathemat-

ical analysis. In Section 5 we will show that this improvement from diversity also occurs when the H matrices represent space-variant projections that cannot be represented as convolutions.

For these computations of the average expected squared error, it is assumed that the *a priori* image set variance is the same as the variance of a continuous uniform distribution from 0 to 256. This range was chosen to make results consistent with the often-used 8-bit-pixel image-display range rather than normalizing the range. However, setting this range does not fix the number of bits per pixel at 8. If noise levels are low enough, the number of significant bits per reconstructed pixel can be increased to provide more information about the imaged scene for later computational analysis. The additional bits would be used to represent fractional pixel values rather than to increase the range of values.

Analysis of the spatial distribution of expected squared-error values for individual pixels found along the diagonal of ξ shows that the expected errors are much larger at the edges of an image tile than inside the tile.[11,17] To present meaningful expected squared-error results, we have assumed that all reconstructed tiles have the edge pixels trimmed off before the average MSE is computed. This would require a reconstruction approach that allowed a small overlap of tile edges when a large image is divided into small tiles for reconstruction.

Figure 5(a) compares the expected squared errors for a reconstruction obtained by two different sets of SI arrays. In each case four different estimators are computed for four different expected noise environments with estimated noise variances, $\hat{\sigma}^2$, equal to 10^{-6}, 10^{-4}, 0.01, and 1.0. The performance is shown for actual noise variances, σ^2, from 10^{-8} to 100.0. Results for three sets of 3×3 SI arrays with $q = 3$ are shown by using dashed curves. Reducing the actual measurement noise variance below 10 brings little improvement to any of the four estimators, which at the resolution of this graph are not distinguishable. In the second case, shown by the solid curves, one 3×3 array of SIs has $q = 3$. The two other SI arrays use a lower resolution with $q = 4$ and $q = 5$. The same actual measurement noise variance is assumed for all three magnifications, and the total number of observations is approximately the same as used in the case with $q = 3$ only. By using three different magnification factors we can improve the performance even though the added systems individually have poorer performance than the SI array with $q = 3$. As the actual noise variance is decreased, the performance of these estimators improves until it is limited by the values of the estimated noise variance, which protects the estimator from propagating large noise components in the reconstruction. It should be noted that integer values for q have been selected to demonstrate the value of magnification diversity with a simple example, but that as long as the q values are significantly different and the ratios of the

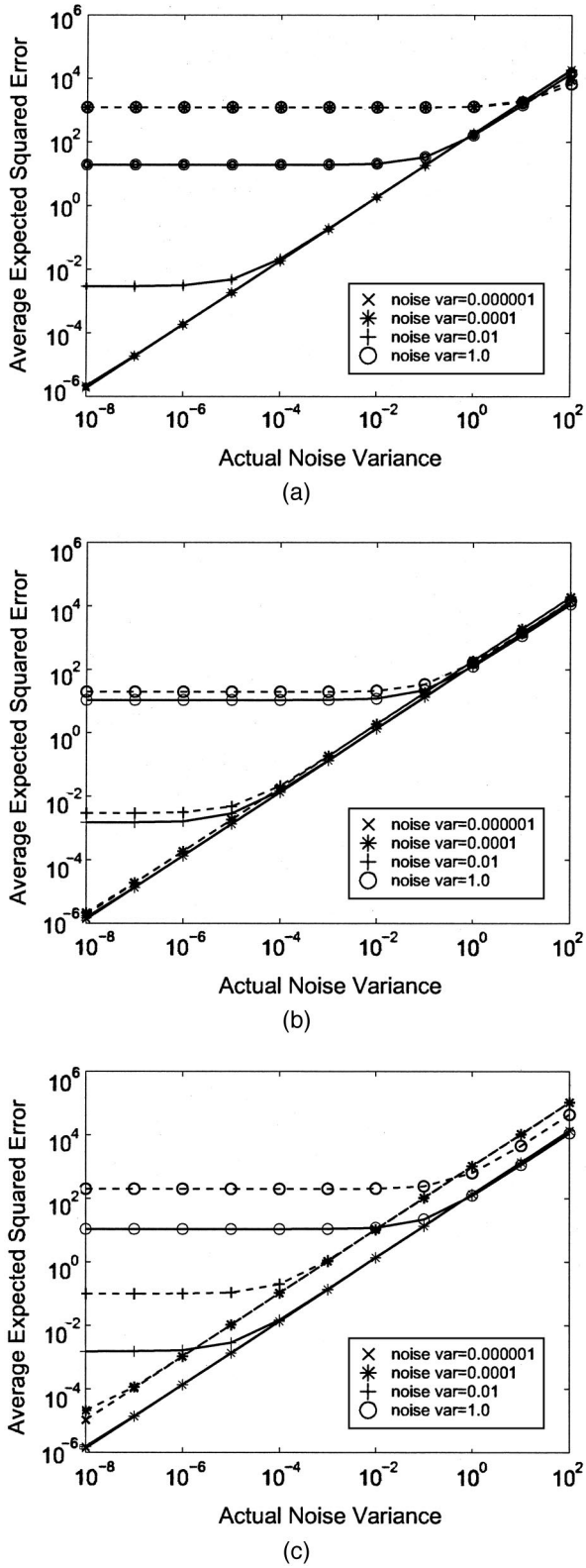

Fig. 5. Expected squared error for (a) 19×19 pixel tile from detectors with $q = 3$ (dashed curves) and detectors with $q = 3$, 4, and 5 (solid curves); (b) 19×19 (dashed curves) and 35×35 pixel tiles (solid curves); (c) detectors with $q = 5$, 6, and 7 (dashed curves) and $q = 3$, 4, and 5 (solid curves).

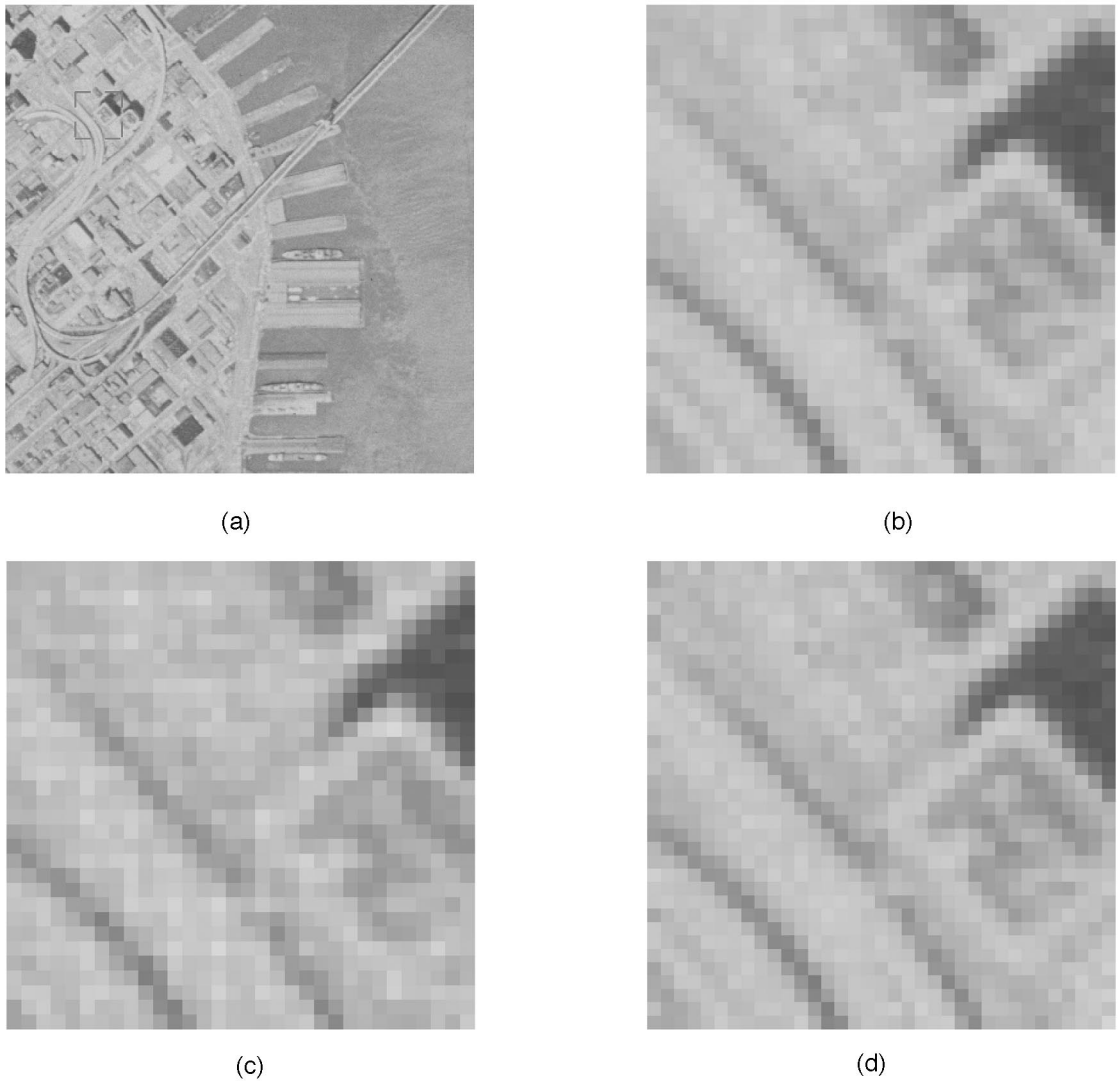

Fig. 6. (a) Image of Bay Bridge (from the University of Southern California Signal and Image Processing Institute Image Database) and (b) 35×35 pixel tile. Reconstructions using a noise variance of 0.01 for (c) $q = 3$ and (d) $q = 3, 4, 5$.

q values are not integers, there is no requirement that the q values be integers.

Figure 5(b) compares the expected squared error for a set of three SI arrays when the size of the tile is roughly doubled to a 35×35 pixel FOV in the source image, shown by solid curves, from a 19×19 pixel FOV, shown by dashed curves. The results for the smaller tile size were also shown in Fig. 5(a) as solid curves. The expected squared error is reduced slightly but not dramatically when the tile size is increased. The same improvement can be obtained from a second stage of computation by making low-rank corrections to the estimate from the smaller tile.[11] This shows that the large computational benefit of reducing the size of the independently reconstructed tiles does not cost much in terms of reconstruction accuracy when magnification diversity is used.

Figure 5(c) compares the expected squared error when the resolution of all the SIs is reduced. Results from a set of SIs with $q = 3$, 4, and 5 and a 35×35 pixel FOV are shown by solid curves here and also in Fig. 5(b). Corresponding results for a set of SIs with $q = 5$, 6, and 7 and a 41×41 pixel FOV are shown by dashed curves. As expected from earlier results, when the value of q increases, the effect of the noise is also increased, and lower measurement noise is needed to achieve the same reconstructed accuracy for higher values of q.

Actual reconstructed tiles from an aerial view of the Bay Bridge are shown in Fig. 6. In Fig. 6(b) a 35×35 pixel tile from the larger image is shown. Using the same number of measurements and the same added noise levels, we reconstructed the image in Fig. 6(c) from the SI arrays with $q = 3$ and the image in Fig. 6(d) with a mixed magnification set of SIs by using $q = 3$, $q = 4$, and $q = 5$. The small features in Fig. 6(d) with low contrast are clearly better defined than in Fig. 6(c), while the high-contrast large edges show only a small improvement in Fig. 6(d) compared with Fig. 6(c). This demonstrates the advantage of reducing the noise level in

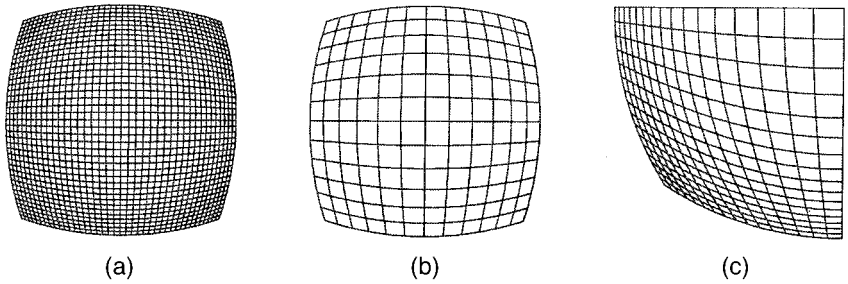

Fig. 7. (a) Projection of a grid through a lens with distortion; (b) grid at the center of the lens; (c) grid near the edge of the lens.

reconstructed images by using magnification diversity when the images will be used for automated segmentation and feature extraction.

5. Effect of Distortion on Reduction of Reconstruction Errors

In a high-performance multilens camera, careful design eliminates most distortion effects, while in a camera with a small single lens, distortion will almost always be evident. For a flat-camera design using arrays of small steerable optical elements, it is possible to calibrate the distortion and use it to achieve some diversity that can improve reconstruction results. An otherwise well-corrected lens that exhibits some distortion can be thought of as having a focal length, and hence magnification, that is a function of field position. Figure 7 shows the projection of thin grid lines through a lens with distortion. Magnified views of the center and the edge are also shown.

For a lens with moderate distortion, the difference between the magnification at the center and at the edge will not be as large as the differences used in Section 4. However, it will be demonstrated that at low noise levels the magnification difference for a fish-eye type of lens is sufficient to improve the reconstruction compared with the use of a distortionless lens. Thus an array of identical SIs with space-variant magnification could be used to achieve optical diversity instead of using arrays of different types of lenses. Subimages could be oriented so that a tile of the source image would be observed through different parts of the lenses, creating a moderate diversity of magnification.

Although previous examples in this paper have used shift-invariant responses, this is not required in the model of Eq. (1) or in the derivation of the estimator in Eq. (2). There are several models for lenses with distortion.[18] A nonparametric model[19] or a rational polynomial model[20] can be used. Light from an image source at position (r, θ) that would fall on a detector at position $(r/q_0, \theta)$ with a distortion-free lens, would fall instead on position (r', θ) for a lens with distortion, where the relationship between r and r' is given by

$$r = \frac{r'/q_0}{[1 - C(r'/q_0)^2]}. \tag{15}$$

The parameter q_0 is the magnification factor at the center of the lens, and parameter C is a distortion coefficient. When $C = 0$ there is no distortion.

The expected squared error for arrays of SIs that use lenses with moderate distortion was computed for two cases. After selecting the values for C and q_0, we computed the H matrices for image tiles viewed near the center of the lens, near the edge of the lens, and at a middle distance from the center. These three SI translations produced matrices analogous to H_{q1}, H_{q2}, and H_{q3}, respectively, used in Section 4. The results shown in Fig. 8 for $q_0 = 3$ are similar to those shown in Fig. 5(a). The dashed curves show the expected errors for the first case when only H_{q1} from the central region of the lens is used. The values of the average expected squared errors are very close to the values observed for the distortionless lens. The solid curves show the estimators for the second case, which uses approximately the same total number of measurements as the first case, but uses all three regions of the lens. The diversity achieved by using different parts of the lens reduces the expected squared error significantly compared with that of the first case as the noise is reduced. The reduction in error is less than the reduction in Fig. 5(a), in which the magnification diversity was greater, but it is still

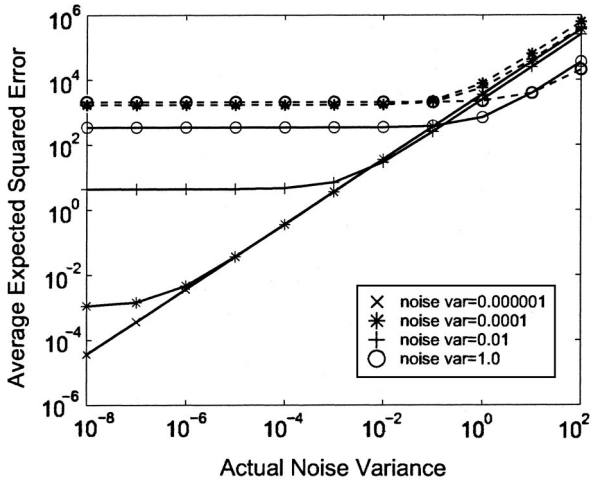

Fig. 8. Performance of a lens with distortion when the tiles are viewed through different parts of the lens (solid curves) versus only near the center (dashed curves).

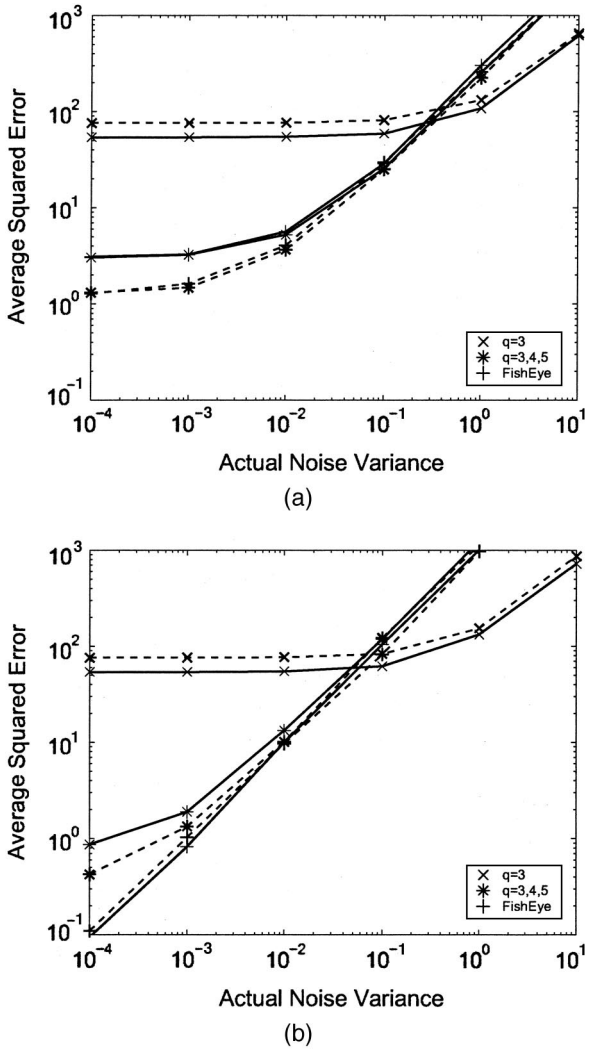

Fig. 9. Simulated performance of a lens with fish-eye-style distortion compared with SI arrays with $q = 3$ only and $q = 3, 4, 5$ on an airplane image (solid curves) and the Bay Bridge image (dashed curves) for noise variances of (a) 1.0 and (b) 0.01.

a significant improvement even for small reductions in measurement noise.

Simulated measurements using the image tiles of the airplane and the Bay Bridge image were used to compute the actual MSE as a function of the variance of the added noise. Figure 9 shows results for 17×17 pixel tiles, with dashed curves representing the lighter lower-contrast Bay Bridge image and solid curves representing the darker higher-contrast airplane image. Results are shown for three cases in each graph. Identical SI arrays with $q = 3$ are compared with magnification diversity from using lenses with distortion and with magnification diversity from using three types of lenses with $q = 3, 4, 5$. For both expected noise environments, the diverse systems outperform the SI arrays with $q = 3$ only. As the actual noise variance increases above the expected value used to compute the MVE, the diverse systems have a higher expected error.

Since it is expected that some distortion will be unavoidable in arrays of small lenses used for a flat-camera design, such as those described in Refs. 7–9, the characteristics of these lenses can be explicitly used in the design of SI geometries to get the potential advantage from the optical diversity due to distortion. The estimators for different regions of the lenses can be computed from calibrated data, and because multiple diverse arrays will be used, small image tiles can be reconstructed independently.

6. Conclusions

Using computational imaging to create high-resolution images from a set of lower-resolution images is desirable to limit the cost of a camera or to construct a camera with a flatter form factor. The analysis of high-resolution image reconstruction from low-resolution images obtained from identical space-invariant imaging systems shows that performance will be limited by a weak response to some spatial frequencies when boundary conditions are not known. A minimum variance estimator (MVE) demonstrates this problem by using computation of expected squared errors and singular-value decompositions.

Acquiring additional images with lower resolution can improve performance more than acquiring additional images at the same resolution, and this relative improvement increases as the overall measurement noise level is reduced. At low noise levels, the number of significant bits per pixel could be increased so that image-processing methods could extract more information from the image source.

The additional images can be obtained with magnification diversity by using imaging components that project the desired scene onto a smaller area of a detector array or by projecting the scene onto the same-sized area of a detector array with a slightly larger sensor size. The improvement is demonstrated through computation of the expected squared error and through simulations by using two test images with varying average intensity levels and varying contrast. A modest magnification diversity can also be obtained from a set of identical imaging systems with a space-variant magnification factor. Subimagers (SIs) could be aligned so that specific tiles of the desired image are captured through different regions of a lens with well-modeled moderate distortion. This is also demonstrated in terms of the expected squared-error computation and simulations.

A second important benefit from effective magnification diversity is that the reconstruction process can be reasonably decoupled compared with the more ill-conditioned problem of reconstruction from identical imaging systems. This will allow parallelization of computation for small image tiles, reduced computation, and improved numerical stability of the result.

The methods presented in this paper can be used to analyze the expected performance of a system with a space-varying response due to combinations of different space-invariant SIs or due to SIs with some well-modeled distortion as long as the system response is adequately modeled by a superposition sum.

References

1. S. Chaudhuri, ed. *Super-Resolution Imaging* (Kluwer Academic, 2001).
2. M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," IEEE Trans. Image Process. **6,** 1646–1658 (1997).
3. S. Baker and T. Kanade, "Limits on super-resolution and how to break them," IEEE Trans. Pattern Anal. Mach. Intell. **24,** 1167–1183 (2002).
4. N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient image superresolution algorithm," IEEE Trans. Image Process. **10,** 573–583 (2001).
5. M. K. Ng and N. K. Bose, "Mathematical analysis of super-resolution methodology," IEEE Signal Process. Mag. **20,** 62–74 (2003).
6. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," IEEE Trans. Image Process. **13,** 1327–1344 (2004).
7. 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).
8. Y. Kitamura, R. Shogenji, K. Yamada, S. Miyatake, M. Miyamoto, T. Morimoto, Y. Masaki, N. Kondou, D. Miyazaki, J. Tanida, and Y. Ichioka, "Reconstruction of a high resolution image on a compound-eye image capturing system," Appl. Opt. **43,** 1719–1727 (2004).
9. M. P. Christensen, M. W. Haney, D. Rajan, S. L. Wood, and S. C. Douglas, "PANOPTES: a thin agile multi-resolution imaging sensor," Presented at the Government Microcircuit Applications and Critical Technology Conference (GOMACTech-05), Las Vegas, Nev., 4–7 April 2005 paper 21.5.
10. M. P. Christensen, V. Bhakta, D. Rajan, S. C. Douglas, S. L. Wood, and M. W. Haney, "Adaptive flat multiresolution multiplexed computational imaging architecture utilizing micromirror arrays to steer subimager field of views," Appl. Opt. **45,** 2884–2892 (2006).
11. S. L. Wood, B. J. Smithson, M. P. Christensen, and D. Rajan, Rerformance of a MVE algorithm for compound eye image reconstruction using lens diversity," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005* (ICASSP' 05) (IEEE, 2005), Vol. II. pp. 593–596.
12. R. N. Bracewell, *Two-Dimensional Imaging* (Prentice-Hall, 1995).
13. A. Macovski, *Medical Imaging Systems* (Prentice-Hall, 1983).
14. T. Kailath, *Linear Systems* (Prentice-Hall, 1980).
15. A. K. Jain, *Fundamentals of Digital Image Processing* (Prentice-Hall, 1989), Chap. 2.
16. R. M. Gray, "On the asympototic eigenvalue distribution of Toeplitz matrices," IEEE Trans. Inf. Theory **IT-18,** 725–730 (1972).
17. S. L. Wood, D. Rajan, M. P. Christensen, S. C. Douglas, and B. J. Smithson, "Resolution improvement for compound eye images through lens diversity," in *Digital Signal Processing Workshop 2004 and the Third IEEE Signal Processing Education Workshop* (IEEE, 2004), pp. 151–155, doi: 10.1109/DSPWS.2004.1437931.
18. M. Born and E. Wolf, *Principles of Optics* (Cambridge U. Press, 1959).
19. X. Ying and Z. Hu, "Distortion correction of fisheye lens using a nonparametric imaging model," in *Proceedings of Asian Conference on Computer Vision* (Asian Federation of Computer Vision Societies 2004), pp. 527–532.
20. C. Brauer-Burchardt and K. Voss, "A new algorithm to correct fish-eye- and strong wide-angle-lens-distortion from single images," in *Proceedings of IEEE International Conference on Image Processing* (IEEE, 2001), pp. 225–228.